Vaasan yliopisto
UNIVERSITY OF VAASA

JOHAN KNIF
BERND PAPE
(Eds)

Contributions to Mathematics, Statistics, Econometrics, and Finance

Essays in Honour of
Professor Seppo Pynnönen

ACTA WASAENSIA 296
STATISTICS 7

FOREWORD

This volume is dedicated to Seppo Pynnönen on the occasion of his 60th birthday and consists of 24 contributions by more than 40 authors in the areas of mathematics, statis-tics, econometrics, and finance. As a consequence of his social and academic skills, Seppo Pynnönen has as a dedicated researcher, teacher, and friend over the years interacted with many people from a wide range of academic areas. This collection provides us an opportunity to express our appreciation and gratitude to him.

Seppo Pynnönen is a very friendly person who is also highly dedicated to all his professional duties such as research, teaching, and administration. Furthermore, he has an ability to stay focused and productive with a positive and inspiring attitude. He is also devoted to his family and keeps up a continuous activity with his sports interest, especially his passion for tennis. Seppo Pynnönen would not travel anywhere without his tennis gear.

We are grateful to all the contributors of this volume for their willingness to submit their work, thus joining us in honoring our dear friend and colleague Seppo Pynnönen. On behalf of all the contributors we would like to say that it is a privilege to have the opportunity to know, collaborate and interact with Seppo Pynnönen and we all look forward to continuing this for many years to come.

We would finally like to thank Tarja Salo for her valuable editorial help with this volume.

Vaasa, January 31, 2014

Johan Knif Bernd Pape

SEPPO PYNNÖNEN 60 YEARS

Seppo Pynnönen was born on March 18, 1954, in Laukaa, Finland and he completed his matriculation exam in Jyväskylä in 1975. He stayed in Jyväskylä until he completed his MSc (1981) and Licentiate (1985) degrees in statistics at the University of Jyväskylä. In 1988 he received his PhD degree in mathematical statistics from the University of Tampere.

The academic career of Seppo Pynnönen started with shorter engagements as an acting lecturer and acting senior lecturer in statistics as well as mathematics and statistics at the universities of Jyväskylä and Vaasa over the period from 1982 to 1985. He finalized his dissertation Testing for additional information in variables in normal theory classification with equal covariance matrices as a research fellow at the Academy of Finland during the years 1986 to 1988. The dissertation was published in Acta Wasaensia No. 23 by the University of Vaasa but he defended his dissertation at the University of Tampere.

In September 1987 he was appointed to the tenure position of senior lecturer in statistics at the University of Vaasa. However, he did not actively continue as a lecturer for long. After a short period in the finance industry, he returned to the academia for several shorter appointments. Over the period from 1990 to 1998, he served as acting associate professor and acting professor in statistics and management sci-ences at the University of Vaasa. It is obvious that these appointments provided him with more time devoted to research than his position as a senior lecturer in statistics. In September 1998 he was appointed to his current position as full ten-ured professor of statistics at the University of Vaasa.

Seppo Pynnönen has a wide research interest. He is primarily concerned with the way statistical modeling and testing are applied. In his publications he has developed new testing and modeling procedures for a wide range of applications. In doing so, he has also shown a genuine interest in the theoretical foundations of the application area in question. This is one of the success factors driving his research.

Seppo Pynnönen has published over 30 articles in refereed international high-ranked scientific journals, over 20 chapters in scientific books and collections, and numerous working and conference papers. The publications cover topics such as variable selection in quadratic discriminant analysis; distributions of linear transformations of residuals from multivariate regressions; measurement and testing in short- as well as long-run event studies; measuring and testing exchange-rate risk exposure; distributional characteristics of risk factors in asset pricing;

measuring calendar effects on asset returns; statistical measurement of stock market reactions to inflation shocks; modeling of the relation between volatility and correlation; and modelling of credit-spread, just to name a few.

He has also applied his statistical knowledge within empirical studies of corporate finance; international stock markets; bond markets; fixed income markets; commodity markets; currency markets; credit markets; markets of foreign direct investments; and even medical studies of osteogenesis imperfecta.

Even though Seppo Pynnönen's main focus in his career has been on research, he has also over the years carried major administrative responsibilities. He has for three different periods been the Head of Department for Mathematics and Statistics at the University of Vaasa for a total of about eight years. After the introduction of the new Finnish university legislation in 2010, he was a member of the board of the first Collegium of the University of Vaasa for three years. Furthermore, for several years he has been active within the Finnish Professors Union.

Seppo Pynnönen has also an international perspective. As a research fellow with funding from the Academy of Finland and with a sabbatical funded by the Finnish Foundations' Professor Pool, he has been visiting Texas A&M University twice. He has actively attended many international scientific conferences and is an active member of, e.g., the Royal Statistical Society, the European Finance Association, the Southern Finance Association, and several national scientific societies. He has also repeatedly served as referee for many high-standard international scientific journals.

One of Seppo Pynnönen's most outstanding traits is his willingness to support and help his colleagues. One good example is his participation in the development of the Masters of Science Program in Computational Finance at the Hanken School of Economics in Vaasa. As one of the first international masters programs in Finland, this program was launched in 1998. Since then Seppo Pynnönen has not only been an active member of the steering group of the program but also actively participated as a teacher in program specific courses. He is also a member of the tenure track steering group at Aalto University.

As part of his profession, Seppo Pynnönen has evaluated applicants for scientific chairs at, e.g, Hanken School of Economics, Helsinki Business School, University of Jyväskylä, and Turku School of Economics. He has, besides having successfully supervised several PhD students at the University of Vaasa, been the official examiner of around 20 PhD dissertations and been the official opponent at close to 15 PhD defense seminars.

Another personal characteristic of Seppo Pynnönen is his ability to take initiative. This characteristic is most obvious within his research projects, but it is also evident from his ability to initiate and manage the organization of scientific workshops. He chaired the organizing committee of the Noon-to-Noon Workshops in 1998 and 2005 and organized the Annual Meeting of the Finnish Statistical Society in 2001.

The work of Seppo Pynnönen has been recognized both nationally and internationally. In 2011 he received an award from the Suomen Arvopaperimarkkinoiden Edistämissäätiö for publishing his paper "Event study testing with cross-sectional correlation of abnormal returns" in one of the three top tier journals in finance. In 2010 he received the best paper award in international finance from the Midwest Finance Association. Furthermore, in the same year he was honored with the 1st degree knight insignia of the Finnish White Rose Orden. This insignia is awarded by the President of Finland.

We have all experienced Seppo Pynnönen as a person with a very positive attitude. No scientific problem seems to be too complicated for him to solve. However, life is sometimes hard and unjust. Seppo Pynnönen has been forced to meet the un-fairness of life too many times. Despite this, with his positive attitude, kindness, and dedication, he has fought back and is today a happy grandfather and one of Finland's most respected academics in his areas of research.

CONTENTS

Foreword V

Seppo Pynnönen 60 years VII

I MATHEMATICS

Ilkka Virtanen
Symmetry between knowing about the future and about the past: A systematic approach applying futuribles and historibles 1

Seppo Hassi
A completion problem for an unbounded nonnegative block operator................ 21

Ka Lok Chu, Simo Puntanen, and George P. H. Styan
Properties of a special symmetric matrix associated with postage stamps 33

II STATISTICS

Erkki P. Liski
Chances of getting pregnant: An estimate using the duration of non-conception with an application to real data.. 45

Wolf-Dieter Richter
Classes of standard Gaussian random variables and their generalizations 53

Jouni Helske and Jukka Nyblom
Improved frequentist prediction intervals for ARMA models by simulation 71

III ECONOMETRICS

Jan Antell and Timo Korkeamäki
Illustrating the effects of cross-sectional correlation on event study results: The Private Securities Ligitation Reform Act of 1995 revisited.......................... 89

Terhi Hagnäs
A sign test of cumulative abnormal returns in event studies based on generalized standardized abnormal returns 105

Markku Lanne and Helmut Lütkepohl
A statistical comparison of alternative identification schemes for monetary policy shocks 137

Niklas Ahlgren and Paul Catani
Tests of cointegration rank with strong persistence in volatility:
An application to the pricing of risk in the long run 153

Leena Kalliovirta, Mika Meitz, and Pentti Saikkonen
Modeling the Euro-USD exchange rate with the Gaussian mixture
autoregressive model 171

Kurt Brännäs
The number of shareholders: Time series modelling and some empirical
results 195

IV FINANCE

Kenneth Högholm, Johan Knif, and Gregory Koutmos
Asymmetric dynamic linkages between returns on banks and other industry
portfolio returns 209

Ali Anari and James Kolari
Fisher, Wicksell, Taylor and the negative interest rate elasticity of
inflation rates 231

Johan Knif, Dimitrios Koutmos, and Gregory Koutmos
Hedge funds: market timing and the dynamics of systematic risk 255

Richard T. Baillie, G. Geoffrey Booth, and Sanders S. Chang
A modern econometric analysis of an ancient exchange rate market 267

Mika Vaihekoski
A note on the calculation of the risk-free rate for tests of asset pricing models
and event studies..291

Kenneth Högholm, Johan Knif, and Tobias Romar
Short term value creation for the bidder: Evidence from Finland 307

Jussi Nikkinen and Sami Vähämaa
Price clustering of stock index derivatives:
The case of thinly traded Finnish markets in the late 1990s..............................329

Ahmad Arslan, Syed Mujahid Hussain, and Jorma Larimo
Foreign direct investment ownership mode strategy and the influences of
pillars in economic freedom..347

Esa Jokivuolle
Reflections on the need for structural reform of banks......................................371

Gunnar Rosenqvist
Arbitrage at the racetrack 381

Jarkko Peltomäki and Emilia Vähämaa
Friday the thirteenth and stock index returns 393

Martti Luoma
Two speculative risk measures in stock markets.. 409

I

MATHEMATICS

SYMMETRY BETWEEN KNOWING ABOUT THE FUTURE AND ABOUT THE PAST: A SYSTEMIC APPROACH APPLYING FUTURIBLES AND HISTORIBLES

Ilkka Virtanen
University of Vaasa

1 Introduction

Knowing about the future has its intrinsic canon of sufficient scientific legitimation. Futurological canon legitimizes beliefs and opinions about the future as knowledge of the future. Instead of considering the future as a single predetermined case, a fan of possible futures, called futuribles is considered as a proper object of futurological conjecture. The manifold conceptualization of the future has a long history from Luis de Molina and others in the 16th century to Bertrand de Jouvenel in the 1950's and 60's. Malaska and Virtanen developed, based on this conceptualization, a general set-theoretic construction, called a theory of futuribles for futures knowledge inquiry. A short summary of the futures manifold and futuribles is presented as the first part of the paper. Before this, the three dimensions of time and their interdepencies are dis-cussed.

The paper further shows that the theory of futuribles can be applied also in history context, to describe and analyze the past and the present. Concepts of historible and presentibles are introduced for that purpose. Examples of historible applications are given in the areas of biological evolution, history of habitation, decision analysis and contrafactual history doctrine.

1.1 Three dimensions of time

A wish to know about the future has been a human intellectual characteristic since Antic Greece and Rome, as we know from the historic stories about the Delphic oracle Sibyl. Human interest in the future can be traced back even to the ancestors of Homo sapiens, as exemplified for instance by Y. Coppens with archaeological findings (Malaska & Virtanen 2009: 65). We can, therefore, say that the future was invented by the emerging consciousness of mind already at the dawn of humankind.

The future is one dimension or part of the total time flow we have already met and still will meet. Consciousness about these three dimensions is typical to humans as presented in a fascinating children's poem "Aika (Time)" (Korolainen 2005: 48-49):

Time

Man has yesterday, today,
man has tomorrow,
grandpa mouse pondered.
And wrinkle in his brow deepened.

Man has all the times,
the mouse has only ***now***, that's tight boundaries.
But – and here grandpa held back:
Is the man really more happy?

Is it really possible to know about the future in a firm and trustworthy way? Scientific knowledge is nothing else than a well-grounded true belief. All sciences from mathematics and natural sciences to social and humanistic sciences stick to this as an epistemological commitment. It means that a subjective belief, intuition or opinion is accepted as an objective knowledge when there is sufficient evidence to convince the others that the belief is true and credible. Knowing about the future makes no exception in this respect.

Knowing about the future is, however, different from knowing about the two other dimensions of the time, the past and the present. Unlike the past or present events, future events do not materialize to our senses, when a desire to know about them appears in our minds. Knowing about the past and present can be grounded on observable factual material evidence whereas conjecturing the future relies on non-factual and intentional data, i.e. mind images and rational conjectures. Pentti Malaska has described the three dimensions of time with a poem "When the time becomes reality" (Malaska & Virtanen 2009: 65; original Finnish version Malaska 1979).

When the time becomes reality

Time flows to the present from two directions,
from the past and from the future.

From the past
as our deeds accomplished and
events materialized
observable to our senses, and

From the future
as our aims and intensions,
objectives targeted, hopes or despairs
experienced by our mind.

The present attracts the times
and moulds them together
as a cosmic black hole,
whereupon they cannot help
but creating our reality.

The three dimensions of time are in close relation with each other. Thinkers have emphasised this relation for centuries. As a Finnish example we can take Michaele Wexionio, a member in the professoriate at the Royal Academy of Turku (the present University of Helsinki) at the time of the Academy's inauguration in 1640. In 1642 he wrote in his doctoral dissertation Discursus Politicus De Prudentia (Wexionio 1642):

Fundamentals of the wisdom are that we choose the good and avoid the bad. In order to get this wisdom we need a threefold ability:

- firstly, memory to analyse the past,
- secondly, understanding to observe the present, and
- thirdly, attention to foresee the future.

1.2 Relativity and duality of time

Since the works by Albert Einstein it has been generally accepted that time is relative by nature. The time passed depends on the observer. This holds good also for history and future. History and future can even be dualistic for each other. This can be exemplified – following the idea presented by a former member of the Academy of Finland, professor Oiva Ketonen – with a hypothetical case from astronomy.

Let us assume that an astronomer observes with a telescope a star which is at a distance of 2500 light years from the earth. Things which the astronomer now observers to happen on the star have happened there 2500 years ago. Things happened since then are historical from the point of view of the star but for the observer on the earth they are events in the star's future, and it is not possible to get any information about them until in the course of time (e.g. about the star's present after 2500 years from now). And vice versa, if there were on the star an observer, who would at the moment (at the earth's present) watch the state of affairs on the earth, the observer could see for example the on-going battle between Athens and Persia on the fields of Marathon. We would have no means to tell the observer what has happened here on the earth after those days.

Our example above has demonstrated that although knowing about the future is different from knowing about the past and the present, these three dimensions of time have much in common and they are closely related. Therefore, it is natural to think that there must be a methodology in the framework of which it is possible to give a common formal presentation for the processes which have been going on in the past and are continuing towards the future through the present.

Malaska and Virtanen (2009) have presented a systemic approach, the theory of futuribles, for describing in a formal way scenarios and future images used in futures studies. They also presupposed that the same approach could be applied to processes already happened in the past. In analogy to futuribles used to analyse the future they called tools for describing the past historibles. A more detailed presentation of historibles was, however, left for further research. The objective of this paper is now to present, using analogy emerging from the theory of futuribles, a formal definition for the historibles and to show how they can be applied in historical contexts, and further to demonstrate the conceptual symmetry between futuribles and historibles.

2 Generic design of futures manifold

Futurible-conception, i.e. the manifold of possible futures instead of a single future, is well accepted in modern futurological inquiry. Growth of the popularity of scenario writing since the 1960's demonstrates this well, as exemplified by the sample of the references in Malaska and Virtanen (2009: 68). Bertrand de Jouvenel coined the term futuribles to a fan of futures in his futurological classic The Art of Conjecture (de Jouvenel 1967). Malaska and Virtanen (2009) utilised the possibilities which this conceptualization offered to futures studies and presented a logical framework for the futurible-conception and called it the theory of futuribles.

This section deals with a formal presentation of futures manifold and futures mapping in the form of generic futures tables, futuribles and futures synopsis. It is based on the work cited above (Malaska and Virtanen 2009). A concise summing up of this work is motivated by the need to present the general methodology earlier applied in futures mapping for inventing the methodology also in history context to describe and analyse the past.

2.1 Futures manifold

Designing a futures map or image starts by identifying the issues which are regarded as vital and relevant in the study; they are called futures variables. Examples of future issues and variable names could be “economic growth”, “export”, “aging rate of population”, “literacy rate”, “dematerialization”, “equality”, “rebound”, “environmental stress”, “energy need”, “material consumption”, “technology development”, and “welfare productivity of GDP” etc. Each issue is itemized into mutually exclusive, alternative possibilities of the issue variety. The items of the issue variety are called values of the variable and the total set of them forms the domain of the variable in the study.

Let the futures variables be denoted by X_i, (i = 1, 2,…, K), where K is the number of identified variables. The domain of the variable Xi is a set of the varieties $\{x_{ij} \mid j=1, 2,\ldots, n_i\}$, where n_i is the number of the different values of X_i.

When an issue is apt to quantitative measurement, the values of the variable are quantities. A future variable may also be measurable only on an ordinal scale, or it may represent plain qualitative aspects of the future on a nominal scale. If the variable has only one value, the variable is called a future constant. For instance, until today the planetary conditions on the Earth have been generally regarded as constant. Nowadays the possibility of an irreversible climate change has transformed that aspect from a future constant to the class of variable. A variable having a domain of a few values only may be taken to serve as a future parameter. The parameter can be used for partitioning the futures space into mutually exclusive sub-spaces. In summary, the futures manifold is defined in the form of formulas (1) to (3).

Let the collection of the future variables Xi be symbolically denoted by the variable set X. We then have

$$X = \{X_i \mid i = 1,\ldots,K\}. \tag{1}$$

The value domains of the variables are

$$X_i = \{x_{ij} \mid j = 1,\ldots,n_i\}, \quad i = 1,\ldots,K. \tag{2}$$

The elementary system defined by (1) and (2) is called a *futures manifold* $\boldsymbol{X}$. It can be interpreted as a K-dimensional coordinate system spanned by the variable set X. The futures manifold X can be symbolically presented as a set of K-dimensional Cartesian points or vectors $\times\boldsymbol{X}_p$:

$$\boldsymbol{X} = \{\times\boldsymbol{X}_p \mid \times\boldsymbol{X}_p \in X_1 \times X_2 \times \ldots \times X_K\}. \tag{3}$$

2.2 Generic table of the futures manifold

The futures manifold $\boldsymbol{X}$, i.e. the system (1) to (3) is possible to represent alternatively in the form of a table. For each future variable X_i a row i of the table is designated and to each value x_{ij} of the variable X_i a cell (i, j) in that row is designated. The resulting table of the manifold is called the *generic table*. The generic table obviously has K rows and n_i cells in the rows. A design of the generic table is illustrated in Table 1.

Table 1. Generic table design for a futures manifold

<table>
<tr><th>Futures variable</th><th colspan="4">Generic table</th><th># cells</th><th>Interpretation of the type of the variable</th></tr>
<tr><td>X_1</td><td colspan="2">x_{11}</td><td colspan="2">x_{12}</td><td>2</td><td>an optional parameter</td></tr>
<tr><td>X_2</td><td>x_{21}</td><td>x_{22}</td><td>x_{23}</td><td>x_{24}</td><td>4</td><td>a variable</td></tr>
<tr><td>X_3</td><td>x_{31}</td><td colspan="2">x_{32}</td><td>x_{33}</td><td>3</td><td>a variable</td></tr>
<tr><td>X_4</td><td>x_{41}</td><td colspan="2">x_{42}</td><td>x_{43}</td><td>3</td><td>a variable</td></tr>
<tr><td>X_5</td><td colspan="4">x_{51}</td><td>1</td><td>a futures constant, background</td></tr>
<tr><td>K=5</td><td colspan="4">M=13</td><td>$\bar{n}$ =2.6</td><td></td></tr>
</table>

In Table 1, the bottom row (variable X_5) has only one value in its domain indicating that the respective issue is a constant futures background and the variable is a future constant. The variables X_2, X_3, and X_4 have four or three values in their domains. They represent conventional future variables with given domains. The first variable X_1 has two values. This variable could be regarded, if relevant, as a future parameter.

2. Technology / Organisation	3. Culture / Values	4. Globalisation	5. Macro economic policies (EMU)	7. Social and employment policies
No major breakthrough. Downsizing. Continuing de-specialisation of Europe in high-tech.	Increasing individualism. Fear of the future.	Globalisation continuing, sectoral resistances, local difficulties.	Broad EMU with limited coordination and no major tensions.	Continuing "decremental" adjustment of social protection.
No major breakthrough. Increasing dualism. Increasing de-specialisation of Europe in high-tech.	Strongly increasing individualism. Social and geographical segregation. Power of lobbies.	Globalisation accelerating. "Borderless world"	Broad EMU with limited coordination and major tensions.	Strong labour market deregulation. Residual welfare state.
Major breakthrough. Europe innovating and/or catching up.	Renaissance of social/ecological awareness. Regions/localities experiments.	Globalisation slowing down, trade conflicts, regional blocks.	Broad EMU with strong coordination.	Strong resistance against welfare state reform.
Major breakthrough. Increasing technologically induced inequality. Europe catching up.	Revolt of the bottom-half against globalisation.	Global crisis	Failure of EMU	Radical reform of welfare state: universalism and individual incentives.
Major breakthrough. Increasing technologically induced inequality. Europe falling behind.				

Figure 1. Generic table of an EU study; the futurible of the non-shaded cells is called the "Laissez faire" future in the study. The table layout is transposed as indicated in the text. (Bertrand et al. 1999, 94)

Figure 1 shows a concrete example of a generic table taken from an EU study (Bertrand et al. 1999). For layout reasons the table is presented in a "transposed form", i.e. the five (K=5) future variables appear horizontally and their values (with 4 to 5 cells) vertically. The non-shaded white cells in the table combined represent a point in the K-dimensional futures space having received a verbal definition of the "Laissez faire" future in the study.

The generic table is a morphological setting of the future "sceneries", i.e. a representation of the possible futures. Each future issue or a variable has multiple varieties such that the rows in the table may have a varying number of cells. The number of the cells in a variable row gives an indication of the coarseness of resolution of the issue presentation. The more cells there are, the finer is the resolution, and vice versa.

Let M denote the total number of cells in the table and $\bar{n}$ the mean number of cells per row. We can then write

$$M = \sum_{i=1}^{K} n_i = K \cdot \bar{n}. \tag{4}$$

Metaphorically, the number of future variables K refers to the *extension of the futures space*, the bigger K the farer the horizon of the space from a centre. The mean number of cells $\bar{n}$ implies the *mean issue resolution*. The total number of cells M, interpreted in (4) as the product of the extension and the mean resolution, indicates the total *expressiveness* of the futures manifold under study.

2.3 Synoptic design of futures mapping

An element of the futures manifold in (3) and its equivalent presentation as a point in the K-dimensional coordinate system is called a synopsis. A synopsis is an exhaustive and exclusive collection of values of the successive variables in the generic table. The synopsis is a design composed of one and only one cell from each variable row of the table. Formally a synopsis F_q is defined by

$$F_q = \left(x_{1q_1}, x_{2q_2}, \ldots, x_{Kq_K}\right),\ q = 1,\ldots,N;\ q_i \in \{1,\ldots,n_i\},\ i = 1,\ldots,K. \tag{5}$$

In Formula (5), N stands for the number of all potential synopses. It depends on the number of the possible values of the variables in their domains according to the multiplication formula (6). There may be some bans which negate the simultaneous presence of some values of distinct variables wherefore the number of feasible synopses may be smaller than the number of all synopses N.

$$N = \prod_{i=1}^{K} n_i = n_1 \times n_2 \times ... \times n_K . \tag{6}$$

For rationalizing the notation the following Dirac's Delta type table D^q is introduced. D^q is a table with the same number of rows and cells and the same format as the generic table. Each cell value of the D^q-table is either 0 or 1 in such a way that each row contains one and only one 1. Let the i^{th} row (i = 1,2,…,K) of the D^q-table be denoted by D^{q_i} and let us further assume that it has its non-zero element in the position $p_i \in \{1,2,\ldots, n_i\}$, i.e. $D^{q_{ip_i}} = 1$ and $D^{q_{ij}} = 1$, when $j \neq p_i$. The table element $D^{q_{ip_i}}$ can be used to pick a cell value x_{ip_i} from address p_i of the futures variable X_i in the generic table. Together all the cells in the D^{q_i}-rows with i = 1,…, K and with p_i = 1,2,…, n_i pick an exhaustive set of the value elements of the future variables that constitutes a synopsis. The Dirac's Delta table thus defines the formal picking of a specific synopsis from the set of all synopses within the generic table. The set of all Dirac's Delta tables is presented by a notation of $\boldsymbol{D} = \{D^q\}$.

With the D^q -table notation a synopsis F_q of $\boldsymbol{X}$ can be presented with scalar product operations (operation denoted by ·) between corresponding rows of the generic table $\boldsymbol{X}$ in (3) and of the Dirac's Delta table D^q:

$$F_q = (D^{q_i} \cdot X_i \,|\, i = 1,2,\ldots,K) = (D^{q_1} \cdot X_1,\ D^{q_2} \cdot X_2,\ldots,\ D^{q_K} \cdot X_K). \tag{7}$$

The operation in (7) results in a vector F_q whose components are scalar products of the row vectors of the tables D^q and $\boldsymbol{X}$. There is one to one correspondence between this result and the previous notations of $\{\times \boldsymbol{X}_p\}$ and $\{F_q\}$.

The set of all synopsises $\{F_q\}$ spanned by the generic table $\boldsymbol{X}$ is called the futures space $\boldsymbol{F}$. With the notation of $\boldsymbol{D}$ the futures space will have a simple expression as a "multiplication" operation (denoted by $\circ$) with the generic table $\boldsymbol{X}$

$$\boldsymbol{F} = \{F_q \,|\, q = 1,\ldots,N\} = \{(D^{q_1} \cdot X_1,\ D^{q_2} \cdot X_2,\ldots,\ D^{q_K} \cdot X_K) \,|\, q = 1,\ldots,N \} = \boldsymbol{D} \circ \mathbf{X} \tag{8}$$

2.4 Futurible – a basic unit of futures mapping

The synopsis concept belongs to the syntactical design of futures mapping. It is a logical form of a possible future. Synopsis and futurible are synonymous equivalents in the sense that futurible is a semantic counterpart of synopsis. Futurible refers to the content, while synopsis gives the logical form in which the content is to be presented. Therefore, the whole set of synopses in (8) also means the fan of the futuribles mapped onto the generic table $\boldsymbol{X}$, and F_q stands for a single futurible.

Each future variable defines an independent dimension of the future into which direction the futures stories can be told and varied within the domain of the variable. The generic table with its K variables spans a K-dimensional futures space, where each futurible represents a map of a possible future "scenery".

It is plausible that certain relations may exist between future variables denying a possibility of some of their values to coexist. In addition, constraints may occur also between futuribles to follow each other. Some futurible may be a necessary condition for another one, and this in turn to yet another one etc., while constraints of another type may deny a succession between futuribles. For instance, the present which in the logical and formal sense – although not semantically – is also a synopsis and a futurible, is a necessary though not sufficient condition for any future to come. The present does not predetermine the course of the successive futuribles, but neither does it leave the course of the future unconstrained. From the synopsis of the present several possibilities are available for futuribles to unfold. Some possible courses of the future may divert from each other irreversibly depending on the different constraints, while other courses may pass partly through the same futuribles. It is, in addition, well- grounded to assume that in the course of the future a given futurible may be reachable from several preceding ones but not from whichever futuribles. A possible chain of futuribles is called a course of the future. Futuribles as well as futures courses may be attached with specific attributes such as probable, desirable, avoidable, non-feasible, or a threat, a utopia or a dystopia.

3 Historibles and presentibles – counterparts to futuribles for the past and the present

As considered in the Introduction, the conception that knowing about the future in a firm and confident way is conceivable, has received a common commitment in modern futures studies. On the other hand, knowing about the future is different from knowing about the past and the present. The latter two can be grounded on observable factual material evidence whereas conjecturing the future relies on non-factual and intentional data, on mind images and rational conjectures. Future is no entity but a continuously unfolding process to be forethought in the mind scenery. In addition, we pointed out that the three dimensions of time, the past, the present, and the future are mutually related. The past and the present make the future possible but they also constraint the future unfolding. The future remembers some of the past and present, but they never fully determine the future course.

At the end of the preceding section we already shortly considered these interdependencies in the terms of futuribles and synopses. We brought forward that the present can in the logical and formal sense be regarded as a synopsis or a futurible which is a necessary though not sufficient condition for any futures to come. The same is equally true for the past. The objective of this section is to present both the past and the present applying the formalism developed for the future and covered by Formulas (1) to (8).

3.1 Design of historibles and presentibles

As already stated, knowing about the past and about the present are – when compared with knowing about the future – are conceptually more similar with each other. Knowing can be grounded on observable factual material evidence. Therefore, the synoptic design of the past and the present can be done uniformly. The design is presented for the past, for the present it is analogous. We call the counterpart for futurible in the design of the past a historible, and a presentible when the present is considered.

Let the issues identified for the past, the history variables, be denoted Y_i, (i = 1, 2,…, L), where L is the number of identified issues. The domain of the history variable Y_i is a set of the varieties $\{y_{ij} \mid j=1, 2,\ldots, l_i\}$, where l_i is the number of the different values of Y_i. It is clear that the term variable has a different meaning in the case of the past (and the present) than in the case of the future. This will be considered more closely by examples later on. Analogously with the elementary system of Formulas (1) and (2), we can define the *histories manifold* as a system of formulas

$$Y = \{Y_i \mid i = 1,\dots, L\}\text{, and} \tag{9}$$

$$Y_i = \{y_{ij} \mid j = 1,\dots, l_i\},\quad i = 1,\dots, L. \tag{10}$$

Again, it can be interpreted as an L-dimensional coordinate system spanned by the variable set Y. The histories manifold, denoted by $\boldsymbol{Y}$, can thus be symbolically presented as a set of L-dimensional Cartesian points or vectors $\times\boldsymbol{Y}_p$:

$$\boldsymbol{Y} = \{\times\boldsymbol{Y}_p \mid \times\boldsymbol{Y}_p \in Y_1\times Y_2 \times\ldots\times Y_L\}. \tag{11}$$

The histories manifold $\boldsymbol{Y}$, i.e. the system of Formulas (9) to (11) is also possible to be represented, following the lines in forming Table 1, in the form of a generic table. The design of the generic table is straightforward and is therefore omitted here.

Equivalently to the future synopsis, a history synopsis is a design composed of one and only one cell from each variable row of the histories manifold table. Formally a history synopsis, H_q, is defined by (12):

$$H_q = \left(y_{1q_1}, y_{2q_2},\dots, y_{Lq_L}\right),\ q = 1,\dots,\Lambda;\ q_i \in \{1,\dots, l_i\},\ i = 1,\dots, L. \tag{12}$$

In Formula (12), Λ stands for the maximum number of all potential synopses and is given by

$$\Lambda = \prod_{i=1}^{L} l_i = l_1 \times l_2 \times\ldots\times l_L. \tag{13}$$

Again, as in the case of future synopses, there may be some bans which negate the simultaneous presence of some values of distinct variables in which case the number of feasible history synopses is smaller than Λ.

3.2 Characteristics of history and present variables

What has been said above indicates that variables in histories and presents manifolds are semantically different from those in the futures manifold. One may even ask if it is possible to think any variability in the present and history courses. Is it quite the contrary true that what has happened in the past is a fact and contains no variability? The answer is naturally yes when the physical process itself is in concern. But our knowledge about the history is imperfect, and along with the progress of research the historical facts may change. As a consequence, when the past is described in the form of a historible or a history synopsis the resulting entity value depends on the time of data on which the creation of the entity value is based. In the following some examples are given to describe the characteristics of variables in histories manifolds, i.e. in historibles and history synopses.

Biological evolution

One of the most revolutionary discoveries in last two centuries' science is the isolation of the DNA molecule and its applications in biology, evolution, history and anthropology, and in various areas of technology. For example, knowledge about the evolution of living plant and animal species on the earth has changed dramatically along with wide-ranging invocation of DNA technology after the 1960's. In history tables, where the variables represent certain issues about the development of living species on the earth, the values of the variables rest highly on the time the knowledge used is from. A historible which is constructed by picking from the history table variable values based on knowledge in the 1950's may be radically different from another historible from the same table when the newest knowledge of 2010's is used. The physical process on the background of the variable values is the same but changes in the knowledge have changed the values. Differences in historibles are a result of progress in science. Increase of knowledge has created new historibles into the histories manifold.

Habitation history of Finland

As an anthropological example we can take the history of habitation of Finland. According to the current view of the habitation history of Finland the first people came to Finland about 10000 – 11000 years ago after the thick continental glacier had drawn back from our country. The first people came from east because most southern and western parts of the country were under water. Afterwards, along with the land rising also people from west settled the southern and western parts of the country. The borderline between these two groups of people follows approximately the border between Sweden and Russia established in Pähkinäsaari

Peace in 1323. It is possible to see some anthropological, linguistic, cultural and religious signs of this historical borderline even in today’s Finland.

Until the latter half of the 20th century the dominant view on the habitation history was that the birthplace of the Finns was solely upon the bend of the river Volga in Eastern Europe. Later archaeological, genetic and linguistic findings have shown, however, that the Finns have ancestors both in the East and in the West. New information has produced a different historible than the old knowledge.

Recent archaeological and geological findings in Ostrobothnia have opened another quite dramatic view for the habitation history of Finland. The origin of this invention is in the discovery and excavation of the Wolf cave in Kristiinankaupunki municipality in Southern Ostrobothnia in the 1990’s.

Wolf Cave (Wolf cave 2013) is a wide horizontal crevice in the primary rock and is named for its location on Wolf Mountain. The cave was formed as a result of erosion, and it is estimated to be more than 2.6 million years old. In the interglacial period, when the sea level was just outside the mouth of the cave, it was filled with layers of sediment and remained untouched until 1996, even though the cave was widely known in the area. The cave opening is 116.5 meters above the current sea level, and the ceiling of the cave is 2.2 meters high at the highest point. It is difficult to precisely determine the size of Wolf Cave because it is still partially filled by sediment layers, but it is estimated to be over 400 m^2. According to the Wolf cave research group, the cave is the only place on earth where evidence of human inhabitancy has been found in a place that was later, during the ice age, covered by a continental glacier. Wolf Cave is northern Europe's oldest known human dwelling site.

The research group (Wolf cave 2013) further presents that they have found in the sediment levels of Wolf Cave evidence of human habitation that includes stone tools, stone chips left from the making of such tools and old hearth remains. Based on the sediment level in which these artefacts were found and age calculations from analysis of pollen samples, these artefacts are estimated to be at least 120 000 years old. This means that the inhabitants must have been Neanderthal men and they have dwelt in the cave prior to the last ice age.

The interpretation of the findings is, however, very controversial, and the conclusions of the research group, although supported by some experts of archaeology and from National Board of Antiquities and Historical Monuments, have not yet gained any common acceptance. But if the claims of the research group turn out to be valid, the habitation history of Finland changes totally.

Both of the examples above show that knowing about the past is not necessarily static. New discoveries and new ways to observe and interpret the vestiges from the past may change our understanding of the times gone. The historical process itself is a definite fact but our knowledge about the process may change along with time. When the past is described in the form of historibles or history synopses a fan of histories may be found. The present knowledge indicates the most relevant and trustworthy historible or synopsis of the history manifold.

Uncertainty of the future – uncertainty in the past and in the present

The following text-book example shows that uncertainty which is linked to the future may have its source deeply in the past. In fact, the uncertainty may not be a physical feature of the future at all. On the contrary, it is a feature of the past – and it has physically born in the past. The example to be presented is the classical Oil drilling problem by Howard Raiffa (Raiffa 1968).

Oil drilling problem

The general problem. An oil wildcatter must decide whether or not to drill at a given site before his option expires. He is uncertain about many things: the cost of drilling, the extent of the oil or gas deposits at the site, the cost of raising the oil, and so forth. He has available the objective records of similar and not-quite-so-similar drillings in this same basin, and he has discussed the peculiar features of this particular deal with his geologist, his geophysicist, and his land agent. He can gain further relevant information (but still not perfect information) about the underlying geophysical structure at this site by conducting seismic soundings. This information, however, is quite costly, and his problem is to decide whether or not to collect this information before he makes his final decision: to drill or not to drill.

Specified problem in a simple form. The oil wildcatter must decide either to drill (act a_1) or not to drill (act a_2). He is uncertain whether the hole is dry (state θ_1), wet (state θ_2), or soaking (state θ_3). His payoffs are given in the following table:

Table 2. Payoffs in the Oil drilling problem

State	Act	
	$\mathbf{a}_1$	$\mathbf{a}_2$
Dry (θ_1)	-\$70000	0
Wet (θ_2)	\$50000	0
Soaking (θ_3)	\$200000	0

We assume here that the cost of drilling is \$70000. The net return of the consequence associated with the (Wet, a_1)- or (θ_2, a_1)-pair is, for example, \$50000, which is interpreted as a return of \$120000 less the \$70000 cost of drilling. Similarly the other figures.

<u>Sample Information.</u> At a cost of \$10000, our wildcatter could take seismic soundings (experiment e_1) which will help determine the underlying geological structure at the site. The soundings will disclose whether the terrain below has (a) no structure (outcome NS) – that's bad, or (b) open structure (outcome OS) – that's so-so, or (c) closed structure (outcome CS) – that's really hopeful. The experts have kindly provided us with the following table, which shows the joint and marginal probabilities.

Table 3. Joint and marginal probabilities associated with seismic soundings

State	Seismic outcome			Marginal probability of state
	No S	Open S	Closed S	
Dry (θ_1)	.300	.150	.050	.500
Wet (θ_2)	.090	.120	.090	.300
Soaking (θ_3)	.020	.080	.100	.200
Marginal probability of seismic outcome	.410	.350	.240	1.000

From the table we can see, for example, that the joint probability of θ_1 (dry) and OS (open structure) is .150; the (marginal) probability of θ_1 is .300 + .150 + .050 = .500; the probability of OS is .150 + .120 + .080 = .350.

Discussion. The future of the oil business depends on many wild cards: the amount of oil at the site, the result of seismic soundings (if applied), and the cost of drilling. To run the business, the wildcatter has many decisions to do: to drill or not to drill, to take or not to take seismic soundings before the drilling decision and, in the case of taking seismic soundings, the way of using this approximate seismic information. To use these elements, it is possible to present the future of the oil business in the form of a futures manifold, as a collection of futuribles or future synopses. Values of the future variables come partly from the wildcatter's decisions, partly from the "decisions of the nature", i.e. due to uncertainties. The use of futuribles does not offer, however, any particularly efficient way to "solve" the wildcatter's decision problem. Its value is rather in opening a view to the future of running the business. It maps all potential outcomes of the future oil business for the wildcatter. As efficient measures to solve the problem Raiffa presents the traditional decision tree analysis and strategy matrix technique.

When we take a closer look at the future variables whose source of variability is in uncertainty, we will find that they differ by nature from each other. The cost of drilling and the result of seismic soundings are pure future variables, the processes will happen and the values of the variables will be determined in the future. But the case is different when the amount of oil at the site is concerned. It is physically no uncertainty in the amount of oil. The amount of oil, if any, has existed at the site already for millions of years. Physically, the amount of oil is rather a history (and a present) than a future variable, and its value has been fixed for a long time ago. The uncertainty in the amount of oil and thus unawareness of the value of the corresponding variable comes from the lack of information available for the decision maker.

Following the physical process in the oil drilling case, the natural way to proceed would thus be to present the past as three historibles with differing amounts of oil as the values of the proper variable. These three historibles lead to three potential presents (possible to be described in the form of three presentibles). Each of these potential presents has futures specific to it. Each historible-presentible chain forms the basis for a separate fan of futuribles.

Contrafactual history conception

The concept of historible offers an elegant way to illustrate the treatment of history by the means of contrafactual history doctrine. In applying this doctrine it is thought that something in the history might have happened otherwise than it really happened. From this alternate starting point a new course of contingencies is generated. This course of contingencies leads to an alternate present for the really existing present. It is, of course, possible to create several such contrafactual histories. Each separate contrafactual history can be presented in the form of a historible.

It is easy to see that this type of historible possesses technically the properties of a futurible. The starting point in the past corresponds the present in futurible formation. The issues considered in a contrafactual historible are not based on observable factual material evidence but on non-factual and intentional data, on mind images and rational conjectures as is the state of affairs for a futurible. Due to the speculative nature of contrafactual history conception it is not considered, however, in more details here.

3.3 Conclusions

The three dimensions of time: the past, the present, and the future were presented in the paper. Interdependencies between these three dimensions were discussed and exemplified with excerpts from literature and philosophy.

Knowing about the future has a different canon of legitimation than that of knowing about the past and present. It can be regarded as more general in the scientific sense because of the intentional characteristic of knowledge of the future. In the paper a logical construction developed by Malaska and Virtanen (2009) and based on a morphological setting called the generic table of the futures manifold was presented together with a syntactic theory of futurible.

The theory of futuribles was utilized to enlarge it to cover also the past and the pre-sent. Analogously to futures manifold the concepts of histories manifold and presents manifolds were developed. A histories (presents) manifold can be presented as a set of historibles (presentibles) or as a set of history (present) synopses. Logically these systemic concepts relating to the different dimensions of time are equivalent, but due to the fact that knowing about the future differs from knowing about the past and the present, the semantic interpretation of the concepts is different. Examples of historible (and presentible) applications were given in the areas of biological evolution, history of habitation, decision analysis and contrafactual history doctrine.

References

Bertrand, G., Michalski, A. & Pench, I.R. (1999). Scenarios Europe 2010. *Five Possible Futures for Europe*. European Commission, Forward Studies Unit.

Jouvenel, Bertrand de (1967). *The Art of Conjecture*. New York, N.Y.: Basic Books.

Korolainen, Tuula (2005). *Kuono kohti tähteä* (Muzzle towards the star), Hämeenlinna: Lasten keskus.

Malaska, Pentti (1979). *Avoimet ja sumeat systeemit* (Open and fuzzy systems). Jyväskylä: Weilin & Göös.

Malaska, Pentti & Virtanen, Ilkka (2009). Theory of futuribles and historibles. *Futura* 1, 64–85.

Raiffa, Howard (1968). *Decision Analysis*. Introductory Lectures on Choices under Uncertainty. New York, N.Y.: Random House.

Wexionio, Michaele (1642). *Discursus politicus de prudentia*. Turku: Royal Academy of Turku.

Wolf cave (2013). Internet pages of the research group (pages read 10.12.2013): http://www.susiluola.fi/eng/wolfcave.php.

A COMPLETION PROBLEM FOR AN UNBOUNDED NONNEGATIVE BLOCK OPERATOR

Seppo Hassi
University of Vaasa

1 Introduction

In this note an unbounded version of a completion problem for a nonnegative block operator is studied. To describe the problem we first recall the completion problem and its solution in the known case of a bounded nonnegative block operator acting on a Hilbert space $\mathcal{H}$.

Let $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ be an orthogonal decomposition of the Hilbert space $\mathcal{H}$ and let A^0 be an incomplete block operator of the form

$$A^0 = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & * \end{pmatrix}, \tag{1}$$

with all entries being bounded operators $A_{ij} \in [\mathcal{H}_j, \mathcal{H}_i]$, $i, j = 1, 2$. If the set

$$\{\, A_{22} \in [\mathcal{H}_2] : \; A = (A_{ij})_{i,j=1}^2 \geq 0 \,\} \tag{2}$$

is nonempty, then $A_{11} \geq 0$ and $A_{21} = A_{12}^*$. The next result provides a complete solution to this completion problem.

Proposition 1. *Let A^0 be a bounded incomplete block operator in the Hilbert space $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ given by* (1). *Assume that $A_{11} \geq 0$ and $A_{21} = A_{12}^*$. Then:*

(i) *There exists a completion $A \in [\mathcal{H}]$ of A^0 with some operator A_{22} if and only if* $\operatorname{ran} A_{12} \subset \operatorname{ran} A_{11}^{1/2}$.

(ii) *In this case the operator $S = A_{11}^{(-1/2)} A_{12}$, where $A_{11}^{(-1/2)}$ denotes the generalized inverse of $A_{11}^{1/2}$, is well defined and $S \in [\mathcal{H}_2, \mathcal{H}_1]$. Moreover, S^*S is the smallest operator in the set* (2).

This result goes back to Shmul'yan (1959) and in the form stated above it can be found from Hassi, Malamud & de Snoo (2004), where it was used as a starting point in establishing a simple description of all contractive selfadjoint extensions of contractive operators as well as for deriving (via Cayley transforms) a complete description of all nonnegative selfadjoint extensions of a nonnegative (not necessary densely defined) operator and, more generally, of an arbitrary nonnegative linear relation.

Initially Shmul'yan used the above result, in particular the minimal solution appearing in Proposition 1, in order to extend the notion of Hellinger integral from the scalar case to the operator-valued case. The Hellinger integral can be defined for completely additive functions $\rho(\cdot)$ and $\sigma(\cdot)$ on a ring $R = \{M\}$ of sets M, when $\rho(M)$ has nonnegative real values and $\sigma(M)$ has complex values and, in addition, $\sigma(\cdot)$ is absolutely continuous with respect to $\rho(\cdot)$. Then the fraction $|\sigma(M)|^2/\rho(M)$ is sub-additive from above and for every $M_0 \in R$ there exists a finite or infinite integral in the sense of Kolmogorov (1930):

$$\int_{M_0} \frac{|\sigma(dM)|^2}{\rho(dM)}, \tag{3}$$

being a certain directed limit of sums

$$\sum_{k=1}^{N} \frac{|\sigma(M_k)|^2}{\rho(M_k)},$$

of finite partitions of M_0 with pairwise disjoint sets $M_k \in R$, $k = 1, \ldots, N$.

The purpose in this note is to study the corresponding completion problem for an unbounded nonnegative block operator A of the form (1), where $A_{11} \geq 0$ is now allowed to be some unbounded densely defined operator, while A_{12} and A_{21} are still bounded operators. It is emphasized that here A_{11} is unbounded, and that it is not assumed to be selfadjoint. This makes the problem harder, as in this case one cannot apply spectral theory or resolvent limits in handing the entry A_{11}, typically appearing via the monotone convergence of increasing sequences $(A_{11} + \varepsilon)^{-1}$ and $A_{21}(A_{11} + \varepsilon)^{-1}A_{12}$, $\varepsilon \downarrow 0$, of bounded selfadjoint operators; cf. Hassi et al. (2004). Therefore, we approach the problem in another way, generalizing and simplifying some ideas going back to Shmul'yan (1959). By combining this with some range description results from Hassi (2004) yields a particularly simple proof for the main result and for a couple of related results.

The main results in this note include a general criterion for the solvability of the stated completion problem, a description of all solutions to this problem when the criterion is satisfied, and extensions of a couple of results from Shmul'yan (1959) that appear also in the construction of the Hellinger operator integral.

2 Linear relations, adjoints and extensions of nonnegative operators

2.1 Some definitions and basic facts

Let T be a linear relation (multi-valued mapping) from the Hilbert space $\mathcal{H}$ to the Hilbert space $\mathcal{K}$. This means that (the graph of) T is a linear subspace of $\mathcal{H} \oplus \mathcal{K}$. A linear relation T is closed if it is a closed subspace of $\mathcal{H} \oplus \mathcal{K}$. Analogous with the case of linear operators $\operatorname{dom} T$, $\ker T$, $\operatorname{ran} T$, and $\operatorname{mul} T$ stand for the domain, kernel, range, and multi-valued part of T, respectively. A linear relation T is (the graph of) a single-valued linear mapping from $\mathcal{H}$ into $\mathcal{K}$ precisely when $\operatorname{mul} T = \{0\}$. The inverse of T is simply defined by $T^{-1} = \{\, \{g, f\} : \{f, g\} \in T \,\}$. The adjoint T^* of T is defined by

$$T^* = \{\, \{h, k\} : (k, f)_{\mathcal{H}} - (h, g)_{\mathcal{K}} = 0 \text{ for all } \{f, g\} \in T \,\}. \tag{4}$$

It is a closed linear relation from $\mathcal{K}$ into $\mathcal{H}$ and it satisfies the usual identities familiar from the case of linear operators. For instance, $\ker T^* = (\operatorname{ran} T)^\perp$, $\operatorname{mul} T^* = (\operatorname{dom} T)^\perp$, and $(T^*)^{-1} = (T^{-1})^* =: T^{-*}$. For a closed linear relation T define $T_\infty = \{\, \{0, g\} \in T\}$ and $T_s = T \ominus T_\infty$, where $\ominus$ refers to the orthogonality in $\mathcal{H} \oplus \mathcal{K}$. This decomposes T orthogonally in $\mathcal{H} \oplus \mathcal{K}$ as follows

$$T = T_s \oplus T_\infty. \tag{5}$$

Here T_s is a closed linear operator with $\operatorname{dom} T_s = \operatorname{dom} T$ and $\operatorname{ran} T_s \subset (\operatorname{mul} T)^\perp = \overline{\operatorname{dom}}\, T^*$, while T_∞ is a closed linear relation with $\operatorname{dom} T_\infty = \{0\}$ and $\operatorname{ran} T_\infty = \operatorname{mul} T$. If in particular $\mathcal{H} = \mathcal{K}$ and T is for instance a closed symmetric relation in $\mathcal{H}$, i.e. $T \subset T^*$, then $\overline{\operatorname{dom}}\, T \subset \overline{\operatorname{dom}}\, T^* = (\operatorname{mul} T)^\perp$. The last inclusion guarantees that the decomposition of $T = T_s \oplus T_\infty$ in (5) becomes orthogonal also in $\mathcal{H} = \overline{\operatorname{dom}}\, T^* \oplus \operatorname{mul} T$. Moreover, T_s is a closed symmetric operator in $\overline{\operatorname{dom}}\, T^*$. By definition, a linear relation T in $\mathcal{H}$ is selfadjoint if $T = T^*$. In particular, a selfadjoint relation is always closed, linear, and symmetric.

The next result in its present general form was proved in Hassi (2004).

Lemma 1. *Let T be a linear relation from $\mathcal{H}$ into $\mathcal{K}$. Then $g' \in \operatorname{ran} T^*$ if and only if there exists $C_{g'} < \infty$, such that*

$$|(f, g')_{\mathcal{H}}| \le C_{g'} \|f'\|_{\mathcal{K}}, \quad \textit{for all } \{f, f'\} \in T. \tag{6}$$

In this case the smallest $C_{g'}$ satisfying (6) *is $C_{g'} = \|g\|_{\mathcal{K}}$ with $\{g, g'\} \in T^*$ and $g \in \overline{\operatorname{ran}}\, T$, i.e., $C_{g'} = \|(T^{-*})_s g'\|_{\mathcal{K}}$.*

Observe, that if $g' \in \operatorname{ran} T^*$ then $\{g, g'\} \in T^*$ with a unique element $g \in \overline{\operatorname{ran}}\, T$. Consequently, there exists a sequence $\{f_n, f_n'\} \in T$ such that $f_n' \to g$ as $n \to \infty$ and now using (4) one gets

$$|(f_n, g')_{\mathcal{H}}| = |(f_n', g)_{\mathcal{K}}| \to \|g\|_{\mathcal{K}}^2.$$

This shows that with the choice $C_{g'} = \|(T^{-*})_s g'\|_{\mathcal{K}}$ the estimate in (6) is sharp.

2.2 *Nonnegative selfadjoint extensions of nonnegative operators and relations*

Let A be a closed nonnegative relation in $\mathcal{H}$. Then A admits an orthogonal decomposition $A = A_s \oplus A_\infty$ as in (5), where the operator part A_s of A is a closed symmetric operator with $\operatorname{dom} A_s = \operatorname{dom} A$ and $\operatorname{ran} A_s \subset (\operatorname{mul} A)^\perp = \overline{\operatorname{dom}}\, A^*$, while A_∞ is a closed selfadjoint relation with $\operatorname{dom} A_\infty = \{0\}$ and $\operatorname{ran} A_\infty = \operatorname{mul} A$. The form domain generated by $A \geq 0$ is the completion of $\operatorname{dom} A$ with respect to the inner product $(f, g) + (f', g) = (f, g) + (A_s f, g)$, where $\{f, f'\}, \{g, g'\} \in A$ and where A_s is the operator part of A. The form domain $\operatorname{dom}[A]$ can be described as follows: $f \in \operatorname{dom}[A]$ if and only if there is a sequence $(f_n) \in \mathcal{H}$, such that

$$f_n \to f, \quad (A_s(f_n - f_m), f_n - f_m) \to 0 \quad (m, n \to \infty). \tag{7}$$

It follows from the First and the Second Representation Theorem (see e.g. Kato (1980)) when extended to nondensely defined closed forms that there is a unique nonnegative selfadjoint relation A_F in $\mathcal{H}$ such that

$$A[f, g] = ((A_F)_s^{1/2} f, (A_F)_s^{1/2} g), \quad f, g \in \operatorname{dom}[A] = \operatorname{dom} A_F^{1/2}.$$

Clearly, A_F is a selfadjoint extension of A in $\mathcal{H}$, the so-called Friedrichs extension of A, which in the densely defined case was introduced by Friedrichs (1934). In fact, A_F is the only selfadjoint extension of A whose domain is contained in $\operatorname{dom}[A]$ and the following alternative description for A_F holds:

$$A_F = \{\, \{f, f'\} \in A^* : \, f \in \operatorname{dom}[A] \,\}.$$

The Kreĭn-von Neumann extension A_K of A, introduced by von Neumann (1930) and Kreĭn (1947), can be defined in the present general setting of linear relations via

$$(A^{-1})_F = (A_K)^{-1}, \quad (A^{-1})_K = (A_F)^{-1}.$$

Hence, A_K^{-1} can be constructed also by means of the representation theorems when applied to the closed form $A^{-1}[f, g]$ with $\operatorname{dom}[A^{-1}] =: \operatorname{ran}[A]$ generated by the inverse A^{-1} of A. In particular, A_K is the only selfadjoint extension of A whose range is contained in $\operatorname{ran}[A]$ and the following description holds:

$$A_K = \{\,\{f,f'\} \in A^* :\ f' \in \operatorname{ran}[A]\,\}.$$

The main result concerning nonnegative selfadjoint extensions of a nonnegative relation A in $\mathcal{H}$ can now be stated. In the densely defined case this result goes back to Kreĭn (1947), for nondensely defined A it was proved by Ando & Nishio (1970), and for nonnegative relations by Coddington & de Snoo (1978).

Theorem 1. *Let A be a closed nonnegative relation in $\mathcal{H}$. Then A_F and A_K are nonnegative selfadjoint extensions of A. Moreover, $\widetilde{A}$ is a nonnegative selfadjoint extension of A in $\mathcal{H}$ if and only if*

$$(A_F+a)^{-1} \le (\widetilde{A}+a)^{-1} \le (A_K+a)^{-1}, \quad a>0.$$

We still need one further preliminary result for proving the main result in the next section. It concerns the descriptions of $\operatorname{ran} A_F^{1/2}$ of the Friedrichs extension A_F of a nonnegative relation A. Again in this general form the result was proved in Hassi (2004) by means of Lemma 1.

Proposition 2. *Let A be a nonnegative linear relation in $\mathcal{H}$ and let A_F be the Friedrichs extension of A. Then*

$$\operatorname{ran} A_F^{1/2} = \left\{\, k \in \mathcal{H} :\ |(f,k)|^2 \le C_k(g,f),\ \forall\,\{f,g\} \in A \textit{ with some } C_k < \infty \,\right\}.$$

Moreover, for every $k \in \operatorname{ran} A_F^{1/2}$ the smallest C_k is given by $C_k = \|h\|^2$ with $\{h,k\} \in A_F^{1/2}$ and $h \in \overline{\operatorname{ran}}\, A_F$, i.e., $C_k = \|((A_F)^{-1/2})_s k\|^2$.

Observe, that the optimal choice of the constant C_k in Proposition 2 can be also expressed with the Moore-Penrose generalized inverse $(A_F)^{(-1/2)}$ of $(A_F)^{1/2}$:

$$C_k = \|(A_F)^{(-1/2)}k\|^2. \tag{8}$$

With this choice of C_k the estimate in Proposition 2 is sharp: this can be seen directly by using the denseness of $\operatorname{dom} A$ in $\operatorname{dom}[A] = \operatorname{dom} A_F^{1/2}$ w.r.t. to the form topology; cf. (7).

3 Solution to the completion problem

In this section a solvability criterion to the stated unbounded completion problem is established and under this condition all bounded completions A_{22} to an incomplete block operator A^0 (see (9) below) are described.

Let $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ be an orthogonal decomposition of the Hilbert space $\mathcal{H}$ and let A^0 be an incomplete block operator of the form

$$A^0 = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & * \end{pmatrix}, \tag{9}$$

where $A_{11} \geq 0$ is assumed to be densely defined, in general non-selfadjoint, operator, while A_{21} and A_{12} are bounded. The completion problem we are interested in is to find a criterion for the existence of a bounded operator A_{22} in $\mathcal{H}_2$ such that the complete 2×2 block operator $A = (A_{ij})_{i,j=1}^2$ becomes nonnegative. If the set

$$\{\, A_{22} \in [\mathcal{H}_2] : \; A = (A_{ij})_{i,j=1}^2 \geq 0 \,\} \tag{10}$$

is nonempty, then necessarily $A_{11} \geq 0$ and $A_{12} = A_{21}^*$ is a bounded and closed operator from $\mathcal{H}_2$ into $\mathcal{H}_1$, while A_{21} is bounded but need not be not closed; one has $\operatorname{dom} A_{21} \supset \operatorname{dom} A_{11}$ with closure $\operatorname{clos} A_{21} = A_{12}^*$ being bounded and everywhere defined on $\mathcal{H}_1$.

The first result provides a general solvability criterion to the above completion problem; it involves the Friedrichs extension of the unbounded operator A_{11}.

Theorem 2. *Let A^0 be an incomplete block operator in $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$, given by* (9)*, and assume that $A_{11} \geq 0$ and $A_{12} = A_{21}^*$. Moreover, let $(A_{11})_F$ be the Friedrichs extension of A_{11} in $\mathcal{H}_1$. Then there exists a nonnegative completion $A = (A_{ij})_{i,j=1}^2$ of A^0 with some bounded operator $A_{22} \in [\mathcal{H}_2]$ if and only if*

$$\operatorname{ran} A_{12} \subset \operatorname{ran} (A_{11})_F^{1/2}. \tag{11}$$

Proof. Assume that there exists a bounded operator A_{22} such that $A = (A_{ij})_{i,j=1}^2$ is a nonnegative completion of A^0. Let $f = f_1 \oplus f_2 \in \mathcal{H}$ with $f_1 \in \operatorname{dom} A_{11}$ and $f_2 \in \mathcal{H}_2$. Moreover, let ξ and ζ be arbitrary complex numbers. Then

$$\begin{aligned} 0 &\leq (A(\xi f_1 + \zeta f_2), \xi f_1 + \zeta f_2) \\ &= \xi\bar{\xi}\,(A_{11}f_1, f_1) + \xi\bar{\zeta}\,(A_{21}f_1, f_2) + \zeta\bar{\xi}\,(A_{12}f_2, f_1) + \zeta\bar{\zeta}\,(A_{22}f_2, f_2) \\ &= \left(\begin{pmatrix} (A_{11}f_1, f_1) & (A_{12}f_2, f_1) \\ (A_{21}f_1, f_2) & (A_{22}f_2, f_2) \end{pmatrix} \begin{pmatrix} \xi \\ \zeta \end{pmatrix}, \begin{pmatrix} \xi \\ \zeta \end{pmatrix} \right), \end{aligned}$$

which shows that the 2×2 scalar matrix appearing in the last equality is nonnegative. Consequently, A is a nonnegative operator in $\mathcal{H}$ if and only if the inequalities

$$(A_{11}f_1, f_1) \geq 0, \;\; (A_{22}f_2, f_2) \geq 0, \;\; |(A_{12}f_2, f_1)|^2 \leq (A_{11}f_1, f_1)\,(A_{22}f_2, f_2) \tag{12}$$

are satisfied for all $f_1 \in \operatorname{dom} A_{11}$ and $f_2 \in \mathcal{H}_2$; here the last inequality uses the symmetry $A_{12} = A_{21}^*$.

It follows from the last inequality in (12) that

$$|(A_{12}f_2, f_1)|^2 \leq \|A_{22}^{1/2}f_2\|^2\,(A_{11}f_1, f_1) \tag{13}$$

and therefore with $C = \|A_{22}^{1/2}f_2\|^2$ one arrives at the inequality

$$|(f_1, A_{12}f_2)|^2 \leq C\,(A_{11}f_1, f_1), \quad \forall f_1 \in \operatorname{dom} A_{11}.$$

According to Proposition 2 this means that $A_{12}f_2 \in \operatorname{ran}(A_{11})_F^{1/2}$ for every $f_2 \in \mathcal{H}$, i.e., the condition (11) is satisfied.

To prove sufficiency, assume that (11) holds. Let $(A_{11})_F^{(-1/2)}$ be the Moore-Penrose generalized inverse of $(A_{11})_F^{1/2}$. Then the operator K given by

$$K := (A_{11})_F^{(-1/2)}A_{12} \tag{14}$$

is well-defined, closed, has domain $\operatorname{dom} K = \mathcal{H}_2$ and, therefore, it is also bounded by the closed graph theorem. We claim that the choice $A_{22} = K^*K$ gives a solution to the stated completion problem, i.e., the corresponding completion $A = (A_{ij})_{i,j=1}^2$ of A^0 is nonnegative. To see this we apply the choice $C_k = \|(A_{11})_F^{(-1/2)}k\|^2$ of the constant in Proposition 2 to the vector $k = A_{12}f_2 \in \operatorname{ran}(A_{11})_F^{1/2}$, $f_2 \in \mathcal{H}_2$; see (8). This means that

$$\begin{aligned} |(f_1, A_{12}f_2)|^2 &\leq \|(A_{11})_F^{(-1/2)}A_{12}f_2\|^2\,(A_{11}f_1, f_1) \\ &= \|Kf_2\|^2\,(A_{11}f_1, f_1) \\ &= (K^*Kf_2, f_2)\,(A_{11}f_1, f_1) \end{aligned} \tag{15}$$

is satisfied for every $f_1 \in \operatorname{dom} A_{11}$ and $f_2 \in \mathcal{H}_2$. This last inequality combined with $(K^*Kf_2, f_2) = \|Kf_2\|^2 \geq 0$ shows that with the bounded selfadjoint operator $A_{22} = K^*K$ all the three inequalities in (12) hold for all $f_1 \in \operatorname{dom} A_{11}$ and $f_2 \in \mathcal{H}_2$. As shown above the inequalities in (12) are equivalent to the completion $A = (A_{ij})_{i,j=1}^2$ with $A_{22} = K^*K$ to be nonnegative. This completes the proof. $\square$

We now proceed by studying the choice $A_{22} = K^*K$ appearing in the proof of Theorem 2 more closely. In fact, the next result shows that when the completion problem (9) is solvable then there is a minimal solution, which is obtained with the choice $A_{22} = K^*K$.

Proposition 3. *Let the assumptions be as in Theorem 2. If the completion problem* (9) *is solvable, i.e. the condition* (11) *is satisfied, then the operator* $A_{22} = K^*K$*, where* $K = (A_{11})_F^{(-1/2)}A_{12}$ *(cf.* (14)*), gives a minimal completion to* A^0 *and all bounded completions* A_{22} *in the set* (10) *are given by the following operator semi-interval:*

$$A_{22} = K^*K + B, \quad 0 \leq B = B^* \in [\mathcal{H}_2].$$

Proof. Assume that A_{22} is a bounded selfadjoint operator such that $A = (A_{ij})_{i,j=1}^2$ is a nonnegative completion of A^0. Then the proof of Theorem 2 shows that A satisfies the inequalities (12) and (13). In particular, $A_{12}f_2 \in \operatorname{ran}(A_{11})_F^{1/2}$ and by Proposition 2 the optimal choice of the constant C_k for $k = A_{12}f_2$ in the inequality (13) is given by (8), i.e.,

$$\|(A_{11})_F^{(-1/2)}A_{12}f_2\|^2 \leq \|A_{22}^{1/2}f_2\|^2$$

or, equivalently, $\|Kf_2\|^2 \leq \|A_{22}^{1/2}f_2\|^2$ holds for all $f_2 \in \mathcal{H}_2$. This last inequality means that $K^*K \leq A_{22}$. Thus, if A_{22} gives a solution then necessarily $A_{22} \geq K^*K$ and, consequently,

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \geq \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & K^*K \end{pmatrix} =: A_{min}.$$

On the other hand, if the completion problem (9) is solvable, then the choice $A_{22} = K^*K$ gives a bounded completion to A^0. Now clearly the completion

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & K^*K + B \end{pmatrix} = A_{min} + \begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix} \geq A_{min} \geq 0$$

is also a solution for every bounded $B \geq 0$. □

Observe that the minimal solution A_{min} can be expressed as a restriction of a nonnegative selfadjoint operator as follows:

$$A_{min} = \begin{pmatrix} (A_{11})_F^{1/2} \\ K^* \end{pmatrix} \begin{pmatrix} (A_{11})_F^{1/2} & K \end{pmatrix} \upharpoonright (\operatorname{dom} A_{11} \oplus \mathcal{H}_2). \tag{16}$$

To see this, note that by the definition of K (see (14)) one has $(A_{11})_F^{1/2}K = A_{12}$ and $\operatorname{clos} A_{21} = A_{12}^* \supset K^*(A_{11})_F^{1/2}$ with $\operatorname{dom} K^*(A_{11})_F^{1/2} = \operatorname{dom}(A_{11})_F^{1/2} \supset \operatorname{dom} A_{11}$ and clearly $(A_{11})_F^{1/2}(A_{11})_F^{1/2} = (A_{11})_F \supset A_{11}$. Finally, the row operator $H = ((A_{11})_F^{1/2} \ K)$ appearing in (16) is closed, which implies by a classical result due to von Neumann (1932) that H^*H is a selfadjoint operator. Therefore,

$$H^*H \supset A_{min} \tag{17}$$

so that H^*H is a nonnegative selfadjoint extension of the operator A_{min}.

We now give some consequences of the above results. The next statement follows immediately from Theorem 2 and Proposition 3.

Corollary 1. *Let $A = (A_{ij})_{i,j=1}^2$ be a block operator, where A_{11} is a densely defined operator in $\mathcal{H}_1$ and the operators A_{12}, A_{21}, and A_{22} are bounded. Then A is* *nonnegative precisely when:*

(i) $A_{11} \geq 0$, $A_{12} = A_{21}^*$;

(ii) $\operatorname{ran} A_{12} \subset \operatorname{ran} (A_{11})_F^{1/2}$;

(iii) $A_{22} \geq K^*K$, *where* $K = (A_{11})_F^{(-1/2)} A_{12}$.

In particular, in this case the operator $A_{22} - K^*K$ *(generalized Schur complement) is bounded, nonnegative, and selfadjoint.*

In the special case that A_{11} is not only symmetric but also selfadjoint, the results in Theorem 2 and Proposition 3 get a simpler form, since then automatically $(A_{11})_F^{1/2} = A_{11}^{1/2}$. Furthermore, in this case the minimal solution A_{min} coincides with the minimal nonnegative selfadjoint extension of the symmetric operator $A_1 = A^0 \upharpoonright \mathcal{H}_1$, i.e., with the Kreĭn-von Neumann extension $(A_1)_K$ of A_1, where $A_1 : \mathcal{H}_1 \to \mathcal{H}$ with $\operatorname{dom} A_1 = \operatorname{dom} A_{11} (\subset \operatorname{dom} A_{21})$ corresponds to the first column of A^0 in (9):

$$A_{min} = (A_1)_K,$$

cf. Theorem 1 stated in the preliminaries.

The next statement specializes Corollary 1 further in the case that $A_{11} = A_{11}^*$.

Corollary 2. *Let* $A = (A_{ij})_{i,j=1}^2$ *be a block operator, where* A_{11} *is a selfadjoint, in general unbounded, operator in* $\mathcal{H}_1$ *while the operators* A_{12}, A_{21}, *and* A_{22} *are bounded. Then* A *is nonnegative (automatically selfadjoint) precisely when:*

(i) $A_{11} \geq 0$, $A_{12} = A_{21}^*$;

(ii) $\operatorname{ran} A_{12} \subset \operatorname{ran} A_{11}^{1/2}$;

(iii) $A_{22} \geq K^*K$, *where* $K = A_{11}^{(-1/2)} A_{12}$.

Furthermore, if $\operatorname{ran} A_{12} \subset \operatorname{ran} A_{11}$ *then the condition* (iii) *takes the following more explicit form:* $A_{22} \geq A_{21} A_{11}^{(-1)} A_{12}$.

Proof. The first assertion follows from Corollary 1; notice that, by the boundedness of the entries A_{12}, A_{21}, and A_{22}, A is selfadjoint precisely when A_{11} is selfadjoint.

To prove the last statement assume that $\operatorname{ran} A_{12} \subset \operatorname{ran} A_{11}$. Then $A_{12} = A_{11}T$, where the operator $T = A_{11}^{(-1)} A_{12} : \mathcal{H}_2 \to \mathcal{H}_1$ is closed, everywhere defined, and hence bounded by the closed graph theorem. Moreover, $K = A_{11}^{1/2} T$ and

$$K^*K \supset (T^* A_{11}^{1/2})(A_{11}^{1/2} T) = T^* A_{12} \supset A_{12}^* A_{11}^{(-1)} A_{12} = A_{12}^* T, \tag{18}$$

and since $\operatorname{dom} A_{12}^*T = \mathcal{H}_2 = \operatorname{dom} K^*K$, all the inclusions in (18) hold as equalities. Thus, $K^*K = A_{12}^*A_{11}^{(-1)}A_{12} = A_{21}A_{11}^{(-1)}A_{12}$ and here the last equality holds due to $\operatorname{dom} A_{11} \subset \operatorname{dom} A_{21}$. $\square$

Corollaries 1 and 2 generalize the well-known criterion for a bounded 2×2 block operator to be nonnegative which in infinite dimensional Hilbert spaces was proved in Shmul'yan (1959, Theorem 1.7).

Observe also that the inclusion $\operatorname{ran} A_{12} \subset \operatorname{ran} A_{11}$ in Corollary 2 holds if $\operatorname{ran} A_{11}$ is closed, which is so in particular if $0 \in \rho(A_{11})$, i.e., if A_{11} is a boundedly invertible selfadjoint operator. In this case the difference $A_{22} - K^*K = A_{22} - A_{21}A_{11}^{-1}A_{12}$ is called a *Schur complement* of $A = (A_{ij})_{i,j=1}^2$. In particular, Corollary 2 extends *Sylvester's criterion*: the bounded operator $A = (A_{ij})_{i,j=1}^2$ with $0 \in \rho(A_{11})$ is nonnegative if and only if $A_{11} \geq 0$, $A_{21} = A_{12}^*$, and $A_{22} - A_{21}A_{11}^{-1}A_{12} \geq 0$.

Next one further result in Shmul'yan (1959, Theorem 1.3) is extended; it is central when extending the Hellinger integral to the operator-valued case. Using the terminology of Shmul'yan (1959) (and the equivalent characterization in Theorem 2) an incomplete block operator A^0 given by (9) is said to be *positive*, if the completion problem admits a solution; that is, if A^0 satisfies the condition (11) in Theorem 2.

Theorem 3. *Let A^0 be an incomplete block operator as in* (9) *which is positive (i.e. the condition* (11) *is satisfied). Then for every element $f_2 \in \mathcal{H}_2$ the following condition is fulfilled:*

$$\sup_{0 \neq f_1 \in \operatorname{dom} A_{11}} \frac{|\,(f_2, A_{21}f_1)|^2}{(A_{11}f_1, f_1)} = (K^*Kf_2, f_2), \tag{19}$$

where $K = (A_{11})_F^{(-1/2)}A_{12}$. In particular, if A^0 is positive then the supremum in the left side of (19) *is finite for every $f_2 \in \mathcal{H}_2$.*

Proof. Since A^0 is positive, the condition (11) is satisfied, and hence by Proposition 3 the operator $A_{22} = K^*K$ gives the minimal solution to the completion problem (9). By Theorem 2 and Proposition 2 for each $k := A_{12}f_2 \in \operatorname{ran}(A_{11})_F^{1/2}$, $f_2 \in \mathcal{H}_2$, and all $f_1 \in \operatorname{dom} A_{11}$, the following estimate, cf. (15),

$$|(f_1, A_{12}f_2)|^2 \leq \|(A_{11})_F^{(-1/2)}A_{12}f_2\|^2(A_{11}f_1, f_1) = (K^*Kf_2, f_2)\,(A_{11}f_1, f_1) \tag{20}$$

holds and, moreover, with this choice of C_k (see (8)) the above estimate becomes sharp; see also the discussion after Proposition 2. This means that

$$\sup_{f_1 \in \operatorname{dom} A_{11}} \frac{|\,(f_1, A_{12}f_2)|^2}{(A_{11}f_1, f_1)} = (K^*Kf_2, f_2),$$

which clearly coincides with (19): observe that if $(A_{11}f_1, f_1) = 0$ then (20) shows that also $A_{21}f_1 = 0$. This completes the proof. □

Conversely, if the supremum on the left side of (19) is finite for every $f_2 \in \mathcal{H}_2$, then it is clear from Proposition 2 and Theorem 2 that A^0 is positive. If, in particular, A_{11} is bounded, this observation gives Shmul'yan (1959, Theorem 1.4).

Following Shmul'yan (1959), let $R = \{M\}$ be a ring of subsets of some space $\mathcal{X}$, let $F_{12}(\cdot)$ be a completely additive set function on R whose values $F_{12}(M)$ are bounded operators from the Hilbert space $\mathcal{H}_2$ to the Hilbert space $\mathcal{H}_1$, let $F_{21}(\cdot) = F_{12}(\cdot)^*$, and let $F_{11}(\cdot)$ be a completely additive operator function on R whose values $F_{11}(M)$ are bounded nonnegative operators in $\mathcal{H}_1$ (or unbounded densely defined nonnegative operators as in (9)). With each $M \in R$ define the corresponding incomplete 2×2 block operator $\mathcal{F}(M)$ via (9). The function $\mathcal{F}(\cdot)$ is called a *completely additive positive system*, if $\mathcal{F}(M)$ is positive; this means that the criterion (11) in Theorem 2 holds for every $M \in R$. Then by using the minimal completion of $\mathcal{F}(M)$, Shmul'yan introduced the *Hellinger operator integral* of the system $\mathcal{F}(\cdot)$ on $M_0 \in R$ as an integral of the countably sub-additive (from below) bounded operator function $K(\cdot)^*K(\cdot)$ (in the sense of Kolmogorov (1930)) and denoted it formally by

$$\int_{M_0} F_{21}(dM) \cdot F_{11}^{-1}(dM) \cdot F_{12}(dM).$$

He showed, for instance, the connection to the scalar Hellinger integral in (3), studied the integrability properties of $F(\cdot)$ in the weak and strong sense, and gave various applications.

As a conclusion it is mentioned that it is possible to extend the main results in this note even further; we will consider such extensions elsewhere and in addition give some further results, consequences, and applications for them.

Acknowledgements: This paper is dedicated to my colleague Seppo Pynnönen on the occasion of his 60th birthday. The author is grateful for financial support from the Emil Aaltonen Foundation for staying on a sabbatical during 2013-2014.

References

Ando, T. & Nishio, K. (1970). Positive selfadjoint extensions of positive symmetric operators. *Tôhoku Mathematical Journal* 22, 65–75.

Coddington, E. & de Snoo, H. (1978). Positive selfadjoint extensions of positive symmetric subspaces. *Mathematische Zeitschrift* 159, 203–214.

Friedrichs, K. (1934). Spektraltheorie halbbeschränkter Operatoren und Anwendung auf die Spektralzerlegung von Differentialoperatoren. *Mathematische Annalen* 109, 465–487.

Hassi, S. (2004). On the Friedrichs and the Kreĭn-von Neumann extension of nonnegative relations. In Contributions to Management Science, Mathematics and Modelling. Essays in Honour of Professor Ilkka Virtanen. *Acta Wasaensia* 122, 37–54.

Hassi, S., Malamud, M. & de Snoo, H. (2004). On Krein's extension theory of nonnegative operators. *Mathematische Nachrichten* 274/275, 40–73.

Kato, T. (1980). *Perturbation theory for linear operators*. Berlin: Springer Verlag.

Kolmogorov, A. (1930). Untersuchungen über den integralbegriff. *Mathematische Annalen* 103, 654–696.

Kreĭn, M. (1947). The theory of selfadjoint extensions of semibounded Hermitian operators and its applications, I. *Matematicheski Sbornik* 20, 431–495.

Shmul'yan, Y. (1959). A Hellinger operator integral. *Matematicheski Sbornik* 49, 381–430.

von Neumann, J. (1930). Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren. *Mathematische Annalen* 102, 49–131.

von Neumann, J. (1932). Über adjungierte Funktionaloperatoren. *Annals of Mathematics (2)* 33, 294–310.

PROPERTIES OF A SPECIAL SYMMETRIC MATRIX ASSOCIATED WITH POSTAGE STAMPS

Ka Lok Chu
Dawson College
Simo Puntanen
University of Tampere
George P. H. Styan
McGill University

1 Introduction

Consider the $n \times n$ real symmetric matrix $\mathbf{A}_{pq}$, where all the elements in the first p and last p rows and columns are equal to $a \neq 0$ and all the elements in the $q \times q$ "centre" are equal to b, with $q = n - 2p$. We note that $\mathbf{A}_{pq}$ is centrosymmetric [Bernstein (2009, p. 181, Def. 3.1.2)] as well as symmetric, and thus is bisymmetric. For example, with $n = 5, p = 1$ and $q = 3$,

$$\mathbf{A}_{13} = \begin{pmatrix} a & a & a & a & a \\ a & b & b & b & a \\ a & b & b & b & a \\ a & b & b & b & a \\ a & a & a & a & a \end{pmatrix}, \tag{1}$$

which we encountered as the pattern of a sheetlet of postage stamps issued by Canada in 1992 to celebrate the 25th anniversary of confederation; see Figure 1.

In this paper we consider the following: with $p \geq 1$ and $q \geq 1$, find

(a) the rank of $\mathbf{A}_{pq}$

(b) the eigenvalues of $\mathbf{A}_{pq}$

(c) a simple necessary and sufficient condition for $\mathbf{A}_{pq}$ to be nonnegative definite

(d) the Moore–Penrose inverse $\mathbf{A}_{pq}^{+}$ of $\mathbf{A}_{pq}$, and

(e) show that the orthogonal projector $\mathbf{A}_{pq}\mathbf{A}_{pq}^{+}$ does not depend on a or b.

This is motivated by Problem 1/SP09 in *Statistical Papers* by Neudecker et al. (2009), who considered similar questions concerning the symmetric $n \times n$ matrix

$$\mathbf{C}_n = \begin{pmatrix} c_1 & c_1 & c_1 & \dots & c_1 \\ c_1 & c_2 & c_2 & \dots & c_2 \\ c_1 & c_2 & c_3 & \dots & c_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_1 & c_2 & c_3 & \dots & c_n \end{pmatrix}, \tag{2}$$

where c_i are real numbers, $i = 1, \dots, n$. Their problem was to find (i) a necessary and sufficient condition for $\mathbf{C}_n$ to be positive definite; (ii) the inverse of $\mathbf{C}_n$ when $\mathbf{C}_n$ is nonsingular; (iii) the determinant of $\mathbf{C}_n$. As Neudecker et al. (2009) note, the matrix $\mathbf{C}_n$ plays an important role in statistics and other areas. A nice application involving the inverse and the determinant of $\mathbf{C}_n$ when $n = 2$ and $n = 3$ appears in the context of conditional Brownian motion by Moyé (2006, Appendix B). For a solution of Problem 1/SP09, see Chu et al. (2011); see also Puntanen et al. (2011, pp. 252–253).

2 Solution

Let us write $\mathbf{A} = \mathbf{A}_{pq}$ and $\mathbf{B} = \frac{1}{a}\mathbf{A}$ $(a \neq 0)$. Then

$$\mathbf{B} = \mathbf{JFK}(\mathbf{JF})', \tag{3}$$

where

$$\mathbf{J} = \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \\ \mathbf{I}_p & \mathbf{0} \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} \mathbf{e}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{e}_q \end{pmatrix}, \quad \mathbf{K} = \begin{pmatrix} 1 & 1 \\ 1 & c \end{pmatrix} \ \text{with } c = \frac{b}{a}, \tag{4}$$

where $\mathbf{e}_p$ is the $p \times 1$ vector with each element equal to 1. It follows at once, since both $\mathbf{J}$ and $\mathbf{F}$ have full column rank, that

$$\operatorname{rank}(\mathbf{A}) = 2 \ \Leftrightarrow \ c \neq 1 \ \Leftrightarrow \ a \neq b; \ \operatorname{rank}(\mathbf{A}) = 1 \ \Leftrightarrow \ c = 1 \ \Leftrightarrow \ a = b. \tag{5}$$

Using (3) we see that the nonzero eigenvalues of $\mathbf{B}$ are the nonzero eigenvalues of

$$\mathbf{F}'\mathbf{J}'\mathbf{J}\mathbf{F}\mathbf{K} = \begin{pmatrix} 2p & 2p \\ q & cq \end{pmatrix}, \tag{6}$$

which are

$$\frac{1}{2}\left(2p + cq \pm \sqrt{(2p+cq)^2 - 8pq(c-1)}\,\right), \tag{7}$$

and so it follows that $\mathbf{B}$ is nonnegative definite if and only if $c \geq 1$; hence $\mathbf{A}$ is nonnegative definite if and only if $b \geq a > 0$.

When $b \neq a$, i.e., $c \neq 1$, then the matrix $\mathbf{K}$ is nonsingular and the Moore–Penrose inverse is

$$\mathbf{A}^+ = \frac{1}{a}\mathbf{B}^+ = \frac{1}{a}\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{K}^{-1}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}', \tag{8}$$

where

$$\mathbf{H} = \mathbf{J}\mathbf{F} = \begin{pmatrix} \mathbf{e}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{e}_q \\ \mathbf{e}_p & \mathbf{0} \end{pmatrix}. \tag{9}$$

Let $\mathbf{E}_{pq} = \mathbf{e}_p\mathbf{e}_q{}'$ denote the $p \times q$ matrix with every element equal to 1. Then

$$\mathbf{A}^+ = \frac{1}{4a^2(c-1)p^2q^2}\begin{pmatrix} cq^2\mathbf{E}_{pp} & -2pq\mathbf{E}_{pq} & cq^2\mathbf{E}_{pp} \\ -2pq\mathbf{E}_{qp} & 4p^2\mathbf{E}_{qq} & -2pq\mathbf{E}_{pq} \\ cq^2\mathbf{E}_{pp} & -2pq\mathbf{E}_{pq} & cq^2\mathbf{E}_{pp} \end{pmatrix}, \tag{10}$$

and the projector is

$$\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+\mathbf{A} = \mathbf{B}\mathbf{B}^+ = \mathbf{B}^+\mathbf{B} = \begin{pmatrix} \frac{1}{2p}\mathbf{E}_{pp} & \mathbf{0} & \frac{1}{2p}\mathbf{E}_{pp} \\ \mathbf{0} & \frac{1}{q}\mathbf{E}_{qq} & \mathbf{0} \\ \frac{1}{2p}\mathbf{E}_{pp} & \mathbf{0} & \frac{1}{2p}\mathbf{E}_{pp} \end{pmatrix}, \tag{11}$$

which does not depend on a or b.

In the special case when $b = a$ the matrix $\mathbf{A} = a\mathbf{E}_{hh}$, with $h = 2p + q$; the Moore–Penrose inverse and projector are

$$\mathbf{A}^+ = \frac{1}{ah^2}\mathbf{E}_{hh}\,, \quad \mathbf{A}\mathbf{A}^+ = \frac{1}{h}\mathbf{E}_{hh}\,; \quad h = 2p + q. \tag{12}$$

3 Illustrations via sheetlets of postage stamps

Our motivation for this problem comes from the sheetlets of postage stamps depicted in Figures 1–5. *Scott* catalogue numbers are as given in the annaual *Scott Standard Postage Stamp Catalogue* [Snee & Kloetzel (2014)]. For an extensive discussion of sheetlets of stamps with a Latin square pattern see Chu et al. (1995) and Loly & Styan (2010a,b,c).

Figure 1. Canada 1992, *Scott* 1446–1447.

The sheetlet of 25 stamps depicted in Figure 1, which may be represented by the matrix $\mathbf{A}_{13}$, was issued by Canada in 1992 to celebrate the 25th anniversary of confederation. The stamp a in the top row shows the insignia for the Order of Canada, Canada's highest civilian order and the centerpiece of the Canadian system of honours created in 1967. Membership in the order is accorded to those who exemplify the order's Latin motto, taken from *Hebrews* 11:16, *desiderantes meliorem patriam,* meaning "They desire a better country". The stamp b in the centre shows a portrait of Daniel Roland Michener (1900–1991), who was the first Governor-General of Canada serving from 1967–1974. He was appointed as such by Elizabeth II, Queen of Canada, on the recommendation of then Prime Minister of Canada Lester Bowles Pearson (1897–1972), to replace Major-General Georges-Philéas Vanier (1888–1967) as viceroy.

Figure 2. Sierra Leone 1990, *Scott* 1213.

We have found no other 5×5 sheetlet with 2 stamps arranged in the pattern defined by $\mathbf{A}_{13}$ but we have found some 3×3 sheetlets with the pattern $\mathbf{A}_{11}$. In particular, the sheetlet shown in Figure 2, where the stamp a in the top row depicts the team from Costa Rica in the World Cup Soccer Championships held in Italy in 1990; the "stamp" b in the centre is a label for "Italia '90: World Cup, Italy 1990". Twenty-four countries participated in this World Cup and Sierra Leone issued a separate stamp a depicting the team from each country, catalogued by *Scott* 1209–1232 in this order: Colombia[1], UAE (United Arab Emirates), South Korea, Cameroons, Costa Rica, Romania, Yugoslavia, Egypt, Netherlands, Uruguay, USSR, Czechoslovakia, Scotland, Belgium, Austria, Sweden, West Germany, England, USA, Eire, Spain, Brazil, Italy, Argentina. The 1990 World Cup was won by West Germany, who beat Argentina 1-0 in the final to win the World Cup for the third time.

Figure 3. Serbia 2004, *Scott* 243.

[1]Colombia is spelled "Columbia" in the selvage, upper left (Figure 2).

Figure 4. Croatia 1991, *Scott* RA22 & 1992, *Scott* 104.

Two other examples of 3×3 sheetlets of stamps with the pattern defined by the matrix $\mathbf{A}_{11}$ are shown in Figures 3 and 5. The sheetlet shown in Figure 3 depicts the theoretical physicist Albert Einstein (1879–1955) and the sheetlet in Figure 5 from Bequia (part of the country of Saint Vincent and the Grenadines) depicts singer and actor Elvis Aaron Presley (1935–1977). The only sheetlets of stamps with the pattern defined by the matrix $\mathbf{A}_{pq}$ that we have found are with $p = q = 1$ (Figures 2, 3, 5) and with $p = 1,\ q = 3$ (Figure 1).

The unusual 6×5 sheetlet shown in Figure 4 features a 3×3 set in the "center" that may be defined by the matrix $\mathbf{A}_{11}$, with the stamp a depicting the Croatian coat of arms and the single "stamp" b in the very center being a label with the Croatian coat of arms in black and white. The bottom-right corner is a set of 4 labels representing an enlarged version of the single stamp a, which appears a total of 25 times in the sheetlet. These stamps were first issued on 1 July 1991 and reissued on 15 January 1992; Croatia declared its independence in 1991 and this was internationally recognized by the European Union and the United Nations on 15 January 1992.

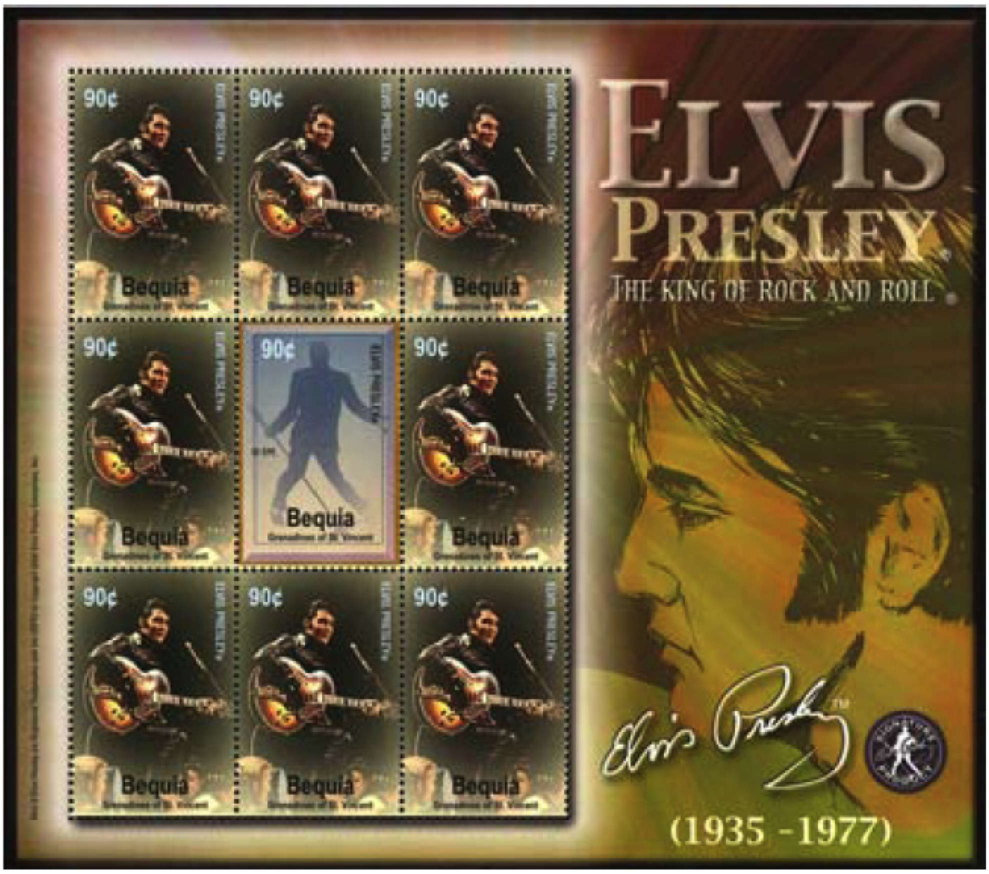


Figure 5. Bequia 2003, *Scott* B15.

4 Acknowledgements

Many thanks go to Peter D. Loly, Evelyn Matheson Styan, and Michel van de Velden for their help. This research has been supported in part by the Natural Sciences and Engineering Research Council of Canada.

References

Bernstein, D. S. (2009). *Matrix Mathematics: Theory, Facts, and Formulas*, Second edition. Princeton University Press.

Chu, K. L., Puntanen, S. & Styan, G. P. H. (2010). Some comments on philatelic Latin squares from Pakistan. *Pakistan Journal of Statistics* 25, 427–471.

Chu, K. L., Puntanen, S. & Styan, G. P. H. (2011). Solution to Problem 1/SP09 "Inverse and determinant of a special symmetric matrix". (Problem proposed by H. Neudecker, G. Trenkler & S. Liu.) *Statistical Papers* 52, 258–260.

Loly, P. D. & Styan, G. P. H. (2010a). Comments on 4×4 philatelic Latin squares. *Chance* 23: 1, 57–62.

Loly, P. D. & Styan, G. P. H. (2010b). Comments on 5×5 philatelic Latin squares. *Chance* 23: 2, 58–62.

Loly, P. D. & Styan, G. P. H. (2010c). Philatelic Latin squares. *Proceedings of the Canadian Society for History and Philosophy of Mathematics* 23, 273–297.

Moyé, L. A. (2006). *Statistical Monitoring of Clinical Trials: Fundamentals for Investigators*. New York: Springer.

Neudecker, H., Trenkler, G. & Liu, S. (2009). Problem 1/SP09: Inverse and determinant of a special symmetric matrix. *Statistical Papers* 52, 258–260.

Puntanen, S., Styan, G. P. H. & Isotalo, J. (2011). *Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty*. Heidelberg: Springer.

Snee, C. & Kloetzel, J. E., eds. (2014). *Scott Standard Postage Stamp Catalogue*. Scott Publishing, Sidney, OH.

II

STATISTICS

CHANCES OF GETTING PREGNANT: AN ESTIMATE USING THE DURATION OF NON-CONCEPTION WITH AN APPLICATION TO REAL DATA

Erkki P. Liski
University of Tampere

1 Introduction

In this study we estimate the chances of getting pregnant from the number of menstrual cycles over which a couple has been trying to conceive. The basic idea of the approach is that a couple's fertility is inherently uncertain and the uncertainty is modeled as a probability distribution for the chance of conceiving in each menstrual cycle. It is common to define infertility as the inability to conceive after a year of trying. In reproductive medicine such rules are considered in deciding what period of non-conception should precede investigation and fertility treatment. For a general discussion on conception rate and fertility see Sozou & Hartshorne (2012).

This study is based on extensive data that were gathered by a questionnaire for the needs of maternity care during the year 1979 in Tampere, Finland. The data contained all mothers who had given birth to a child during the year 1979 in Tampere, 2096 births in all (Liski (1983)). The data were collected by nurses working in different maternity care centers of Tampere. The mothers were interviewed in connection of their conventional visits to the maternity care center. In this context, the nurse asked for the number of months over which a couple had been trying to conceive. All mothers could not, or did not want to answer the question. However, 1744 mothers answered (352 observation were lost).

A measure of reproduction in this study is the number of menstrual cycles required to achieve pregnancy which is assumed to follow a geometric distribution with parameter p. The value of the parameter p is a couple's probability of achieving a pregnancy in a new cycle. The probability p can be considered as the intrinsic conception rate of a couple. The conception process is like tossing a coin repeatedly until a success is achieved. The couples are not identical but they vary in characteristics which have a bearing on the probability of achieving a pregnancy in a new cycle. This variation among couples means that the probability p varies among the couples. We say, for example, that couple a is more fertile than couple b if the success probability p_a of couple a is greater than the success probability p_b of couple b. Then, over a fixed number of cycles, couple a is more likely to conceive

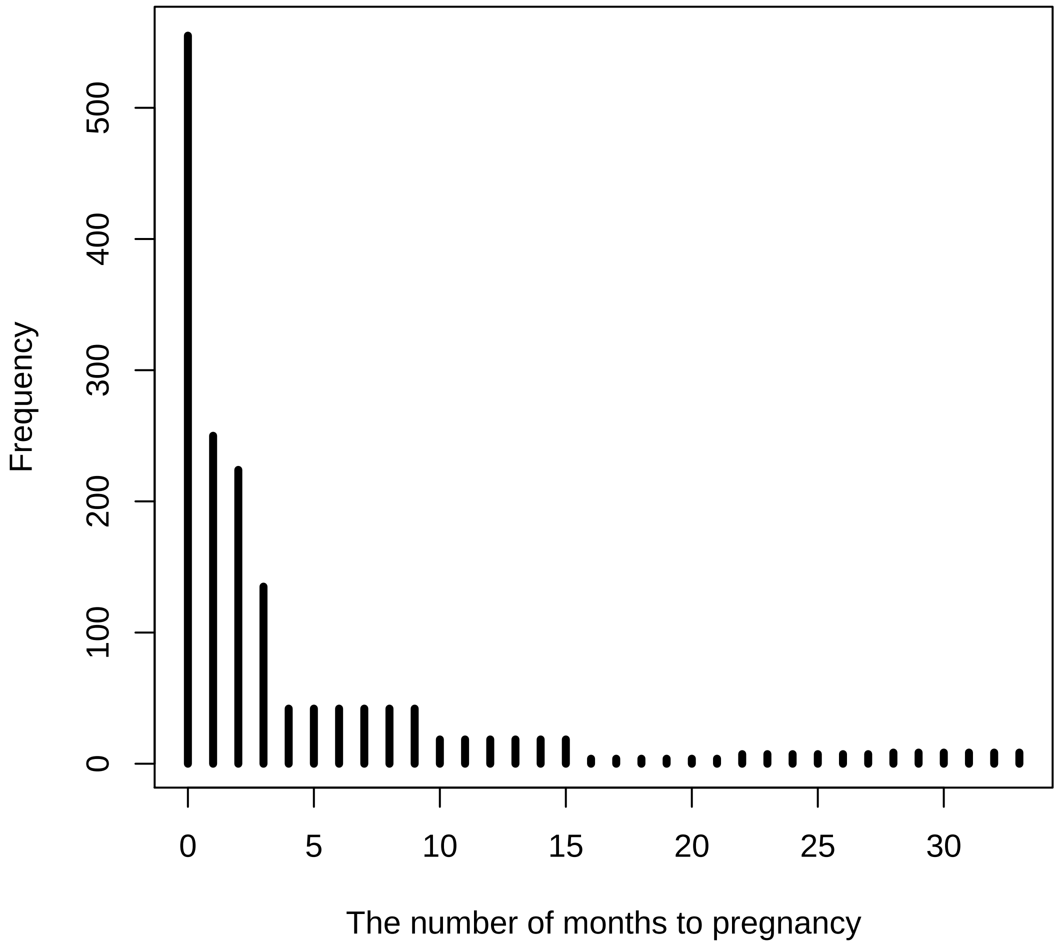


Figure 1. The distribution of the number of menstrual cycles to pregnancy i.e. the number of trials before the success. Data on 1744 mothers who have given birth to a child in 1979 in Tampere.

than couple b, but it remains possible that couple b will conceive and couple a will not. Thus overall variation in outcomes results from a combination of two kinds of randomness: the Bernoulli process where the waiting time follows the geometric distribution with parameter p, and variation among couples in their underlying fertility, i.e. variation of p.

2 Beta-Geometric Model for Cycles to Pregnancy

Let X denote the number of "failed" trials required until "success". In this study the random variable X is the number of menstrual cycles required for conception. It is assumed that $X|p$ follows the geometric distribution with parameter p. Then, the probability mass function of X is

$$P(X = x|p) = (1-p)^{x-1}p,\ x = 0, 1, 2, \ldots. \tag{1}$$

Suppose that data are available on m couples (In our data $m = 1744$) and the intrinsic conception rate takes the values $p_1, p_2, \ldots, p_m$ with relative frequences (weights)

$$w_1, w_2, \ldots, w_m,\ \text{where } w_j > 0 \text{ and } \sum_{j=1}^{m} w_j = 1.$$

Then the probability mass function of m-component finite mixture from m different geometric distributions (1) with success probabilities $p_1, p_2, \ldots, p_m$ is

$$P_w(X = x) = \sum_{j=1}^{m} w_j(1-p_j)^{x-1}p_j,\ x = 0, 1, 2, \ldots.$$

Since p may take any value in the parameter space $0 < p < 1$ one can assume underlying continuous distributions with probability density $f(p)$ for p in $0 < p < 1$. Then the continuous mixture of geometric distributions, i.e. the marginal distribution of the waiting time X to success, is

$$P_f(X = x) = \int_0^1 (1-p)^{x-1}pf(p)\,dp,\ x = 0, 1, 2, \ldots. \tag{2}$$

A convenient distribution for p is the beta distribution, because it is the natural conjugate family for the geometric distribution DeGroot (1970). The beta family is sufficiently flexible and produces a convenient mixed distribution, namely, the beta-geometric distribution. Thus we assume that p follows the beta distribution with probability density function given by

$$f(p; \alpha, \beta) = \frac{1}{B(\alpha, \beta)}p^{\alpha-1}(1-p)^{\beta-1},\ 0 < p < 1, \tag{3}$$

where

$$B(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

is the beta function and $\Gamma(\alpha)$ is the gamma function $\Gamma(\alpha) = \displaystyle\int_0^\infty t^{\alpha-1}e^{-t}dt$.

The mean of the beta distribution (3) is

$$\mathrm{E}(p) = \frac{\alpha}{\alpha+\beta}. \tag{4}$$

The marginal distribution of X is

$$\begin{aligned} P(X=x) &= \int_0^1 (1-p)^{x-1} p f(p;\alpha,\beta)\,dp \\ &= \frac{1}{B(\alpha,\beta)} \int_0^1 p^{\alpha}(1-p)^{x+\beta-1} \\ &= \frac{B(\alpha+1, x+\beta)}{B(\alpha,\beta)}. \end{aligned} \tag{5}$$

The mean and variance of the beta-geometric random variable X are (Johnson & Kotz (1969))

$$\mathrm{E}(X) = \frac{\beta}{(\alpha-1)}$$

and

$$\mathrm{Var}(X) = \frac{\alpha\beta(\alpha+\beta-1)}{(\alpha-1)^2(\alpha-1)}.$$

Liski (1983) applied the beta-geometric model in the analyses of data on the number of menstrual cycles to pregnancy for a total of 1744 women. Weinberg & Gladen (1986) analyzed fecundability data for a total of 586 women.

3 Estimation

It is clear from Table 1 that the waiting time to pregnancy heavily depends on the mothers age. For example, the mean waiting time in the age group $35-46$ is over three times longer than in the age group $15-20$. Here the waiting time to pregnancy is the number of menstrual cycles needed to archieve pregnancy. We have estimated the parameters α and β of the waiting time distribution (5) by using the method of moments (Table 2). The parameters are estimated separately for each age group. Note that in Table 1 the waiting time to pregnancy is estimated from the empirical distribution whereas in Table 2 the waiting time is estimated using the theoretical model (5). The empirical distribution (cf. Figure 1) is available only in the form of grouped frequences. This is due to the fact that mothers do not remember exactly the waiting time to pregnancy. This is especially the case when the waiting time is long, for example over six months.

Table 1. Mean, standard deviation, median, 25% and 75% percentiles of the number of menstrual cycles to pregnancy for a total of 1744 women by age. Data from Liski (1983).

				Percentiles		
Age	Number of women	Mean	Std.	25%	50%	25%
$15-20$	97	2.16	3.36	1.50	2.00	4.29
$21-23$	272	3.07	5.19	1.61	2.67	5.30
$24-27$	593	4.55	8.66	1.85	3.21	5.95
$28-30$	368	5.83	11.36	1.98	3.46	7.09
$31-34$	306	6.54	14.18	1.87	3.48	7.11
$35-46$	108	7.00	12.27	1.76	3.35	7.15
All	1744	4.86	10.19	1.79	3.19	6.64

Using these estimates we obtain estimates for the waiting time distributions and for the distribution of per-cycle probability of success in each age group. Thus the dependence of fertility on mothers age can be demonstrated.

Table 2. Estimates of the parameters α and β by age for the distribution (5) from the data summarized in Table 1. Mean of the the conception rate distribution (3) and mean of the waiting time (wt) to pregnancy (5) by age.

Age	$15-20$	$21-23$	$24-27$	$28-30$	$31-34$	$35-46$	All
α	5.05	3.74	3.01	2.89	2.65	3.18	2.80
β	8.72	8.42	9.15	11.04	10.79	14.23	8.90
Mean of p	0.37	0.31	0.25	0.21	0.20	0.18	0.24
Mean of wt	2.15	3.07	4.55	5.84	6.54	6.53	4.94

The data were obtained retrospectively, starting from a pregnancy in each case. Liski (1983) analyzed this fecundability data for a total of 1744 women. Thus the data do not contain couples whose per-cycle probability of success is zero. In this data the estimated median of the per-cycle probability p is 0.22, i.e. 78% of the couples have the per-cycle probability p less than 0.5.

4 Discussion

As couples get older, their fertility declines. In particular, increasing female age is associated with a decreasing conception rate. Female age, therefore, may be expected to have a bearing on how many cycles of attempted natural conception need to elapse before medical investigation and treatment is appropriate. The data from Liski (1983) provide information on mother's age and on the number of cycles

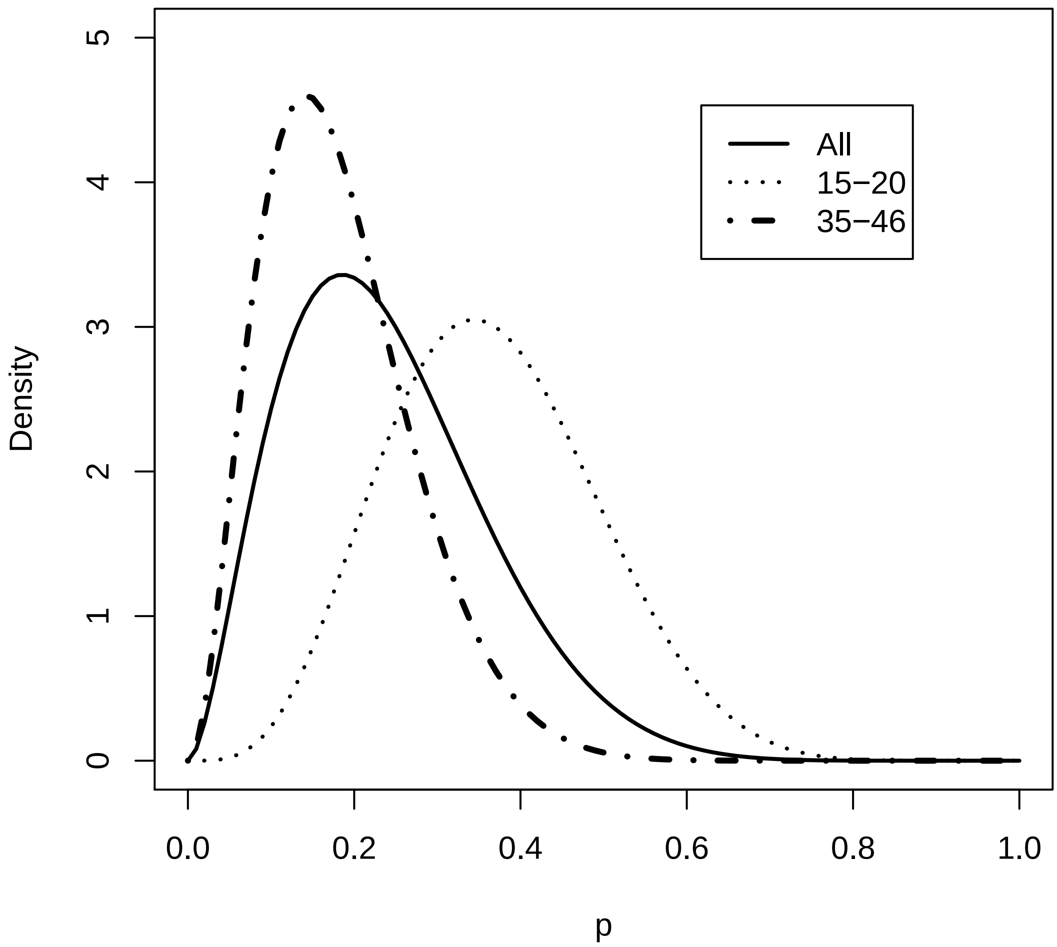


Figure 2. The estimated distribution of the per-cycle probability of success for the age groups $15-20$, $35-46$ and all 1744 couples of age $15-46$.

needed by couples to archieve pregnancy. The intrinsic conception rate of a couple, i.e. their probability of conception per cycle p is basically unknown. There is a considerable variation of conception rate between couples. In our beta-geometric modeling framework, this uncertainty is described by a probability distribution for the conception rate p. This distribution will depend on the population from which the couple is drawn. Especially, the conception rate depends on mothers age which can be clearly seen from Table 2. It should be noted that in our data all women had succeeded to become pregnant. Thus the fertility of the couples in this data is higher than in the whole population.

What can be deduced about a couple's fertility from the duration of their attempt to conceive? For clinical decision-making the question is this: How many cycles of non-conception is a sufficient indicator that a couple needs investigation and treatment. This problem requires a probabilistic analysis. For example, in the age group $15-20$ only 2% of the mothers have the waiting time to pregnancy at least one year whereas 16% of the mothers in the group $35-46$ have the waiting time to pregnancy at least one year. There are costs associated with investigating and treating couples who may have conceived without medical assistance. On the other hand, delaying the treatment of infertile couples may result in worse outcomes, or they may even lose the chance to have their own genetic child.

References

Sozou, P.D. & Hartshorne, G.M. (2012). Time to pregnancy: A computational method for using the duration of non-conception for predicting conception. *PLoS ONE* 7(10).

Liski, E.P. (1983). Application of a model for the waiting time before pregnancy. Department of Mathematical Sciences, University of Tampere, Finland. Report A 95.

Johnson, N.L. & Kotz, S. (1969). *Discrete Distributions*. New York: John Wiley & Sons.

DeGroot, M.H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

Weinberg, P. & Gladen, B. C. (1986). The Beta-geometric distribution applied to comparative fecundability studies. *Biometrics* 42, 547–560.

CLASSES OF STANDARD GAUSSIAN RANDOM VARIABLES AND THEIR GENERALIZATIONS

Wolf-Dieter Richter
University of Rostock

1 Introduction

Numerous results on the skewed normal distribution and its generalizations were derived during the last decade. Already the authors in Genton (2004) list various representations of random variables following such a distribution. In the case of univariate skewed distributions, these representations typically make use of bivariate random vectors. A unified geometric approach to different such representations of the one-dimensional skewed normal distribution and its generalizations is given in Günzel, Richter, Scheutzow, Schicker & Venz (2012). This approach is based on a representation of the Gaussian law which was originally derived in Richter (1985) and several subsequent papers for the purposes of large deviation theory. This geometric measure representation was extended in Richter (1991) to spherical distributions. Basic introductions into the area of geometric measure theory and related fields are given in Federer (1969), Nachbin (1976), Morgan (1984), Wijsman (1984), Barndorff-Nielsen, Blaesild & Eriksen (1989), Schindler (2003), Kallenberg (2005), Krantz & Parks (2008), and in Muirhead (1982), Eaton (1983), Farrel (1985), and Richter (2009). A certain uniquely defined measure on the Borel σ-field $\mathcal{B}$ on the Euclidean sphere, the so called uniform distribution on $\mathcal{B}$, plays a fundamental role for these considerations. Several authors have exploited properties of this distribution for different purposes. Seppo Pynnönen (2013) demonstrates how to use the uniform distribution on a Stiefel manifold for dealing with a fundamental problem in statistics. He derives the distributions of linear combinations of internally studentized ordinary least squares residuals of multivariate regression analysis.

The geometric representation of the Gaussian law in Richter (1985) exploits this uniform distribution with the help of the so called intersection percentage function(ipf). The idea behind the definition of this function stems from the very old method of measuring the content of an area or a body by comparing it with a well studied one. Analyzing the ipf, it was shown in Günzel et al. (2012) that the univariate skewed normal distribution and any of its spherical generalizations is closely connected with measuring intersections of two half planes with the help of a bivariate normal or spherical distribution, respectively. This result was extended in

Richter & Venz (2014) to the higher dimensional situation. There are other k-variate distributions which are connected with measuring other types of subsets of R^n with $n > k$. As just to mention a few of such multivariate cases, we recall that Student and Fisher distributions and their generalizations may be studied from a geometric measure theoretical point of view by measuring one- and two-sided cones having their apex in the origin, see Richter (1991, 1995, 2007 and 2009). The Student distribution is also connected with considerations on non-linearly transformed cone type sets, see Richter (1995) and Ittrich & Richter (2005), and noncentral χ^2- and Fisher-distributions are connected with balls and cones having their center or apex outside the origin, respectively, see Ittrich, Krause & Richter (2000) and Krause & Richter (2004). These and several other examples show that geometric measure representations apply in a great variety of situations. For some more two-dimensional results, we refer to Kalke, Richter & Thauer (2013) and Müller & Richter (2014).

The bivariate Gaussian measure geometric representation will be used in the present paper to unify and to extend the proofs of two seemingly different results in Shepp (1964) on normal functions of normal random variables, see also Cohen (1981), Baringhaus, Henze & Morgenstern (1988) and Bansal, Hamedani, Key, Volkmer, Zhang & Behboodian (1999) . It will turn out from Theorem 3 that both results are just special cases of a more general representation formula for the univariate standard Gaussian law. Actually, we construct a class of standard Gaussian distributed random variables including known cases as special cases. This class will be essentially enlarged in Theorem 4 to the class $\Phi(\Phi_{0_2,I_2})$ of univariate standard Gaussian random variables which are derived from bivariate standard Gaussian vectors. The spherical extensions of these results in Theorems 5 and 6 can be viewed as well as generalizations of a result derived in Arellano-Valle (2001).

Let (X, Y) denote a random vector taking its values in R^2. If (X, Y) follows the two-dimensional standard Gaussian law, we shall write

$$(X, Y) \sim \Phi_{0_2,I_2}. \tag{1}$$

The cumulative distribution function (cdf) of the standard Gaussian law on the real line will be denoted by Φ. The following theorem was in part repeatedly proved in Shepp (1964), Cohen (1981), Baringhaus et al. (1988) and Bansal et al. (1999) by exploiting stable distribution theory, proving McLaurin series expansions, using coordinate transformation or Laplace transformation, exploiting a representation of the densities of chi-distributed random variables and various other techniques.

Theorem 1 *[A] If the random vector* (X, Y) *satisfies assumption* (1) *then*

$$P(\frac{2XY}{\sqrt{X^2+Y^2}} < w) = \Phi(w),\ w \in R. \tag{2}$$

[B] If relation (2) *holds then assumption* (1) *is fulfilled.*

Methods from geometric measure theory will be used in Section 3 to reprove part *[A]* of this theorem and a slightly adapted version of this new proof will enable us to even generalize part *[A]* of the theorem for large classes of random variables. In words close to those in Silvermann (2000), this demonstrates the actually given value of reproving. For the corresponding results we refer to Theorems 3 and 4.

If (X, Y) follows a two-dimensional spherical distribution with density generator (dg) h, i.e. if the probability density function (pdf) $f_{(X,Y)}$ of (X, Y) is

$$f_{(X,Y)}(x, y) = h(||(x, y)||^2),\ (x, y) \in R^2$$

where $||.||$ denotes the Euclidean norm in R^2, we shall write

$$(X, Y) \sim \Phi_{h;0_2,I_2}. \tag{3}$$

The cdf of any marginal distribution of $\Phi_{h;0_2,I_2}$ is an univariate spherical distribution and is denoted throughout this note by Φ_h. According to Fang, Kotz & Ng (1990), its density φ_h satisfies the representation

$$\varphi_h(w) = 2\int\limits_0^\infty h(z^2 + w^2)dz,\ w \in (-\infty,\ \infty),$$

and the cdf itself will be called the spherical marginal h-generalization Φ_h of Φ. The next theorem is a generalization of the first one and was proved in Arellano-Valle (2001) using analytical methods which are based upon a stochastic representation for spherically distributed random vectors. Such representations go back to Kelker (1970), Johnson & Kotz (1970), Cambanis, Huang & Simons (1981), Anderson & Fang (1990), Fang et al. (1990), and Fang & Shang (1990). To be more concrete, properties of uniformly distributed vectors are combined in Arellano-Valle (2001) with certain relations from trigonometry.

Theorem 2 *[A] If* (X, Y) *satisfies assumption* (3) *then*

$$P(\frac{2XY}{\sqrt{X^2+Y^2}} < w) = \Phi_h(w), w \in R. \tag{4}$$

[B] If relation (4) *holds then assumption* (3) *is fulfilled.*

The rest of this note is organized as follows. We state generalizations of Theorem 1 dealing with Gaussian distributions in Section 2.1 and generalizations of Theorem 2 dealing with spherical distributions in Section 2.2. Theorem 6 will be the main result of this note. In Section 3 we provide geometric measure theoretical reproofs of known results from Section 1. The proofs of the results in Section 2 will be based upon the reproofs outlined in Section 3 and will be given in the final Section 4.

2 Main results

2.1 Classes of standard normally distributed functions of bivariate standard Gauss vectors

Let us consider the random variable $S(X,Y) = \frac{\lambda_1 X^2+\lambda_2 XY+\lambda_3 Y^2}{\sqrt{\mu_1 X^2+\mu_2 XY+\mu_3 Y^2}}$ which is a measurable function of the random vector (X,Y).

Theorem 3 *If (1) holds then*

$$P(S(X,Y) < w) = \Phi(w),\ w \in R \tag{5}$$

for all coefficients $\lambda_i, \mu_i,\ i \in \{1,2,3\}$, *satisfying*

$$\lambda_1 = 2a_{11}a_{21},\ \lambda_2 = 2(a_{11}a_{22} + a_{12}a_{21}),\ \lambda_3 = 2a_{12}a_{22} \tag{6}$$

and

$$\mu_1 = a_{11}^2 + a_{21}^2,\ \mu_2 = 2(a_{11}a_{12} + a_{21}a_{22}),\ \mu_3 = a_{12}^2 + a_{22}^2 \tag{7}$$

where $O = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ *is an orthogonal matrix.*

The following two examples describe those random variables which were considered under the normality assumption for (X,Y) already in Shepp (1964), Cohen (1981), Baringhaus et al. (1988) and Bansal et al. (1999).

Example 1 *If* $O = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ *then* $S(X,Y) = \frac{2XY}{\sqrt{X^2+Y^2}}$.

Example 2 *If* $O = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ *then* $S(X,Y) = \frac{X^2-Y^2}{\sqrt{X^2+Y^2}}$. *Notice that the orthogonal matrix O used here describes an anticlockwise rotation around the origin of* R^2 *through the angle of 45 degrees.*

Let us recall that it was shown already in Shepp (1964) by exploiting stable distribution theory and in Cohen (1981) by using McLaurin expansion that under (1) the random variable $\frac{2XY}{\sqrt{X^2+Y^2}}$ follows the standard Gaussian law. Some elementary proofs of this result were given in Baringhaus et al. (1988) and a characterization of the standard Gaussian law by this property was derived in Bansal et al. (1999). The proof of Theorem 3 will be based upon an invariance property of the bivariate Gaussian law and will be given in Section 4.

The aim of our following consideration is to significantly enlarge the class of univariate random variables being standard Gaussian distributed if assumption (1) is fulfilled. To this end, we denote the Euclidean circle of radius r and having its center in the origin of R^2 by

$$C(r) = \{(x,y) \in R^2 : x^2 + y^2 = r^2\},\ r > 0.$$

Lemma 1 *Let the Borel sets* A *and* B *from* R^2 *satisfy the equation*

$$B \cap C(r) = O(r)[A \cap C(r)] \text{ for almost all } r > 0$$

where each $O(r)$ *is an orthogonal matrix. Then*

$$\Phi_{0_2,I_2}(B) = \Phi_{0_2,I_2}(A).$$

The method of measuring subsets of R^2 resulting from this lemma will be called the bivariate standard Gaussian measure indivisiblen method. This method reflects in a generalized sense the ancient ideas of Cavalieri and Torricelli and was basically established in Richter (1985) and some subsequent papers within certain considerations on multivariate large deviation probabilities. Moreover, it was exploited in several papers for studying the Gaussian measure of sets from certain statistically well motivated classes of sets, e.g., in Richter (1995), Ittrich et al. (2000), Krause & Richter (2004) and Ittrich & Richter (2005) as just to mention a few of them. The next definition is in the spirit of such work. It makes use of the notion of the Euclidean circle ipf of a Borel set A from R^2 which is defined as

$$\mathcal{F}(A,r) = \frac{1}{2\pi r} l(A \cap C(r)),\ r > 0$$

where l means the Euclidean arc-length.

Definition 1 *A random variable $T : R^2 \to R$ belongs to the class IpfRep of random variables if the Euclidean circle ipfs of its sublevel sets*

$$A(w) = \{(x,y) \in R^2 : T(x,y) < w\},\ w \in R,$$

allow the joint representation

$$\mathcal{F}(A(w),r) = I_{[0,w]}(r)I_{[0,\infty)}(w) + I_{[|w|,\infty)}(r)[\frac{1}{2} + \frac{sign(w)}{2\pi}\arccos(1 - \frac{2w^2}{r^2})].$$

Example 3 *It follows immediately from the proof in Section 3.1 that, under the assumptions (6) and (7), $S(X,Y) \in$ IpfRep.*

Definition 2 *If a bivariate random vector (X,Y) satisfies the assumption (1) and, for a function $T : R^2 \to R$, the random variable $T(X,Y)$ follows the univariate standard Gaussian distribution then we say that $T(X,Y)$ belongs to the class $\Phi(\Phi_{0_2,I_2})$ of univariate standard Gaussian random variables derived from a bivariate standard Gaussian vector.*

Theorem 4 *If the random variable T satisfies the assumption $T \in$ IpfRep then $T(X,Y) \in \Phi(\Phi_{0_2,I_2})$.*

In other words, if there holds (1) and $T \in IpfRep$ then

$$P(T(X,Y) < w) = \Phi(w),\ w \in R.$$

Example 4 *It is well known that the random variables $T(X,Y) = X$ and $T(X,Y) = Y$ belong to the class $\Phi(\Phi_{0_2,I_2})$ and that their cdfs may be determined by measuring half planes. Figure 2 indicates that the Φ_{0_2,I_2}-value of a half plane might be composed by the Φ_{0_2,I_2}-values of two quadrants from R^2 arising in the proof of Theorem 3. This provides formally a reproof of the well known result on marginal distributions, using the method of Section 3.*

Remark 1 *An example of a certain non-linear transformation of a cone which has the same ipf as the cone itself was introduced in Richter (1995) and further developed in Ittrich & Richter (2005) for the purposes of non-linear regression analysis. In such cases, Lemma 1 applies. Notice that one can construct different types of statistics from the class IpfRep in an analogous way. For more examples of how this method works, we refer to Richter (1995).*

According to Lemma 1, the Euclidean circle ipfs of the random variable's T sub-level sets became the main tool of investigation in this note. Making use of it, we are now in a position to further generalize the results of Theorems 3 and 4. This will be done in the next section.

2.2 *Classes of functions of two-dimensional spherical random vectors following the univariate spherical marginal h-generalization Φ_h of Φ*

In the present section we state far-reaching generalizations of the results presented in the previous section. To be concrete, results will be given under the assumption (3) upon the random vector (X, Y) being much more general than those under assumption (1). Assumption (3) allows both heavy and light distribution tails of the bivariate vector distribution. Examples of different types of density generating functions can be found, e.g., in Fang et al. (1990) and in Kalke et al. (2013). First, we give a generalization of Theorem 3.

Theorem 5 *If (X, Y) satisfies (3) then $P(S(X, Y) < w) = \Phi_h(w), w \in R$ for all coefficients $\lambda_i, \mu_i, i \in \{1, 2, 3\}$ fulfilling conditions (6) and (7), respectively.*

At the same time as this theorem generalizes Theorem 3, it generalizes part A of Theorem 2. The following definition extends Definition 2.

Definition 3 *If a bivariate random vector (X, Y) satisfies assumption (3) and, for a function $T : R^2 \to R$, the random variable $T(X, Y)$ follows the univariate spherical marginal h-generalization Φ_h of Φ then we say that $T(X, Y)$ belongs to the class $\Phi_h(\Phi_{h;0_2,I_2})$ of univariate random variables following a spherical marginal h-generalization Φ_h of Φ derived from a bivariate spherical distribution.*

The final and main result of this note follows immediately the line of the previous results and their proofs in Sections 3 and 4.

Theorem 6 *If $T \in IpfRep$ then $T(X, Y) \in \Phi_h(\Phi_{h;0_2,I_2})$.*

That is, if (3) and $T \in IpfRep$ then $P(T < w) = \Phi_h(w), w \in R$.

3 Reproofs

3.1 Geometric measure theoretical proof of Theorem 1 [A]

In this section, we shall give a geometric measure theoretical reproof of Theorem 1*[A]*. Reproofs are of special interest in the area of didactics of mathematics. They help to make mathematical relationships more clear. For a discussion of the general value of reproofs, see Silvermann (2000). Sometimes, reproofs open new perspectives for proving either sharper or more general versions of known results. Actually, the latter is the case in the present note. To be more concrete, our reproofs of known results in Section 3 will be the basic parts in Section 4 of the proofs of the main results from Section 2. Let

$$W = \frac{2XY}{\sqrt{X^2+Y^2}}.$$

We consider now the two-dimensional Borel set

$$A(w) = \{(x,y) \in R^2 : \frac{2xy}{\sqrt{x^2+y^2}} < w\},\, w \in R.$$

There holds

$$P(W < w) = \Phi_{0_2,I_2}(A(w)), w \in R.$$

The probability $P(W < w)$ can be splitted as

$$P(W < w) = I_{(-\infty,0)}(w)P(sign(X) \neq sign(Y),\, W < w)$$

$$+I_{[0,\infty)}(w)[\frac{1}{2} + P(sign(X) = sign(Y),\, W < w)],\, w \in R.$$

The shaded areas in Figures 1-3 illustrate the sets to be measured in R^2 with Φ_{0_2,I_2} or $\Phi_{h;0_2,I_2}$ for representing the cdf of W under the assumption (1) or (3), respectively.

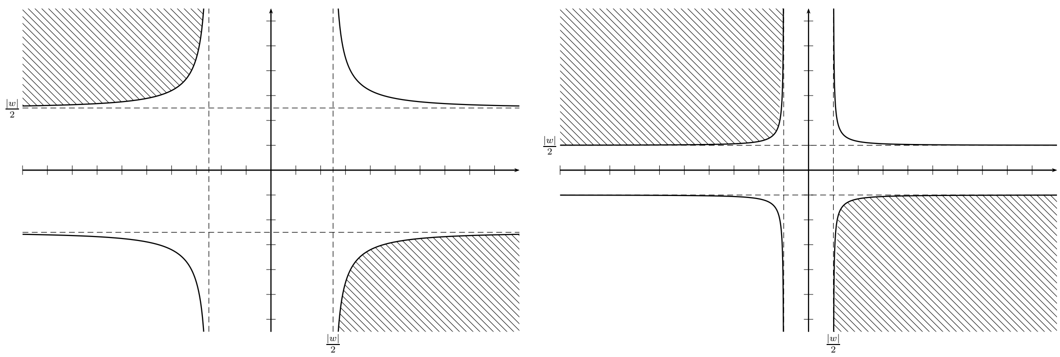
The p-functional $|.|_p : R^2 \to [0,\infty)$ which is defined by

$$|(x,y)|_p := (|x|^p + |y|^p)^{1/p},\, (x,y) \in R^2$$

is a norm if $p \geq 1$ and, according to Moszyńska & Richter (2012), an antinorm if $p \in (0,1)$ and a semi-antinorm if $p < 0$. Making use of this functional, we get

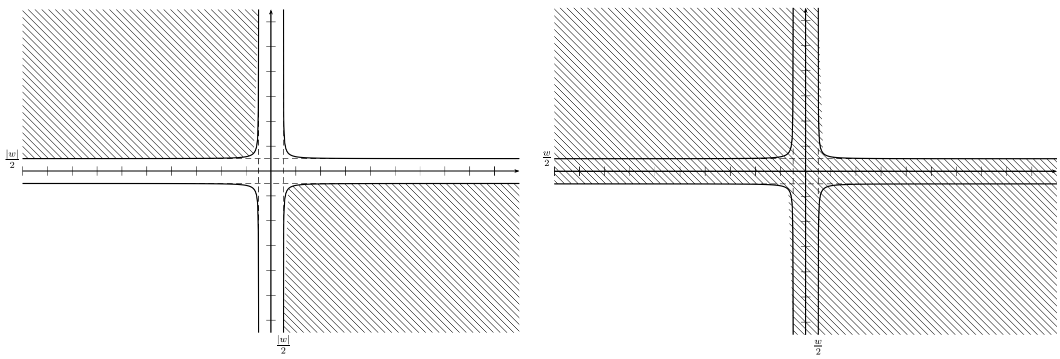
$$P(W < w) = I_{(-\infty,0)}(w)P(sign(X) \neq sign(Y),\, |(X,Y)|_{-2} > -\frac{w}{2})$$

$$+I_{[0,\infty)}(w)[\,\frac{1}{2} + P(sign(X) = sign(Y),\, |(X,Y)|_{-2} < \frac{w}{2})\,].$$



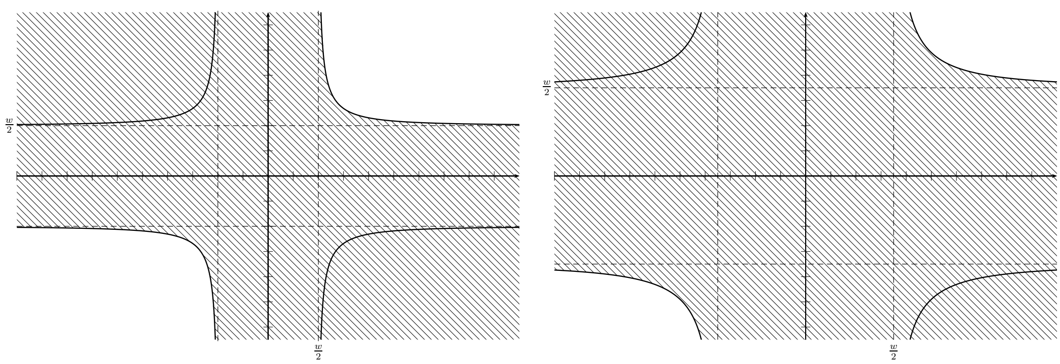
(a) w=-5 (b) w=-2

Figure 1. Values from the left tail of the cdf of T are the Gaussian or spherical measure of the shaded areas.



(a) w=-1 (b) w=1

Figure 2. Values from the central region of the cdf of T. Skip over the case $w = 0$ being closely connected with the case of a half space.



(a) w=4 (b) w=7

Figure 3. Values from the right tail of the cdf of T are Φ_{0_2,I_2} or $\Phi_{h;0_2,I_2}$-measure values of the shaded areas.

Let the sectors of the canonical fan in R^2 be denoted according to the anticlockwise enumeration by $C_1, ..., C_4$ and let $\mathcal{F}(A(w), r), r > 0$ be the Euclidean circle ipf of the set $A(w)$. Then

$$\mathcal{F}(A(w), r) = \sum_{i=1}^{4} \mathcal{F}(A(w) \cap C_i, r).$$

In the case $w > 0$,

$$\mathcal{F}(A(w) \cap C_2, r) = \mathcal{F}(A(w) \cap C_4, r) = \frac{1}{4}, \forall r > 0,$$

$$\mathcal{F}(A(w) \cap C_1, r) = \mathcal{F}(A(w) \cap C_3, r) = \frac{1}{4}, 0 \leq r \leq w$$

and

$$\mathcal{F}(A(w), r) = \frac{1}{2} + 2\mathcal{F}(A(w) \cap C_1, r),\ r > w.$$

Moreover,

$$A(w) \cap C_1 \cap C(r) = P(\frac{rw}{2}) \cap C_1 \cap C(r) \tag{8}$$

where

$$P(t) = \{(x, y) : xy < t\}, t \in R.$$

In a certain sense, calculations needed for considering the ipf of the set $A(w)$ which is generated by the statistic W are transformed into calculations needed for considering the ipf of the set $P(\frac{rw}{2})$ which is generated by the random variable $X \cdot Y$. It follows from the results in Kalke et al. (2013) that the restriction of the ipf of the set $P(t), t > 0$ to the set $P(t) \cap C_1$ allows the representation

$$\mathcal{F}(P(t) \cap C_1, r) = \frac{\alpha}{\pi} I_{(\sqrt{2|t|}, \infty)}(r) + \frac{1}{4} I_{[0, \sqrt{2|t|}\,]}(r)$$

where

$$\alpha = \frac{1}{4} \arccos(1 - \frac{8t^2}{r^4}).$$

With $t = \frac{rw}{2}$ and $\alpha = \frac{1}{4} \arccos(1 - \frac{2w^2}{r^2})$, we have that

$$\mathcal{F}(A(w), r) = (\frac{1}{2} + \frac{1}{2\pi} \arccos(1 - \frac{2w^2}{r^2}))\, I_{[\,w, \infty)}(r) + I_{[0,\,w]}(r). \tag{9}$$

Now, the geometric measure representation formula in Richter (1985 and 1995), see also Kalke et al. (2013), applies such that

$$P(W < w) = \Phi_{0_2, I_2}(A(w))$$

$$= \int_0^w re^{-r^2/2}dr + \int_w^\infty re^{-r^2/2}[\frac{1}{2} + \frac{1}{2\pi}\arccos(1 - \frac{2w^2}{r^2})]dr$$

$$= 1 - e^{-w^2/2} + \frac{1}{2}\int_w^\infty re^{-r^2/2}dr + \frac{1}{2\pi}\int_w^\infty re^{-r^2/2}\arccos(1 - \frac{2w^2}{r^2})dr$$

$$= 1 - \frac{1}{2}e^{-w^2/2} + \frac{1}{2\pi}\int_w^\infty re^{-r^2/2}\arccos(1 - \frac{2w^2}{r^2})dr.$$

It follows immediately that

$$\frac{d}{dw}P(W < w)$$

$$= \frac{1}{2}we^{-w^2/2} + \frac{1}{2\pi}\int_w^\infty re^{-r^2/2}(-1)\frac{-4w/r^2}{\sqrt{1 - (1 - \frac{2w^2}{r^2})^2}}dr - \frac{1}{2\pi}we^{-w^2/2}\arccos(1 - \frac{2w^2}{w^2})$$

$$= \frac{1}{\pi}\int_w^\infty re^{-r^2/2}\frac{dr}{\sqrt{r^2 - w^2}}.$$

On changing variables $r^2 - w^2 = z^2$, we get $dz = \frac{r\,dr}{\sqrt{r^2 - w^2}}$ and

$$\frac{d}{dw}P(W < w) = \phi_{0,1}(w).$$

The following calculations will show that the latter equation is also true if $w < 0$. In this case,

$$\mathcal{F}(A(w) \cap C_1, r) = \mathcal{F}(A(w) \cap C_3, r) = 0, \forall r > 0,$$

$$\mathcal{F}(A(w) \cap C_2, r) = \mathcal{F}(A(w) \cap C_4, r) = 0, 0 \leq r \leq w$$

and

$$\mathcal{F}(A(w), r) = 2\mathcal{F}(A(w) \cap C_2, r),\ \forall r > w.$$

Further,

$$A(w) \cap C_2 \cap C(r) = P(\frac{wr}{2}) \cap C_2 \cap C(r) \tag{10}$$

$$= \{(x, y) \in R^2 : x < 0, y > 0, x^2 + y^2 = r^2, \frac{2(-x)y}{r} > -w\}.$$

It follows from the results in Kalke et al. (2013) that the restriction of the ipf of the set $P(t), t < 0$ to the set $P(t) \cap C_2$ allows the representation

$$\mathcal{F}(P(t) \cap C_2, r) = (\frac{1}{4} - \frac{\alpha}{\pi})\, I_{(\sqrt{2|t|},\infty)}(r),\ r > 0.$$

With $t = \frac{rw}{2}$, we conclude that

$$\mathcal{F}(A(w), r) = 2\,(\frac{1}{4} - \frac{\alpha}{\pi})\, I_{[|w|,\infty)}(r). \tag{11}$$

It follows from (9) and (11) that $W \in$ IpfRep. The geometric measure representation applies, hence

$$P(W < w) = \Phi_{0_2,I_2}(A(w))$$

$$= \int\limits_{|w|}^{\infty} re^{-r^2/2} 2[\frac{1}{4} - \frac{1}{4\pi}\arccos(1 - \frac{2w^2}{r^2})]dr$$

$$= \frac{1}{2}e^{-w^2/2} - \frac{1}{2\pi}\int\limits_{|w|}^{\infty} re^{-r^2/2}\arccos(1 - \frac{2w^2}{r^2})dr.$$

This yields

$$\frac{d}{dw}P(W < w) = -\frac{w}{2}e^{-w^2/2} + \frac{1}{2\pi}\int\limits_{|w|}^{\infty} re^{-r^2/2}\frac{4w/r^2}{\sqrt{1-(1-\frac{2w^2}{r^2})^2}}dr$$

$$+\frac{1}{2\pi}we^{-w^2/2}\arccos(1-2) = \frac{1}{\pi}\int\limits_{|w|}^{\infty} re^{-r^2/2}\frac{dr}{\sqrt{r^2 - w^2}}.$$

Changing variables as in the case $w > 0$, the result follows immediately.

3.2 Geometric measure theoretical proof of Theorem 2[A]

Because of the equation

$$P(W < w) = \Phi_{h;0_2,I_2}(A(w))$$

this proof follows basically the line of the preceding one until that point where we proved that $W \in$ IpfRep. Now, the geometric measure representation formula for the spherical distribution with dg h in Richter (1991) applies. Notice that the integral

$$I_h = \int\limits_0^{\infty} r\, h(r^2)dr$$

occuring in that formula for an arbitrary dg h equals $\frac{1}{2\pi}$. Hence, for $w > 0$,

$$P(W < w) = \Phi_{h;0_2,I_2}(A(w))$$

$$= 2\pi \left(\frac{1}{4\pi} + \frac{1}{2} \int_{0}^{w} rh(r^2)dr + \frac{1}{2\pi} \int_{w}^{\infty} rh(r^2) \arccos(1 - \frac{2w^2}{r^2})dr \right).$$

It follows that

$$\frac{d}{dw} P(W < w) = \pi w h(w^2)$$

$$+ \int_{w}^{\infty} rh(r^2) \frac{4w/r^2}{\sqrt{1 - (1 - \frac{2w^2}{r^2})^2}} dr - wh(w^2) \arccos(-1)$$

$$= 2 \int_{w}^{\infty} \frac{r}{\sqrt{r^2 - w^2}} h(r^2) dr.$$

If $w < 0$ then

$$P(W < w) = 2\pi \int_{|w|}^{\infty} rh(r^2) 2[\frac{1}{4} - \frac{1}{4\pi} \arccos(1 - \frac{2w^2}{r^2})] dr$$

$$= \pi \int_{|w|}^{\infty} rh(r^2)dr - \int_{|w|}^{\infty} rh(r^2) \arccos(1 - \frac{2w^2}{r^2})dr.$$

Hence,

$$\frac{d}{dw} P(W < w) = -\pi w h(w^2) + \int_{|w|}^{\infty} rh(r^2) \frac{4w/r^2}{\sqrt{1 - (1 - \frac{2w^2}{r^2})^2}} dr + wh(w^2) \arccos(1-2)$$

$$= 2 \int_{|w|}^{\infty} \frac{r}{\sqrt{r^2 - w^2}} h(r^2) dr = 2 \int_{0}^{\infty} h(z^2 + w^2) dz.$$

4 Proofs of the main results

The basic idea is to further exploit the geometric measure representations of the two-dimensional Gaussian and spherical distribution laws in the proofs of Theorems 4 up to 6. The proof of Theorem 3, however, more directly refers to known results.

Proof of Theorem 3 It follows from the well known invariance properties of the bivariate standard Gaussian law that

$$P(\begin{pmatrix} X \\ Y \end{pmatrix} \in A(w)) = P(O\begin{pmatrix} X \\ Y \end{pmatrix} \in A(w))$$

for any orthogonal matrix O. Hence,

$$\Phi(w) = P(\frac{2XY}{\sqrt{X^2+Y^2}} < w)$$

$$= P(\frac{2(a_{11}X + a_{12}Y)(a_{21}X + a_{22}Y)}{\sqrt{(a_{11}X + a_{12}Y)^2 + (a_{21}X + a_{22}Y)^2}} < w)$$

$$= P(\frac{\lambda_1 X^2 + \lambda_2 XY + \lambda_3 Y^2}{\sqrt{\mu_1 X^2 + \mu_2 XY + \mu_3 Y^2}})$$

where the coefficients λ_i and $\mu_i, i = 1, 2, 3$ satisfy the equations (6) and (7), respectively.

Proofs of Theorems 4,5 and 6

Looking through again the reproofs of Theorems 1 and 2 and the final proof of Theorem 3, Definition 1 and Lemma 1 apply. Theorems 4, 5 and 6 are now immediate conclusions from the consideration in Section 3. These proofs show that if we use geometric measure representations in the proofs of Theorems 1 and 2 then the proofs of the generalizations in Theorems 4, 5 and 6 become quite short. However, some of the earlier proofs of, e.g., Theorem 1 are actually shorter than our reproof in Section 3.1.

Acknowledgement The author thanks Seppo Pynnönen for the very insightful discussions during authors stay 2012 in Vaasa which were among several other things also closely connected with the basic notion of the uniform distribution. This notion is dealt with in the present note just in a very special case while the much more advanced interest of Seppo is in Pynnönen (2013) in Stiefel manifolds. The author also thanks Thomas Dietrich for drawing the figures in PSTricks.

References

Anderson, T.W. & Fang, K.T. (1990). On the theory of multivariate elliptically contoured distributions and their applications. In *Statistical inference in elliptically contoured and related distributions*. New York: Allerton Press, 1–23.

Arellano-Valle, R.B. (2001). On some characterizations of spherical distributions. *Statist. Probab. Lett.* 54, 227–232.

Bansal, N., Hamedani, C.G., Key, E.S., Volkmer, H., Zhang, H. & Behboodian, J. (1999). Some characterizations of the normal distribution. *Statistics & Probability Letters* 42, 393–400.

Baringhaus, L., Henze, N. & Morgenstern, D. (1988). Some elementary proofs of the normality of $XY/(X^2 + Y^2)^{1/2}$ when X and Y are normal. *Statist. Probab. Lett.* 42, 393–400.

Barndorff-Nielsen, O.E., Blaesild, P. & Eriksen, P.S. (1989). *Decomposition and Invariance of Measures, and Statistical Transformation Models*. LN Statistics. Springer-Verlag.

Cambanis, S., Huang, S. & Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis* 11, 368–385.

Cohen, Jr., E.A. (1981). A note on normal functions of normal random variables. *Comput. Math. Applic.* 7, 395–400.

Eaton, M.L. (1983). *Multivariate Statistics, a Vector Space Approach*. New York: Wiley.

Fang, K.T., Kotz, S. & Ng, K.W. (1990). *Symmetric multivariate and related distributions*. New York: Chapman and Hall.

Fang, K.T. & Shang, Y.T. (1990). *Generalized Multivariate Analysis*. Berlin: Springer.

Farrel, R.H. (1985). *Multivariate Calculation. Use of the Continuous Groups*. Berlin: Springer.

Federer (1969). *Geometric Measure Theory*. Berlin - Heidelberg - New York: Springer Verlag.

Genton (Ed.) (2004). *Skew-elliptical distributions and their applications. A journey beyond normality*. CRC, Boca Raton, Fl: Chapman and Hall.

Günzel, T., Richter, W.-D., Scheutzow, S., Schicker, K. & Venz, J. (2012). Geometric approach to the skewed normal distribution. *Journal of Statistical Planning and Inference* 142, 3209–3224. DOI doi.org/10.1016/j/jspi.2012.06.009.

Ittrich, C., Krause, D. & Richter, W.-D. (2000). Probabilities and large quantiles of noncentral generalized Chi-Square distributions. *Statistics* 34, 53–101.

Ittrich, C. & Richter, W.-D. (2005). Exact tests and confidence regions in nonlinear regression. *Statistics* 39: 1, 13–42.

Johnson, N.L. & Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions. Bd. 2.* New York: Wiley.

Kalke, S., Richter, W.-D. & Thauer, F. (2013). Linear combinations, products and ratios of simplicial or spherical variates. *Communications in Statistics: Theory and Methods* 42: 3, 505–527. DOI 10.1080/0361 0926.2011.

Kallenberg, O. (2005). *Probabilistic Symmetries and Invariance Principles*. New York: Springer.

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhya, Ser. A* 32, 419–430.

Krantz, S.G. & Parks, H.R. (2008). *Geometric Integration Theory*. Boston, Basel,Berlin: Birkhäuser.

Krause, D. & Richter, W.-D. (2004). Exact probabilities of correct classifications for uncorrelated repeated measurements from elliptically contoured distributions. *Journal of Multivariate Analysis* 89, 36–69.

Morgan, F. (1984). *Geometric Measure Theory*. San Diego: Academic Press.

Moszyńska, M. & Richter, W.-D. (2012). Reverse triangle inequality. Antinorms and semi-antinorms. *Studia Scientarum Mathematicarum Hungarica* 49: 1, 120–138. DOI 10.1556/SScMath.49.2012.1.1192.

Muirhead, R. (1982). *Aspects of Multivariate Statistical Theory*. New York: Wiley.

Müller, K. & Richter, W.-D. (2014). Exact extreme value, product, and ratio distributions under non-standard assumptions. *A StA Advances in Statistical Analysis* , Submitted.

Nachbin, L. (1976). *The Haar Integral*. Huntington, N.Y.: Krieger Publication Co.

Pynnönen, S. (2013). Distribution of an arbitrary linear transformation of internally Studentized residuals of multivariate regression with elliptical errors. *Journal of Multivariate Analysis* 107, 40–52.

Richter, W.-D. (1985). Laplace-Gauss integrals, Gaussian measure asymptotic behaviour and probabilities of moderate deviations. *Zeitschrift für Analysis und ihre Anwendungen* 4, 257–267.

Richter, W.-D. (1991). Eine geometrische Methode in der Stochastik. *Rostocker Mathematisches Kolloquium* 44, 63–72.

Richter, W.-D. (1995). A geometric approach to the Gaussian law. In V. Mammitzsch & H. Schneeweiß (Eds.) *Symposia Gaussiana, Conference B*. Berlin: Walter de Gruyter and Co., 25–45.

Richter, W.-D. (2007). Generalized spherical and simplicial coordinates. *J. Math. Anal. Appl.* 336, 1187–1202. DOI 10.1016/j.jmaa.2007.03.047.

Richter, W.-D. (2009). Continuous $l_{n,p}$-symmetric distributions. *Lithuanian Mathematical Journal* 49: 1, 93–108.

Richter, W.-D. & Venz, J. (2014). Geometric representations of multivariate skewed elliptically contoured distributions. Submitted .

Schindler, W. (2003). *Measures with Symmetry Properties*. LNM 1805. Berlin-Heidelberg: Springer Verlag.

Shepp, L. (1964). Normal functions of normal random variables. *SIAM Review* 6, 459–460.

Silvermann, H. (2000). The value of reproving. *PRIMUS* 4: 2, 151–154.

Wijsman, R. (1984). *Invariant Measures on Groups and their Use in Statistics*. IMS Lecture Notes-Monograph Series 14.

IMPROVED FREQUENTIST PREDICTION INTERVALS FOR ARMA MODELS BY SIMULATION

Jouni Helske and Jukka Nyblom
University of Jyväskylä

1 Introduction

In a traditional approach to time series forecasting, prediction intervals are usually computed as if the chosen model were correct and the parameters of the model completely known, with no reference to the uncertainty regarding the model selection and parameter estimation. The parameter uncertainty may not be a major source of prediction errors in practical applications, but its effects can be substantial if the series is not too long. The problems of interval prediction are discussed in depth in Chatfield (1993, 1996) and Clements & Hendry (1999).

Several proposals have been made for improving prediction intervals when parameters are estimated. One group of solutions focus on finding a more accurate prediction mean squared error in the presence of estimation; e.g. see Phillips (1979), Fuller & Hasza (1981), Ansley & Kohn (1986), Quenneville & Singh (2000), and Pfeffermann & Tiller (2005). Both analytic and bootstrap approaches are tried. Barndorff-Nielsen & Cox (1996) give general results for prediction intervals in the presence of estimated parameters. These results are further developed for time series models by Vidoni (2004, 2009). Bootstrap solutions are given by several authors; see for example Beran (1990), Masarotto (1990), Grigoletto (1998), Kim (2004), Pascual, Romo & Ruiz (2004), Clements & Kim (2007), Kabaila & Syuhada (2008), and Rodriguez & Ruiz (2009).

Here we show how to take into account the parameter uncertainty in a fairly simple way under autoregressive moving average (ARMA) models. We construct prediction intervals having approximately correct frequentist coverage probability, i.e. an average coverage probability over the realizations is approximately correct under the true parameter values. Due to the uncertainty in parameter estimation, the traditional plug-in method usually provides prediction intervals with average coverage probabilities falling below the nominal level. Our proposed method is based on Bayesian approach. Therefore the coverage probability is exactly correct if one is ready to accept the chosen prior distribution. But our aim is to find such priors that yield approximately correct coverage probabilities also in the frequentist sense. As a computational device the fairly simple importance sampling is employed in posterior calculations. The method is an extension of the approach proposed by Helske & Nyblom (2013) for pure autoregressive models. The paper is organized as follows. Sections 2 and 3 derive general results, and section 4 applies them to ARMA models. Section 5 discusses prior distributions. Section 6 compares the plug-in method to Bayesian solutions by means of simulation experiments. Section 7 presents an application to real data. Section 8 concludes.

2 The model

We start with a fairly general linear model and later apply the results to ARMA models. Assume that the observations $y_1, \ldots, y_n$ are stacked in a vector $\boldsymbol{y}$ satisfying the model

$$\boldsymbol{y} \,|\, \boldsymbol{\psi}, \sigma, \boldsymbol{\beta} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{V}_\psi), \tag{1}$$

where $\boldsymbol{X}$ is the $n \times k$ matrix of fixed regressors with rows $\boldsymbol{x}_t' = (x_{t1}, \ldots, x_{tk})$, and $\sigma^2 \boldsymbol{V}_\psi$ is the covariance matrix depending on the parameters $(\psi_1, \ldots, \psi_r)' = \boldsymbol{\psi}$. We assume that $\boldsymbol{X}$ is of full rank k. The error vector is defined as $\boldsymbol{\epsilon} = \boldsymbol{y} - \boldsymbol{X\beta}$. Plainly $\boldsymbol{\epsilon} \sim \mathrm{N}(0, \sigma^2 \boldsymbol{V}_\psi)$. Next recall the well known identity

$$\begin{aligned}(\boldsymbol{y} - \boldsymbol{X\beta})' \boldsymbol{V}_\psi^{-1} (\boldsymbol{y} - \boldsymbol{X\beta}) &= (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\psi)' \boldsymbol{V}_\psi^{-1} (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\psi) \\ &\quad + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_\psi) \boldsymbol{X}' \boldsymbol{V}_\psi^{-1} \boldsymbol{X} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_\psi),\end{aligned}$$

where

$$\widehat{\boldsymbol{\beta}}_\psi = (\boldsymbol{X}' \boldsymbol{V}_\psi^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{V}_\psi^{-1} \boldsymbol{y}.$$

The estimate $\widehat{\boldsymbol{\beta}}_\psi$ is the generalized least squares estimate for $\boldsymbol{\beta}$ when $\boldsymbol{\psi}$ is known. Define also

$$S_\psi^2 = (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\psi)' \boldsymbol{V}_\psi^{-1} (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\psi).$$

Then the likelihood can be written as

$$\begin{aligned} p(\boldsymbol{y} \,|\, \boldsymbol{\psi}, \boldsymbol{\beta}, \sigma) &= (2\pi)^{-\frac{n}{2}} \sigma^{-n} |\boldsymbol{V}_\psi|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} (\boldsymbol{y} - \boldsymbol{X\beta})' \boldsymbol{V}_\psi^{-1} (\boldsymbol{y} - \boldsymbol{X\beta}) \right) \\ &= (2\pi)^{-\frac{n}{2}} \sigma^{-n} |\boldsymbol{V}_\psi|^{-\frac{1}{2}} \exp\left(-\frac{S_\psi^2}{2\sigma^2} \right) \\ &\quad \times \exp\left(-\frac{1}{2\sigma^2} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_\psi) \boldsymbol{X}' \boldsymbol{V}_\psi^{-1} \boldsymbol{X} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_\psi) \right). \end{aligned}$$

Although our main purpose is to derive frequentist prediction intervals, we use the Bayes approach in their construction. Therefore, assume now that the parameters $\boldsymbol{\beta}$, σ and $\boldsymbol{\psi}$ are random and have a joint prior distribution. Moreover, $\boldsymbol{\psi}$ is independent from $\boldsymbol{\beta}$ and σ with $(\boldsymbol{\beta}, \log\sigma)$ having the improper uniform prior distribution. Let $p(\boldsymbol{\psi})$ be the prior of $\boldsymbol{\psi}$. Then the joint prior is of the form $p(\boldsymbol{\psi})/\sigma$. These assumptions lead to the joint posterior density

$$\begin{aligned}
p(\boldsymbol{\beta}, \boldsymbol{\psi}, \sigma \,|\, \boldsymbol{y}) \;&\propto\; p(\boldsymbol{\psi})\sigma^{-n-1}|\boldsymbol{V}_{\boldsymbol{\psi}}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'\boldsymbol{V}_{\boldsymbol{\psi}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})\right) \\
&\quad \times \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\boldsymbol{X}'\boldsymbol{V}_{\boldsymbol{\psi}}^{-1}\boldsymbol{X}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\right) \\
&\propto\; p(\boldsymbol{\psi})|\boldsymbol{V}_{\boldsymbol{\psi}}|^{-\frac{1}{2}}\sigma^{-(n-k+1)} \exp\left(-\frac{S_{\boldsymbol{\psi}}^2}{2\sigma^2}\right) \qquad (2) \\
&\quad \times\sigma^{-k} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\boldsymbol{X}'\boldsymbol{V}_{\boldsymbol{\psi}}^{-1}\boldsymbol{X}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\right). \qquad (3)
\end{aligned}$$

Let us factorize the posterior as

$$p(\boldsymbol{\psi}, \sigma, \boldsymbol{\beta} \,|\, \boldsymbol{y}) \;=\; p(\boldsymbol{\psi} \,|\, \boldsymbol{y})p(\sigma \,|\, \boldsymbol{\psi}, \boldsymbol{y})p(\boldsymbol{\beta} \,|\, \boldsymbol{\psi}, \sigma, \boldsymbol{y}).$$

The formula (2)–(3) yield the conditional posteriors

$$\begin{aligned}
\boldsymbol{\beta} \,|\, \boldsymbol{\psi}, \sigma, \boldsymbol{y} \;&\sim\; N\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\psi}}, \sigma^2(\boldsymbol{X}'\boldsymbol{V}_{\boldsymbol{\psi}}^{-1}\boldsymbol{X})^{-1}\right), \\
\left.\frac{S_{\boldsymbol{\psi}}^2}{\sigma^2}\,\right|\, \boldsymbol{\psi}, \boldsymbol{y} \;&\sim\; \chi^2(n-k).
\end{aligned}$$

For $\boldsymbol{\psi}$, the marginal posterior is

$$p(\boldsymbol{\psi} \,|\, \boldsymbol{y}) \;\propto\; p(\boldsymbol{\psi})|\boldsymbol{V}_{\boldsymbol{\psi}}|^{-\frac{1}{2}}|\boldsymbol{X}'\boldsymbol{V}_{\boldsymbol{\psi}}^{-1}\boldsymbol{X}|^{-\frac{1}{2}}S_{\boldsymbol{\psi}}^{-(n-k)}, \qquad (4)$$

whenever the right side is integrable. In section 4, $\boldsymbol{\psi}$ and the related covariance matrix $\boldsymbol{V}_{\boldsymbol{\psi}}$ are specified through an appropriate ARMA model.

3 Bayesian prediction intervals

Assume that the future observations $y_{n+1}, y_{n+2}, \ldots$ come from the same model (1) with known values $\boldsymbol{x}_{n+1}, \boldsymbol{x}_{n+2}, \ldots$. Let

$$\begin{aligned}
E(y_{n+h} \,|\, \boldsymbol{y}, \boldsymbol{\beta}, \sigma, \boldsymbol{\psi}) \;&=\; \hat{y}_{n+h|n}(\boldsymbol{\beta}, \boldsymbol{\psi}) \qquad (5) \\
\operatorname{var}(y_{n+h} \,|\, \boldsymbol{y}, \boldsymbol{\beta}, \sigma, \boldsymbol{\psi}) \;&=\; \sigma^2 v_{n+h|n}^2(\boldsymbol{\psi}). \qquad (6)
\end{aligned}$$

Then

$$y_{n+h} \,|\, \boldsymbol{y}, \boldsymbol{\beta}, \sigma, \boldsymbol{\psi} \sim N(\hat{y}_{n+h|n}(\boldsymbol{\beta}, \boldsymbol{\psi}), \sigma^2 v_{n+h|n}^2(\boldsymbol{\psi})), \quad h = 1, 2, \ldots,$$

where for simplicity of notation the dependence on $\boldsymbol{x}_{n+1}, \ldots, \boldsymbol{x}_{n+h}$ is not explicitly shown. Then the Bayesian prediction intervals boils down to computing posterior probabilities of the form

$$P(y_{n+h} \leq b|\boldsymbol{y}) = E\left[\Phi\left(\frac{b - \hat{y}_{n+h|n}(\boldsymbol{\beta}, \boldsymbol{\psi})}{\sigma v_{n+h|n}(\boldsymbol{\psi})}\right) \,\middle|\, \boldsymbol{y}\right],$$

where $E(\cdot \,|\, \boldsymbol{y})$ refers to expectation with respect to the posterior distribution of $(\boldsymbol{\beta}, \sigma, \boldsymbol{\psi})$.

In practice the computation is accomplished by simulation. Suppose that we have the maximum likelihood estimate $\widehat{\boldsymbol{\psi}}$ and its approximate large sample covariance matrix $\widehat{\boldsymbol{\Sigma}}$. Then we employ the following importance sampling for computing prediction intervals:

(i) Draw $\boldsymbol{\psi}_j$ from $N(\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\Sigma}})$, and compute the weight

$$w_j = \frac{p(\boldsymbol{\psi}_j \mid \boldsymbol{y})}{g(\boldsymbol{\psi}_j)},$$

where $p(\boldsymbol{\psi}_j \mid \boldsymbol{y})$ is defined in (4) and

$$g(\boldsymbol{\psi}_j) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\psi}_j - \widehat{\boldsymbol{\psi}})'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\psi}_j - \widehat{\boldsymbol{\psi}})\right).$$

(ii) Draw $q_j \sim \chi^2(n-k)$ independently from $\boldsymbol{\psi}_j$, and let $\sigma_j^2 = S^2_{\boldsymbol{\psi}_j}/q_j$.

(iii) Draw $\boldsymbol{\beta}_j \sim N(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\psi}_j}, \sigma_j^2(\boldsymbol{X}'\boldsymbol{V}^{-1}_{\boldsymbol{\psi}_j}\boldsymbol{X})^{-1})$.

(iv) Repeat (i)–(iii) independently for $j = 1, \ldots, N$.

(v) Compute the weighted average

$$\bar{P}_N(b) = \frac{\sum_{j=1}^N w_j \Phi\left(\frac{b - \hat{y}_{n+h|n}(\boldsymbol{\beta}_j, \boldsymbol{\psi}_j)}{\sigma_j v_{n+h|n}(\boldsymbol{\psi}_j)}\right)}{\sum_{j=1}^N w_j}. \qquad (7)$$

(vi) Find the values b_α and $b_{1-\alpha}$ such that $\bar{P}_N(b_\alpha) = \alpha$ and $\bar{P}_N(b_{1-\alpha}) = 1 - \alpha$. When N is large $(b_\alpha, b_{1-\alpha})$ yields a prediction interval with coverage probability $1 - 2\alpha$.

4 Regression with ARMA errors

The regression model with ARMA errors is defined by the equations

$$\begin{aligned} y_t &= \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + \epsilon_t, & (8)\\ \epsilon_t &= \phi_1 \epsilon_{t-1} + \ldots + \phi_p \epsilon_{t-p} + \xi_t + \theta_1 \xi_{t-1} + \ldots + \theta_q \xi_{t-q}, & (9) \end{aligned}$$

where ξ_t are independent for all t and drawn from $N(0, \sigma^2)$. Thus, the process $\{\epsilon_t\}$ is ARMA(p, q) that we assume stationary and invertible. This is a special case of the model in section 2 with $\boldsymbol{\psi}' = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$. Let $r = \max(p, q+1)$. For notational convenience we add zeros to either autoregressive or moving average parameters such that we have $\phi_1, \ldots \phi_r$ and $\theta_1, \ldots, \theta_{r-1}$. Of course, if $r = 1$ there are no moving average parameters. Following Durbin & Koopman (2001, pp. 46–47) the model (8)–(9) can be put into a state space form as

$$\begin{aligned} y_t &= \boldsymbol{z}_t' \boldsymbol{\alpha}_t, & (10)\\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{T}\boldsymbol{\alpha}_t + \boldsymbol{R}\xi_{t+1}, & (11) \end{aligned}$$

where $\boldsymbol{z}_t' = (\boldsymbol{x}_t', 1, 0, \ldots, 0)$,

$$\boldsymbol{\alpha}_t = \begin{pmatrix} \boldsymbol{\beta}_t \\ \epsilon_t \\ \phi_2 \epsilon_{t-1} + \ldots + \phi_r \epsilon_{t-r+1} + \theta_1 \xi_t + \ldots + \theta_{r-1} \xi_{t-r+2} \\ \vdots \\ \phi_r \epsilon_{t-1} + \theta_{r-1} \xi_t \end{pmatrix},$$

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0}' & \phi_1 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \\ \mathbf{0}' & \phi_{r-1} & 0 & & 1 \\ \mathbf{0}' & \phi_r & 0 & \cdots & 0 \end{pmatrix}, \quad \boldsymbol{R} = \begin{pmatrix} \mathbf{0} \\ 1 \\ \theta_1 \\ \vdots \\ \theta_{r-1} \end{pmatrix}.$$

Note that this formulation implies that actually $\boldsymbol{\beta}_t$ is constant $\boldsymbol{\beta}$. The initial distribution for $\boldsymbol{\alpha}_1$ is $N(\mathbf{0}, \boldsymbol{P}_1)$ with

$$\boldsymbol{P}_1 = \begin{pmatrix} \kappa \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{pmatrix}, \tag{12}$$

where $\kappa \boldsymbol{I}$ corresponds to $\boldsymbol{\beta}_1$, and $\boldsymbol{\Gamma}$ is the covariance matrix of the stationary ARMA component of $\boldsymbol{\alpha}_t$.

Let $\boldsymbol{T}_{\boldsymbol{\phi}}$ and $\boldsymbol{R}_{\boldsymbol{\theta}}$ be the blocks of $\boldsymbol{T}$ and $\boldsymbol{R}$, respectively, related to the ARMA process. Then $\boldsymbol{\Gamma}$ satisfies $\boldsymbol{\Gamma} = \boldsymbol{T}_{\boldsymbol{\phi}}\boldsymbol{\Gamma}\boldsymbol{T}'_{\boldsymbol{\phi}} + \boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{R}'_{\boldsymbol{\theta}}$ and is given by

$$\mathrm{vec}(\boldsymbol{\Gamma}) = (\boldsymbol{I} - \boldsymbol{T}_{\boldsymbol{\phi}} \otimes \boldsymbol{T}_{\boldsymbol{\phi}})^{-1}\mathrm{vec}(\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{R}'_{\boldsymbol{\theta}}),$$

see Durbin & Koopman (2001, p. 112). The $\mathrm{vec}(\cdot)$ notation stands for the column-wise transformation of a matrix to a vector.

The initial distribution for $\boldsymbol{\beta}_1$ is actually defined through the limit $\kappa \to \infty$ which corresponds to the improper constant prior for $\boldsymbol{\beta}$ assumed in section 2. Durbin & Koopman (2001, Ch. 5) gives the updating formulas under this assumption called diffuse initialization. Thus, the Kalman filter together with the diffuse initialization automatically yields the values

$$\begin{aligned} E(\boldsymbol{\beta}_{n+1} \mid \boldsymbol{y}, \sigma, \boldsymbol{\psi}) &= \widehat{\boldsymbol{\beta}}_{\boldsymbol{\psi}}, \\ \mathrm{cov}(\boldsymbol{\beta}_{n+1} \mid \boldsymbol{y}, \sigma, \boldsymbol{\psi}) &= \sigma^2(\boldsymbol{X}'\boldsymbol{V}_{\boldsymbol{\psi}}^{-1}\boldsymbol{X})^{-1}. \end{aligned}$$

Additionally the Kalman filter gives the prediction errors

$$e_{t|t-1} = y_t - E(y_t \mid y_1, \ldots, y_{t-1}, \sigma, \boldsymbol{\psi}), \quad t = 1, \ldots, n,$$

and their variances

$$\mathrm{var}(e_{t|t-1}) = \mathrm{var}(y_t \mid y_1, \ldots, y_{t-1}, \sigma, \boldsymbol{\psi}) = \sigma^2 v_{t|t-1}^2, \quad t = 1, \ldots, n.$$

Due to the improper uniform prior of $\boldsymbol{\beta}$, i.e. the diffuse initialization, some variances $v_{t|t-1}^2 \to \infty$, as $\kappa \to \infty$ (Durbin & Koopman, 2001, sect. 5.2.1). Let $F = \{t \mid v_{t|t-1}^2 \text{ is finite}, t = 1, \ldots, n\}$. Then given $\boldsymbol{\psi}$ we have, by the results of Durbin & Koopman (2001, sect. 7.2.1), that

$$\begin{aligned} \sum_{t \in F} \frac{e_{t|t-1}^2}{v_{t|t-1}^2} &= S_{\boldsymbol{\psi}}^2, \\ \prod_{t \in F} v_{t|t-1}^2 &= |\boldsymbol{V}_{\boldsymbol{\psi}}|^{-\frac{1}{2}} |\boldsymbol{X}'\boldsymbol{V}_{\boldsymbol{\psi}}^{-1}\boldsymbol{X}|^{-\frac{1}{2}}. \end{aligned}$$

Because $\boldsymbol{X}$ is of rank k, the number of finite variances is $n - k$. We have now all elements for the algorithm of section 3 except the prior $p(\boldsymbol{\psi})$ that is discussed in the next section.

5 Jeffreys's rule for priors

Good candidates for the prior meeting our purposes is found by Jeffreys's rule which leads to the square root of the determinant of the Fisher information matrix. Apart from an additive constant, the log-likelihood is here

$$\ell(\boldsymbol{\beta},\sigma,\boldsymbol{\psi}) = -n\log\sigma - \frac{1}{2}\log|\boldsymbol{V}_{\psi}| - \frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{V}_{\psi}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}).$$

A straightforward calculation gives the information matrix

$$\begin{aligned}
\boldsymbol{I}(\boldsymbol{\beta},\sigma,\boldsymbol{\psi}) &= \begin{bmatrix} \frac{1}{\sigma^2}(\boldsymbol{X}'\boldsymbol{V}_{\psi}^{-1}\boldsymbol{X}) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{2n}{\sigma^2} & \frac{1}{\sigma}\boldsymbol{I}_{21}'(\boldsymbol{\psi}) \\ \boldsymbol{0} & \frac{1}{\sigma}\boldsymbol{I}_{21}(\boldsymbol{\psi}) & \boldsymbol{I}_{22}(\boldsymbol{\psi}), \end{bmatrix}, \\
[\boldsymbol{I}_{21}(\boldsymbol{\psi})]_i &= \operatorname{trace}\left(\boldsymbol{V}_{\psi}^{-1}\frac{\partial \boldsymbol{V}_{\psi}}{\partial\psi_i}\right), \quad i=1,\ldots,r \\
[\boldsymbol{I}_{22}(\boldsymbol{\psi})]_{ij} &= \frac{1}{2}\operatorname{trace}\left(\boldsymbol{V}_{\psi}^{-1}\frac{\partial \boldsymbol{V}_{\psi}}{\partial\psi_i}\boldsymbol{V}_{\psi}^{-1}\frac{\partial \boldsymbol{V}_{\psi}}{\partial\psi_j}\right), \quad i,j=1,\ldots,r.
\end{aligned}$$

Hence,

$$|\boldsymbol{I}(\boldsymbol{\beta},\sigma,\boldsymbol{\psi})|^{\frac{1}{2}} = \frac{1}{\sigma^{k+1}}\left|\boldsymbol{X}'\boldsymbol{V}_{\psi}^{-1}\boldsymbol{X}\right|^{\frac{1}{2}}\left|\boldsymbol{I}_{22}(\boldsymbol{\psi})-(2n)^{-1}\boldsymbol{I}_{21}(\boldsymbol{\psi})\boldsymbol{I}_{21}(\boldsymbol{\psi})'\right|^{\frac{1}{2}}. \quad (13)$$

Because we want the joint prior to be of the form $p(\boldsymbol{\psi})/\sigma$, we insert $k=0$ in (13) and define

$$p(\boldsymbol{\psi}) \propto \left|\boldsymbol{X}'\boldsymbol{V}_{\psi}^{-1}\boldsymbol{X}\right|^{\frac{1}{2}}\left|\boldsymbol{I}_{22}(\boldsymbol{\psi})-(2n)^{-1}\boldsymbol{I}_{21}(\boldsymbol{\psi})\boldsymbol{I}_{21}(\boldsymbol{\psi})'\right|^{\frac{1}{2}}. \quad (14)$$

With this specification $p(\boldsymbol{\psi})/\sigma$ is called here the exact joint Jeffreys prior. Note that this prior depends on the sample size n. The approximate joint prior of the same form is obtained with

$$p(\boldsymbol{\psi}) \propto \left|\boldsymbol{X}'\boldsymbol{V}_{\psi}^{-1}\boldsymbol{X}\right|^{\frac{1}{2}}|\boldsymbol{J}_{\psi}|^{\frac{1}{2}}, \quad (15)$$

where

$$\boldsymbol{J}_{\psi} = \lim_{n\to\infty} n^{-1}\left(\boldsymbol{I}_{22}(\boldsymbol{\psi})-(2n)^{-1}\boldsymbol{I}_{21}(\boldsymbol{\psi})\boldsymbol{I}_{21}(\boldsymbol{\psi})'\right).$$

Substituting either (14) or (15) to (4) we find that the determinant $|\boldsymbol{X}'\boldsymbol{V}_{\psi}^{-1}\boldsymbol{X}|$ cancels.

Box et al. (2008, Ch. 7) gives useful results for the ARMA(p,q) models. We find that $\boldsymbol{J}_{\psi}^{-1}/n$ is the large sample covariance matrix of the maximum likelihood estimate $\widehat{\boldsymbol{\psi}}$. In the pure AR model we have $|\boldsymbol{V}_{\psi}| = |\boldsymbol{J}_{\psi}|$, although the matrices are different. For the pure MA models the same determinant equation is approximately true, but the same does not apply to the mixed models. The marginal Jeffreys priors are obtained by dropping off the factor $\left|\boldsymbol{X}'\boldsymbol{V}_{\psi}^{-1}\boldsymbol{X}\right|^{\frac{1}{2}}$ in (14) and (15).

The numerical evaluation of the posteriors involves the determinant $|\boldsymbol{V}_{\psi}|$, the inverse $\boldsymbol{V}_{\psi}^{-1}$ and the partial derivatives of $\boldsymbol{V}_{\psi}$. For short series the determinant and the inverse can be calculated directly. For longer series we can use the formulas provided by Lin & Ho (2008). The partial derivatives can be found recursively as follows. Recall the state space representation (10)–(11) and the initial covariance matrix $\boldsymbol{\Gamma}$ in (12). Due to stationarity of the process $\{\boldsymbol{\alpha}_t\}$ we find that $\operatorname{cov}(\boldsymbol{\alpha}_{t+s}, \boldsymbol{\alpha}_t) = \boldsymbol{T}^s\boldsymbol{P}_1$, where the block $\boldsymbol{T}_{\boldsymbol{\phi}}^s\boldsymbol{\Gamma}$ corresponds the autocovariance matrix of the ARMA process. The position $(1,1)$ of this matrix shows $\operatorname{cov}(y_{t+s}, y_t)$. We find the partial derivatives recursively for the autoregressive parameters

$$\frac{\partial(\boldsymbol{T}_{\boldsymbol{\phi}}^s\boldsymbol{\Gamma})}{\partial\phi_j} = \frac{\partial\boldsymbol{T}_{\boldsymbol{\phi}}}{\partial\phi_j}\boldsymbol{T}_{\boldsymbol{\phi}}^{s-1}\boldsymbol{\Gamma} + \boldsymbol{T}_{\boldsymbol{\phi}}\frac{\partial(\boldsymbol{T}_{\boldsymbol{\phi}}^{s-1}\boldsymbol{\Gamma})}{\partial\phi_j}, \quad s = 1, 2, \ldots.$$

For moving average parameters we have

$$\frac{\partial(\boldsymbol{T}_{\boldsymbol{\phi}}^s\boldsymbol{\Gamma})}{\partial\theta_j} = \boldsymbol{T}_{\boldsymbol{\phi}}^s\frac{\partial\boldsymbol{\Gamma}}{\partial\theta_j}, \quad s = 1, 2, \ldots.$$

Because $\boldsymbol{\Gamma}$ satisfies $\boldsymbol{\Gamma} = \boldsymbol{T}_{\boldsymbol{\phi}}\boldsymbol{\Gamma}\boldsymbol{T}_{\boldsymbol{\phi}}' + \boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{R}_{\boldsymbol{\theta}}'$, we find by differentiating on both sides that

$$\begin{aligned}
\frac{\partial\boldsymbol{\Gamma}}{\partial\phi_j} &= \boldsymbol{T}_{\boldsymbol{\phi}}\frac{\partial\boldsymbol{\Gamma}}{\partial\phi_j}\boldsymbol{T}_{\boldsymbol{\phi}}' + \frac{\partial\boldsymbol{T}_{\boldsymbol{\phi}}}{\partial\phi_j}\boldsymbol{\Gamma}\boldsymbol{T}_{\boldsymbol{\phi}}' + \boldsymbol{T}_{\boldsymbol{\phi}}\boldsymbol{\Gamma}\frac{\partial\boldsymbol{T}_{\boldsymbol{\phi}}'}{\partial\phi_j}, \\
\frac{\partial\boldsymbol{\Gamma}}{\partial\theta_j} &= \boldsymbol{T}_{\boldsymbol{\phi}}\frac{\partial\boldsymbol{\Gamma}}{\partial\theta_j}\boldsymbol{T}_{\boldsymbol{\phi}}' + \frac{\partial\boldsymbol{R}_{\boldsymbol{\theta}}}{\partial\theta_j}\boldsymbol{R}_{\boldsymbol{\theta}} + \boldsymbol{R}_{\boldsymbol{\theta}}\frac{\partial\boldsymbol{R}_{\boldsymbol{\theta}}'}{\partial\theta_j}.
\end{aligned}$$

which implies that

$$\begin{aligned}
\operatorname{vec}\left(\frac{\partial\boldsymbol{\Gamma}}{\partial\phi_j}\right) &= (\boldsymbol{I} - \boldsymbol{T}_{\boldsymbol{\phi}}\otimes\boldsymbol{T}_{\boldsymbol{\phi}})^{-1}\operatorname{vec}\left(\frac{\partial\boldsymbol{T}_{\boldsymbol{\phi}}}{\partial\phi_j}\boldsymbol{\Gamma}\boldsymbol{T}_{\boldsymbol{\phi}}' + \boldsymbol{T}_{\boldsymbol{\phi}}\boldsymbol{\Gamma}\frac{\partial\boldsymbol{T}_{\boldsymbol{\phi}}'}{\partial\phi_j}\right), \\
\operatorname{vec}\left(\frac{\partial\boldsymbol{\Gamma}}{\partial\theta_j}\right) &= (\boldsymbol{I} - \boldsymbol{T}_{\boldsymbol{\phi}}\otimes\boldsymbol{T}_{\boldsymbol{\phi}})^{-1}\operatorname{vec}\left(\frac{\partial\boldsymbol{R}_{\boldsymbol{\theta}}}{\partial\theta_j}\boldsymbol{R}_{\boldsymbol{\theta}} + \boldsymbol{R}_{\boldsymbol{\theta}}\frac{\partial\boldsymbol{R}_{\boldsymbol{\theta}}'}{\partial\theta_j}\right).
\end{aligned}$$

6 Simulation experiments for ARMA models

Recall that our primary goal is to improve frequentist coverage probabilities in interval prediction. For that purpose we have conducted simulation experiments to find out the benefits of the Bayesian approach especially in relation to the standard plug-in method. The latter method yields the well known intervals

$$\hat{y}_{n+h|n}(\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\beta}}) \pm z_\alpha\hat{\sigma}v_{n+h|n}(\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\beta}}), \quad \hat{\sigma} = S^2/(n-k), \tag{16}$$

see (5) and (6).

In all simulations the length of the time series is 50, and the regression part consists of the constant term $\beta_1 = \beta$ only, i.e. $\boldsymbol{X} = (1, \ldots, 1)'$. The affine linear transformation on the observations $y_i \mapsto a + cy_i$ yields the same transformation on the limits $b_\alpha \mapsto a + cb_\alpha$ in item (vi) of section 3. Therefore we can set in simulations, without loss of generality, $\sigma = 1$, and $\beta = 0$. We simulate 5000 replicates from a given ARMA process with fixed coefficients, and from each realization we estimate the parameters by maximum likelihood, and compute the prediction intervals using the plug-in method (16) as well as the Bayesian interval from the formula (7) with $N = 100$. Because the main variation in simulations is between series, the sample size in computing the prediction interval need not be large. Because in simulation we know all the parameters we can compute the frequentist conditional coverage probability

$$P(b_\alpha \leq y_{n+h} \leq b_{1-\alpha} \,|\, \boldsymbol{y}, \beta = 0, \sigma = 1, \boldsymbol{\psi}),$$

where $\boldsymbol{\psi}$ specifies the parameters used in a simulation, and the limits $b_\alpha, b_{1-\alpha}$ are fixed. Averaging these probabilities over the 5000 replications of $\boldsymbol{y}$ from the same model, gives us a good estimate of the frequentist coverage probability

$$P(b_\alpha \leq y_{n+h} \leq b_{1-\alpha} \,|\, \beta = 0, \sigma = 1, \boldsymbol{\psi}),$$

where all $y_{n+h}, b_\alpha, b_{1-\alpha}$ are random. This frequentist coverage probability is used when we compare the plug-in method and the five different Bayesian methods. The joint priors $p(\boldsymbol{\psi})/\sigma$ used in the experiment are defined through $p(\boldsymbol{\psi})$ as follows:

- Uniform prior $p(\boldsymbol{\psi}) \propto 1$.
- Approximate joint Jeffreys's prior $p(\boldsymbol{\psi}) \propto |\boldsymbol{X}'\boldsymbol{V}_{\boldsymbol{\psi}}^{-1}\boldsymbol{X}|^{\frac{1}{2}}|\boldsymbol{J}_{\boldsymbol{\psi}}|^{\frac{1}{2}}$.
- Approximate marginal Jeffreys's prior $p(\boldsymbol{\psi}) \propto |\boldsymbol{J}_{\boldsymbol{\psi}}|^{\frac{1}{2}}$.
- Exact joint Jeffreys's prior

$$p(\boldsymbol{\psi}) \propto \left|\boldsymbol{X}'\boldsymbol{V}_{\boldsymbol{\psi}}^{-1}\boldsymbol{X}\right|^{\frac{1}{2}} \left|\boldsymbol{I}_{22}(\boldsymbol{\psi}) - (2n)^{-1}\boldsymbol{I}_{21}(\boldsymbol{\psi})\boldsymbol{I}_{21}(\boldsymbol{\psi})'\right|^{\frac{1}{2}}.$$

- Exact marginal Jeffreys's prior

$$p(\boldsymbol{\psi}) \propto \left|\boldsymbol{I}_{22}(\boldsymbol{\psi}) - (2n)^{-1}\boldsymbol{I}_{21}(\boldsymbol{\psi})\boldsymbol{I}_{21}(\boldsymbol{\psi})'\right|^{\frac{1}{2}}.$$

All the five priors above are constrained onto the the stationarity and invertibility regions. Figure 1 shows the coverage probabilities of one step ahead prediction intervals for ARMA(1,1) processes with varying values of ϕ and θ. In all cases the

Bayesian methods are superior to the plug-in method, and the differences between priors are rather small. The drop in the curves occurs in the neighborhood of $\phi + \theta = 0$ which corresponds to the white noise process, i.e. the parameters are then unidentified. Also the nearly white noise processes yield unstable estimates for ϕ and θ.

The Figure 2 shows the results for the ten step ahead predictions, where again the plug-in method stays below the nominal level in all cases. On the other hand, the coverage probabilities of the Bayesian method is somewhat over the nominal level in most cases, except when the autoregressive parameter ϕ is near the bounds of the stationary region. Also the variation between different priors is somewhat larger here than in the one step ahead predictions. In most cases the uniform prior is the closest to the nominal level. The variation due to the moving average part is smaller here than in the one step ahead predictions.

In Figure 3 the coverage probabilities of ARMA(2,1) processes are shown, with varying parameter values and forecast horizon ranging from one to ten. Cases where $\phi_1 = -1.4$ correspond to alternating autocorrelation function, and in these cases coverage probabilities are usually higher than in non-alternating cases ($\phi_1 = 1.4$). Also, uniform stationary prior seems to perform slightly worse than Jeffreys's priors. Again in all cases the Bayesian methods are superior to the plug-in method. In non-alternating cases the marginal Jeffreys priors seem to give higher coverages than the joint versions, but in alternating cases the difference is negligible. Overall, Bayesian methods perform relatively well.

7 Predicting the number of Internet users

As an illustration, we apply our method to the series of the number of users logged on to an Internet server each minute over 100 minutes. The data is previously studied by Makridakis et al. (1998) and Durbin & Koopman (2001). The former authors fitted ARMA(3,0) to the differenced series, whereas the latter ones preferred ARMA(1,1) for the same series. We use here the first 84 differences for model fitting, and then compute the prediction intervals for the next 15 time points. The Akaike information criterion suggests ARMA(1,1) as the best model. The estimated ARMA coefficients are $\hat{\phi} = 0.65, \hat{\theta} = 0.49$. The additional two estimates are $\hat{\beta} = 0.84$, and $\hat{\sigma}^2 = 10.07$. The complete time series with the simulated 90% prediction intervals are shown in Figure 4, together with median estimates which are computed by setting $\alpha = 0.5$ in the Bayesian calculations. For the plug-in method, the mean is used. These simulations are based on 100,000 replicates. As the differences between exact and approximate versions of Jeffreys's prior turns out to

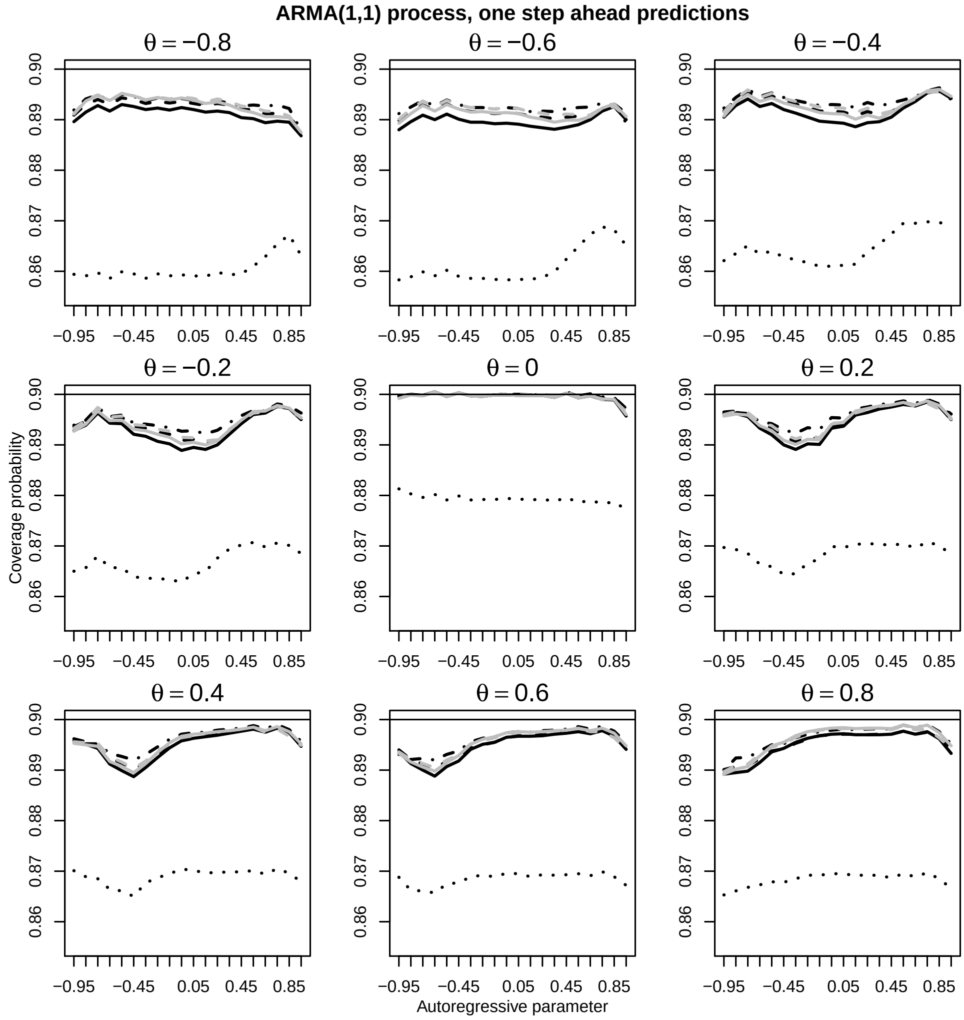


Figure 1. Coverage probabilities of one step ahead prediction intervals for ARMA(1,1) processes. The lines are: black dotted line = plug-in method, the solid black line = approximate joint Jeffreys's prior, the solid gray line = exact joint Jeffreys's prior, the dashed black line = approximate marginal Jeffreys's prior, the dashed gray line = exact marginal prior, the dot-and-dash line = uniform stationary prior.

be negligible, only approximate versions are shown. However, difference between joint and marginal priors is evident: marginal priors give substantially larger upper bounds for the prediction intervals. The upper bounds given by uniform prior is between the different Jeffreys priors, whereas the plug-in gives much smaller upper bounds than any of simulated intervals. On the lower bounds, differences are smaller.

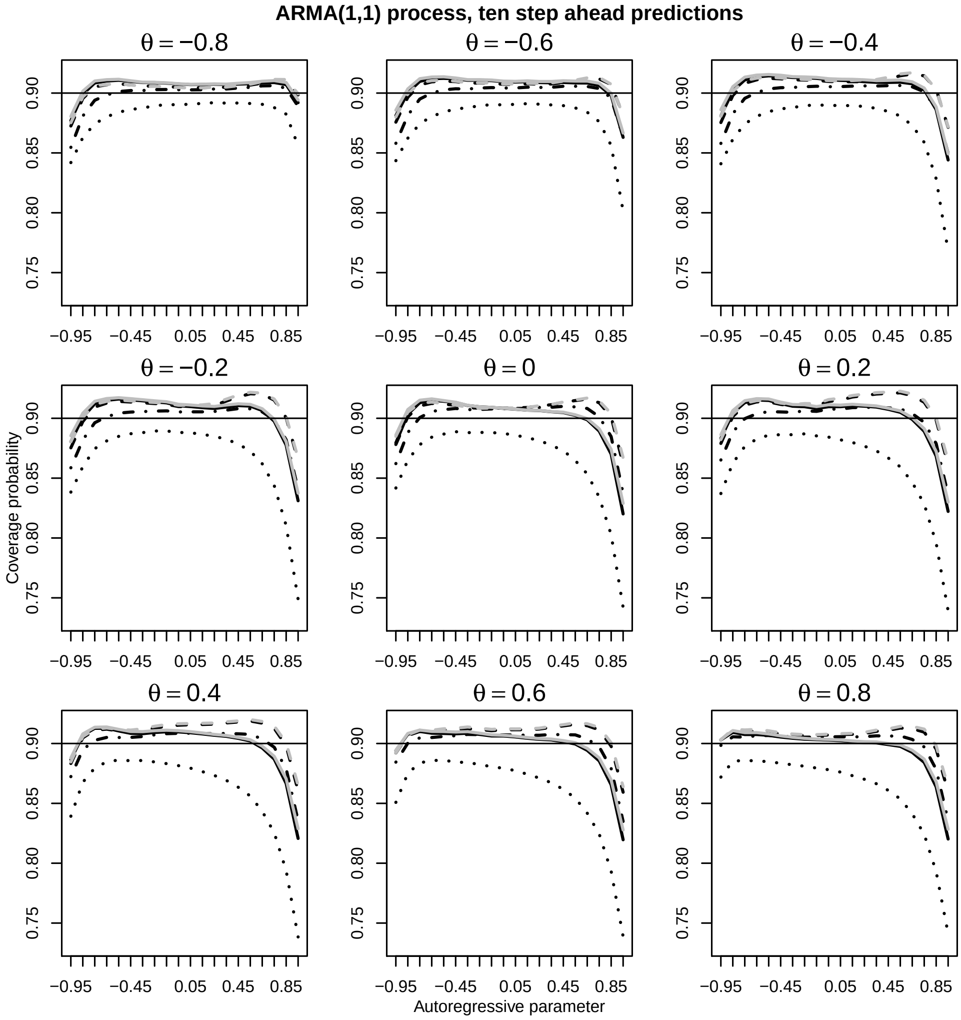


Figure 2. Coverage probabilities of ten step ahead prediction intervals for ARMA(1,1) processes. The lines are: black dotted line = plug-in method, the solid black line = approximate joint Jeffreys's prior, the solid gray line = exact joint Jeffreys's prior, the dashed black line = approximate marginal Jeffreys's prior, the dashed gray line = exact marginal prior, the dot-and-dash line = uniform stationary prior.

Given that the estimated model is correct, we can compute the average coverage probabilities of the intervals. These are given in Table 1 when the forecast horizon $h = 15$. The prediction limits and their standard errors are also given. The reported mean coverage probabilities are based on 10,000 series replicates. Within each replicate 100 values are used in (7) for the Bayesian prediction interval.

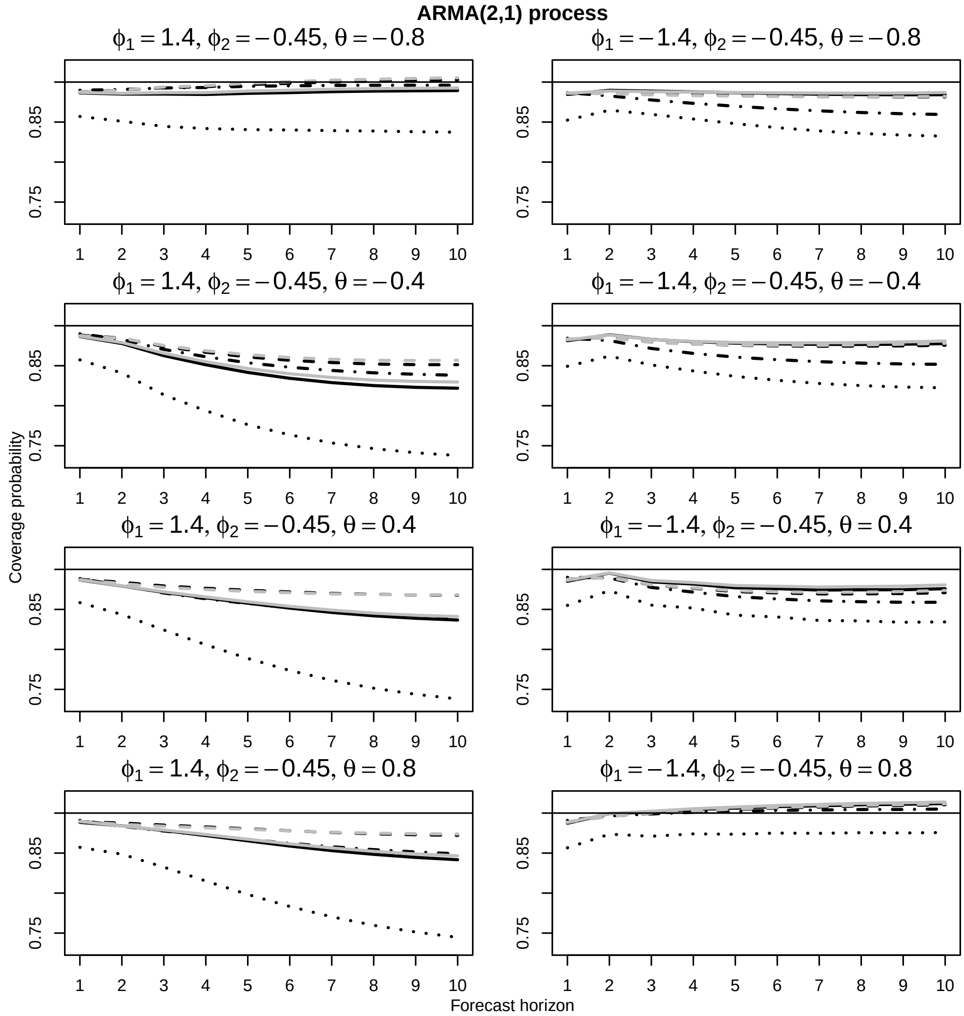


Figure 3. Coverage probabilities of the prediction intervals of varying step sizes for ARMA(1,1) processes. The lines are: black dotted line = plug-in method, the solid black line = approximate joint Jeffreys's prior, the solid gray line = exact joint Jeffreys's prior, the dashed black line = approximate marginal Jeffreys's prior, the dashed gray line = exact marginal prior, the dot-and-dash line = uniform stationary prior.

8 Discussion

In this paper we have extended the importance sampling approach presented in Helske & Nyblom (2013) from AR models to general ARMA models, and studied the effect of different prior choices on the coverage probabilities using simulated and real data. Extension of this approach to integrated ARMA models is straightforward. As may be inferred from sections 2 and 3, our method could be applied also to models outside the ARIMA framework. Compared to Markov Chain Monte

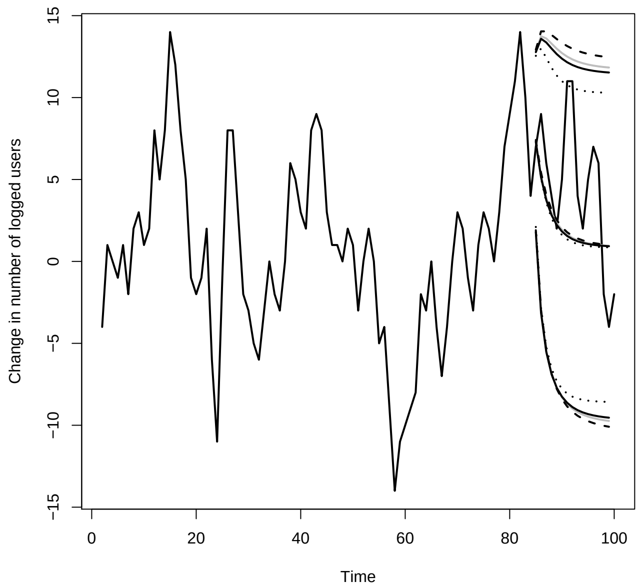


Figure 4. The prediction bands for the change of the number of users logged on to the Internet during the last 15 minutes. The lines are the black dotted line = the traditional plug-in method, the solid black line = approximate joint Jeffreys's prior, the dashed black line = approximate marginal Jeffreys's prior, the solid gray line = uniform stationary prior.

Table 1. Coverage probabilities and prediction limits for the Internet series with forecast horizon $h = 15$ and the nominal coverage probability of 0.9.

	Uniform	Joint	Marginal	Plug-in
Coverage	0.906	0.900	0.914	0.866
$\hat{b}_\alpha$	-9.73	-9.54	-10.09	-8.57
s.e.($\hat{b}_\alpha$)	0.02	0.02	0.06	–
$\hat{b}_{1-\alpha}$	11.83	11.53	12.46	10.29
s.e.($\hat{b}_{1-\alpha}$)	0.02	0.01	0.02	–

Carlo methods, we argue that method presented here is more straightforward to implement and understand, and it could also be computationally cheaper as we are only sampling the model parameters, not the future observations itself. Although we do not need to concern ourselves with the convergence problems of MCMC methods, careful checking of obtained importance weights is still needed. For example if the estimated model parameters are near the boundary of the stationary region with large variance, most of the weights can be zero due to the stationary constraint and there can be few simulated parameters with very large weights which dominate the whole sample. On the other hand, this should also be visible in the standard errors of the prediction limits, which are easily obtained during prediction interval computation.

Our simulation studies show that a simple uniform prior with stationarity and invertibility constraints performs relatively well in most cases. As the uniform prior is computationally much cheaper than the different versions of Jeffreys's prior, we feel that it could be used as a default prior in practical cases. In addition, a similar check as in section 7 regarding the average coverage probabilities can give further information on the accuracy of the adopted prior.

References

Ansley, C.F. & Kohn, R. (1986). Prediction Mean Squared Error for State Space Models With Estimated Parameters. *Biometrika* 73, 467–473.

Barndorff-Nielsen, O.E. & Cox, D.R. (1996). Prediction and Asymptotics. *Bernoulli* 2, 319–340.

Beran, R. (1990). Calibrating Prediction Regions. *Journal of the American Statistical Association* 85, 715–723.

Box, G.E.P., Jenkins, G.M. & Reinsel, G.C. (2008). *Time Series Analysis: Forecasting and Control*. Fourth edition. Hoboken: Wiley.

Chatfield, C. (1993). Calculating Interval Forecasts. *Journal of Business & Economic Statistics* 11, 121–135.

Chatfield, C. (1996). Model Uncertainty and Forecast Accuracy. *Journal of Forecasting* 15, 495–508.

Clements, M.P. & Hendry, D.F. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge: The MIT Press.

Clements, M.P. & Kim, J.H. (2007). Bootstrap Prediction Intervals for Autoregressive Time Series. *Computational Statistics & Data Analysis* 51, 3580–3594.

Durbin, J. & Koopman, S.J. (2001). *Time Series Analysis by State Space Methods*. New York: Oxford University Press.

Fuller, W.A. & Hasza, D.P. (1981). Properties of Predictors for Autoregressive Time Series. *Journal of the American Statistical Association* 76, 155–161.

Grigoletto, M. (1998). Bootstrap Prediction Intervals for Autoregressions: Some Alternatives. *International Journal of Forecasting* 14, 447–456.

Helske, J. & Nyblom, J. (2013). Improved Frequentist Prediction Intervals for Autoregressive Models by Simulation. Submitted.

Kabaila, P. & Syuhada, K. (2008). Improved Prediction Limits for AR(p) and ARCH(p) processes. *Journal of Time Series Analysis* 29, 213–223.

Kim, J.H. (2004). Bootstrap Prediction Intervals for Autoregression Using Asymptotically Mean-Unbiased Estimators. *International Journal of Forecasting* 20, 85–97.

Lin, T.I. & Ho, H.J. (2008). A simplified approach to inverting the autocovariance matrix of a general ARMA(p, q) process. *Statistics and Probability Letters* 78, 36–41.

Makridakis, S., Wheelwright, S.C. & Hyndman, R.J. (1998). *Forecasting: Methods and Applications*. Third edition. New York: Wiley.

Masarotto, G. (1990). Bootstrap Prediction Intervals for Autoregressions. *International Journal of Forecasting* 6, 229–239.

Pascual, L., Romo, J. & Ruiz, E. (2004). Bootstrap Predictive Inference for ARIMA Processes. *Journal of Time Series Analysis* 25, 449–465.

Pfeffermann, D. & Tiller, R. (2005). Bootstrap Approximation to Prediction MSE for State-Space Models With Estimated Parameters. *Journal of Time Series Analysis* 26, 893–916.

Phillips, P.C.B. (1979). The Sampling Distribution of Forecasts From a First-Order Autoregression. *Journal of Econometrics* 9, 241–261.

Quenneville, B. & Singh, A.C. (2000). Bayesian Prediction Mean Squared Error for State Space Models With Estimated Parameters. *Journal of Time Series Analysis* 21, 219–236.

Rodriguez, A. & Ruiz, E. (2009). Bootstrap Prediction Intervals in State-Space Models. *Journal of Time Series Analysis* 30, 167–178.

Vidoni, P. (2004). Improved Prediction Intervals for Stochastic Process Models. *Journal of Time Series Analysis* 25, 137–154.

Vidoni, P. (2009). A Simple Procedure for Computing Improved Prediction Intervals for Autoregressive Models. *Journal of Time Series Analysis* 30, 577–590.

III

ECONOMETRICS

ILLUSTRATING THE EFFECTS OF CROSS-SECTIONAL CORRELATION ON EVENT STUDY RESULTS: THE PRIVATE SECURITIES LITIGATION REFORM ACT OF 1995 REVISITED

Jan Antell and Timo Korkeamäki
Hanken School of Economics

1 Introduction

We study the effects of method choice on the event study results related to a well-studied regulatory change that took place in the U.S. in 1995. Fast flow of information related to a Presidential veto and its subsequent overturn, paired with a well-defined set of most affected industries, make the Public Securities Litigation Reform Act of 1995 an interesting case study. Consistent with prior simulation studies, we find that event-induced variance and cross-sectional dependence have a marked effect on event study results in a case with severe event date clustering. We also report significant differences among different methods to account for cross-sectional correlation.

Event study methods have become the work horse of empirical finance research, and much of what we know, in particular of corporate finance, is based on event study evidence. While MacKinlay (1997) notes early uses of the method already in the 1930s, the increased availability of daily stock returns since the 1970s has been an important factor in making event study the method of choice in many finance inquiries.[1] According to Kothari & Warner (2007), 212 papers in the *Journal of Finance* make use of the event study methodology between 1975 and 2000. Further, Kolari & Pynnönen (2010, online appendix) list 75 event studies in leading finance journals that have accounted for event-inflated variance and cross-sectional correlation.

As the event study methodology is providing important evidence regarding financial phenomena, it is of utmost importance for finance researchers to have a solid understanding of the strengths and the pitfalls of the method. Along with event study's popularity, a much-needed literature has grown on the methodological

[1] Fama, et al. (1969) is often viewed as the inaugural short-term event study in finance.

aspects of event studies. Thanks to those influential works, our knowledge has greatly increased over the years regarding the size and power properties of tests used in event studies, and the effects of issues such as event-induced variance and cross-sectional association between events[2]

Most of the previous studies on the statistical properties of the event study methodology make use of simulations, where the method's ability to capture a known economic effect chosen by the researcher is measured. In this paper, we have chosen a different path. We use an actual event with widespread effects on stock returns, as we re-visit a well-studied regulatory change that took place in the U.S. in 1995. In December 1995, the U.S. congress enacted on the Private Securities Litigation Reform Act (PSLRA), which would limit investors' ability to sue firms in securities fraud cases. However, President Clinton used his veto power on December 19, 1995 to overturn the legislation. The House of Representatives and the Senate subsequently voted to override the Presidential veto on December 20 and December 22, respectively. The fast flow of opposite types of information in this case provide an excellent opportunity for researchers to study the economic effects of the legislation. Indeed, a number of previous studies have considered the effects of the PSLRA on stock prices in industries that are most disposed to securities litigation, namely computers, electronics, pharmaceuticals/biotechnology, and retailing.[3]

As far as we can tell, the previous studies on the PSLRA fail to account for event-induced variance, which is of special concern in event studies with event date clustering, especially in those concerning regulatory changes, as they affect all sample firms simultaneously (Binder, 1985). The issue is an important one, as Harrington & Shrider (2007) show that problems with event-induced variance intensify in presence of cross-sectional correlation in the effects, and Kolari & Pynnönen (2010) provide evidence that even low levels of such correlation have a marked effect on inferences drawn from an event study. Obviously, in events with common event days, abundant potential sources for cross-sectional correlation exist. The use of a limited number of industry portfolios is also likely to increase the cross-sectional correlation within the sample. All three above-mentioned studies on the PSLRA use the so called portfolio method, which is suggested by Jaffe (1974) and Campbell, et al. (1997) as a solution to the cross-sectional dependence

---

[2] See e.g. Brown & Warner (1980, 1985), Boehmer, et al. (1991), and Kolari & Pynnönen (2010, 2011).

[3] For prior studies on PSLRA, see Spiess & Tkac (1997), Johnson, et al. (2000), and Ali & Kallapur (2001). For analysis of litigation risk in the aforementioned industries, see Francis, et al. (1994).

problems arising from extreme event date clustering in cases such as regulatory changes[4]. However, as Kolari & Pynnönen (2010) note, the portfolio method suffers from low power. In this study, we use the Kolari & Pynnönen (2010) Adj-BMP test statistic to account for cross-sectional correlation in the effects of the PSLRA events, and contrast our findings with those obtained from the alternative methods.

2 The Private Securities Litigation Reform Act of 1995

The main purpose of the suggested PSLRA was to limit frivolous securities fraud law suits. A bipartisan view at the time was that the balance between deterring securities fraud and assuring that the private securities litigation process was not used abusively was severely tilted, and speculative securities class action suits were common (Phillips & Miller, 1996). Class action suits on securities fraud typically originate from alleged corporate misstatements that have resulted in losses to those investors who have bought the shares at artificially high prices. Prior to the PSLRA, the plaintiffs could file a law suit without identifying any specific corporate actions that had misled the investors. In order to curb frivolous filings, the PSLRA included provisions that required plaintiffs to specify the facts behind their claims. Furthermore, if the suit was later to be found frivolous, the plaintiffs would face penalties, such as paying the other parties' attorneys' fees (Ali & Kallapur, 2001). Also, the provisions of the PSLRA would limit liability of those defendants with a limited role in the alleged misconduct (King & Schwartz, 1997), whereas previously, large claims that were disproportionate in comparison to their part of the blame had been addressed to "deep-pocket defendants", such as accounting firms.

A common problem with event studies into law reforms is that regulatory changes take time, and their content tends to be intensely discussed, both publicly and privately, during the process. The PSLRA makes no exception to this, as Avery (1996) points to a "long and winding road" towards the proposal. Ali & Kallapur (2001) consider 18 legislative events during 1995, leading into the Presidential veto on December 20. Their portfolio of high litigation firms exhibits statistically significant returns on only six of those days. The relative lack of findings during

[4] A simple way to account for event-induced variance in such setting would be to use heteroskedasticity-robust standard errors in hypothesis testing. However, Harrington & Shrider (2007) report that such consideration was not common at the time. Accordingly, none of the three previous empirical inquiries into the stock reactions upon the PSLRA mentions any corrections for heteroskedasticity.

the progress towards the PSLRA is likely to be due to the gradual learning of the regulation among market participants. However, the events around the Presidential veto in late December 1995 are likely to be less perfectly anticipated, and the significant event study evidence that is reported in prior studies on the reform supports that view.

While the sudden changes in the legislative process provide an interesting laboratory to study the economic effects of the PSLRA reform, some of the opposite effects unfortunately coincide, which challenges the clean identification in our study, and the previous studies alike. After the Congressional approval of the PSLRA, President Clinton received the bill on December 6, 1995. Clinton would have to sign the bill by December 19, 1995, or it would automatically become law. While the President was initially prepared to sign the bill, at the end, he vetoed it less than one hour before the deadline on December 19, 1995 (Johnson, et al., 2000). The House overrode his veto the next day, so that information regarding both the Presidential veto and the House override reached the market on the same trading day. However, rumors of eventual veto surfaced already on December 18, 1995, for which reason we follow prior studies into the PSLRA, and consider that date as our first event date[5]. All three previous studies on the reform find the December 20, 1995 abnormal returns to be positive and significant in the sample of high litigation industries, whereas December 18, 1995 abnormal returns are systematically negative.

3 Estimation and testing

When testing for event effects, the test statistic should account at least for (A) cross-sectional heteroskedasticity, (B) event-induced variance inflation, and (C) cross-sectional correlation. In the following treatment, we start by defining the models for abnormal returns and mean abnormal returns, and continue by stating three models for testing event effects in an increasing order of complication. Later, we also consider two additional models for reference.

We use continuously compounded abnormal returns in excess return format. They are defined as

[5] Both Johnson, et al. (2000), and Spiess & Tkac (1997) mention that on Friday, December 15, President Clinton had a dinner with William Lerach, who was not only a well-known lawyer behind numerous securities fraud cases, and a vocal opponent of the PSLRA, but also a significant donor to the Democratic party. They speculate that the dinner affected Clinton’s view on the PSLRA.

$$AR_{it}^{e} = R_{it}^{e} - \mathrm{E}\left(R_{it}^{e}\right), \quad i = 1,\dots,N, \quad t = 1,\dots,T_1,\dots,\tau,\dots,T_2, \tag{1}$$

where AR_{it}^{e} is the abnormal return in excess of the risk-free rate of return for event i, R_{it}^{e} is the observed excess return, and $\mathrm{E}(R_{it}^{e})$ is the expected return. The estimation window ends at T_1, the event day is given by τ, and the whole sample period ends at T_2. We estimate the expected return with two alternative specifications, i.e., the Fama–French (1993) three-factor model (henceforth FF3) and the traditional one-factor market model:

$$E\left(R_{it}^{e}\right) = \alpha + \beta_{i,M} R_{mt}^{e} + \beta_{i,SMB} R_{SMB,t} + \beta_{i,HML} R_{HML,t}, \quad t = 1,\dots,T_1, \tag{2}$$

$$E\left(R_{it}^{e}\right) = \alpha + \beta_{i,M} R_{mt}^{e}, \quad t = 1,\dots,T_1, \tag{3}$$

where R_{mt}^{e} is the excess return on the market, R_{SMB} is the return on a portfolio with a long position in small company stocks and a short position in large-cap stocks, and R_{HML} is the return on a portfolio with a long position in high book-to-market stocks and a short position in low book-to-market stocks.

The effect of the event can be measured by the average abnormal return on the event day (day τ=0), which is defined as

$$AAR_0 = \frac{1}{N}\sum_{i=1}^{N} AR_{i0}\,. \tag{4}$$

Another measure is given by the average standardized abnormal return, defined as

$$ASAR_0 = \frac{1}{N}\sum_{i=1}^{N} \frac{AR_{i0}}{s_i}, \tag{5}$$

where s_i is the time-series standard deviation over the estimation period. We prefer the specification in Equation (5), as standardizing alleviates problems with cross-sectional heteroskedasticity (A). The economic effect of the event should be assessed from the easy-to-understand average abnormal return, while the average standardized abnormal return is used for statistical significance testing.

The first statistic is the classic test calculating the standard error from the estimation period, in our paper labelled the *standardized residuals test* (*SRT*):

$$SRT = \frac{ASAR_{0,\mathrm{PC}}}{\frac{1}{N}\sqrt{\sum_{i=1}^{N}\left(s_{i,Pat}^{2}\right)}}, \tag{6}$$

where $s_{i,Pat}^2$ is an estimate for the time-series standard deviation of the estimation period abnormal returns, incorporating a degrees of freedom correction as in Patell (1976) that stems from using an estimated standard deviation instead of the true standard deviation. Without the correction the denominator of Equation (6) simplifies to $1/\sqrt{N}$. In $ASAR_{0,\mathrm{PC}}$, a prediction error correction has been made on the individual standardized abnormal returns. The correction arises from using regression parameters from the estimation period on predicting the normal return in the event period (see for example Campbell, et al. (1997), p. 159). Generally, the corrections are fairly small. Following the Lindeberg–Feller central limit theorem, *SRT* follows asymptotically the standard normal distribution**.**

While *SRT* accounts for cross-sectional heteroskedasticity (A), the statistic overstates the significance in presence of an event-induced increase in variance (B). To alleviate the problem, Boehmer, Musumeci and Poulsen (1991) combine the standardized residuals test and the cross-sectional approach to a statistic we label *BMP*:

$$BMP = \frac{ASAR_{0,PE}}{s_{ASAR_{0,PE}} / \sqrt{N}}, \tag{7}$$

where $ASAR_{0,PE}$ is corrected with the prediction error correction, and $s_{ASAR_{0,PE}}$ is an estimate of the cross-sectional standard deviation. *BMP* is *t*-distributed with *N*–1 degrees of freedom.

Neither *SRT* nor *BMP* accounts for cross-sectional correlation (C) among the abnormal returns, which is likely to plague evaluations concerning individual industries or studies that exhibit event-day clustering. Kolari & Pynnönen (2010) propose a correction to the BMP test that adjusts for cross-sectional correlation. We label the test KP, while Kolari & Pynnönen (2010) denote it as Adj–BMP:

$$KP = \frac{ASAR_{0,PE} \cdot \sqrt{N}}{s_{ASAR_{0,PE}} \cdot \left(1 + (N-1)\bar{r}\right)} = BMP \cdot \sqrt{\frac{1-\bar{r}}{1+(N-1)\bar{r}}}, \tag{8}$$

where $\bar{r}$ is the average correlation among the abnormal returns over the estimation period. If $\bar{r}$ is zero, KP simplifies back to BMP. The KP test is robust against all malign properties (A), (B) and (C) we mention above. *KP* is *t*-distributed with *N*–1 degrees of freedom.

For reference, we also use the crude dependence adjustment (CDA) set forth by Brown & Warner (1980, 1985):

$$CDA = \frac{ASAR_0}{\sqrt{\frac{1}{T_1 - 1}\sum_{t=1}^{T_1}\left(ASAR_t - \frac{1}{T_1}\sum_{t=1}^{T_1} ASAR_t\right)^2}}. \tag{9}$$

The standard error of *CDA* is simply the standard deviation of the time-series of average standardized abnormal returns over the estimation period. While this test accounts for cross-sectional correlation, it fails to account for event induced variance. *CDA* is t-distributed with T_1–1 degrees of freedom.

Finally, we also consider the portfolio method and use a regression model. The model is given by

$$R^e_{p,t} = \alpha + \beta R^e_{m,t} + \sum_{j=1}^{3}\gamma_j D_{j,t} + \varepsilon_{p,t}, \quad t = 1, \ldots, T_2, \tag{10}$$

where $R^e_{p,t}$ is the continuously compounded excess return on an equally-weighted portfolio over the whole sample period, $R^e_{m,t}$ the corresponding excess return on the market portfolio, and $D_{j,t}$ are indicator variables that take the value one on each of the event days and zero otherwise. The dates are defined in the next section. The parameters γ_j capture the event effects, which we then test both using OLS standard errors, and standard errors that are robust to heteroskedasticity and autocorrelation as in Andrews (1991), labelled *HACSE*. With event dates heavily clustered in calendar time, the *CDA* test, in which abnormal returns are estimated first, and the significance of the average abnormal return is then tested, is fairly similar to the portfolio method, in which the portfolio of returns is formed first, and the abnormal returns are then extracted. The main difference in our application is that the former uses standardized abnormal returns while the latter is based on non-standardized returns.

4 Data and results

In accordance with previous studies on the PSLRA, we use CRSP as our data source for daily stock returns. The factor returns for the Fama–French three-factor model are retrieved from Kenneth French’s online data library. We limit our analysis to industries that are indicated as having high litigation risk in Francis, et al. (1994), namely computers (SIC codes 3570-3577 and 7370-7374), electronics (SIC codes 3600-3674), pharmaceuticals/biotechnology (SIC codes 2833-2836 and 8731-8734), and retailing (SIC codes 5200-5961). In the tests where firm-level abnormal returns are estimated, our estimation period starts on 3 January

1995 and ends 1 December 1995, and it thus comprises of 233 observations. There are slight variations between our paper and the prior studies on the PSLRA, regarding both estimation period and sample selection. In applying the portfolio method, the prior studies include the returns for the entire calendar year of 1995, in other words the time after the events is also included. We follow their choice in our portfolio tests. As the portfolio returns are tabulated on a day-by-day basis, Ali & Kallapur (2001) use a sample that varies in size throughout the estimation period due to missing returns. On the other hand, Johnson, et al. (2000) require their firms to "have a complete 1995 daily returns data". As we estimate abnormal returns for each individual firm separately, we require a minimum of 50 observations during the estimation period, with no missing values during the last ten days of the estimation window. Nevertheless, by using these filters, we retain a sample that is very similar in size to the prior studies. In comparison to Ali & Kallapur (2001), whose range of number of firms per industry is indicated in parentheses, our sample consists of 562 (492-579 in Ali & Kallapur, 2001) firms in computers, 472 (430-484) firms in electronics, 74 (74-79) firms in pharmaceuticals, and 434 (441-450) firms in retail.[6] Based on the flow of events around the Presidential veto and its overturn, as described in Section 2, we consider the following event dates: 18 December 1995, 20 December 1995 and 22 December 1995.

Table 1 shows the descriptive statistics for equally-weighted portfolios for each of the four industries, and for the combined whole sample portfolio, as well as for the value-weighted market portfolio. The time period is 3 January 1995 to 29 December 1995, combining the estimation period, the event-dates and the post-event period, and thus totaling 252 observations. The returns are continuously compounded excess returns in percentage format. The annualized mean returns are fairly high, ranging from 13.0 percent to 52.2 percent and averaging at 38.8 percent, being almost twice the market return of 19.8 percent. The volatilities are fairly low, between 7.6 percent and 13.0 percent. Note, however, that the volatilities for individual companies are much larger on average (not reported). The values for skewness and excess kurtosis suggest that the daily returns are not normally distributed. In fact, Jarque-Bera (1987) tests for normality (not reported) reject the null hypothesis of normally distributed returns for all series.

[6] Ali & Kallapur (2001) note that Johnson et al. (2000) include SIC code 2830 in their pharmaceutical sample, which results in a significantly larger set of 191 firms. They further speculate that the difference in sample selection accounts for the fact that Johnson et al. (2000) findings regarding the pharmaceutical industry's reaction to the Senate override on 12/22/1995 deviate from those reported in Ali & Kallapur (2001) and Spiess & Tkac (1997). Johnson, et al. (2000) also separate Hardware and Software firms from the set that is very similar to our Computers sub-sample, and they do not consider the Retail industry in their analysis.

Table 1. Descriptive Statistics. The time period is 3 January 1995 to 29 December 1995. All returns are continuously compounded, in percentage format and in excess of the risk-free rate of return. The industry portfolios are equally-weighted while the market portfolio is value-weighted.

$T = 252$	Mean (annualized %)	Volatility (annualized %)	Minimum (in percent)	Maximum (in percent)	Skewness	Excess kurtosis
Computers	52.22	13.00	-3.95	1.80	-1.53	4.86
Electronics	47.97	12.59	-3.40	2.12	-1.30	3.40
Pharmaceuticals	32.65	10.19	-3.27	2.04	-0.75	3.20
Retail	13.04	7.62	-1.91	1.15	-0.59	0.98
Whole Sample	38.84	10.33	-3.03	1.50	-1.45	4.23
Market portfolio	19.75	7.42	-1.78	1.35	-0.42	1.44

We begin analyzing our data by having a first look at the time series of volatility in our sample. Figure 1 shows the cross-sectional standard deviation of the standardized abnormal returns based on the Fama-French three-factor model, and the five-day rolling standard deviation of the returns for our equally-weighted whole sample portfolio. Figure 1 clearly indicates that the time series of our sample exhibits heteroskedasticity. Part of that heteroskedasticity is likely due to the increased cross-sectional dependence during the event period, also indicated in Figure 1. Figure 1 thus serves as a further motivation for re-visiting the stock reactions to the PSLRA.

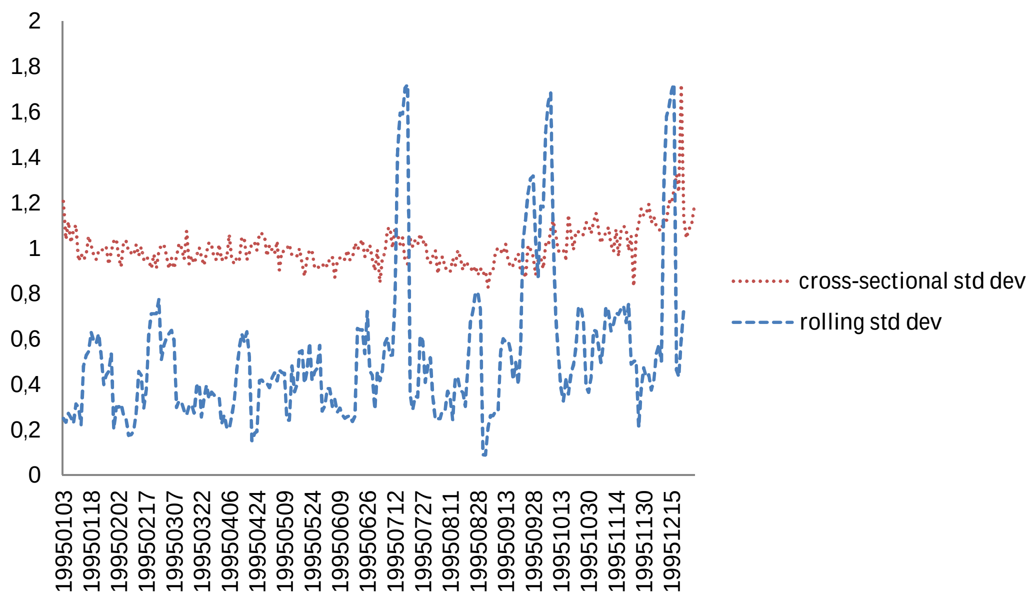


Figure 1. Daily standard deviation of the cross section and the 5-day rolling time series

Table 2. Tests for event effects on three event days for four industries. FF3 refers to the Fama–French (1993) three-factor model (2) and Market to the one-factor market model (3). SRT is the standardized residuals test (6), BMP the Boehmer et al. (1991) test (7), KP the Kolari & Pynnönen (2010) test (8), and CDA the crude dependence adjustment test (9). Pf. Method refers to the model indicated in Equation (10). All tests are t-tests. HACSE is the Andrews (1991) heteroskedasticity and autocorrelation consistent estimator for the standard error. * refers to statistical significance at 5 percent level, and ** refers to 1 percent.

Panel *A*. Computers	18 Dec 1995		20 Dec 1995		22 Dec 1995	
N = 562	FF3	Market	FF3	Market	FF3	Market
Abnormal return	-0.731	-1.908	-0.093	1.498	0.162	0.540
Std abnormal return	-0.122	-0.428	-0.056	0.338	0.007	0.103
SRT	-2.868**	-10.088**	-1.320	7.979**	0.170	2.430*
BMP	-2.230*	-7.913**	-1.031	6.279**	0.161	2.330*
KP	-1.424	-2.392*	-0.644	1.898	0.103	0.704
CDA	-1.789	-3.184**	-0.819	2.503*	0.108	0.765
Pf. Method OLS		-3.335**		2.873**		1.068
Pf. Method OLS HACSE		-7.132**		15.732**		14.080**
Panel *B*. Electronics	18 Dec 1995		20 Dec 1995		22 Dec 1995	
N = 472	FF3	Market	FF3	Market	FF3	Market
Abnormal return	0.198	-0.877	-0.584	0.906	-0.224	0.126
Std abnormal return	0.113	-0.176	-0.110	0.276	-0.032	0.058
SRT	2.452*	-3.802**	-2.384*	5.960**	-0.682	1.263
BMP	2.012*	-2.261*	-2.009*	5.130**	-0.606	1.144
KP	1.152	-1.084	-1.150	1.706	-0.347	0.380
CDA	1.312	-1.317	-1.284	2.050*	-0.384	0.436
Pf. Method OLS		-1.608		1.666		0.324
Pf. Method OLS HACSE		-3.735**		9.566**		4.988**
Panel *C*. Pharmaceuticals	18 Dec 1995		20 Dec 1995		22 Dec 1995	
N = 74	FF3	Market	FF3	Market	FF3	Market
Abnormal return	-1.894	-2.378	-0.075	0.764	0.168	0.350
Std abnormal return	-0.381	-0.522	-0.037	0.170	0.193	0.240
SRT	-3.265**	-4.466**	-0.313	1.451	1.653	2.054*
BMP	-2.768**	-3.658**	-0.228	0.989	1.555	1.945
KP	-2.135*	-2.674*	-0.177	0.723	1.200	1.422
CDA	-2.231*	-3.064**	-0.220	1.029	1.235	1.461
Pf. Method OLS		-4.066**		1.387		0.667
Pf. Method OLS HACSE		-14.856**		12.470**		9.821**
Panel *D*. Retail	18 Dec 1995		20 Dec 1995		22 Dec 1995	
N = 434	FF3	Market	FF3	Market	FF3	Market
Abnormal return	-0.456	-0.889	-0.088	0.859	0.117	0.310
Std abnormal return	-0.097	-0.225	-0.077	0.197	-0.054	0.003
SRT	-2.019*	-4.662**	-1.592	4.088**	-1.121	0.061
BMP	-1.665	-3.860**	-1.281	3.375**	-0.917	0.051
KP	-1.127	-1.954	-0.867	1.709	-0.621	0.026
CDA	-1.284	-2.326*	-1.014	2.109*	-0.747	0.032
Pf. Method OLS		-2.179**		2.427*		1.131
Pf. Method OLS HACSE		-7.672**		20.501**		16.136**

Panel *E*. Whole sample	18 Dec 1995		20 Dec 1995		22 Dec 1995	
N = 1542	FF3	Market	FF3	Market	FF3	Market
Abnormal return	-0.425	-1.328	-0.241	1.100	0.030	0.338
Std abnormal return	-0.055	-0.298	-0.077	0.271	-0.013	0.067
SRT	-2.159*	-11.641**	-3.022**	10.580**	-0.520	2.632**
BMP	-1.744	-9.494**	-2.420*	8.585**	-0.461	2.365*
KP	-0.926	-2.227*	-1.285	2.014*	-0.245	0.555
CDA	-1.135	-2.875**	-1.588	2.613**	-0.273	0.650
Pf. Method OLS		-3.023**		2.719**		0.961
Pf. Method OLS HACSE		-6.809**		15.355**		13.647**

Table 2 shows the results both for the individual industries and for the whole sample. Both non-standardized and standardized abnormal returns are reported. In the interest of space, we only report test statistics that refer to the standardized tests. Along with the *t*-statistics for the *SRT*, *BMP*, *KP*, and *CDA* tests, we also report the *t*-statistics obtained from the regressions of portfolio time series, using Equation (10).

Several interesting observations arise from Table 2. Looking at the whole sample, the results for 18 December are consistently negative and statistically significant across different methods for the market model. This is consistent with the prior studies mentioned above, with one exception. While Ali & Kallapur (2001) also find a negative reaction to the December 18 veto rumors when observing the "conventional *p*-value", they note the Jain (1986) finding that significance levels from the market model tend to be overstated when the market return is large on the event date. They mitigate the problem, by using a randomization method, which results in a *p*-value of only 0.16.

Our findings regarding December 20, 1995 indicate a positive and significant stock reaction when using the market model. Variability exists among industries and the Pharmaceuticals exhibit significant results only if the *Portfolio HACSE* method is considered. Finally, the December 22 results are weaker, albeit statistically significant with a number of samples and test statistics, such as the *SRT* and the *BMP* tests for the whole sample. Weaker findings are consistent with prior studies, as Johnson, et al. (2000) is the only one of the three studies to report statistically significant findings for the whole sample on that date.

Another interesting observation can be made regarding the differences between the market model and the FF3 results. In numerous instances, the two methods provide opposite findings. For example, the December 20 findings for the whole sample are positive and significant when the market model is used, but when the Fama-French factors are included in the estimation of abnormal returns, the results turn negative, and in the case of *SRT* and *BMP* statistics even statistically significant. It is also interesting to note the average correlation used in the *KP* test. While the average correlation when using the FF3 model for the whole sample is 0.00165 (not reported), it is 0.01102 for the market model. This explains the larger corrections to the *t*-values when using the market model instead of the FF3 model.

Finally, we also observe large and seemingly systematic variation across test statistics. First, in comparison to the standardized residual test, the results tend to get statistically weaker when event-induced variance is accounted for, using the *BMP* method. However, the effect is generally moderate, which is somewhat surprising, given the large shift in the standard deviation around the event days, indicated in Figure 1. When the cross-sectional dependence is further accounted for by using the *KP* method, the *t*-values clearly decrease. This is intuitive, as the cross-sectional dependence may exist within an industry that is affected by the PSLRA, even after controlling for market-wide effects. The results for *CDA*, also accounting for cross-sectional correlation, are in line with those of *KP*. The correction is not as strong as the CDA test does not account for event-inflated variance. In a case such as ours, the critical question is whether individual firms should be considered as independent observations, or if part of the cross-sectional dependence is represented by the industry reaction to the news.

It is also worth noting that while the portfolio method with standard OLS gives results that are very comparable to the traditional abnormal return results, accounting for heteroskedasticity and autocorrelation with the *HACSE* correction appears to boost the statistical significance to an extreme.[7] Given the popularity of the portfolio method, as indicated by Kolari & Pynnönen (2010, online appendix), the inconsistency between the *BMP* adjustment to the traditional abnormal returns, and the *HACSE* correction to portfolio abnormal returns is interesting.

7 Conclusions

In this paper we study the effects of the Private Securities Litigation Reform Act (PSLRA), a regulatory change that took place in the U.S. in 1995, on four industries inclined to be affected by the reform. Despite some minor differences in sample selection procedures between our paper and the prior studies on the stock reactions to the PSLRA, our evidence is very similar to the previous results when

[7] When we use the White (1980) standard errors only, instead of accounting also for autocorrelation with HACSE, we obtain *t*-statistics of similar magnitude (not reported). Note that Harrington & Shrider (2007) advocate the use of the regression method with White standard errors to account for event induced variance.

either the traditional event study methodology or the portfolio method with regular OLS *t*-statistics is used. We show, however, that the results are very sensitive to the choice of return generating model, and the choice between the market model and the FF3 model can even result in a change of the sign of the coefficient. We also show that correcting for event-inflated variance with the *BMP* method has a moderate effect on the significance of the results. However, when correcting for cross-sectional association between the abnormal returns as in Kolari & Pynnönen (2010), the significance of the results clearly decreases, showing that failing to account for cross-sectional effects may lead to spurious conclusions. The use of standard errors that are robust to heteroskedasticity and autocorrelation in the portfolio setting results in a very large upward shift in *t*-statistics. This is a puzzling result, and calls for further research.

Finally, we would like to congratulate Seppo Pynnönen on his 60th birthday. All the best!

References

Ali, A. & Kallapur, S. (2001). Securities price consequences of the Private Securities Litigation Reform Act of 1995 and related events. *Accounting Review* 76, 431–460.

Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Avery, J. (1996). Securities litigation reform: The long and winding road to Private Securities Litigation Reform Act of 1995. *The Business Lawyer* 51, 335–378.

Binder, J. (1985). Measuring the effects of regulation with stock price data. *Rand Journal of Economics* 16, 167–183.

Boehmer, E., Musumeci, J. & Poulsen, A. (1991). Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics* 30, 253–272.

Brown, S.J. & Warner, J.B. (1980). Measuring security price performance. *Journal of Financial Economics* 8, 205–258.

Brown, S.J. & Warner, J.B. (1985). Using daily stock returns: The case of event studies. *Journal of Financial Economics* 14, 3–31.

Campbell, J.Y., Lo, A.W. & MacKinlay, A.C. (1997). The econometrics of financial markets. Princeton, NJ.: Princeton University Press.

Fama, E., Fisher, L., Jensen, M. & Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review* 10, 1–21.

Fama, E. & French, K.R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3–56.

Francis, J., Philbrick, D. & Schipper, K. (1994). Shareholder litigation and corporate disclosures. *Journal of Accounting Research* 32, 137–164.

Harrington, S.E. & Shrider, D.G. (2007). All events induce variance: Analyzing abnormal returns when effects vary across firms. *Journal of Financial and Quantitative Analysis* 42, 229–256.

Jaffe, J.F. (1974). The effect of regulatory changes on insider trading. *Bell Journal of Economics and Management Science* 5:1, 93-121.

Jain, P. (1986). Analyses of the distribution of security market model prediction errors for daily returns data. *Journal of Accounting Research* 24, 76–96.

Jarque, C.M. & Bera, A.K. (1987). A test for normality of observations and regression residuals. *International Statistics Review/Revue Internationale de Statistique* 55, 163–172.

Johnson, M., Kasznik, R. & Nelson, K.(2000). Shareholder wealth effects of the Private Securities Litigation Reform Act of 1995. *Review of Accounting Studies* 5, 217–233.

King, R.R. & Schwartz, R. (1997). The Private Securities Litigation Reform Act of 1995: A discussion of three provisions. *Accounting Horizons* 11, 92–106.

Kolari, J.W. & Pynnönen, S. (2011). Nonparametric rank tests for event studies. *Journal of Empirical Finance* 18, 953–971.

Kolari, J.W. & Pynnönen, S. (2010). Event study testing with cross-sectional correlation of abnormal returns. *Review of Financial Studies* 23, 3996–4025.

Kolari, J. W. & Pynnönen, S. (2010). Online appendix to accompany event study testing with cross-sectional correlation of abnormal returns. *Review of Financial Studies* 23.

Kothari, S.P. & Warner, J.B. (2007). Econometrics of event studies. In Eckbo, B.E. (ed.), *Handbooks of Corporate Finance: Empirical Corporate Finance*, Elsevier/North-Holland.

MacKinlay, A.C. (1997). Event studies in economics and finance. *Journal of Economic Literature* 35, 13–39.

Patell, J.M. (1976). Corporate forecasts of earnings per share and stock price behaviour: Empirical tests. *Journal of Accounting Research* 14, 246–276.

Phillips, R.M. & Miller, G.C. (1996). The Private Securities Litigation Reform Act of 1995: Rebalancing litigation risks and rewards for class action plaintiffs, defendants and lawyers. *The Business Lawyer* 51, 1009–1069.

Spiess, K. & Tkac, P. (1997). The Private Securities Litigation Reform Act of 1995: The stock market casts its vote. *Managerial and Decision Economics* 18, 545–561.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.

A SIGN TEST OF CUMULATIVE ABNORMAL RETURNS IN EVENT STUDIES BASED ON GENERALIZED STANDARDIZED ABNORMAL RETURNS

Terhi Hagnäs
University of Vaasa

1 Introduction

Researchers use event study methods to measure stock price reactions against events and many event studies rely on parametric test statistics. Standardized parametric event study tests presented by Patell (1976) and Boehmer et al. (1991) (BMP) have been more popular than conventional nonstandardized tests in testing abnormal security price performance, because of their better power properties. Harrington & Shrider (2007) have argued that, in short-horizon testing of mean abnormal returns, tests that are robust against cross-sectional variation in the *true* abnormal return should always be used. They have found that the BMP test statistic is a good candidate for a robust parametric test in conventional event studies.[1] Although many event studies rely on parametric test statistics, a disadvantage of parametric statistics is that they embody detailed assumptions about the probability distribution of returns. Nonparametric statistics do not usually require as stringent assumptions about return distributions as parametric tests. (e.g., Cowan (1992)).

The sign tests are nonparametric tests often used in event studies. Also nonparametric procedures like the sign tests can be misspecified, if an incorrect assumption about the data is imposed. For example Brown & Warner (1980) and Brown & Warner (1985), and Berry et al. (1990) have demonstrated that a sign test assuming an excess return median of zero is mis-specified. Corrado & Zivney (1992) have introduced a sign test based on standardized excess returns that does not assume a median of zero, but instead uses a sample excess return median to calculate the sign of an event date excess return. The results of simulation experiments presented in Corrado & Zivney (1992) indicate that their sign test provides reliable and well-specified inferences in event studies. They have also reported that their version of

[1]Conventional event studies are defined as those focusing only on mean stock price effects. Other types of event studies include (for example) the examination of return variance effects (Beaver (1968) and Patell (1976)), trading volume (Beaver (1968) and Campbell & Wasley (1996)), accounting performance (Barber & Lyon (1997)) and earnings management procedures (Dechow et al. (1995) and Kothari et al. (2005)).

the sign test is better specified than the ordinary t-test and has a power advantage over the ordinary t-test in detecting small levels of abnormal performance.

The parametric tests derived by Patell and BMP can be applied to testing cumulative abnormal returns (CARs) over multiple day windows. Corrado & Zivney (1992) have derived the sign test only for testing one-day abnormal returns (ARs). Kolari & Pynnönen (2011) have derived a nonparametric rank test of CARs, which is based on generalized standardized abnormal returns (GSARs). They have found that their rank test has superior (empirical) power relative to popular parametric tests both at short and long CAR-window lengths. Their test statistic has also been shown to be robust to abnormal return serial correlation and event-induced volatility. Kolari & Pynnönen (2011) have also suggested that GSARs derived by them can be used to extend the sign test in Corrado & Zivney (1992) for testing CARs. Hence, in an effort to overcome previous pitfalls in the test statistics, and thereby provide more powerful test methods for common practice in event studies, we present a new sign test statistics (SIGN-GSAR-T) and its modified version (SIGN-GSAR-Z) based on GSARs. These statistics can be used equally well for testing simple day ARs and CARs.

Cowan (1992) has also derived a sign test for testing CARs and his test is called generalized sign test. The generalized sign test compares the proportion of positive ARs around an event to the proportion from a period unaffected by the event. In this way the generalized sign test takes account of a possible asymmetric return distribution under the null hypothesis. Cowan (1992) has reported that the generalized sign test is well specified for event windows of one to eleven days. He has also reported that the test is powerful and becomes relatively more powerful as the length of the CAR-window increases.

In empirical simulations, the new sign test statistics presented in this paper are compared with the generalized sign test derived by Cowan (1992), the rank test derived by Kolari & Pynnönen (2011) as well as the parametric tests derived by Patell and BMP, and the ordinary t-test. Our results show that especially the test statistic SIGN-GSAR-T has several advantages over previous testing procedures. First, it is robust against a certain degree of cross-correlation caused by event day clustering. For example, according to Kolari & Pynnönen (2010) it is well known that event studies are prone to cross-sectional correlation among abnormal returns when the event day is the same for sample firms. For this reason the test statistics cannot assume independence of abnormal returns. They have also shown that even when cross-correlation is relatively low, event-date clustering is serious in terms of over-rejecting the null hypothesis of zero average abnormal returns, when it is true. Also we report that when the event-dates are clustered, all the examined test statistics, except the test statistic SIGN-GSAR-T and the rank test derived by Kolari

& Pynnönen (2011), over-reject the null hypothesis both for short and long CAR-windows. Second, the test statistic SIGN-GSAR-T seems to be robust to the event-induced volatility. Third, it proves to have also good empirical power properties. Thus, the SIGN-GSAR-T test procedure makes available a nonparametric test for general application to the mainstream of event studies.

2 The sign of the GSAR

In forthcoming theoretical derivations, the following explicit assumption is made:

Assumption 1 *Stock returns r_{it} are weak white noise continuous random variables with*

$$\begin{aligned} E[r_{it}] &= \mu_i \textit{ for all } t, \\ \text{var}[r_{it}] &= \sigma_i^2 \textit{ for all } t, \\ \text{cov}[r_{it}, r_{is}] &= 0 \textit{ for all } t \neq s, \end{aligned} \tag{1}$$

and where i refers to the i^{th} stock and t and s are time indexes.

Let AR_{it} represent the abnormal return of security i on day t, and let day $t = 0$ indicate the event day.[2] The days $t = T_0 + 1, T_0 + 2, \ldots, T_1$ represent the estimation period days relative to the event day, and the days $t = T_1 + 1, T_1 + 2, \ldots, T_2$ represent event window days, again relative to the event day. Furthermore L_1 represents the estimation period length and L_2 represents the event period length. Standardized abnormal returns are defined as

$$\text{AR}'_{it} = \text{AR}_{it}/S(\text{AR}_i), \tag{2}$$

where $S(\text{AR}_i)$ is the standard deviation of the regression prediction errors in the abnormal returns computed as in Campbell et al. (1997).

The cumulative abnormal return (CAR) from day τ_1 to τ_2 with $T_1 < \tau_1 \leq \tau_2 \leq T_2$ is defined as

$$\text{CAR}_{i,\tau_1,\tau_2} = \Sigma_{t=\tau_1}^{\tau_2} \text{AR}_{it}, \tag{3}$$

and the time period from τ_1 to τ_2 is often called a CAR-window or a CAR-period. Then the corresponding standardized cumulative abnormal return (SCAR) is defined as

$$\text{SCAR}_{i,\tau_1,\tau_2} = \frac{\text{CAR}_{i,\tau_1,\tau_2}}{S(\text{CAR}_{i,\tau_1,\tau_2})}, \tag{4}$$

[2]There are different ways to define the abnormal returns (AR_{it}). One quite often used method is to use market model to estimate the abnormal returns. In section 4 we present how the abnormal returns can be calculated with the help of the market model.

where $S(\text{CAR}_{i,\tau_1,\tau_2})$ is the standard deviation of the CARs adjusted for forecast error (see Campbell et al. (1997)). Under the null hypothesis of no event effect both AR'_{it} and $\text{SCAR}_{i,\tau_1,\tau_2}$ are distributed with mean zero and (approximately) unit variance.

In order to account for the possible event-induced volatility Kolari & Pynnönen (2011) re-standardize the SCARs like Boehmer et al. (1991) with the cross-sectional standard deviation to get re-standardized SCAR

$$\text{SCAR}^*_{i,\tau_1,\tau_2} = \frac{\text{SCAR}_{i,\tau_1,\tau_2}}{S(\text{SCAR}_{\tau_1,\tau_2})}, \tag{5}$$

where

$$S(\text{SCAR}_{\tau_1,\tau_2}) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\text{SCAR}_{i,\tau_1,\tau_2} - \overline{\text{SCAR}}_{\tau_1,\tau_2})^2} \tag{6}$$

is the cross-sectional standard deviation of $\text{SCAR}_{i,\tau_1,\tau_2}$s and

$$\overline{\text{SCAR}}_{\tau_1,\tau_2} = \frac{1}{n}\sum_{i=1}^{n}\text{SCAR}_{i,\tau_1,\tau_2}. \tag{7}$$

Again $\text{SCAR}^*_{i,\tau_1,\tau_2}$ is a zero mean and unit variance random variable. The generalized standardized abnormal returns (GSARs) are defined similar to Kolari & Pynnönen (2011):

Definition 1 *The generalized standardized abnormal return (GSAR) is defined as*

$$\text{GSAR}_{it} = \begin{cases} \text{SCAR}^*_{i,\tau_1,\tau_2}, & \textit{in CAR-period} \\ \text{AR}'_{it}, & \textit{otherwise}, \end{cases} \tag{8}$$

where $\text{SCAR}^*_{i,\tau_1,\tau_2}$ *is defined in equation (5) and* AR'_{it} *is defined in equation (2).*

Thus the CAR-window is considered as one point in time in which the GSAR equals the re-standardized cumulative abnormal return defined in equation (5), and for other points in time GSAR equals the usual standardized abnormal returns defined in equation (2).

The time indexing is redefined such that the CAR-window of length $\tau_2 - \tau_1 + 1$ is squeezed into one observation with time index $t = 0$. Thus, considering the standardized cumulative abnormal return as one observation, in the testing procedure there are again $L_1 + 1$ observations of which the first L_1 are the estimation period (abnormal) returns and the last one is the cumulative return.

Kolari & Pynnönen (2011) have suggested that the GSARs can be used to extend the sign test in Corrado & Zivney (1992) for testing CARs. This can be achieved by defining the sign of the GSAR like:

Definition 2 *The sign of the generalized standardized abnormal return $GSAR_{it}$ is*

$$G_{it} = \text{sign}[\text{GSAR}_{it} - \text{median}(\text{GSAR}_i)], \tag{9}$$

where $\text{sign}(x)$ *is equal to +1, 0, -1 as x is* > 0, $= 0$ *or* < 0.

If $T = L_1 + 1$ is even, the corresponding probabilities for the sign of the GSAR for values +1, 0 and -1 are

$$\text{Pr}[G_{it} = 1] = \text{Pr}[G_{it} = -1] = \frac{1}{2} \tag{10}$$

and

$$\text{Pr}[G_{it} = 0] = 0. \tag{11}$$

If $T = L_1 + 1$ is odd, the corresponding probabilities for the sign of the GSAR for values +1, 0 and -1 are

$$\text{Pr}[G_{it} = 1] = \text{Pr}[G_{it} = -1] = \frac{T-1}{2T} \tag{12}$$

and

$$\text{Pr}[G_{it} = 0] = \frac{1}{T}. \tag{13}$$

The expectations, variances and covariances of the sign of GSAR are presented in Appendix A for even and odd T, and summarized in Proposition 1.

Proposition 1 *The expectation for the sign of the GSAR defined in (9) is*

$$E[G_{it}] = 0 \tag{14}$$

for T being even or odd. Furthermore the variance and covariance of the sign of the GSAR are

$$var[G_{it}] = \begin{cases} 1, & \textit{for even } T \\ \frac{T-1}{T}, & \textit{for odd } T \end{cases} \tag{15}$$

and

$$cov[G_{it}, G_{is}] = \begin{cases} -\frac{1}{T-1}, & \textit{for even } T \\ -\frac{1}{T}, & \textit{for odd } T. \end{cases} \tag{16}$$

Furthermore i=1,...,n and t≠s.

3 The test statistics SIGN-GSAR-T and SIGN-GSAR-Z

The null hypothesis of no mean event effect, reduces to

$$H_0 : \mu = 0, \tag{17}$$

where μ is the expectation of the (cumulative) abnormal return. As Kolari & Pynnönen (2011) suggested, we introduce a new sign test statistic (called hereafter SIGN-GSAR-T), which can be used for testing the presented null hypothesis. The test statistic SIGN-GSAR-T is defined as

$$t_{\text{SGT}} = \frac{Z_1\sqrt{T-2}}{\sqrt{T-1-Z_1^2}}, \tag{18}$$

where

$$Z_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} G_{i0}/S(G), \tag{19}$$

with

$$S(G) = \sqrt{\frac{1}{T}\sum_{t\in\mathcal{T}}(\frac{1}{\sqrt{n_t}}\sum_{i=1}^{n_t} G_{it})^2}, \tag{20}$$

in which n_t is the number of nonmissing returns in the cross-section of n-firms on day t and $\mathcal{T} = \{T_0 + 1, \ldots, T_1, 0\}$. The Z_1 statistic in equation (19) is the sign test derived by Corrado & Zivney (1992) for testing single event-day abnormal returns. Instead of using Z_1 in equation (19) as a test statistic, we use t_{SGT} in equation (18), because it has better properties e.g. the asymptotic distribution of t_{SGT} is the t-distribution.

Proofs of the Theorem 1 and Theorem 2 regarding the asymptotic distributions of Z_1 and the test statistic SIGN-GSAR-T defined in equations (19) and (18), respectively, are presented in Appendix B for both cases T being even and odd.

Theorem 1 (Asymptotic distribution of Z_1): *For a fixed T, under the assumption of cross-sectional independence, the density function of the asymptotic distribution of the test statistic Z_1 defined in equation* (19) *when $n \to \infty$, is*

$$f_{Z_1}(z) = \frac{\Gamma\left[(T-1)/2\right]}{\Gamma\left[(T-2)/2\right]\sqrt{(T-1)\pi}}\left(1 - \frac{z^2}{T-1}\right)^{\frac{1}{2}(T-2)-1}, \tag{21}$$

for $|z| \le \sqrt{T-1}$ and zero elsewhere, where $\Gamma(\cdot)$ is the Gamma function.

Thus, Theorem 1 implies that $(Z_1)^2/(T-1)$ is asymptotically Beta distributed with parameters $1/2$ and $(T-2)/2$.

Corrado & Zivney (1992) conjecture that for sufficiently large sample size, the Central Limit Theorem implies that the distribution of Z_1 should converge to normality. By Theorem 1 we can conclude that the asymptotic normality holds only if also T is large enough. This follows from the fact that in equation (21)

$$\left(1 - \frac{z^2}{T-1}\right)^{\frac{1}{2}(T-2)-1} \to e^{-\frac{1}{2}z^2} \tag{22}$$

and the normalizing constant

$$\frac{\Gamma\left[(T-1)/2\right]}{\Gamma\left[(T-2)/2\right]\sqrt{(T-1)\pi}} \to 1/\sqrt{2\pi} \tag{23}$$

as $T \to \infty$, implying the limiting $N(0,1)$-distribution.

Theorem 2 (Asymptotic distribution of the test statistic SIGN-GSAR-T): *Under the assumptions of Theorem* 1*,*

$$t_{\text{SGT}} = Z_1\sqrt{\frac{T-2}{T-1-(Z_1)^2}} \xrightarrow{d} t_{T-2}, \tag{24}$$

as $n \to \infty$, where Z_1 is defined in equation (19)*, $\xrightarrow{d}$ denotes convergence in distribution, and t_{T-2} denotes the Student t-distribution with $T-2$ degrees of freedom.*

Given that the t-distribution approaches the $N(0,1)$-distribution as the degrees of freedom $T-2$ increases, also the null distribution of the test statistic t_{SGT} approach the standard normal distribution as $T \to \infty$.

Remark 1 *Using facts about statistics based on signs (see Appendix A), it is easy to show that*

$$\text{var}[\overline{G_0}] = \begin{cases} \frac{1}{n}, & \textit{for even } T \\ \frac{T-1}{nT} \approx \frac{1}{n}, & \textit{for odd } T, \end{cases} \tag{25}$$

where $\overline{G_0} = \frac{1}{n}\sum_{i=1}^{n} G_{i0}$. Thus, under the assumption that $\text{var}[\overline{G_0}] = 1/n$*, a useful test statistic for the null hypothesis (17) is*

$$t_{\text{SGZ}} = \frac{\overline{G_0}}{\sqrt{\text{var}[\overline{G_0}]}} = \overline{G_0}\sqrt{n}, \tag{26}$$

for which the null distribution converges rapidly to the standard normal distribution, $N(0,1)$, as the number of firms increases. We henceforth refer to this statistic as SIGN-GSAR-Z.

The simplicity of the test statistic SIGN-GSAR-Z makes it an attractive alternative to the test statistic SIGN-GSAR-T. This is particularly the case when the event days across the sample firms are not clustered. However, in the presence of event day clustering, which causes cross-sectional correlations between the returns, the SIGN-GSAR-T can be expected to be much more robust than the SIGN-GSAR-Z test statistic.

Asymptotic Distributions: Cross-Sectional Dependence (Clustered Event Days)

Cross-sectional dependence due to clustered event days (the same event days across the firms) changes materially the asymptotic properties of the test statistics and in particular those statistics that do not account for the cross-sectional dependence.

As stated in Lehmann (1999) it is still frequently true that the asymptotic normality holds provided that the average cross-correlation, $\overline{\rho}_n$, tends to zero rapidly enough such that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n}\rho_{ij}\to\gamma \tag{27}$$

as $n\to\infty$.

In financial applications this would be the case if there are a finite number of firms in each industry and the return correlations between industries were zero. In fact this is a special case of so called *m*-independence. Generally, a sequence of random variables X_1, X_2,..., is said to be *m*-independent, if X_i and X_j are independent if $|i-j|>m$. In cross-sectional analysis this would mean that the variables can be ordered such that when the index difference is larger than *m*, the variables are independent. (See Kolari & Pynnönen (2011)).

In such a case, we can show in the same manner as in Kolari & Pynnönen (2011) that the result in (27) holds. More precisely, assuming that for any fixed t, G_{it} defined in equation (9) are *m*-independent, $i=1,2,...,n$, $(n>m)$, the correlation matrix of $G_{1t},...,G_{nt}$ is band-diagonal such that all ρ_{ij} with $|i-j|>m$ are zeros. In such a correlation matrix there are $m(2n-m-1)$ nonzero correlations in addition to the n ones on the diagonal. Thus, in the double summation (27) there are $m(2n-m-1)$ non-zero elements, and it can be written such that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\rho_{ij}=\frac{m(2n-m-1)}{n}\widetilde{\rho}_n\to\gamma, \tag{28}$$

where $\widetilde{\rho}_n$ is the average of the $m(2n-m-1)$ cross-correlations in the band-diagonal correlations matrix and $\gamma=2m\widetilde{\rho}$ is a finite constant with $\widetilde{\rho}=\lim_{n\to\infty}\widetilde{\rho}_n$ and $2m=\lim_{n\to\infty}m(2n-m-1)/n$.

Thus, under the *m*-independence the asymptotic distribution of the test statistic

SIGN-GSAR-Z is

$$t_{\text{SGZ}} \to N(0, 1 + \gamma). \tag{29}$$

This implies that the test statistic SIGN-GSAR-Z is not robust to cross-sectional correlation of the return series. Typically $\gamma > 0$, which means that t_{SGZ} will tend to over-reject the null hypothesis.

The limiting distributions of the test statistic SIGN-GSAR-T turns out to apply also under *m*-independence. This follows from the fact that, if the asymptotic normality holds under the *m*-independence such that the limiting correlation effect is $1 + \gamma$, then using the scaled variables, $G_{it}/\sqrt{1+\gamma}$, in place of the original variables, all the results in Theorem 1 and Theorem 2 follow, because in Z_1 defined in (19) and t_{SGT} defined in (18) are invariant to the scaling of the observations (the zero-one sign of the GSARs). Therefore, the theoretical derivation indicates that when the event-dates are clustered, the test statistic SIGN-GSAR-T behave better than the test statistic SIGN-GSAR-Z.

4 Simulation design

In this section we present the simulation design, which is used to examine the empirical behaviour of the test statistics SIGN-GSAR-T and SIGN-GSAR-Z. As for example Kolari & Pynnönen (2011) have concluded, the optimality of a test can be judged on the basis of size and power. Within a class of tests of given size (Type I error probability), the one that has the maximum power (minimum Type II error probability) is the best. A testing procedure is robust, if the Type I error rate is not affected by real data issues such as non-normality, event-induced volatility, auto-correlation and cross-correlation of returns. Consequently, the aim of our simulations is to focus on the robustness and power properties of the tests. Non-normality, auto-correlation, and other data issues are captured in the simulation by using actual return data instead of artificially simulated data. Event-induced volatility effects are investigated by introducing volatility change within the event period, and the effect of cross-sectional correlation is examined by setting the same event day in the return series for each firm in the sample.

4.1 Sample construction

We use the well-known simulation approach presented by Brown & Warner (1980), and widely used in several other methodological studies (e.g., Brown & Warner (1985), Corrado (1989), Cowan (1992), Campbell & Wasley (1993) and Cowan & Sergeant (1996)). From the data base 1,000 portfolios each of 50 stocks are constructed with replacements. Each time a stock is selected, a hypothetical event date is randomly generated and the event day is denoted as day "0". The results are reported for event day $t = 0$ abnormal return $\text{AR}(0)$ and for cumulative abnormal returns $\text{CAR}(-1, +1)$, $\text{CAR}(-5, +5)$ and $\text{CAR}(-10, +10)$. The estimation period is comprised of 239 days prior to the event period, hence days from -249 to -11. The event period is comprised of 21 days, hence days from -10 to +10. Therefore, the estimation period and event period altogether comprises of 260 days. In order for a return series to be included, no missing returns are allowed in the last 30 days from -19 to +10.

In earlier studies (e.g., Charest (1978), Mikkelson (1981), Penman (1982) and Rosenstein et al. (1990)) it has been found that the event period standard deviation is about 1.2 to 1.5 times the estimation period standard deviation. Therefore, the increased volatility is introduced by multiplying the cumulated event period returns by a factor $\sqrt{c}$ with values $c = 1.5$ for an approximate 20 percent increased volatility, $c = 2.0$ for an approximate 40 percent increased volatility and $c = 3.0$ for an approximate 70 percent increased volatility due to the event effect.[3] To add realism the volatility factors c are generated for each stock based on the following uniform distributions $U[1, 2]$, $U[1.5, 2.5]$ and $U[2.5, 3.5]$. This generates on average the variance effects of 1.5, 2.0 and 3.0. Furthermore for the no volatility effect experiment $c = 1.0$ is fixed.

For investigating the power properties a similar method as for example Campbell & Wasley (1993) is used. Hence, for single-day event period [$\text{AR}(0)$] the abnormal performance is artificially introduced by adding the indicated percentage (a constant) to the day-0 return of each security. While, in the multiday setting [$\text{CAR}(-1, +1)$, $\text{CAR}(-5, +5)$ and $\text{CAR}(-10, +10)$], abnormal performance is introduced by selecting one day of the CAR-period at random and adding the particular level of abnormal performance to that day's return. By this we aim to mimic the real situations, where there can be the information leakage and delayed adjustment. That is, if the markets are inefficient, information may leak before the event, which shows up as abnormal behaviour before the event day. Delays in the event information show up as abnormal return behaviour after the event day.

We also study the effect of event-date clustering on the test statistics. The effect of event-date clustering is examined by constructing 1,000 portfolios each of 50 stocks again from the data base, but all stocks in the portfolio have exactly the same event date.

[3]Because $\sqrt{1.5} \approx 1.2$, $\sqrt{2.0} \approx 1.4$ and $\sqrt{3.0} \approx 1.7$.

4.2 Abnormal return model

The abnormal behaviour of security returns can be estimated through the market model

$$r_{it} = \alpha_i + \beta_i r_{mt} + \epsilon_{it}, \tag{30}$$

where again r_{it} is the return of stock i at time t, r_{mt} is the market index return at time t and ϵ_{it} is a white noise random component, which is not correlated with r_{mt}. The resulting ARs are obtained as differences of realized and predicted returns on day t in the event period

$$\text{AR}_{it} = r_{it} - (\hat{\alpha}_i + \hat{\beta}_i r_{mt}), \tag{31}$$

where the parameters are estimated from the estimation period with ordinary least squares. According to Campbell et al. (1997) the market model represents a potential improvement over the traditional constant-mean-return model, because by removing the portion of the return that is related to variation in the market's return, the variance of the AR is reduced. This can lead to increased ability to detect event effects.

4.3 Test statistics

Next we present the test statistics, which are used in the simulations. The ordinary t-test (ORDIN) is defined as

$$t_{\text{ORDIN}} = \frac{\overline{\text{CAR}}_{\tau_1,\tau_2}}{S(\overline{\text{CAR}}_{\tau_1,\tau_2})}, \tag{32}$$

where

$$\overline{\text{CAR}}_{\tau_1,\tau_2} = \frac{1}{n}\sum_{i=1}^{n} \text{CAR}_{i,\tau_1,\tau_2}, \tag{33}$$

in which $\text{CAR}_{i,\tau_1,\tau_2}$ is defined in equation (3) and $S(\overline{\text{CAR}}_{\tau_1,\tau_2})$ is the standard error of the average cumulative abnormal return $\overline{\text{CAR}}_{\tau_1,\tau_2}$ adjusted for the prediction error (see again Campbell et al. (1997)). The ordinary t-test statistic is asymptotically $N(0,1)$-distributed under the null hypothesis of no event effect.

Patell (1976) test statistic (PATELL) is

$$t_{\text{PATELL}} = \sqrt{\frac{n(L_1-4)}{L_1-2}}\,\overline{\text{SCAR}}_{\tau_1,\tau_2}, \tag{34}$$

where $\overline{\text{SCAR}}_{\tau_1,\tau_2}$ is the average of the standardized CAR defined in equation (7), and L_1 is again the length of the estimation period. Also the test statistic derived by Patell is asymptotically $N(0,1)$-distributed under the null hypothesis.

The Boehmer et al. (1991) test statistics (BMP) is

$$t_{\text{BMP}} = \frac{\overline{\text{SCAR}}_{\tau_1,\tau_2}\sqrt{n}}{S(\text{SCAR}_{\tau_1,\tau_2})}, \tag{35}$$

where again $S(\text{SCAR}_{\tau_1,\tau_2})$ is the cross-sectional standard deviation of SCARs defined in (6), and $\overline{\text{SCAR}}_{\tau_1,\tau_2}$ is defined in equation (7). Also the test statistic t_{BMP} is asymptotically $N(0,1)$-distributed under the null hypothesis.

We follow Kolari & Pynnönen (2011) and define the demeaned standardized abnormal ranks of the GSARs as

$$U_{it} = \text{Rank}(\text{GSAR}_{it})/(T+1) - 1/2, \tag{36}$$

where $i = 1, ..., n$ and $t \in \mathcal{T} = \{T_0 + 1, ..., T_1, 0\}$ is the set of time indexes including the estimation period for $t = T_0 + 1, ..., T_1$ and to the CAR for $t = 0$, with $T_0 + 1$ and T_1 the first and last observation on the estimation period, and $T = L_1 + 1 = T_1 - T_0 + 1$ is the total number of observations with L_1 estimation period returns and the one CAR. Then the generalized rank test statistic (GRANK) is defined as

$$t_{\text{GRANK}} = Z_2\sqrt{\frac{T-2}{T-1-Z_2^2}}, \tag{37}$$

where

$$Z_2 = \frac{\overline{U}_0}{S_{\overline{U}}} \tag{38}$$

with

$$S_{\overline{U}} = \sqrt{\frac{1}{T}\sum_{t\in\mathcal{T}}\frac{n_t}{n}\overline{U}_t^2} \tag{39}$$

and

$$\overline{U}_t = \frac{1}{n_t}\sum_{i=1}^{n_t} U_{it}. \tag{40}$$

Furthermore n_t is the number of valid GSARs available at point in time t, $t \in \mathcal{T} = \{T_0 + 1, ..., T_1, 0\}$, $T = T_1 - T_0 + 1$ is the number of observations, and $\overline{U}_0$ is the mean $\overline{U}_t$ for $t = 0$ (CAR). According to Kolari & Pynnönen (2011) the asymptotic distribution of the test statistic GRANK is Student t-distribution with $T-2$ degrees of freedom. Again given that the t-distribution approaches the $N(0,1)$-distribution as the degrees of freedom $T - 2$ increases, also the null distribution of the test statistic t_{GRANK} approach the standard normal distribution as $T \to \infty$.

The generalized sign test statistic presented by Cowan (1992) is

$$t_{\text{COWAN}} = \frac{w - n\widehat{p}}{\sqrt{n\widehat{p}(1-\widehat{p})}}, \tag{41}$$

where w is the number of stocks in the event window for which the CAR is positive and n is again the number of the stocks. Furthermore

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{t=T_0+1}^{T_1} S_{it}, \tag{42}$$

where m_i is the number of non-missing returns in the estimation period for security-event i and

$$S_{it} = \begin{cases} 1 & \text{if } \mathrm{AR}_{it} > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{43}$$

According to Cowan (1992) the generalized sign test statistic (SIGN-COWAN) is asymptotically $N(0,1)$-distributed under the null hypothesis.

4.4 The data

The data in this simulation design consists of daily closing prices of 1,500 U.S. stocks that make up the S&P 400, S&P 500, and S&P 600 indexes. S&P 400 covers the mid-cap range of stocks, S&P 500 the large-cap range of stocks and S&P 600 the small-cap range of stocks. The five percent of the stocks having the smallest trading volume are excluded. Therefore, 72 stocks from S&P 600, two stocks from S&P 400 and one stock from S&P 500 are excluded. The sample period spans from the beginning of July, 1991 to October 31, 2009. S&P 400 index was launched in June in 1991, which is why the sample period starts in the beginning of July, 1991. Official holidays and observance days are excluded from the data. By using actual (rather than artificial) stock returns in repeated simulations, a reliable and realistic view about the comparative real data performance of the test statistics in true applications is attained. The returns are defined as log-returns

$$r_{it} = \log(P_{it}) - \log(P_{it-1}), \tag{44}$$

where P_{it} is the closing price for stock i at time t.

5 Empirical results

5.1 Sample statistics

Table 1 reports sample statistics from $1,000$ simulations for the event day abnormal returns and for the cumulative abnormal returns: CAR$(-1,+1)$, CAR$(-5,+5)$ and CAR$(-10,+10)$. It also reports sample statistics for the test statistics for AR(0), CAR$(-1,+1)$, CAR$(-5,+5)$ and CAR$(-10,+10)$. Under the null hypothesis of no even effect test statistics ORDIN, PATELL, BMP, SIGN-COWAN and SIGN-GSAR-Z should be approximately $N(0,1)$-distributed. Strictly speaking, the asymptotic distributions of GRANK and SIGN-GSAR-T should be t -distributions with $T-2$ degrees of freedom. However, with $T-2$ equal to 238, the normal approximation should be valid and so the null distributions of the test statistics GRANK and SIGN-GSAR-T approach the standard normal distribution. Hence, we can conclude that under the null hypothesis of no event effect all the test statistics should have zero mean and (approximately) unit variance.

Considering only the single abnormal returns AR(0) in Panel A of Table 1, it can be noted that the means of all the test statistics are statistically close to zero. For example (in absolute value) the largest mean of -0.024 for the PATELL statistic is only 1.113 standard errors away from zero. In longer CAR-windows the means of the test statistics, albeit small, start to deviate significantly away from the theoretical value of zero. Considering on the 3-day CARs, CAR$(-1,+1)$, in Panel B of Table 1, we see that only the means for PATELL and BMP deviate significantly away from zero. While, considering on the 11- and 21-day CARs, CAR$(-5,+5)$ and CAR$(-10,+10)$, in Panels C and D of Table 1, it is noticable that means for almost all the test statistics deviate significantly away from the theoretical value. Nonetheless, it can be seen that the means of the test statistics PATELL and BMP deviate more rapidly and clearly from the theoretical value of zero than the means of the other test statistics. Furthermore, it can be seen that the mean of the test statistic GRANK seems to deviate more slowly from the theoretical value of zero than the means of the other test statistics. Importantly, all standard deviations of the test statistics are close to their theoretical values of unity.

Table 1. Sample statistics from 1,000 simulations. (Superscripts a, b and c correspond to the significance levels 0.10, 0.05 and 0.01.)

Panel A AR(0): AR(0)	Mean	Med.	Std.	Skew.	Kurt.	Min.	Max.
AR(0), %	0.004	-0.008	0.413	-0.082	1.018	-1.688	1.641
ORDIN	0.008	-0.019	1.053	-0.079	0.701	-3.878	3.694
PATELL	-0.024	-0.036	1.113	-0.193	1.170	-6.178	3.837
BMP	-0.013	-0.033	1.000	-0.013	0.146	-4.000	3.777
GRANK	0.002	-0.010	0.974	0.056	0.071	-3.518	3.375
SIGN-COWAN	-0.002	-0.042	0.958	0.059	-0.120	-3.475	2.999
SIGN-GSAR-T	-0.016	0.000	0.990	0.041	-0.206	-2.630	2.997
SIGN-GSAR-Z	-0.016	0.000	1.082	0.037	-0.226	-2.828	3.111
Panel B: CAR$(-1,+1)$	Mean	Med.	Std.	Skew.	Kurt.	Min.	Max.
CAR$(-1,+1)$, %	-0.010	-0.029	0.671	-0.019	0.146	-2.288	2.096
ORDIN	-0.018	-0.040	0.988	-0.028	0.306	-3.759	3.329
PATELL	-0.067^{b}	-0.085	1.077	0.133	0.113	-3.380	4.059
BMP	-0.054^{a}	-0.088	1.023	0.159	0.083	-3.208	3.856
GRANK	-0.001	0.011	1.021	0.088	0.189	-3.332	3.963
SIGN-COWAN	0.042	0.108	1.020	0.063	0.127	-3.475	3.832
SIGN-GSAR-T	0.028	0.000	1.036	0.050	0.082	-3.225	3.610
SIGN-GSAR-Z	0.031	0.000	1.128	0.038	0.045	-3.677	3.960
Panel C: CAR$(-5,+5)$	Mean	Med.	Std.	Skew.	Kurt.	Min.	Max.
CAR(-5,+5), %	-0.076	-0.027	1.269	-0.114	0.346	-5.178	4.455
ORDIN	-0.060^{b}	-0.020	0.959	-0.183	0.433	-3.977	3.441
PATELL	-0.132^{c}	-0.108	1.107	-0.034	0.363	-4.005	3.992
BMP	-0.113^{c}	-0.117	1.036	0.088	0.157	-3.417	3.603
GRANK	0.016	0.062	1.038	0.011	0.113	-3.073	3.334
SIGN-COWAN	0.085^{b}	0.102	0.973	-0.028	-0.081	-2.764	3.334
SIGN-GSAR-T	0.065^{b}	0.000	0.993	0.018	-0.005	-2.804	3.284
SIGN-GSAR-Z	0.071^{b}	0.000	1.084	0.030	0.034	-3.111	3.677
Panel D: CAR$(-10,+10)$	Mean	Med.	Std.	Skew.	Kurt.	Min.	Max.
CAR(-10,+10), %	-0.056	-0.029	1.800	-0.136	0.018	-5.442	4.845
ORDIN	-0.038	-0.015	0.967	-0.225	0.149	-3.332	2.749
PATELL	-0.130^{c}	-0.108	1.105	-0.287	0.852	-5.208	4.129
BMP	-0.100^{c}	-0.117	1.042	0.011	0.033	-3.092	3.759
GRANK	0.071^{b}	0.063	1.053	-0.049	0.282	-3.645	3.942
SIGN-COWAN	0.180^{c}	0.162	1.013	-0.052	0.295	-3.296	3.536
SIGN-GSAR-T	0.148^{c}	0.241	0.997	-0.059	0.370	-3.275	3.423
SIGN-GSAR-Z	0.158^{c}	0.283	1.090	-0.088	0.413	-3.677	3.677

Table 2. Cramer-von Mises tests of the distributions. ORDIN [Eq.(32)], PATELL [Eq.(34)], BMP [Eq.(35)], SIGN-COWAN [Eq.(41)] and SIGN-GSAR-Z [Eq. (26)] are tested against the standard normal distribution whereas GRANK [Eq.(37)] and SIGN-GSAR-T [Eq.(18)] are tested against the Student t-distribution with 238 $(= T-2)$ degrees of freedom. Superscripts a and b correspond to the significance levels 0.05 and 0.01.

Test statistic	AR(0)	CAR(-1,+1)	CAR(-5,+5)	CAR(-10,+10)
ORDIN	0.054	0.218	0.350	0.196
PATELL	0.164	0.795^b	1.488^b	1.104^b
BMP	0.066	0.625	1.277^b	0.985^b
GRANK	0.074	0.029	0.143	0.541^a
SIGN-COWAN	0.136	0.270	0.916^b	2.994^b
SIGN-GSAR-T	0.361	0.387	0.855^b	2.400^b
SIGN-GSAR-Z	0.918^b	1.006^b	1.288^b	2.871^b

5.2 Empirical distributions

Table 2 reports Cramer-von Mises normality tests for ORDIN, PATELL, BMP, SIGN-COWAN and SIGN-GSAR-Z, and Cramer-von Mises tests for GRANK and SIGN-GSAR-T against a t-distribution with 238 $(= T - 2)$ degrees of freedom. Departures from normality (t-distribution for GRANK and SIGN-GSAR-T) of the statistics are typically not statistically significant for the AR(0) and CAR$(-1, +1)$, i.e., in the short CAR-windows. Only the normality of the test statistic PATELL is rejected for CAR$(-1, +1)$ and the test statistic SIGN-GSAR-Z for both AR(0) and CAR$(-1, +1)$. In the long CAR-windows (11 and 21 days) the normality is rejected for almost every test statistic. The results indicate that particularly for short CAR-windows a sample size of $n = 50$ series seems to be large enough to warrant the asymptotic t-distribution for SIGN-GSAR-T.

In Figure 1 empirical quantiles of the test statistic SIGN-GSAR-T are displayed from 1,000 simulations against theoretical quantiles of the test statistic SIGN-GSAR-T for AR(0), CAR$(-1, +1)$, CAR$(-5, +5)$ and CAR$(-10, +10)$. Only the statistic SIGN-GSAR-T is considered, because it is derived in this paper and because Cramer-von Mises tests reject the normality of the test statistic SIGN-GSAR-Z for both short and long CAR-windows. On the vertical axis of Figure 1 are the Student t quantiles with $T-2 = 238$ degrees of freedom and on the horizontal axis are the test statistics SIGN-GSAR-T. If the statistic follow the theoretical distribution depicted on the vertical axis, the plots should be close to the 45 degree diagonal line. According to Figure 1 the empirical distributions of the test statistics SIGN-

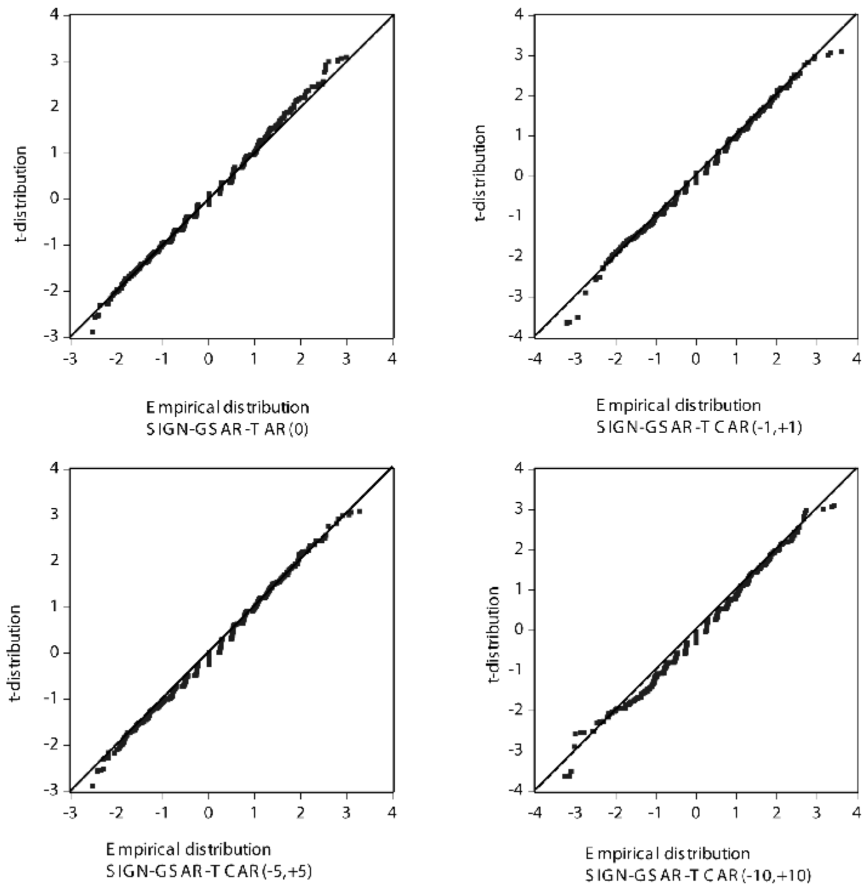


Figure 1. The Q-Q plots of the test statistic SIGN-GSAR-T

GSAR-T and Student t-distributions seem to match quite well, because the plots lie quite well on a straight line.

5.3 Rejection rates

The first three columns in the first part of the Table 3 report the lower tail, upper tail and two-tailed rejection rates (type I errors) at the 5 percent level under the null hypothesis of no event mean effect with no event-induced volatility. Almost all rejection rates are close to the nominal rate of 0.05 for short CAR-windows of AR(0) and CAR$(-1,+1)$. Only PATELL statistic tends to over-reject the null hypothesis for the two-tailed tests and SIGN-GSAR-Z statistics tends to over-reject for left and right tail tests as well as two-tailed tests. For the longer CAR-windows of CAR$(-5,+5)$ and CAR$(-10,+10)$ again all the other test statistics except PATELL, BMP, SIGN-COWAN and SIGN-GSAR-Z reject close to the nominal rate with rejection rates that are well within the approximate 99 percent confidence interval of $[0.032, 0.068]$. For the longer CAR-windows the PATELL

Table 3. Lower tail, upper tail and two-tailed rejection rates. The table reports the lower tail, upper tail and two-tailed rejection rates (type I errors) at the 5 percent level under the null hypothesis of no event mean effect. The 99 percent confidence interval around the 0.05 rejection rate is $[0.032, 0.068]$. $c = 1.5/2.0/3.0$ correspond to an event induced volatility increase of approximately 20%, 40%, and 70% respectively (no increase for $c = 1.0$).

	c=1.0			c=1.5		
Test statistic	Left	Right	Two-tail	Left	Right	Two-tail
Panel A: AR(0)						
ORDIN	0.050	0.055	0.058	0.085	0.091	0.113
PATELL	0.064	0.057	0.070	0.109	0.103	0.132
BMP	0.048	0.042	0.045	0.051	0.044	0.041
GRANK	0.044	0.048	0.047	0.041	0.048	0.046
SIGN-COWAN	0.040	0.044	0.037	0.040	0.044	0.037
SIGN-GSAR-T	0.048	0.041	0.042	0.046	0.044	0.041
SIGN-GSAR-Z	0.087	0.071	0.085	0.084	0.073	0.086
Panel B: CAR(-1,+1)						
ORDIN	0.052	0.044	0.047	0.085	0.090	0.108
PATELL	0.067	0.058	0.076	0.112	0.092	0.139
BMP	0.053	0.054	0.054	0.051	0.052	0.051
GRANK	0.049	0.051	0.052	0.052	0.050	0.055
SIGN-COWAN	0.051	0.062	0.055	0.051	0.062	0.055
SIGN-GSAR-T	0.056	0.057	0.058	0.056	0.057	0.059
SIGN-GSAR-Z	0.082	0.091	0.105	0.083	0.092	0.105
Panel C: CAR(-5,+5)						
ORDIN	0.052	0.033	0.042	0.092	0.065	0.098
PATELL	0.080	0.054	0.080	0.130	0.093	0.141
BMP	0.066	0.046	0.060	0.063	0.044	0.057
GRANK	0.056	0.054	0.058	0.053	0.050	0.058
SIGN-COWAN	0.042	0.045	0.042	0.042	0.045	0.042
SIGN-GSAR-T	0.041	0.053	0.039	0.041	0.053	0.038
SIGN-GSAR-Z	0.076	0.079	0.091	0.076	0.080	0.090
Panel D: CAR(-10,+10)						
ORDIN	0.053	0.033	0.048	0.090	0.078	0.099
PATELL	0.086	0.050	0.075	0.125	0.089	0.145
BMP	0.067	0.047	0.069	0.065	0.050	0.066
GRANK	0.053	0.063	0.064	0.053	0.063	0.061
SIGN-COWAN	0.032	0.073	0.060	0.032	0.073	0.060
SIGN-GSAR-T	0.032	0.063	0.054	0.033	0.066	0.065
SIGN-GSAR-Z	0.044	0.092	0.089	0.046	0.094	0.093
				(*continues*)		

Table 3. *(continued)*

	c=2.0			c=3.0		
Test statistic	Left	Right	Two-tail	Left	Right	Two-tail
Panel A: AR(0)						
ORDIN	0.131	0.127	0.162	0.174	0.188	0.261
PATELL	0.136	0.136	0.194	0.183	0.180	0.291
BMP	0.051	0.043	0.045	0.048	0.042	0.044
GRANK	0.042	0.046	0.048	0.043	0.046	0.045
SIGN-COWAN	0.040	0.044	0.037	0.040	0.044	0.037
SIGN-GSAR-T	0.048	0.042	0.042	0.046	0.042	0.042
SIGN-GSAR-Z	0.087	0.074	0.086	0.087	0.070	0.085
Panel B: CAR(-1,+1)						
ORDIN	0.113	0.127	0.165	0.163	0.165	0.250
PATELL	0.149	0.128	0.190	0.203	0.164	0.293
BMP	0.054	0.053	0.054	0.054	0.055	0.054
GRANK	0.051	0.048	0.054	0.048	0.052	0.053
SIGN-COWAN	0.051	0.062	0.055	0.051	0.062	0.055
SIGN-GSAR-T	0.057	0.056	0.057	0.056	0.056	0.058
SIGN-GSAR-Z	0.081	0.090	0.105	0.083	0.091	0.104
Panel C: CAR(-5,+5)						
ORDIN	0.127	0.083	0.148	0.173	0.132	0.227
PATELL	0.166	0.112	0.200	0.217	0.149	0.299
BMP	0.072	0.043	0.060	0.067	0.041	0.062
GRANK	0.057	0.051	0.058	0.055	0.053	0.057
SIGN-COWAN	0.042	0.045	0.042	0.042	0.045	0.042
SIGN-GSAR-T	0.041	0.053	0.039	0.041	0.053	0.038
SIGN-GSAR-Z	0.076	0.079	0.091	0.076	0.080	0.090
Panel D: CAR(-10,+10)						
ORDIN	0.053	0.033	0.048	0.090	0.078	0.099
PATELL	0.086	0.050	0.075	0.125	0.089	0.145
BMP	0.067	0.047	0.069	0.065	0.050	0.066
GRANK	0.053	0.063	0.064	0.053	0.063	0.061
SIGN-COWAN	0.032	0.073	0.060	0.032	0.073	0.060
SIGN-GSAR-T	0.032	0.063	0.054	0.033	0.066	0.065
SIGN-GSAR-Z	0.044	0.092	0.089	0.046	0.094	0.093

tends to over-reject in addition of the two-tailed tests also on the lower tail. The BMP statistic tends to somewhat over-reject the null hypothesis for two-tailed test for CAR$(-10,+10)$ and the SIGN-COWAN statistic tends to over-reject the null hypothesis for CAR$(-10,+10)$ for the upper tail test. The SIGN-GSAR-Z statistic over-rejects the null hypothesis again for left and right tailed tests as well as two-tailed tests. It seems that the tails of the test statistic SIGN-GSAR-Z are fat, which may be the reason why the Cramer-von Mises test rejects the normality of the test statistic SIGN-GSAR-Z in every case.

The remainder of Table 3 report the rejection rates under the null hypothesis in the cases where the event-induced variance is present. ORDIN and PATELL tests over-reject when the variance increases, which is a well-known outcome. At the highest factor of $c = 3.0$ the type I errors for both ORDIN and PATELL are in the range from 0.2 to 0.3 in two-tailed testing, that is, five to six times the nominal rate. The SIGN-GSAR-Z statistic over-rejects the null hypothesis again for left and right tail tests as well as two-tailed tests. Note that because test statistic SIGN-COWAN takes account only of the sign of the difference between AR and zero, and not for example the sign of the difference between AR and its median, the event-induced volatility does not have an impact on the rejection rates of the test statistics SIGN-COWAN. Hence, the test statistics BMP, GRANK, SIGN-COWAN and SIGN-GSAR-T seem to be the best options in the cases where the event induced volatility is present.

5.4 Power of the tests

5.4.1 Non-clustered Event Days

The power results of the test statistics for two-tailed tests are shown in panels A to D of Table 4 and graphically depicted in Figures 2 to 5.[4] The zero abnormal return line in each panel of Table 4 indicates the type I error rates and replicates the columns 4 and 7 in each panel of Table 3. The remaining lines of Table 4 indicate the rejection rates for the respective ARs shown in the first column.

There are three outstanding results. First, at all levels of ARs (positive or negative), ORDIN, which is based on non-standardized returns is materially less powerful than the other test statistics that are based on standardized returns. Second, the test statistic GRANK seems to be one of the most powerful tests for shorter CAR-windows as well as for the longer CAR-windows. Third, the test statistic SIGN-GSAR-T seems to be somewhat less powerful than the test statistic SIGN-COWAN.

[4]Nevertheless, Figures 2–5 do not include the test statistics PATELL and SIGN-GSAR-Z, because they over-reject the null hypothesis.

Table 4. Non-clustered event days: Powers of the test statistics. The zero abnormal return line indicates the type I error rates. The other lines indicate the rejection rates for the respective ARs shown in the first column.

Panel A: AR(0)							
AR	ORDIN	PATELL	BMP	GRANK	SIGN-COWAN	SIGN-GSAR-T	SIGN-GSAR-Z
−3.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000
−2.0	0.996	1.000	0.999	1.000	1.000	0.997	0.999
−1.0	0.720	0.960	0.909	0.958	0.912	0.856	0.910
±0.0	**0.058**	**0.070**	**0.045**	**0.047**	**0.037**	**0.042**	**0.085**
+1.0	0.729	0.942	0.899	0.969	0.928	0.886	0.928
+2.0	0.994	0.999	0.996	0.999	1.000	1.000	1.000
+3.0	1.000	1.000	0.999	1.000	1.000	1.000	1.000
Panel B: CAR(-1,+1)							
AR	ORDIN	PATELL	BMP	GRANK	SIGN-COWAN	SIGN-GSAR-T	SIGN-GSAR-Z
−3.0	0.994	1.000	0.999	1.000	0.999	0.998	1.000
−2.0	0.983	0.965	0.971	0.984	0.957	0.909	0.954
−1.0	0.313	0.595	0.572	0.656	0.528	0.446	0.575
±0.0	**0.047**	**0.076**	**0.054**	**0.052**	**0.055**	**0.058**	**0.105**
+1.0	0.297	0.535	0.514	0.638	0.572	0.479	0.611
+2.0	0.834	0.981	0.958	0.993	0.971	0.928	0.964
+3.0	0.987	1.000	0.999	1.000	0.999	0.995	0.998
Panel C: CAR(-5,+5)							
AR	ORDIN	PATELL	BMP	GRANK	SIGN-COWAN	SIGN-GSAR-T	SIGN-GSAR-Z
−3.0	0.633	0.912	0.892	0.913	0.782	0.716	0.799
−2.0	0.326	0.619	0.614	0.628	0.481	0.438	0.550
−1.0	0.120	0.244	0.239	0.233	0.155	0.146	0.221
±0.0	**0.042**	**0.080**	**0.060**	**0.058**	**0.042**	**0.039**	**0.091**
+1.0	0.085	0.182	0.171	0.234	0.195	0.175	0.254
+2.0	0.312	0.535	0.538	0.659	0.572	0.490	0.614
+3.0	0.616	0.867	0.843	0.920	0.865	0.791	0.856
Panel D: CAR(-10,+10)							
AR	ORDIN	PATELL	BMP	GRANK	SIGN-COWAN	SIGN-GSAR-T	SIGN-GSAR-Z
−3.0	0.355	0.680	0.670	0.655	0.471	0.422	0.551
−2.0	0.184	0.379	0.378	0.368	0.223	0.212	0.299
−1.0	0.088	0.165	0.154	0.141	0.077	0.067	0.139
±0.0	**0.048**	**0.075**	**0.069**	**0.064**	**0.060**	**0.054**	**0.089**
+1.0	0.069	0.126	0.118	0.161	0.150	0.124	0.203
+2.0	0.162	0.316	0.320	0.398	0.349	0.300	0.403
+3.0	0.351	0.609	0.598	0.699	0.613	0.539	0.637

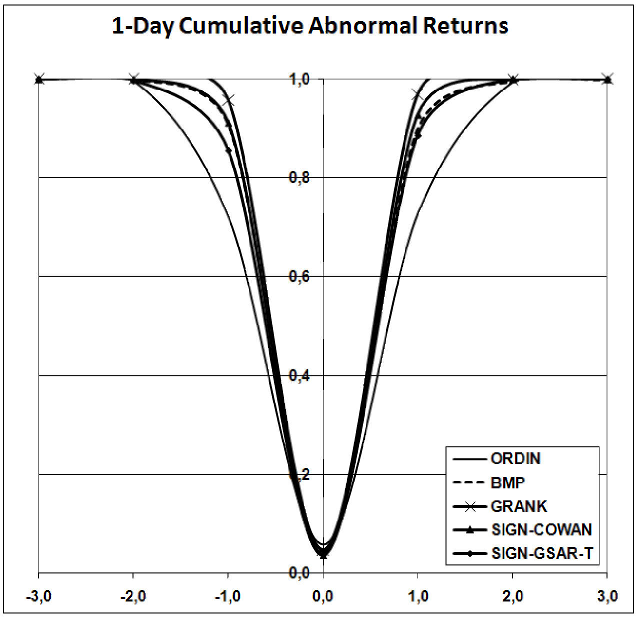


Figure 2. The power results for AR(0)

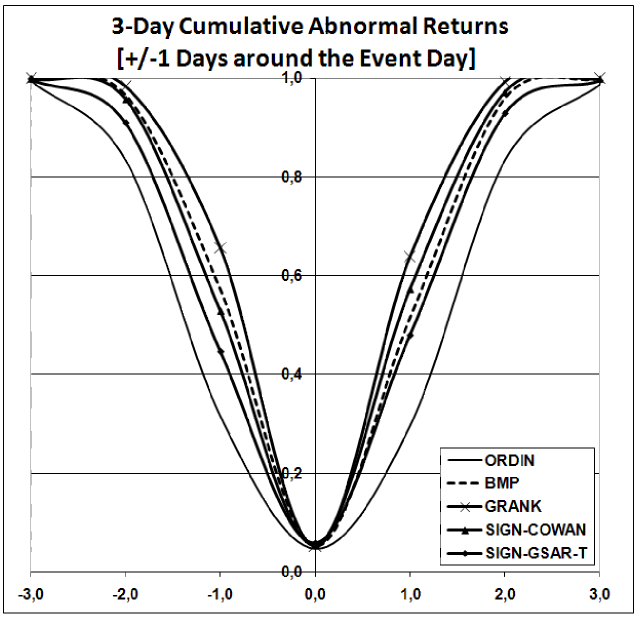


Figure 3. The power results for CAR$(-1, +1)$

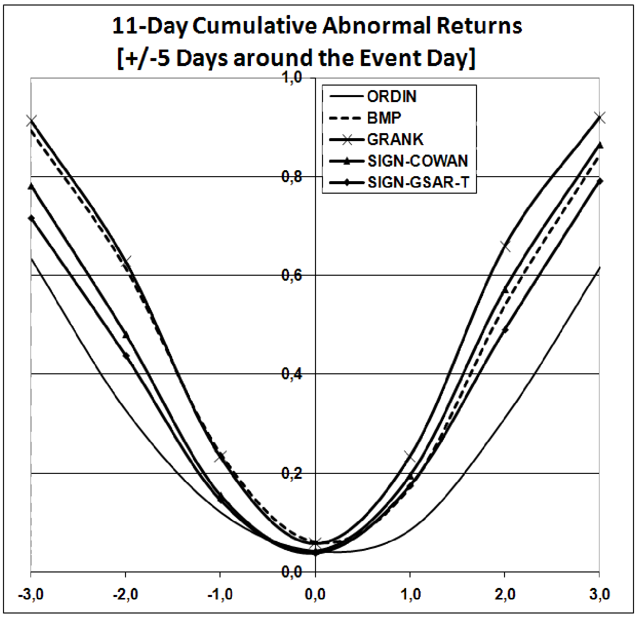


Figure 4. The power results for CAR$(-5, +5)$

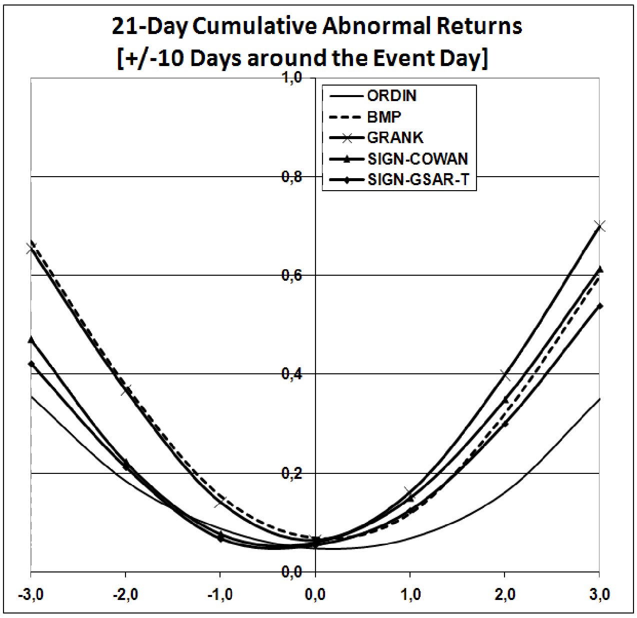


Figure 5. The power results for CAR$(-10, +10)$

5.4.2 Clustered event days

Table 5 reports the type I error and power results of the tests with clustered event-days. The zero abnormal return line in each panel again indicates the type I error rates at the 5 percent level under the null hypothesis of no event mean effect.

Table 5. Clustered event days: Powers of the test statistics. The zero abnormal return line indicates the type I error rates. The other lines indicate the rejection rates for the respective ARs shown in the first column. In each simulation round a single event day is randomly selected and held constant for each of the 50 securities sampled *without* replacement in the next step.

Panel A: AR(0)							
AR	ORDIN	PATELL	BMP	GRANK	SIGN-COWAN	SIGN-GSAR-T	SIGN-GSAR-Z
−3.0	0.995	0.998	0.996	0.990	0.997	0.985	0.995
−2.0	0.968	0.988	0.982	0.953	0.981	0.936	0.980
−1.0	0.692	0.858	0.844	0.716	0.851	0.627	0.852
±0.0	**0.177**	**0.203**	**0.216**	**0.055**	**0.190**	**0.055**	**0.244**
+1.0	0.690	0.838	0.828	0.714	0.839	0.618	0.853
+2.0	0.965	0.989	0.980	0.960	0.985	0.933	0.987
+3.0	0.999	0.999	0.996	0.992	0.998	0.987	0.999
Panel B: CAR(-1,+1)							
AR	ORDIN	PATELL	BMP	GRANK	SIGN-COWAN	SIGN-GSAR-T	SIGN-GSAR-Z
−3.0	0.928	0.976	0.956	0.923	0.954	0.874	0.960
−2.0	0.750	0.895	0.880	0.779	0.858	0.680	0.863
−1.0	0.397	0.563	0.572	0.380	0.529	0.290	0.586
±0.0	**0.202**	**0.244**	**0.258**	**0.080**	**0.215**	**0.059**	**0.254**
+1.0	0.421	0.575	0.582	0.382	0.594	0.323	0.618
+2.0	0.774	0.893	0.880	0.805	0.902	0.745	0.905
+3.0	0.950	0.981	0.967	0.940	0.969	0.909	0.976
Panel C: CAR(-5,+5)							
AR	ORDIN	PATELL	BMP	GRANK	SIGN-COWAN	SIGN-GSAR-T	SIGN-GSAR-Z
−3.0	0.591	0.775	0.779	0.603	0.689	0.456	0.730
−2.0	0.405	0.554	0.577	0.348	0.480	0.240	0.532
−1.0	0.244	0.322	0.355	0.140	0.280	0.104	0.337
±0.0	**0.197**	**0.221**	**0.247**	**0.083**	**0.204**	**0.062**	**0.257**
+1.0	0.278	0.350	0.363	0.182	0.331	0.137	0.392
+2.0	0.450	0.586	0.609	0.425	0.594	0.325	0.623
+3.0	0.654	0.797	0.801	0.662	0.810	0.545	0.806
Panel D: CAR(-10,+10)							
AR	ORDIN	PATELL	BMP	GRANK	SIGN-COWAN	SIGN-GSAR-T	SIGN-GSAR-Z
−3.0	0.442	0.583	0.638	0.387	0.486	0.244	0.587
−2.0	0.327	0.442	0.462	0.227	0.337	0.136	0.381
−1.0	0.247	0.295	0.319	0.128	0.227	0.072	0.280
±0.0	**0.206**	**0.239**	**0.249**	**0.087**	**0.214**	**0.068**	**0.257**
+1.0	0.237	0.287	0.315	0.145	0.300	0.104	0.339
+2.0	0.339	0.451	0.476	0.266	0.447	0.190	0.469
+3.0	0.468	0.614	0.626	0.443	0.619	0.330	0.639

Consistent with earlier results (e.g., Kolari & Pynnönen (2010)), test statistics like ORDIN, PATELL and BMP are prone to material over-rejection of the true null hypothesis of no event effect. The results reported in Table 5 indicate that also the test statistic SIGN-COWAN and SIGN-GSAR-Z are prone to material over-rejection of the true null hypothesis of no event effect. According to Table 5 test statistics GRANK and SIGN-GSAR-T are much more robust to cross-correlation caused by event day clustering. However, a notable distinction of the power results in Table 5 of these statistics compared to those in Table 4 is that the powers tend to be discernibly lower in the clustered case. This is due to the information loss caused by cross-correlation. The problem is discussed in more detail in Kolari & Pynnönen (2010).

To summarize, the derived test statistic SIGN-GSAR-T as well as the GRANK statistic are quite robust against clustered event days. In addition the well-established asymptotic properties of SIGN-GSAR-T, its robustness against event-induced volatility, and competitive power properties make it a valuable testing procedure in event studies.

6 Conclusion

We have proposed the nonparametric sign tests SIGN-GSAR-T and SIGN-GSAR-Z based on GSARs. These tests extend the single day sign test statistic presented by Corrado & Zivney (1992) to efficient testing of CARs. We have also derived the theoretical asymptotic distributions of the statistics when the estimation period is finite. The proposed testing procedure based on SIGN-GSAR-T, in particular, has advantages of being well specified under the null hypothesis of no event mean effect and being robust to event-induced volatility and cross-correlation (clustered event days) of the returns. Simulation results with actual stock returns also show that the SIGN-GSAR-T test statistic has good empirical power properties. Our results suggest the use of the test statistic SIGN-GSAR-T particularly in the cases where the event days are clustered.

References

Barber, B. & Lyon J. (1997). Detecting long-run abnormal returns: The empirical power and specification test statistics. *Journal of Financial Economics* 81, 175–213.

Beaver, W. (1968). The information content of annual announcements. *Journal of Accounting Research Supplement* 6, 67–92.

Berry, M., Gallinger G. & Henderson, G. (1990). Using daily stock returns in event studies and the choice of parametric versus nonparametric test statistics. *Quarterly Journal of Business and Economics* 29, 70–85.

Boehmer, E., Musumeci J. & Poulsen, A. (1991). Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics* 30, 253–272.

Brown, S. & Warner, J. (1980). Measuring security price performance. *Journal of Financial Economics* 8, 205–258.

Brown, S. & Warner, J. (1985). Using daily stock returns: The case of event studies. *Journal of Financial Economics* 14, 3–31.

Campbell, C. & Wasley, C. (1993). Measuring security price performance using daily NASDAQ returns. *Journal of Financial Economics* 33, 73–92.

Campbell, C. & Wasley, C. (1996). Measuring abnormal trading volume for samples of NYSE/ASE and NASDAQ securities using parametric and non-parametric test statistics. *Review of Quantitative Finance and Accounting* 6, 309–326.

Campbell, J., Lo, A. & MacKinlay, A. (1997). *The Econometrics of Financial Markets*. Princeton: Princeton University Press.

Charest, G. (1978). Dividend information, stock returns, and market efficiency. *Journal of Financial Economics* 6, 265–296.

Corrado, C. (1989). A nonparametric test for abnormal security price performance in event studies. *Journal of Financial Economics* 23, 385–395.

Corrado, C. & Zivney, T. (1992). The specification and power of the sign test in event study hypothesis tests using daily stock returns. *Journal of Financial and Quantitative Analysis* 27, 465–478.

Cowan, A. (1992). Nonparametric event study tests. *Review of Quantitative Finance and Accounting* 2, 343–358.

Cowan, A. & Sergeant, A. (1996). Trading frequency and event study test specification. *Journal of Banking & Finance* 20, 1731–1757.

Dechow, P. Sloan, R. & Sweeney, A. (1995). Detecting earnings management. *Accounting Review* 70, 3–42.

Harrington, S. & Shrider, D. (2007). All events induce variance: Analyzing abnormal returns when effects vary across firms. *Journal of Financial and Quantitative Analysis* 42, 229–256.

Kolari, J. & Pynnönen, S. (2010). Event study testing with cross-sectional correlation of abnormal returns. *The Review of Financial Studies* 23, 3996–4025.

Kolari, J. & Pynnönen, S. (2011). Nonparametric rank tests for event studies. *Journal of Empirical Finance* 18, 953–971.

Kothari, S., Leone, A. & Wasley, C. (2005). Performance-matched discretionary accruals. *Journal of Accounting and Economics* 39, 163–197.

Lehmann, E. (1999). *Elements of Large-Sample Theory*, Springer-Verlag, New York.

Mikkelson, W. (1981). Convertible calls and security returns. *Journal of Financial Economics* 9, 237–264.

Patell, J. (1976). Corporate forecasts of earnings per share and stock price behavior: Empirical test. *Journal of Accounting Research* 14, 246–276.

Penman, S. (1982). Insider trading and the dissemination of firms forecast information. *Journal of Business* 55, 479–503.

Pynnönen, S. (2010). *Distribution of Linear Transformations of Internally Studentized Least Squared Residuals.* Available at SSRN: http://ssrn.com/abstract=1691997.

Rosenstein, S. & Wyatt, J. (1990). Outside directors, board independence, and shareholders wealth. *Journal of Financial Economics* 26, 175–191.

Appendices

Appendix 1. The properties of the sign of the GSAR

We derive the theoretical expectation and variance of G_{it} as well as the theoretical covariance between G_{it} and G_{is}, $t \neq s$, $t, s = 1, \ldots, T$, in both of the cases $T = L_1 + 1$ being even and odd.

Using equations from (10) to (13) it is straightforward to see that

$$E[G_{it}] \quad = \quad 0 \tag{45}$$

and

$$\text{var}[G_{it}] = \begin{cases} 1, & \text{for even } T \\ \frac{T-1}{T}, & \text{for odd } T. \end{cases} \tag{46}$$

Again, if $t \neq s$, it is straightforward to verify the following probabilities

$$\Pr[G_{it}G_{is} = 1] = \begin{cases} \frac{\frac{T}{2}-1}{T-1}, & \text{for even } T \\ \frac{T-3}{2T}, & \text{for odd } T, \end{cases} \tag{47}$$

$$\Pr[G_{it}G_{is} = 0] = \begin{cases} 0, & \text{for even } T \\ \frac{2}{T}, & \text{for odd } T \end{cases} \tag{48}$$

and

$$\Pr[G_{it}G_{is} = -1] = \begin{cases} \frac{\frac{T}{2}}{T-1}, & \text{for even } T \\ \frac{T-1}{2T}, & \text{for odd } T. \end{cases} \tag{49}$$

Furthermore for T being even

$$\text{cov}[G_{it}, G_{is}] = E[G_{it}G_{is}] = -\frac{1}{T-1} \tag{50}$$

and for T being odd

$$\text{cov}[G_{it}, G_{is}] = E[G_{it}G_{is}] = -\frac{1}{T}. \tag{51}$$

Appendix 2. The asymptotic distributions of Z_1 and SIGN-GSAR-T

The following Lemmas are utilized in the proofs of Theorem 1 and Theorem 2. Proofs of these Lemmas can be obtained as special cases from Pynnönen (2010).

Lemma 1 *Define*

$$x = \mathbf{Q}y, \tag{52}$$

where $\mathbf{Q}$ *is a* $T \times T$ *idempotent matrix of rank* $r \leq T$ *and* $y = (y_1, \ldots, y_T)'$ *is a vector of independent* $N(0,1)$ *random variables, such that* $y \sim N(0, \mathbf{I})$*, where* $\mathbf{I}$ *is a* $T \times T$ *identity matrix. Furthermore, let* m *be a* T *component column vector of real numbers such that* $m'\mathbf{Q}m > 0$*. Then*

$$z_m = \frac{m'x/\sqrt{m'\mathbf{Q}m}}{\sqrt{x'x/r}} \tag{53}$$

has the distribution with density function

$$f_{z_m}(z) = \frac{\Gamma(r/2)}{\Gamma\left[(r-1)/2\right]\sqrt{r\,\pi}}\left(1 - \frac{z^2}{r}\right)^{\frac{1}{2}(r-1)-1}, \tag{54}$$

where $|z| < \sqrt{r}$*, and zero otherwise, and where* $\Gamma(\cdot)$ *is the gamma function.*

Lemma 2 *Under the assumptions of Lemma* 1

$$t_m = z_m\sqrt{\frac{r-1}{r-z_m^2}} \tag{55}$$

is distributed as the Student t*-distribution with* $r-1$ *degrees of freedom.*

Proof of the Theorem 1: The proof of the theorem is adapted from Kolari and Pynnönen (2011). In order to derive the asymptotic distribution of the Z_1 defined in equation (19), the G_{it}s defined in (9) are collected to a column vector $G_i = (G_{i,T_0+1}, G_{i,T_0+2}, \ldots, G_{i,T_1}, G_0)'$ of $T = T_1 - T_0 + 1$ components, where the prime denotes transpose and $i = 1, \ldots, n$ with n the number of series. Then by assumption the random vectors G_is are independent and, by Proposition 1, identically distributed random vectors with zero means and identical equicorrelation covariance matrices such that

$$E\left[G_i\right] \;=\; 0 \tag{56}$$

and

$$\text{cov}\left[G_i\right] = \begin{cases} (1-\varrho)I + \varrho\iota\iota', & \text{for even } T \\ \frac{T-1}{T}\left[(1-\varrho)I + \varrho\iota\iota'\right], & \text{for odd } T. \end{cases} \tag{57}$$

Again $i = 1, \ldots, n$, where ι is a vector of T ones, I is a $T \times T$ identity matrix, and

$$\varrho = -\frac{1}{T-1}. \tag{58}$$

Thus, the covariance matrix in (57) becomes

$$\Sigma = \operatorname{cov}[G_i] = \begin{cases} \frac{T}{T-1}\left(\mathbf{I} - \frac{1}{T}\iota\iota'\right), & \text{for even } T \\ \left(\mathbf{I} - \frac{1}{T}\iota\iota'\right), & \text{for odd } T. \end{cases} \tag{59}$$

It should be noted that the matrix $\mathbf{I} - T^{-1}\iota\iota'$ is an idempotent matrix of rank $T-1$, which implies that Σ is singular in both of the cases for T being even or odd.

However, because G_is are independent with zero means and finite covariance matrices (59), the Central Limit Theorem applies such that

$$\sqrt{n}\,\bar{G} \xrightarrow{d} \left(\frac{T}{T-1}\right)^{\frac{1}{2}} x, \tag{60}$$

when T is even and

$$\sqrt{n}\,\bar{G} \xrightarrow{d} x, \tag{61}$$

when T is odd, as $n \to \infty$, where

$$x \sim N(0, \mathbf{Q}) \tag{62}$$

with the (idempotent) singular covariance matrix

$$\mathbf{Q} = \mathbf{I} - \frac{1}{T}\iota\iota', \tag{63}$$

and in (60) and (61), $\bar{G} = (\bar{G}_{T_0+1}, \ldots, \bar{G}_{T_1}, \bar{G}_0)'$ with

$$\bar{G}_t = \frac{1}{n}\sum_{i=1}^{n} G_{it}, \tag{64}$$

where $t \in \{T_0+1, \ldots, T_1, 0\}$. Note that the sum of $G_{i,t}$ over the time index t is zero for all $i = 1, \ldots, n$, i.e., $\iota' G_i = 0$ for all $i = 1, \ldots, n$, which implies that $\iota'\bar{G} = 0$.

Let ι_0 be a column vector of length $T = T_1 - T_0 + 1$ with one in position in the event day $t = 0$ and zeros elsewhere. In terms of the T-vector $\bar{G}$ and under the assumption that $n_t = n$ for all $t \in \{T_0+1, \ldots, T_1, 0\}$, we can write the Z_1-statistic in equation (19) as

$$Z_1 = \frac{\iota_0'\bar{G}}{\sqrt{\bar{G}'\bar{G}/T}} = \frac{\iota_0'\bar{G}/\sqrt{(T-1)/T}}{\sqrt{\bar{G}'\bar{G}/(T-1)}}. \tag{65}$$

Defining in Lemma 1

$$m = \iota_0 \tag{66}$$

and

$$\mathbf{Q} = \mathbf{I} - \frac{1}{T}\iota\iota', \tag{67}$$

we obtain

$$m'\mathbf{Q}m = \frac{(T-1)}{T}, \tag{68}$$

such that the ratio z_m in (53) becomes

$$z_m = \frac{\iota_0' x/\sqrt{(T-1)/T}}{\sqrt{x'x/(T-1)}}, \tag{69}$$

the distribution of which, after arranging term, has the density function,

$$f_{z_m}(z) = \frac{\Gamma\left[(T-1)/2\right]}{\Gamma\left[(T-2)/2\right]\sqrt{(T-1)\pi}}\left(1 - \frac{z^2}{T-1}\right)^{\frac{1}{2}(T-3)} \tag{70}$$

for $|z| < \sqrt{T-1}$ and zero elsewhere.

Because of the convergence results in (60) and (61) and that the function

$$h(\bar{G}) = \frac{\iota_0'\bar{G}/\sqrt{(T-1)/T}}{\sqrt{\bar{G}'\bar{G}/(T-1)}} \tag{71}$$

is continuous, the continuous mapping theorem implies $h(\bar{G}) \xrightarrow{d} h(x)$. That is,

$$Z_1 = \frac{\iota_0'\bar{G}/\sqrt{(T-1)/T}}{\sqrt{\bar{G}'\bar{G}/(T-1)}} \xrightarrow{d} \frac{\iota_0' x/\sqrt{(T-1)/T}}{\sqrt{x'x/(T-1)}} = z_m, \tag{72}$$

which implies that the density function of the limiting distribution of Z_1 for fixed T, as $n \to \infty$, is of the form defined in equation (70), completing the proof of Theorem 1.

Proof of the Theorem 2: By the proof of Theorem 1, $Z_1 \xrightarrow{d} z_m$, where z_m is defined in equation (69) with $r = T-1$. Because the function $g(z) = z\sqrt{(T-2)/(T-1-z^2)}$ is continuous, for $|z| < \sqrt{T-1}$, the continuous mapping theorem implies $Z_{SGT} = g(Z_1) \xrightarrow{d} g(z_m)$. That is,

$$Z_{\mathrm{SGT}} \xrightarrow{d} z_m\sqrt{\frac{T-2}{T-1-z_m^2}}, \tag{73}$$

where the distribution of the right hand side expression is by Lemma 2 the t-distribution with $T-2$ degrees of freedom, completing the proof of Theorem 2.

A STATISTICAL COMPARISON OF ALTERNATIVE IDENTIFICATION SCHEMES FOR MONETARY POLICY SHOCKS

Markku Lanne
*University of Helsinki and CREATES**
Helmut Lütkepohl
DIW Berlin and Freie Universität Berlin†

1 Introduction

In a standard structural vector autoregressive (SVAR) approach, it is typical that just enough restrictions are imposed to just-identify the structural shocks. Clearly, any assumptions regarding certain restrictions imposed on a model may be incorrect and imposing false restrictions may lead to biased results and wrong conclusions. Therefore the desire to impose as few restrictions as possible is understandable. The drawback is, however, that different just-identified models cannot be compared by statistical tests. Comparisons therefore often rely on plausibility checks. For example, one may prefer one model to another one because the impulse responses of the former model have more plausible shapes.

Monetary policy is an active research area where different SVAR models coexist. For example, Christiano, Eichenbaum & Evans (1999) (henceforth CEE) review a number of identification schemes for SVAR models which have been used in the related literature to specify monetary policy shocks and analyze their impact on the economy. In some of these schemes, the structural restrictions just-identify the monetary policy shocks. Thus, in the standard setup, the different schemes cannot be checked against the data with statistical tests but a decision on which scheme to use is based on the subjective opinion of specific researchers. For example, CEE suppose that certain reactions to monetary policy shocks are widely accepted in the profession and therefore only identification schemes for monetary policy shocks should be considered which imply these widely accepted responses of the variables. Even if one accepts this strategy, it may not lead to a unique set of shocks.

Therefore, in this study we will apply different approaches that use certain statisti-

*The first author acknowledges financial support from CREATES (DNRF78) funded by the Danish National Reseach Foundation, and from the Academy of Finland.
†The final part of the research for this paper was conducted while the second author was Bundesbank Professor at the Freie Universität Berlin. Support from Deutsche Forschungsgemeinschaft (DFG) through SFB649 “Economic Risk” is gratefully acknowledged.

cal properties of the data for generating additional identifying information for the structural shocks. The first approach of this kind utilizes the change in the volatility of the shocks. It was proposed by Rigobon (2003), Rigobon & Sack (2003) and Lanne & Lütkepohl (2008). It has been argued in the related literature that there has been a moderation of marcoeconomic fluctuations in the mid 1980s and our first approach uses this feature to identify monetary policy shocks. In this approach the change in the covariance structure of the model is assumed to have occurred at some prespecified point in time. An alternative approach was proposed by Lanne & Lütkepohl (2010). They show how a nonnormal distribution of the residuals can be used as identifying information. In particular, they assume a mixed normal distribution of the residuals and show how this feature of the data can be utilized for identification purposes. Again statistical tests can be applied to check the normality of the data and if rejected allowing for a more general class of distributions is a natural next step.[1] We will consider these two different approaches to compare a set of different identification schemes for US monetary policy.

More precisely, in our empirical analysis we use the same variables as CEE: log of real aggregate output (Y_t) and the log of its deflator (P_t), the smoothed change in an index of sensitive commodity prices ($PCOM_t$), the log of nonborrowed reserves plus extended credit (NBR_t), the log of total reserves (TR_t), the federal funds rate (FF_t) and the log of M1 (M_t). We use monthly US data only while CEE consider both monthly and quarterly data. As monthly data is available for most of the variables of interest here, it is worth utilizing the additional information in the more frequently observed series. Only output is not available in monthly form and therefore proxies as in CEE are used for this variable and its deflator (see Section 5). Our sampling period is 1965M7 - 1995M6 which is also the sampling period used by CEE. Although longer time series are now available, it may be worth considering exactly the same data as CEE to ensure that differences in the results are driven by the different methods used rather than different data.

Our study is structured as follows. In the next section the VAR model setup and the different sets of identifying restrictions considered by CEE will be presented. In Section 3 the 'statistical' identification strategies are presented. Estimation of the structural models is discussed in Section 4 and the results of the empirical analysis are presented in Section 5. Conclusions are drawn in Section 6.

[1]The latter approach can be generalized by allowing the covariance structure to be governed by Markov switching, as discussed by Lanne et al. (2010). However, in this paper we confine ourselves to the simple special case of a mixed normal error distribution that seems sufficient for identification in our empirical application.

2 The Model Setup

CEE consider a K-dimensional reduced form VAR(p) model of the type

$$Z_t = Dd_t + A_1 Z_{t-1} + \cdots + A_p Z_{t-p} + u_t, \tag{1}$$

where d_t is a deterministic term with coefficient matrix D, the A_j's $(j = 1, \ldots, p)$ are $(K \times K)$ coefficient matrices and u_t is a white noise error term. They partition the K-dimensional vector of observable variables Z_t as

$$Z_t = \begin{bmatrix} X_{1t} \\ S_t \\ X_{2t} \end{bmatrix}, \tag{2}$$

where X_{1t} is $(k_1 \times 1)$, S_t is (1×1) and X_{2t} is $(k_2 \times 1)$. The vector X_{1t} contains variables whose contemporaneous values appear in the monetary authority's information set, i.e., variables orthogonal to the monetary policy shock. The variables in X_{2t} only appear with a lag in the information set and S_t is the monetary authority's policy instrument.

The structural shocks are usually obtained from the reduced form residuals by a linear transformation. If only specific structural shocks are of interest, it suffices to find a linear transformation of u_t which delivers these particular shocks of interest and possibly specifies the other shocks in an arbitrary way. In our case the monetary shocks are of primary interest. CEE show that a block triangular transformation $\varepsilon_t = A_0 u_t$, with

$$A_0 = \begin{bmatrix} \underset{(k_1 \times k_1)}{a_{11}} & \underset{(k_1 \times 1)}{0} & \underset{(k_1 \times k_2)}{0} \\ \underset{(1 \times k_1)}{a_{21}} & \underset{(1 \times 1)}{a_{22}} & \underset{(1 \times k_2)}{0} \\ \underset{(k_2 \times k_1)}{a_{31}} & \underset{(k_2 \times 1)}{a_{32}} & \underset{(k_2 \times k_2)}{a_{33}} \end{bmatrix} \tag{3}$$

fully identifies the monetary policy shocks as the (k_1+1)th component of ε_t. Notice that the other elements of A_0 may be chosen such that $A_0^{-1} A_0^{-1\prime} = \Sigma_u$ and, hence, $\varepsilon_t \sim (0, I_K)$.

CEE consider three alternative schemes for just-identifying the monetary policy shocks:

1. *FF policy shock*: $X_{1t} = (Y_t, P_t, PCOM_t)'$, $X_{2t} = (NBR_t, TR_t, M_t)'$ and the federal funds rate is the policy instrument, that is, $S_t = FF_t$. This identification scheme is motivated by arguments presented in Bernanke & Blinder (1992), Sims (1986, 1992) and others.

2. *NBR policy shock*: $X_{1t} = (Y_t, P_t, PCOM_t)'$, $X_{2t} = (FF_t, TR_t, M_t)'$ and $S_t = NBR_t$. This scheme is based on work, for instance, of Christiano & Eichenbaum (1992).

3. *NBR/TR policy shock*: $X_{1t} = (Y_t, P_t, PCOM_t, TR_t)'$, $X_{2t} = (FF_t, M_t)'$ and $S_t = NBR_t$. CEE attribute this identification scheme to Strongin (1995).

As mentioned earlier, in a standard SVAR setting the implied zero restrictions on A_0 suffice to just-identify the monetary policy shocks. They do not provide over-identifying restrictions, however, which could be tested against the data. In the next section it will be explained how such over-identifying information can be obtained from the statistical properties of u_t which are usually not taken into account in a standard SVAR analysis.

3 Statistical Approaches for the Identification of Shocks

3.1 Identification via Heteroskedasticity

Our first approach to identify the structural shocks via specific statistical properties of the data assumes that there is at least one change in the volatility of the residuals and, hence, the residuals of the basic model (2.1) are heteroskedastic. As mentioned earlier, this approach has been used in SVAR analyses by Rigobon (2003), Rigobon & Sack (2003) and Lanne & Lütkepohl (2008). In the first paper the relationship between the returns on different bonds is analyzed. Rigobon & Sack (2003) use the idea of identification via heteroskedasticity to investigate the relation between monetary policy and the stock market. Finally, Lanne & Lütkepohl (2008) use this devise to compare different identification schemes for US monetary policy. The latter study is closely related to the one presented in the empirical section of the present paper. The model setup and sample period are slightly different, however.

To introduce the idea, let us assume that there is a single change in the volatility of the variables during the sample period. Hence, suppose we have a sample of size T, $Z_1, \ldots, Z_T$, and there is a change in the volatility of the shocks during the sample period, say in period T_B, so that

$$E(u_t u_t') = \begin{cases} \Sigma_1 & \text{for } t = 1, \ldots, T_B - 1, \\ \Sigma_2 & \text{for } t = T_B, \ldots, T. \end{cases} \tag{4}$$

From matrix theory it is well-known that the covariance matrices Σ_1 and Σ_2 can be diagonalized simultaneously, that is, there exists a $(K \times K)$ matrix W and a diagonal matrix $\Psi = \text{diag}(\psi_1, \ldots, \psi_K)$ with positive diagonal elements ψ_i, $i =$

$1, \ldots, K$, such that $\Sigma_1 = WW'$ and $\Sigma_2 = W\Psi W'$ (e.g., Lütkepohl (1996, Section 6.1.2)). Here the diagonal elements of Ψ reflect the changes in the variances of the shocks after the possible change in volatility has occurred. In fact, a change in volatility has occurred if the ψ_i's are different from one. Lanne & Lütkepohl (2008) show that W is unique except for sign changes if all ψ_i's are distinct and ordered in some way. Thus, if we choose $A_0^{-1} = W$, we get uniqueness of the shocks $\varepsilon_t = A_0 u_t$ (except for changes in sign) without the need for further identifying assumptions. Choosing $A_0 = W^{-1}$, the structural shocks have identity covariance matrix in the first regime and Ψ in the second regime, that is,

$$E(\varepsilon_t \varepsilon_t') = \begin{cases} I_K & \text{for } t = 1, \ldots, T_B - 1, \\ \Psi & \text{for } t = T_B, \ldots, T. \end{cases}$$

Hence, they are orthogonal in both regimes. In turn, requiring that they are orthogonal in both regimes suffices to identify our shocks uniquely if all ψ_i's are distinct. Notice that orthogonality of the structural shocks is a standard assumption in SVAR analysis.

The requirement that all ψ_i's are distinct is satisfied if the changes in volatility are not proportional in all variables. Even if the volatility in one of the shocks does not change at all, that is, one of the ψ_i's may be unity, the ψ_i's may of course be all distinct and this is the essential requirement for uniqueness of A_0. Any other restrictions for A_0, as for example formulated in Section 2, then become over-identifying. Since the change in variance is a testable assumption, we do not have to rely exclusively on information from economic theory or other sources to ensure identification of the structural shocks. Instead we can use the information in the data and apply statistical procedures to obtain identification.

To obtain uniqueness of W and, hence, of A_0 we may, for example, order the ψ_i's from smallest to largest. If restrictions are imposed on A_0, they may only be compatible with one specific ordering of the ψ_i's and not necessarily with an ordering according to size. Therefore, in estimations with restrictions on A_0 we do not impose any specific ordering on the ψ_i's but let the data decide which one is best. This is no problem in principle because local identification is ensured for any ordering of the ψ_i's, provided they are all distinct.

The fact that it is always possible to reverse the signs of all elements in a single column of $W = A_0^{-1}$ without affecting the likelihood is not a problem either in the present context because for asymptotic theory to work we only need local identification which is ensured despite the sign changes. For practical purposes changing all signs in a column of W just means to consider a negative shock if the shock is positive originally or vice versa.

It is also possible, of course, to accommodate more than one change in volatility. If there are $n+1$ different regimes and the covariances in the different regimes are WW', $W\Psi_1W'$, ..., $W\Psi_nW'$, where the Ψ_i's are all diagonal matrices, uniqueness of W (up to sign) is ensured, for example, if the diagonal elements in only one of the Ψ_i matrices are all distinct and we can again choose $A_0 = W^{-1}$.

So far we have assumed that only the residual covariance matrix changes and the other VAR parameters remain constant. This assumption was made for convenience because it ensures identification of the shocks. Obviously, statistical tests may be used to check the constancy of the other parameters as well. CEE and Bernanke & Mihov (1998b) argue that structural changes found by other authors in the data set underlying our empirical study may have been due to a change in the residual covariances only and not to a change in the whole dynamic structure. While identification of the shocks can also be achieved if other parameters vary as well, the impulse responses may be affected if there are changes in the other parameters. Such changes would therefore complicate our analysis.

If uniqueness of A_0 is ensured by residual heteroskedasticity, then all the restrictions from economic theories are over-identifying and, hence, can be tested. In particular, the just-identifying restrictions discussed in the previous section can be tested and we will do so in the empirical section. In other words, it can be checked whether they are compatible with orthogonality of the shocks across the different regimes. In the next subsection we will present another approach to serve the same purpose.

3.2 Mixed Normal Residuals

In our analysis of the reduced form VAR model for the time series described in Section 1, we will find strong evidence that the errors are not normally distributed. Hence, it makes sense to specify a more general distribution. A quite general class of distributions is given by a mixture of two normal distributions. Therefore we assume that u_t is a mixture of two serially independent normal random vectors such that

$$u_t = \begin{cases} e_{1t} \sim \mathcal{N}(0, \Sigma_1) \text{ with probability } \gamma, \\ e_{2t} \sim \mathcal{N}(0, \Sigma_2) \text{ with probability } 1-\gamma. \end{cases} \tag{5}$$

Here $\mathcal{N}(0, \Sigma)$ denotes a multivariate normal distribution with zero mean and covariance matrix Σ. The $(K \times K)$ covariance matrices Σ_1 and Σ_2 are assumed to be distinct and the mixture probability γ, $0 < \gamma < 1$, is a parameter of the model. If $\Sigma_1 = \Sigma_2$, u_t has a $\mathcal{N}(0, \Sigma_1)$ distribution and γ is not identified. Therefore we assume $\Sigma_1 \neq \Sigma_2$. We note that u_t has mean zero and covariance matrix $\Sigma_u = \gamma\Sigma_1 + (1-\gamma)\Sigma_2$, that is, $u_t \sim (0, \gamma\Sigma_1 + (1-\gamma)\Sigma_2)$.

This model was proposed by Lanne & Lütkepohl (2010) who also show that a $(K \times K)$ matrix W and a diagonal matrix $\Psi = \text{diag}(\psi_1, \ldots, \psi_K)$, $\psi_i > 0$ $(i = 1, \ldots, K)$, exist such that $\Sigma_1 = WW'$ and $\Sigma_2 = W\Psi W'$. Thus, we may parameterize Σ_u as

$$\Sigma_u = W(\gamma I_K + (1-\gamma)\Psi)W'. \tag{6}$$

If all ψ_i's are distinct, then, for a given ordering of the ψ_i's, the matrix W in this decomposition is unique except that all signs of a column may be reversed. If we choose

$$A_0^{-1} = W(\gamma I_K + (1-\gamma)\Psi)^{1/2} \tag{7}$$

so that $\Sigma_u = A_0^{-1}A_0^{-1\prime}$, this decomposition of Σ_u is the unique one (apart from sign changes) which diagonalizes both Σ_1 and Σ_2 and also, of course, Σ_u. In other words, if we think of the two normal distributions which are mixed in (5) as representing two different regimes, the structural shocks are uncorrelated in both regimes. Hence, as in the heteroskedastic model, any additional restrictions on A_0 are over-identifying and testable. Note that this model differs from the previous one in that the allocation of time periods to the regimes is governed by a random mechanism whereas in Section 3.1 we have assumed that the regime changes occur at fixed, prespecified time points. Also, the residual distribution in (5) is not heteroskedastic.

4 Estimation

For our mixed normal model, maximum likelihood (ML) estimation is in principle the method of choice because we have made an assumption regarding the distribution of the residuals. Also for the heteroskedastic model using Gaussian ML estimation is useful because it results in estimators with desirable asymptotic properties even if the actual residual distribution is non-Gaussian. For both models the likelihood function and its normal equations are nonlinear in the parameters, however. Therefore maximizing the full likelihood function may be a formidable task if the dimension of the process, K, and/or the VAR order, p, are large, as in the case of our empirical example. Because the VAR coefficients can be estimated consistently by equation-wise OLS with standard asymptotic properties, it is in fact possible to estimate the structural parameters based on a "concentrated likelihood function" where the VAR coefficients are replaced by their OLS estimators. The resulting estimation methods for the two models of interest here will be discussed next. We denote by $\hat{u}_t = Z_t - \hat{D}d_t - \hat{A}_1 Z_{t-1} - \cdots - \hat{A}_p Z_{t-p}$ the residuals from estimating the reduced form VAR model (1) by equation-wise OLS.

4.1 Heteroskedastic Residuals

Let us focus on the case of two regimes with different residual covariance matrices as in (4) with $\Sigma_1 = WW'$ and $\Sigma_2 = W\Psi W'$ and define

$$\widetilde{\Sigma}_1 = \frac{1}{T_B - 1}\sum_{t=1}^{T_B-1} \hat{u}_t\hat{u}_t' \quad \text{and} \quad \widetilde{\Sigma}_2 = \frac{1}{T - T_B + 1}\sum_{t=T_B}^{T} \hat{u}_t\hat{u}_t'.$$

Replacing the VAR parameter estimators in the Gaussian log-likelihood function by their OLS estimators gives a "concentrated log likelihood function" of the form

$$\begin{aligned}\log L_H = &-\frac{T_B - 1}{2}\left(\log\det(WW') + \text{tr}\left\{\widetilde{\Sigma}_1(WW')^{-1}\right\}\right)\\ &-\frac{T - T_B + 1}{2}\left(\log\det(W\Psi W') + \text{tr}\left\{\widetilde{\Sigma}_2(W\Psi W')^{-1}\right\}\right). \qquad (8)\end{aligned}$$

Maximizing this function gives estimators $\widetilde{W}$ and $\widetilde{\Psi}$ of W and Ψ, respectively. Note, however, that these estimators are not full ML estimators even if the true residual distribution is Gaussian because the OLS estimators of the VAR coefficients from (1) are not ML estimators. They do not account for the heteroskedasticity in the residuals. We will use the estimators of the structural parameters obtained from maximizing (8) and refer the reader to Lanne & Lütkepohl (2008) for further discussion of the estimation procedure and the properties of the estimators.

Any over-identifying restrictions imposed on the structural parameters can be tested by likelihood ratio (LR) type tests. Thus, if the identification schemes considered in Section 2 imply over-identifying restrictions we can check them against the data by LR type tests based on optimizing the objective function (8) with and without restrictions. Clearly, the resulting tests are not really LR tests because they are based on maximizing the pseudo concentrated likelihood in (8) rather than the fully maximized likelihood function. Still it can be seen from the discussion in Lanne & Lütkepohl (2008) that they have the usual asymptotic properties of standard LR tests. Therefore we may use a χ^2 distribution with as many degrees of freedom (df) as there are zero restrictions imposed on A_0, provided all ψ_i's are distinct. If the latter condition is not satisfied, the asymptotic distribution of our pseudo LR test will still be χ^2 under general conditions. The number of df may be smaller than the number of zeros placed on A_0, however. In other words, our tests may be conservative when used with critical values from a χ^2 distribution with as many df as there are zero restrictions on A_0.

It is straightforward to extend the estimation procedure to the case of more than two regimes. We will not present the details to save space. In the empirical analysis models with up to three regimes will be used. It may also be worth noting that Rigobon (2003) has shown for his slightly less general setup that, under suitable conditions, the time invariant parameters may be estimated consistently under usual assumptions even if the break times are fixed incorrectly.

4.2 Mixed Normal Residuals

For the mixed normal model (5) with $\Sigma_1 = WW'$ and $\Sigma_2 = W\Psi W'$ we estimate the parameters γ, Ψ and W by maximizing the pseudo concentrated likelihood function

$$L_{MN}\left(W, \Psi, \gamma\right) = \prod_{t=1}^{T} \hat{f}_{t-1}(Z_t), \tag{9}$$

where

$$\begin{aligned} \hat{f}_{t-1}\left(Z_t\right) &= \gamma \det(W)^{-1} \exp\left\{-\frac{1}{2}\hat{u}_t'(WW')^{-1}\hat{u}_t\right\} \\ &\quad + (1-\gamma)\det(\Psi)^{-1/2}\det(W)^{-1}\exp\left\{-\frac{1}{2}\hat{u}_t'(W\Psi W')^{-1}\hat{u}_t\right\}. \end{aligned}$$

Regarding tests of over-identifying restrictions for the structural parameters the same applies as for the heteroskedastic model. Thus, both model types allow us to test the restrictions for the A_0 parameters presented in Section 2 if at least some of the diagonal elements of Ψ are different and the model is a valid representation of the DGP.

5 Empirical Analysis

In the empirical analysis we use monthly US data for the period 1965M7 - 1995M6 which corresponds to the sample period used by CEE.[2] A similar sample period was also used in studies by Bernanke & Mihov (1998a) and Lanne & Lütkepohl (2008). Our sample size is 360. We use nonfarm payroll employment as proxi for aggregate output and the implicit deflator of personal consumption expenditure as proxi for its deflator, as in CEE. The reduced form model is a 7-dimensional VAR(12) with an intercept.

The literature on the monetary transmission mechanism in the US presents evidence for a number of possible structural breaks during our sample period and in particular

[2]The data were obtained from L. Christiano's homepage http://www.faculty.econ.northwestern.edu/faculty/christiano/research.htm.

changes in the volatility of the shocks are diagnosed by different authors. Thus, our heteroskedastic model may be justified. Moreover, we have applied tests for nonnormality to the residuals of our model and have found clear evidence against Gaussian residuals. Therefore, considering a more general distribution class such as the mixed normals seems also reasonable. In the following we will consider both types of modelling assumptions. Thereby we will also be able to study the robustness of our main results with respect to the identifying assumptions.

5.1 Results for Heteroskedastic Models

As mentioned earlier, different dates of possible volatility changes were found in the literature for similar models during our sample period. In the following we will only consider changes in 1979M10 and 1984M2. Bernanke & Mihov (1998a) and CEE agree on these dates. According to Bernanke & Mihov (1998a) the choice of these two break dates is based on a "combination of historical and statistical evidence." Moreover, Lanne & Lütkepohl (2008) present further statistical evidence for changes in the residual covariances of models similar to ours in these two months. Notice that the intermediate period 1979M10 - 1984M2 roughly corresponds to the Volcker era which is often regarded as special as far as monetary policy is concerned.

There is some disagreement in the literature regarding the type of structural break. Our assumption of a heteroskedastic model is supported by Bernanke & Mihov (1998b) and CEE. Assuming changes only in the disturbance covariance matrices and, hence, in the volatility of the structural shocks is not uncommon in the related literature (see, for example, Sims & Zha (2006)). In summary, our heteroskedastic model and our assumptions regarding the timing of changes in the volatility are not unconventional and have been confirmed by a variety of methods and authors.

For illustrative purposes and to check the robustness of our results we consider a model with just one change in the residual covariance in 1984M2 and one with two changes in 1979M10 and 1984M2. The evidence for a change in 1984M2 was somewhat stronger than for 1979M10 in the study by Lanne & Lütkepohl (2008). It is therefore plausible to use 1984M2 as break date if only one break is considered.

The estimated ψ_i's for both models are presented in Table 1. These parameters are of particular interest because the shocks are fully identified by assuming orthogonality in both regimes if all the ψ_i's are distinct. Taking into account the standard errors in Table 1, it is not clear that all the ψ_i's are distinct in the two models. On the other hand, the estimates present strong evidence that at least some of the ψ_i's in each of the two models are different. That result is in fact sufficient to test the

Table 1. Estimation Results for Parameters of VAR(12) Models with Heteroskedastic Errors for Sampling Period 1965M7–1995M6

	1 Regime		2 Regimes			
Parameter	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
ψ_1	0.8000	0.1334	0.8735	0.2761	0.4623	0.0801
ψ_2	1.2588	0.2034	1.2161	0.4405	0.7458	0.1284
ψ_3	0.4927	0.0740	0.9791	0.2845	0.7992	0.1397
ψ_4	0.3063	0.0487	5.2158	1.1729	0.6916	0.1179
ψ_5	0.3866	0.0686	1.5875	0.4137	1.8507	0.3110
ψ_6	1.6341	0.2919	1.7791	0.5160	1.4448	0.2481
ψ_7	0.7189	0.1131	0.6639	0.3452	0.3619	0.0757

structural restrictions from Section 2. Notice that if some of the ψ_i's are distinct, at least some of the structural restrictions can be tested. Using the LR type tests mentioned in Section 4 for checking the restrictions, the number of degrees of freedom of the asymptotic χ^2 distributions may be lower than the number of zeros imposed by the different identification schemes. Hence, it would reduce the estimated p-values. Thus, the p-values of our tests based on the assumption that all ψ_i's are distinct would actually be conservative.

In Table 2 results for both models are presented with p-values based on the assumption of distinct ψ_i's. Given that these p-values are conservative, any model that can be rejected on the basis of the p-value in Table 2 can also be rejected if some of the ψ_i's are equal.

The restrictions $a_{12} = 0$ and $a_{13} = 0$ imply that the monetary policy shock is orthogonal to the elements of X_{1t}; a_{12} corresponds to the direct effect of S_t on X_{1t} and a_{13} to the indirect effect via the impact of the shock on X_{2t}. The restriction $a_{23} = 0$ is derived from the assumption that the monetary policy authority does not see X_{2t} when setting S_t. For a correct identification scheme none of the null hypotheses in Table 2 should be rejected. This is clearly not the case for the NBR/TR scheme. It is rejected in both models, with one or two changes in covariance. At least the p-values for the tests of $a_{12} = a_{13} = a_{23} = 0$ are smaller than 1% and, hence, the NBR/TR scheme is clearly rejected even with our conservative tests. In fact, if two regime changes are allowed for, there are even more rejections and, hence, the evidence against the NBR/TR scheme is quite strong in our setup.

For the NBR identification scheme the situation is also quite clear in the model with two structural changes because both $a_{13} = 0$ and $a_{12} = a_{13} = a_{23} = 0$ produce very small p values below 1% and are, hence, rejected at common significance levels. The situation is different, however, if only one change in covariance is allowed for. In that case, $a_{12} = a_{13} = a_{23} = 0$ is the only restriction that can be rejected at the 10% level in the NBR scheme. Given that the model with two breaks is more

Table 2. LR Type Tests of Identification Schemes Based on Heteroskedastic Models

Regime Change in 1984M2

	H_0	df	mean loglik	LR	p-value
NBR/TR	$a_{12} = 0$	4	5.3441	3.4104	0.4916
	$a_{13} = 0$	8	5.3337	10.6070	0.2250
	$a_{23} = 0$	2	5.3485	0.3271	0.8491
	$a_{12} = a_{13} = a_{23} = 0$	14	5.2895	41.3633	0.0002
NBR	$a_{12} = 0$	3	5.3481	0.5916	0.8984
	$a_{13} = 0$	9	5.3388	7.0505	0.6319
	$a_{23} = 0$	3	5.3484	0.4176	0.9366
	$a_{12} = a_{13} = a_{23} = 0$	15	5.3164	22.6687	0.0914
FF	$a_{12} = 0$	3	5.3479	0.7726	0.8560
	$a_{13} = 0$	9	5.3362	8.8949	0.4470
	$a_{23} = 0$	3	5.3477	0.8700	0.8327
	$a_{12} = a_{13} = a_{23} = 0$	15	5.3320	11.7833	0.6954

Regime Change in 1979M10 and 1984M2

	H_0	df	mean loglik	LR	p-value
NBR/TR	$a_{12} = 0$	4	5.4155	26.3575	2.6802e–5
	$a_{13} = 0$	8	5.4233	20.9426	0.0073
	$a_{23} = 0$	2	5.4525	0.6334	0.7286
	$a_{12} = a_{13} = a_{23} = 0$	14	5.3660	60.8736	8.3617e–8
NBR	$a_{12} = 0$	3	5.4516	1.2528	0.7404
	$a_{13} = 0$	9	5.4119	28.8631	0.0007
	$a_{23} = 0$	3	5.4514	1.3433	0.7189
	$a_{12} = a_{13} = a_{23} = 0$	15	5.3991	37.7858	0.0010
FF	$a_{12} = 0$	3	5.4494	2.7770	0.4273
	$a_{13} = 0$	9	5.4402	9.1872	0.4202
	$a_{23} = 0$	3	5.4512	1.5451	0.6719
	$a_{12} = a_{13} = a_{23} = 0$	15	5.4284	17.4070	0.2951

credible and given that our tests are potentially asymptotically conservative, these results present considerable evidence also against the NBR identification scheme.

The situation is quite different for the FF identification scheme. Here none of the p-values is even close to a reasonable significance level for a usual test. In fact, all p-values are bigger than 20%. Thus, the FF scheme is the only one which can stand up against the data in our setup. Since our tests are potentially conservative if not all ψ_i's are distinct, it may well be that our test has not enough power to show that even this scheme is not compatible with the data. However, if any one of the three identification schemes is consistent with the data, it is the FF scheme, at least in our testing framework.

One lesson to be learned from this exercise is that the model setup has a substantial impact on the results. Ignoring one of the changes in the residual covariance matrix can make a substantial difference. Nevertheless there is also a considerable robustness in our results. Even if the change in 1979M10 is ignored, the results point in the same direction as in the model which allows for two changes. A sufficiently critical interpretation of the p-values would result in the same overall conclusions in both models.

Our general result is to some extent in line with the findings of Lanne & Lütkepohl (2008)[3]. Using a slightly different setup in which they also test restrictions on the deeper parameters of the monetary models, they find statistical evidence against all the models. However, the evidence against the FF model is not as strong as that against the other models. The latter result is in line with our tests presented in Table 2.

5.2 Results for Model with Mixed Normal Residuals

As mentioned earlier, there is substantial statistical evidence against Gaussian residuals. Therefore fitting a mixed normal distribution to the residuals using the method described in Section 4 becomes a plausible alternative to the approach used in the previous subsection. The estimated ψ_i's are given in Table 3. Clearly, taking into account the estimated standard errors, there is strong evidence that at least some ψ_i's are distinct. Therefore we proceed under this assumption in the following. Again, our tests of restrictions for the structural parameters may be conservative if some of the ψ_i's are in fact identical.

[3]The article contains a slight error which is corrected in an Erratum available on Markku Lanne's web pages at http://blogs.helsinki.fi/lanne/.

Table 3. Estimation Results for Parameters of VAR(12) Models with Mixed Normal Errors for Sampling Period 1965M7 - 1995M6

Parameter	Estimate	Std Err
ψ_1	0.3400	0.0887
ψ_2	0.6440	0.1581
ψ_3	1.6064	0.4045
ψ_4	2.0717	0.5018
ψ_5	3.1103	0.7102
ψ_6	5.1202	1.1495
ψ_7	7.0656	1.5924
γ	0.8014	0.0441

Table 4. LR Type Tests of Identification Schemes Based on Mixed Normal Model

	H_0	df	mean loglik	LR	p-value
NBR/TR	$a_{12} = 0$	4	5.3722	19.2931	0.0007
	$a_{13} = 0$	8	5.3953	3.2086	0.9206
	$a_{23} = 0$	2	5.3976	1.5938	0.4507
	$a_{12} = a_{13} = a_{23} = 0$	14	5.3707	20.3441	0.1197
NBR	$a_{12} = 0$	3	5.3847	10.6070	0.0141
	$a_{13} = 0$	9	5.3628	25.8216	0.0022
	$a_{23} = 0$	3	5.3974	1.7330	0.6296
	$a_{12} = a_{13} = a_{23} = 0$	15	5.3625	26.0513	0.0375
FF	$a_{12} = 0$	3	5.3979	1.4059	0.7042
	$a_{13} = 0$	9	5.3832	11.6023	0.2367
	$a_{23} = 0$	3	5.3987	0.8352	0.8410
	$a_{12} = a_{13} = a_{23} = 0$	15	5.3745	17.6575	0.2811

Interestingly, the results of the pseudo LR tests presented in Table 4 are fully in line with those from the heteroskedastic model. The NBR/TR and NBR models are strongly rejected because some of the p-values are smaller than 1%, whereas the FF model can not be rejected at common significance levels. Thus, letting the data decide on the allocation of regimes rather than fixing the change dates as in the heteroskedastic model, produces basically the same conclusions regarding the different identification schemes for monetary policy shocks. Thus, our results are overall quite robust to variations in our identifying assumptions.

6 Conclusions

In this study we have compared three identification schemes for monetary policy shocks which cannot be tested in a standard SVAR framework because in that setting there are no over-identifying restrictions. We utilize the fact that the underlying reduced form VAR model has a potentially changing covariance structure and that the residuals are clearly nonnormal. These data features allow us to get additional identifying information and enable us to test the identification schemes for the monetary policy shocks against the data. Only one of the three identification schemes is not rejected in this framework. More precisely, a scheme where monetary shocks are induced via the federal funds rate is the only one which cannot be rejected in our framework. This result is robust with respect to the specific statistical setup used. It is obtained for both the heteroskedastic model and a model with mixed normal residuals.

References

Bernanke, B.S. & Blinder, A. (1992). The federal funds rate and the channels of monetary transmission. *American Economic Review* 82, 901–921.

Bernanke, B.S. & Mihov, I. (1998a). Measuring monetary policy. *Quarterly Journal of Economics* 113, 869–902.

Bernanke, B.S. & Mihov, I. (1998b). The liquidity effect and long-run neutrality. *Carnegie-Rochester Conference Series on Public Policy* 49, 149–194.

Christiano, L.J. & Eichenbaum, M. (1992). Identification and the liquidity effect of a monetary policy shock. In A. Cukierman, Z. Hercowitz & L. Leiderman (Eds.)

Political Economy, Growth, and Business Cycles. Cambridge, MA: MIT Press. 335–370.

Christiano, L.J., Eichenbaum, M. & Evans, C. (1999). Monetary policy shocks: What have we learned and to what end? In J.B. Taylor & M. Woodford (Eds) *Handbook of Macroconomics, 1A.* Amsterdam: Elsevier. 65–148.

Lanne, M. & Lütkepohl, H. (2008). Identifying Monetary Policy Shocks via Changes in Volatility. *Journal of Money, Credit and Banking* 40, 1131–1149.

Lanne, M. & Lütkepohl, H. (2010). Structural vector autoregressions with non-normal residuals. *Journal of Business & Economic Statistics* 28, 159–168.

Lanne, M., Lütkepohl, H. & Maciejowska, K. (2010). Structural vector autoregressions with Markov switching. *Journal of Economic Dynamics & Control* 34, 121–131.

Lütkepohl, H. (1996). *Handbook of Matrices*. Chichester: John Wiley & Sons.

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics* 85, 777–792.

Rigobon, R. & Sack, B. (2003). Measuring the reaction of monetary policy to the stock market. *Quarterly Journal of Economics* 118, 639–669.

Sims, C.A. (1986). Are forecasting models usable for policy analysis? *Quarterly Review, Federal Reserve Bank of Minneapolis* 10, 2–16.

Sims, C.A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review* 36, 975–1000.

Sims, C.A. & Zha, T. (2006). Were there regime switches in U.S. monetary policy? *American Economic Review* 96, 54–81.

Strongin, S. (1995). The identification of monetary policy disturbances: Explaining the liquidity puzzle. *Journal of Monetary Economics* 35, 463–498.

TESTS OF COINTEGRATION RANK WITH STRONG PERSISTENCE IN VOLATILITY: AN APPLICATION TO THE PRICING OF RISK IN THE LONG RUN

Niklas Ahlgren, Paul Catani
Hanken School of Economics

1 Introduction

Financial econometrics, in particular the empirical analysis of prices and returns in financial markets, is an area of research to which Seppo Pynnönen has made many important contributions. We pay our tribute to Seppo Pynnönen by contributing with a chapter combing two themes in financial econometrics. The first theme is the pricing of credit risk in the long run. The second theme is tests of cointegration rank with strong persistence in volatility, in the form of autoregressive conditional heteroskedasticity (ARCH) in the series. ARCH effects are pervasive in financial time series in general, and in the first differences, or returns, of series like credit default swap (CDS) prices and credit spreads in particular.

A credit default swap (CDS) is a credit derivative which provides a bondholder with protection against the risk of default by the company. The equivalence of CDS prices and credit spreads was derived by Duffie (1999). Blanco et al. (2005) test for a long-run relation between CDS prices and credit spreads using cointegration. Many researchers have followed their lead, and there is by now an extensive literature on testing for an equilibrium relation between CDS prices and credit spreads. The empirical studies typically find that the CDS prices and credit spreads are cointegrated for some but not all companies in the sample. Blanco et al. discuss some theoretical reasons for rejecting cointegration between CDS prices and credit spreads. However, several econometric issues arise which are related to the time series properties of CDS prices and credit spreads. First, is it plausible that failure to find support for a long-run relation between CDS prices and credit spreads is caused by a large stationary root in cointegrated systems of CDS prices and credit spreads, so that the tests for cointegration have low power? Second, does an ARCH effect in the series make asymptotic tests for cointegration unreliable? Third, can bootstrap tests be used instead of asymptotic tests, with the purpose of improving the properties of the tests? Fourth, what are the effects of bootstrapping on the size and power of tests for cointegration between CDS prices and credit spreads? In this paper we conduct Monte Carlo simulation experiments designed to answer these questions.

The paper is organised as follows. In Section 2 asymptotic, bootstrap and wild bootstrap (WB) tests of cointegration rank are presented formally. In Section 3 the power of the tests are analysed in situations designed to resemble CDS prices and credit spreads, and the results are presented in the form of simulated power functions. In Section 4 cointegration between CDS prices and credit spreads is tested for a subsample of companies, and in Section 5 the power of the tests is investigated by simulating the data. In Section 6 conclusions are presented.

2 Bootstrap and Wild Bootstrap Tests of Cointegration Rank

We consider the p-dimensional vector autoregressive (VAR) model in error correction form:

$$\Delta \mathbf{X}_t = \mathbf{\Pi X}_{t-1} + \sum_{i=1}^{k-1} \mathbf{\Gamma}_i \Delta \mathbf{X}_{t-i} + \boldsymbol{\alpha\rho}'\mathbf{D}_t + \boldsymbol{\phi}\mathbf{d}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T, \quad (1)$$

where the errors $\boldsymbol{\varepsilon}_t$ are IID$(\mathbf{0}, \mathbf{\Omega})$, and the initial observations are assumed to be fixed. Here $\boldsymbol{\alpha\rho}'\mathbf{D}_t + \boldsymbol{\phi}\mathbf{d}_t$ denotes the deterministic part of the model.

The matrix $\mathbf{\Pi}$ has reduced rank, $\mathbf{\Pi} = \boldsymbol{\alpha\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $p \times r$ matrices of rank $r < p$, the number r being the cointegration rank. The cointegrating vectors are $\boldsymbol{\beta}$. The model for the deterministic terms is $\boldsymbol{\alpha\rho}'\mathbf{D}_t + \boldsymbol{\phi}\mathbf{d}_t$, where $\boldsymbol{\rho}$ is $r \times 1$. We assume that the deterministic terms correspond to a restricted constant, where $\mathbf{D}_t = 1$ and $\mathbf{d}_t = 0$.

We denote the model with rank r by $H(r)$ and the model with rank p by $H(p)$. The likelihood ratio (LR) statistic for testing cointegration rank r against p is (Johansen (1996))

$$Q_{r,T} = -T \sum_{i=r+1}^{p} \ln(1 - \widehat{\lambda}_i), \quad (2)$$

where the eigenvalues $\widehat{\lambda}_1 > \cdots > \widehat{\lambda}_p$ are the p largest solutions to the eigenvalue problem $|\lambda \mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}| = 0$, where $\mathbf{S}_{ij} = T^{-1}\sum_{t=1}^{T} \mathbf{R}_{it}\mathbf{R}'_{jt}$, $i, j = 0, 1$, and $\mathbf{R}_{0t}$ and $\mathbf{R}_{1t}$ are $\Delta \mathbf{X}_t$ and $(\mathbf{X}_{t-1}, 1)'$ corrected for $\Delta \mathbf{X}_{t-1}, \dots, \Delta \mathbf{X}_{t-k+1}$.

Bootstrap algorithms for the LR test of cointegration rank are proposed in Swensen (2006), Swensen (2009) and Cavaliere et al. (2012a). We use the algorithm of Cavaliere et al. (2012a), which recursively generates bootstrap observations $\mathbf{X}^*_{r,t}$

from

$$\Delta \mathbf{X}^*_{r,t} = \widehat{\boldsymbol{\alpha}}^{(r)} \widehat{\boldsymbol{\beta}}^{(r)\prime} \mathbf{X}^*_{r,t-1} + \sum_{i=1}^{k-1} \widehat{\boldsymbol{\Gamma}}^{(r)}_i \Delta \mathbf{X}^*_{r,t-i} + \widehat{\boldsymbol{\alpha}}^{(r)} \widehat{\boldsymbol{\rho}}^{(r)\prime} \mathbf{D}_t + \boldsymbol{\varepsilon}^*_{r,t}, \tag{3}$$

where the parameter estimates $\widehat{\boldsymbol{\alpha}}^{(r)}$, $\widehat{\boldsymbol{\beta}}^{(r)}$, $\widehat{\boldsymbol{\rho}}^{(r)}$ and $\widehat{\boldsymbol{\Gamma}}^{(r)}_1, \ldots, \widehat{\boldsymbol{\Gamma}}^{(r)}_{k-1}$ are the restricted estimates under $H(r)$ of the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\rho}$, and $\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{k-1}$, respectively. The bootstrap errors $\boldsymbol{\varepsilon}^*_{r,t}$ are obtained by resampling from the restricted residuals $\widehat{\boldsymbol{\varepsilon}}_{r,t}$ under $H(r)$.

Algorithm (Bootstrap LR test of cointegration rank (Cavaliere et al. 2012b)).

1. Estimate model (1) under $H(r)$ using Gaussian quasi maximum likelihood estimation (QMLE) to obtain the restricted estimates $\widehat{\boldsymbol{\alpha}}^{(r)}$, $\widehat{\boldsymbol{\beta}}^{(r)}$, $\widehat{\boldsymbol{\rho}}^{(r)}$ and $\widehat{\boldsymbol{\Gamma}}^{(r)}_1, \ldots, \widehat{\boldsymbol{\Gamma}}^{(r)}_{k-1}$ and the restricted residuals $\widehat{\boldsymbol{\varepsilon}}_{r,t}$.
2. Check that the characteristic polynomial $|\widehat{\boldsymbol{\Pi}}^{(r)}(z)| = 0$, where $\widehat{\boldsymbol{\Pi}}^{(r)}(z) = (1-z)\mathbf{I}_p - \widehat{\boldsymbol{\alpha}}^{(r)} \widehat{\boldsymbol{\beta}}^{(r)\prime} z - \sum_{i=1}^{k-1} \widehat{\boldsymbol{\Gamma}}^{(r)}_i (1-z) z^i$, has $p-r$ roots equal to 1 and all other roots outside the unit circle.

   If the condition is satisfied, proceed to step 3.
3. Generate the bootstrap sample recursively from (3) initialised at $\mathbf{X}^*_{r,t} = \mathbf{X}_t$, $t = 1, \ldots, k$, where the bootstrap errors $\boldsymbol{\varepsilon}^*_{r,t}$ are independent draws with replacement from the residuals $\widehat{\boldsymbol{\varepsilon}}_{r,t}$.
4. Compute the bootstrap LR statistic $Q^{\mathrm{B}}_{r,T}$ from the bootstrap sample $\mathbf{X}^*_{r,1}, \ldots, \mathbf{X}^*_{r,T}$ using (2). Define the bootstrap p-value as $p^*_{r,T} = 1 - G^*_{r,T}(Q_{r,T})$, where $G^*_{r,T}(\cdot)$ denotes the conditional (on the original data) cumulative distribution function (CDF) of $Q^{\mathrm{B}}_{r,T}$.
5. The bootstrap test of $H(r)$ against $H(p)$ at the level α rejects $H(r)$ if $p^*_{r,T} \leq \alpha$.

In the empirical application in Section 4 and simulations in Section 5 we use the wild bootstrap (WB) in addition to the IID bootstrap when the errors are (conditionally) heteroskedastic. The WB errors are generated as $\boldsymbol{\varepsilon}^*_{r,t} = \widehat{\boldsymbol{\varepsilon}}_{r,t} w_t$ in step 3 of the bootstrap algorithm, where $\{w_t\}^T_{t=1}$ is a sequence of random variables taking values 1 and -1 with probabilities 0.5.

Cavaliere et al. (2012b) establish the validity of the asymptotic, bootstrap and WB tests under the assumption of the existence of 4th moments of the errors. They show that all tests are valid under conditional heteroskedasticity. If the errors are unconditionally heteroskedastic, the WB test is valid, but the asymptotic and IID bootstrap tests are not (Cavaliere et al. 2012a).

3 Simulated Power Functions

We simulate the power functions for the case of a bivariate system ($p = 2$) and a single cointegrating vector ($r = 1$). The DGP used in the simulations is a VAR(2) process,

$$\Delta \mathbf{X}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{X}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots T, \tag{4}$$

where $\boldsymbol{\alpha} = (a, 0,)'$, $\boldsymbol{\beta} = (1, 0)'$ and

$$\boldsymbol{\Gamma}_1 = \begin{pmatrix} \gamma & \delta \\ \delta & \gamma \end{pmatrix}.$$

The cointegration rank is $r_0 = 0$ when $a = 0$ and $r_0 = 1$ when $a < 0$. The value $a = 0$ therefore gives size of the tests and the values $a < 0$ give power of the tests. In the simulations we have set $\boldsymbol{\rho} = \mathbf{0}$. The parameters γ and δ are not of particular interest for our purposes (see Cavaliere et al. 2012b). In the simulations we set $\gamma = \delta = 0$, in which case $\boldsymbol{\Gamma}_1 = \mathbf{0}$, and then the DGP reduces to a VAR(1) model. The roots of the companion matrix are therefore 1 and $1 + a$. The estimated model is a VAR(2) model with a restricted constant.

For the errors we consider the constant conditional correlation generalised autoregressive conditional heteroskedasticity (CCC-GARCH) model of Bollerslev (1990):

$$\boldsymbol{\varepsilon}_t = \mathbf{D}_t \mathbf{z}_t,$$

where $\mathbf{D}_t = \text{diag}(h_{1t}^{1/2}, h_{2t}^{1/2})$ is a diagonal matrix of conditional standard deviations of $\boldsymbol{\varepsilon}_t$ and $\mathbf{z}_t \sim \text{NID}(\mathbf{0}, \mathbf{P})$ and $\mathbf{P} = (\rho_{ij})$ is a positive definite covariance matrix with ones on the main diagonal. We focus on the CCC-GARCH$(1, 1)$ model with

$$\mathbf{h}_t = \mathbf{a}_0 + \mathbf{A}_1 \boldsymbol{\varepsilon}_{t-1}^{(2)} + \mathbf{B}_1 \mathbf{h}_{t-1},$$

where $\boldsymbol{\varepsilon}_t^{(2)} = (\varepsilon_{1t}^2, \varepsilon_{2t}^2)'$, $\mathbf{h}_t = (h_{1t}, h_{2t})'$ is a (2×1) vector of conditional variances of $\boldsymbol{\varepsilon}_t$, $\mathbf{a}_0$ is a (2×1) vector of positive constants, and $\mathbf{A}_1$ and $\mathbf{B}_1$ are (2×2) parameter matrices which are diagonal with positive diagonal elements. The parameter values are contained in Table 1. In the baseline case of DGP 1, $\mathbf{A}_1 = \mathbf{B}_1 = \mathbf{0}$, and then $\boldsymbol{\varepsilon}_t = \mathbf{z}_t$, $\mathbf{z}_t \sim \text{NID}(\mathbf{0}, \mathbf{I}_2)$. DGP 2 is characterised by very strong persistence in volatility ($a_{ii} + b_{ii} = 0.98$). DGP 3 is characterised by moderate persistence in volatility in the first equation ($a_{11} + b_{11} = 0.85$) and very strong persistence in volatility in the second equation ($a_{22} + b_{22} = 0.999$). The stationarity condition for the CCC-GARCH$(1,1)$ model is $\lambda(\mathbf{\Gamma_C}) < 1$, where λ is the modulus of the largest eigenvalue of a certain matrix $\mathbf{\Gamma_C}$ (see e.g. He & Teräsvirta (2004)). The CCC-GARCH$(1,1)$ processes of DGPs 2 and 3 satisfy the condition for weak and strict stationarity. He and Teräsvirta (2004: 908) give a result concerning the existence of the 4th moment matrix of $\boldsymbol{\varepsilon}_t$. The 4th moment condition is satisfied when $\lambda(\mathbf{\Gamma_{C\otimes C}}) < 1$, where λ is the modulus of the largest eigenvalue of a certain matrix $\mathbf{\Gamma_{C\otimes C}}$. DGP 2 satisfies the condition (the largest eigenvalue is 0.973). DGP 3 does not satisfy the condition (the largest eigenvalue is 1.059). The values for ρ are $\rho = 0$ and 0.5. We report the full results for $\rho = 0$ in the CCC-GARCH models of DGPs 2 and 3, and comment on the results for $\rho = 0.5$.

The series lengths are $T = 250, 500$ and 1000. The number of Monte Carlo replications is $M = 100000$. The computations and simulations are performed in `R` (R Development Core Team (2012)), version $2.15.2$. We use the `ccgarch` package of Nakatani (2013), version $0.2.0$-2, for simulating the CCC-GARCH$(1,1)$ models and checking the 4th moment condition.

Table 1. Parameter values of the DGPs for the errors

DGP			
	DGP 1		
$\mathbf{a}_0 = \begin{pmatrix} 0.15 \\ 0.15 \end{pmatrix}$	$\mathbf{A}_1 = \mathbf{0}$	$\mathbf{B}_1 = \mathbf{0}$	$\rho = 0$
	DGP 2		
$\mathbf{a}_0 = \begin{pmatrix} 0.15 \\ 0.15 \end{pmatrix}$	$\mathbf{A}_1 = \begin{pmatrix} 0.08 & 0 \\ 0 & 0.08 \end{pmatrix}$	$\mathbf{B}_1 = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix}$	$\rho = 0, 0.5$
	DGP 3		
$\mathbf{a}_0 = \begin{pmatrix} 0.15 \\ 0.15 \end{pmatrix}$	$\mathbf{A}_1 = \begin{pmatrix} 0.35 & 0 \\ 0 & 0.175 \end{pmatrix}$	$\mathbf{B}_1 = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.824 \end{pmatrix}$	$\rho = 0, 0.5$

The power functions of the bootstrap and WB LR tests of cointegration rank are simulated using the procedure in Davidson & MacKinnon (2006) for estimating the power of a bootstrap test. Denote by B the number of bootstrap replications. The procedure involves simulating one bootstrap replication instead of B bootstrap replications in each Monte Carlo replication, i.e. $B = 1$, and computing $2M$ LR statistics instead of $M(B+1)$ LR statistics, which substantially reduces the computational burden. The procedure was implemented by Ahlgren & Antell (2013) for estimating the power of bootstrap LR tests of cointegration rank.

Table 2 summarises the simulated size of the asymptotic test (denoted $Q_{r,T}$), bootstrap test (denoted $Q^{\mathrm{B}}_{r,T}$) and WB test (denoted $Q^{\mathrm{WB}}_{r,T}$). The nominal significance level is 5%. The $Q_{r,T}$ test is slightly oversized and the $Q^{\mathrm{B}}_{r,T}$ test is slightly undersized when the errors are generated by DGPs 1 and 2. However, when the errors are generated by DGP 3, both $Q_{r,T}$ and $Q^{\mathrm{B}}_{r,T}$ are oversized, and the size distortions increase with the series length. The $Q^{\mathrm{WB}}_{r,T}$ test, on the other hand, has size close to the nominal 5% level in all DGPs for the errors. The simulations show that $Q_{r,T}$ and $Q^{\mathrm{B}}_{r,T}$ are not valid in DGP 3, for which the condition for the existence of the 4th moment matrix of ε_t is violated. The $Q^{\mathrm{WB}}_{r,T}$ test is correctly sized although it lacks theoretical justification.

Table 2. Simulated size of the asymptotic, bootstrap and wild bootstrap LR tests of cointegration rank. The nominal significance level is 5%.

	$Q_{r,T}$	$Q^{\mathrm{B}}_{r,T}$	$Q^{\mathrm{WB}}_{r,T}$	$Q_{r,T}$	$Q^{\mathrm{B}}_{r,T}$	$Q^{\mathrm{WB}}_{r,T}$	$Q_{r,T}$	$Q^{\mathrm{B}}_{r,T}$	$Q^{\mathrm{WB}}_{r,T}$
	$T=250$			$T=500$			$T=1000$		
DGP 1	0.060	0.031	0.047	0.057	0.031	0.049	0.056	0.033	0.049
DGP 2	0.073	0.043	0.047	0.073	0.044	0.049	0.068	0.043	0.050
DGP 3	0.118	0.076	0.048	0.122	0.086	0.049	0.126	0.092	0.050

The power functions are simulated for 100 points of a between 0 and -0.4 for $T = 250$ in Figure 1, 0 and -0.2 for $T = 500$ in Figure 2 and 0 and -0.1 for $T = 1000$ in Figure 3. The values are chosen in order to zoom in the power functions on points close to 1, which correspond to large stationary roots $1 + a$. As we zoom in, we clearly see the effect of ARCH errors on the power of the tests when there is a large stationary root in the system. The power functions are unadjusted and show the power of the tests at the nominal 5% level. For easy reference, Table 3 summarises the power of the tests when the largest stationary root is equal to 0.98, 0.95 and 0.90.

In the baseline case of DGP 1 with normal errors, the asymptotic $Q_{r,T}$ test has higher power than the bootstrap $Q^{\mathrm{B}}_{r,T}$ and WB $Q^{\mathrm{WB}}_{r,T}$ tests, while $Q^{\mathrm{WB}}_{r,T}$ has higher power than $Q^{\mathrm{B}}_{r,T}$. For example, assume that we have $T = 500$ observations and the largest stationary root is equal to 0.98. Then $Q_{r,T}$ has power 12.0%, $Q^{\mathrm{B}}_{r,T}$ has power 7.8% and $Q^{\mathrm{WB}}_{r,T}$ has power 10.4%. The corresponding powers for $T = 1000$ are 36.4%, 26.0% and 34.4%, respectively. Obtaining high power when the largest root is equal to 0.98 requires a sample of more than $T = 1000$ observations, which corresponds to more than four years of daily observations.

In DGP 2 with very strong persistence in volatility, the asymptotic $Q_{r,T}$ test is more powerful than the bootstrap $Q^{\mathrm{B}}_{r,T}$ and WB $Q^{\mathrm{WB}}_{r,T}$ tests. However, the size distortion of $Q_{r,T}$ also increases compared to the baseline case with normal errors. For example, assume that we have $T = 500$ observations and the largest stationary root is equal to 0.95. Then $Q_{r,T}$ has power 56.1%, $Q^{\mathrm{B}}_{r,T}$ has power 43.0% and $Q^{\mathrm{WB}}_{r,T}$ has power 45.7%. The corresponding powers for $T = 1000$ are 98.6%, 96.7% and 97.5%,

Table 3. Simulated power of the asymptotic, bootstrap and wild bootstrap LR tests of cointegration rank when the largest root is equal to 0.98, 0.95 and 0.9. The nominal significance level is 5%.

	$Q_{r,T}$			$Q^{\mathrm{B}}_{r,T}$			$Q^{\mathrm{WB}}_{r,T}$		
DGP	1	2	3	1	2	3	1	2	3
T				Largest root $= 0.98$					
250	0.066	0.081	0.124	0.040	0.049	0.083	0.051	0.049	0.047
500	0.120	0.142	0.210	0.078	0.099	0.155	0.104	0.099	0.082
1000	0.364	0.395	0.466	0.260	0.295	0.377	0.344	0.320	0.225
				Largest root $= 0.95$					
250	0.166	0.190	0.243	0.103	0.120	0.165	0.139	0.127	0.098
500	0.538	0.561	0.597	0.404	0.430	0.482	0.499	0.457	0.318
1000	0.993	0.986	0.961	0.978	0.967	0.934	0.991	0.975	0.849
				Largest root $= 0.90$					
250	0.531	0.548	0.579	0.380	0.411	0.440	0.477	0.428	0.313
500	0.991	0.984	0.965	0.971	0.959	0.934	0.987	0.969	0.852
1000	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000	0.998

respectively. In DGP 3 with very strong persistence in volatility, we observe large size distortion in both $Q_{r,T}$ and $Q^{\mathrm{B}}_{r,T}$. The $Q^{\mathrm{WB}}_{r,T}$ test is the only test with the correct size in DGP 3. The $Q_{r,T}$ test has the highest rejection probability, followed by the $Q^{\mathrm{B}}_{r,T}$ test, while the $Q^{\mathrm{WB}}_{r,T}$ test has the lowest rejection probability. For large stationary roots the rejection probabilities of $Q_{r,T}$ are seen to be about twice as large as those of $Q^{\mathrm{WB}}_{r,T}$. For example, assume that we have $T = 1000$ observations and the largest stationary root is equal to 0.98. Then $Q_{r,T}$ has power 46.6%, $Q^{\mathrm{B}}_{r,T}$ has power 37.7% and $Q^{\mathrm{WB}}_{r,T}$ has power 22.5%. It should be noted that $Q^{\mathrm{WB}}_{r,T}$ is the only test with the correct size in DGP 3, though.

Changing the correlation coefficient to $\rho = 0.5$ has little effect on the size of the tests, with the exception that the size distortions of $Q_{r,T}$ and $Q^{\mathrm{B}}_{r,T}$ increase in DGP 3. However, there is an increase in power of all tests. For example, when the errors follow DGP 2, $T = 500$ and the largest stationary root is equal to 0.95, $Q_{r,T}$ has power 76.8%, $Q^{\mathrm{B}}_{r,T}$ has power 65.6% and $Q^{\mathrm{WB}}_{r,T}$ has power 67.7%.

The simulations show that all tests have low power if there is a large stationary root in the system and very strong persistence in volatility. In DGP 1 with normal errors, the power of the tests is almost 100% for $T = 1000$ when the largest stationary root is equal to 0.95, and the same is true for $T = 500$ when the largest stationary root is equal to 0.90. In DGPs 2 and 3 with very strong persistence in volatility, obtaining a test with power close to 100% when the largest stationary root is equal to 0.95 requires $T = 1000$ observations, and when the largest stationary root is equal to 0.90 requires $T = 500$ observations. Obtaining high power when the largest stationary root is equal to 0.98 requires a series length of more than $T = 1000$ observations.

The simulations of Lee & Tse (1996) produce high powers for the asymptotic $Q_{r,T}$ test with very strong persistence in volatility, because in their simulations the largest stationary root is equal to 0.80. The conclusion of Ahlgren & Antell (2013) that ARCH errors do not have a large effect on the power of tests of cointegration rank should be qualified when there is a large stationary root in the system and very strong persistence in volatility. Cavaliere et al. (2012b) simulate the power also when unconditional heteroskedasticity is present in the errors. They find that the asymptotic and bootstrap tests are oversized, whereas the WB test has the correct size. The WB test has lower power when there is unconditional heteroskedasticity compared to the case with IID errors.

4 Credit Default Swap Prices

A credit default swap (CDS) is a credit derivative which provides a bondholder with protection against the risk of default by the company. If a default occurs, the holder is compensated for the loss by an amount which equals the difference between the par value of the bond and its market value after the default. The CDS price is the annualised fee (expressed as a percentage of the principal) paid by the protection buyer. Denote by p_t^{CDS} the CDS price and p_t^{CS} the credit spread on a risky bond over the risk-free rate. The basis is the difference between the CDS price and bond spread:

$$s_t = p_t^{\text{CDS}} - p_t^{\text{CS}}.$$

Duffie (1999) derived the equivalence of the CDS price and credit spread. If the two markets price credit risk equally in the long run, then the prices should be equal, so that the basis $s_t = 0$.

Because p_t^{CDS} and p_t^{CS} are $I(1)$ series, the non-arbitrage relation can be tested as an equilibrium relation in the cointegrated VAR model. The vector $\mathbf{X}_t$ with the value 1 appended is $\mathbf{X}_t = (p_t^{\text{CDS}}, p_t^{\text{CS}}, 1)'$. The financial theory posits that $\mathbf{X}_t$ is cointegrated with cointegrating vector $\boldsymbol{\beta} = (1, -1, c)'$, so that $\boldsymbol{\beta}'\mathbf{X}_t = p_t^{\text{CDS}} - p_t^{\text{CS}} + c$ is a cointegrating relation. In theory $c = 0$, but in practice it may be different from zero.

The equivalence of CDS prices and credit spreads of US and European investment-grade companies has been tested in the cointegrated VAR model and it has been found that a parity relation holds for some but not all companies (see e.g. Blanco et al. (2005), Zhu (2006), Dötz (2007) and Forsbäck (2012)). We take a subsample

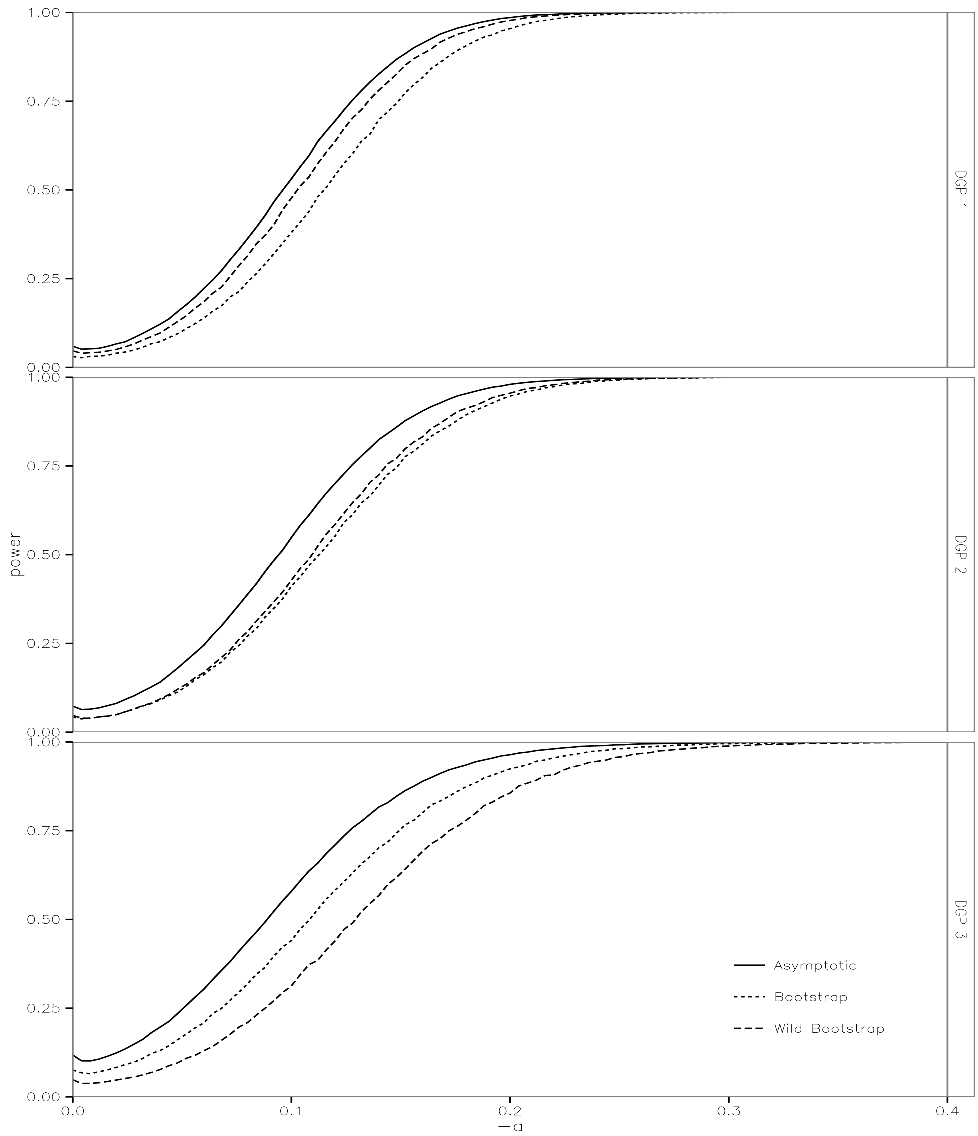


Figure 1. Simulated power functions of the LR tests of cointegration rank for $T = 250$

of the companies in Table 1 of Blanco et al.[1] The companies in our subsample are Bank of America, Citigroup, Goldman Sachs, Barclays Bank and Vodafone, the first three of which are US and the remaining two European companies. We use 5-year maturity CDS prices and credit spreads from Datastream. The data are daily observations from 1 January 2009 to 31 January 2012, and the number of daily observations for each company is $T = 804$.

[1]The data used by Blanco et al. (2005) include 33 companies altogether (referred to as reference entities), which itself is only a small cross section of all US and European reference entities.

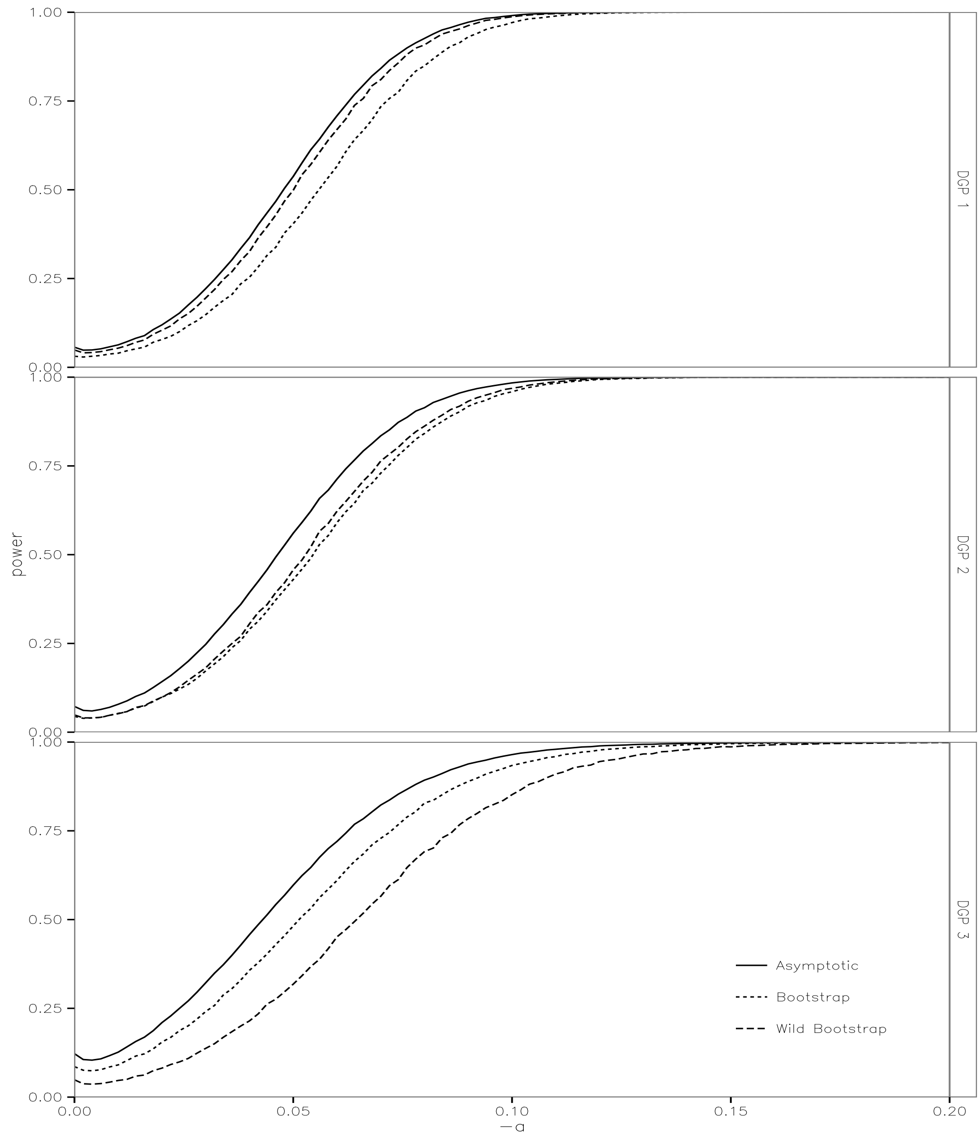


Figure 2. Simulated power functions of the LR tests of cointegration rank for $T = 500$

We use the Schwarz (SC) and Hannan–Quinn (HQ) information criteria to determining the lag length in the VAR models for p_t^{CDS} and p_t^{CS}. The lag length $p = 2$ is selected for Bank of America, $p = 3$ for Citigroup, Goldman Sachs and Vodafone, and $p = 4$ for Barclays Bank. The estimated models include a restricted constant and dummy variables listed in Table 5 to account for outliers. Based on tests for autocorrelation (not reported), we conclude that the VAR models provide good descriptions of the data.

Table 4 reports tests for heteroskedastic and autoregressive conditional heteroskedastic (ARCH) errors. The tests are all significant at the 1% level, with the exception of the ARCH test for the equation for p_t^{CS} for Bank of America. Thus there is strong

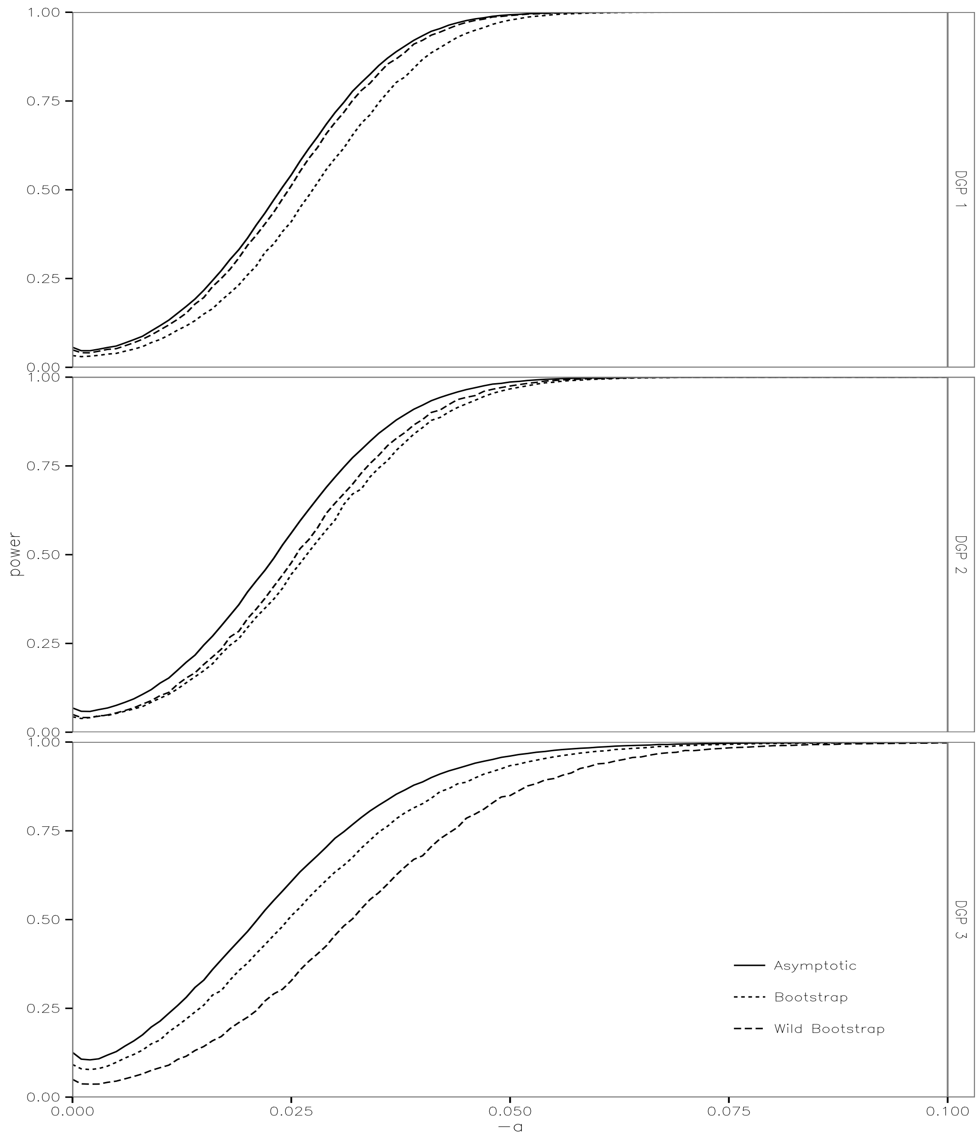


Figure 3. Simulated power functions of the LR tests of cointegration rank for $T = 1000$

evidence of both types of heteroskedasticity in the errors.

The largest roots of the companion matrix of the VAR models are reported in Table 5. The largest roots with cointegration rank $r = 1$ imposed are between 0.975 and 0.987, with the exception of Citigroup for which the largest root is 0.904.

The eigenvalues, likelihood ratio statistics, asymptotic, bootstrap and WB p-values are given in Table 5. The p-values of the asymptotic test are computed from the distribution approximations in Doornik (1998). The number of bootstrap replications is $B = 100000$. The asymptotic $Q_{r,T}$ test rejects $r = 0$ and accepts $r = 1$, with the exception of Goldman Sachs, and then p_t^{CDS} and p_t^{CS} are cointegrated. For

Table 4. LM tests for heteroskedastic and autoregressive conditional heteroskedastic (ARCH) errors in the VAR models for the CDS prices and credit spreads. The table reports the F-statistics and p-values. HET is the White test for heteroskedastic errors. ARCH(2) is the test for autoregressive conditional heteroskedastic errors. All statistics are F.

	HET		ARCH(2)	
VAR(2) model for Bank of America				
p_t^{CDS}	17.954	0.000	9.194	0.000
p_t^{CS}	2.967	0.003	2.764	0.064
System	8.875	0.000	—	—
VAR(3) model for Citigroup				
p_t^{CDS}	9.218	0.000	19.728	0.000
p_t^{CS}	4.348	0.000	7.348	0.000
System	7.207	0.000	—	—
VAR(3) model for Goldman Sachs				
p_t^{CDS}	9.047	0.000	22.784	0.000
p_t^{CS}	6.091	0.000	13.745	0.000
System	5.439	0.000	—	—
VAR(4) model for Barclays Bank				
p_t^{CDS}	9.572	0.000	21.075	0.000
p_t^{CS}	11.062	0.000	17.353	0.000
System	6.643	0.000	—	—
VAR(3) model for Vodafone				
p_t^{CDS}	8.173	0.000	7.420	0.000
p_t^{CS}	16.494	0.000	24.731	0.000
System	10.635	0.000	—	—

Goldman Sachs, the $Q_{r,T}$ test fails to reject $r = 0$. The bootstrap $Q^{\text{B}}_{r,T}$ test rejects $r = 0$ for Citigroup, Barcalys Bank and Vodafone. The bootstrap p-value for Bank of America is 6.7% and for Goldman Sachs 46.3%. The WB $Q^{\text{WB}}_{r,T}$ test rejects $r = 0$ only for Citigroup. Observe the large differences between the bootstrap and WB p-values in Table 5. We conjecture that this discrepancy is caused by unconditional heteroskedasticity in the errors. In that case the bootstrap test is not valid but the WB test is valid (Cavaliere et al. (2012b)).

The financial theory posits further that the cointegrating vector is of the form $\boldsymbol{\beta} = (1, -1, c)'$, and possibly $\boldsymbol{\beta} = (1, -1, 0)'$. Table 6 reports the results of LR tests (Johansen (1996)) for the restrictions on $\boldsymbol{\beta}$ implied by the theory. In order to test the restrictions, we have imposed the cointegration rank $r = 1$ on all models, including the model for Goldman Sachs for which the tests of cointegration rank failed to reject $r = 0$. We find that the restrictions are accepted for Bank of America and Goldman Sachs, but rejected for Citigroup, Barclays Bank and Vodafone.

Table 5. The largest roots of the companion matrix, the eigenvalues, likelihood ratio statistics, asymptotic, bootstrap and WB p-values for the CDS prices and credit spreads. The bootstrap and WB p-values are denoted by 'B' and 'WB', respectively. The VAR models contain dummy variables taking the value 1 for the date in question and 0 otherwise: 9 March 2009, 8 August 2011 and 9 August 2011 for Bank of America, 24 February 2009 and 9 April 2009 for Citigroup, 8 April 2009 for Goldman Sachs, 4 June 2009 and 14 September 2011 for Barclays Bank, and 21 April 2009, 5 June 2009 and 10 May 2010 for Vodafone.

r	Largest roots		$\widehat{\lambda}_{r+1}$	$Q_{r,T}$	$p_{r,T}$	${p_{r,T}}^{\mathrm{B}}$	${p_{r,T}}^{\mathrm{WB}}$
	VAR(2) model for Bank of America						
0	1.0000	1.0000	0.0217	22.514	0.022	0.067	0.196
1	1.0000	0.9789	0.0062	4.949	0.299	0.310	0.473
	VAR(3) model for Citigroup						
0	1.0000	1.0000	0.0549	46.298	0.000	0.000	0.014
1	1.0000	0.9038	0.0014	1.086	0.925	0.944	0.974
	VAR(3) model for Goldman Sachs						
0	1.0000	1.0000	0.0102	12.754	0.392	0.463	0.588
1	1.0000	0.9866	0.0057	4.569	0.345	0.347	0.400
	VAR(4) model for Barclays Bank						
0	1.0000	1.0000	0.0291	26.804	0.004	0.008	0.064
1	1.0000	0.9754	0.0040	3.202	0.554	0.561	0.603
	VAR(3) model for Vodafone						
0	1.0000	1.0000	0.0238	22.224	0.025	0.046	0.263
1	1.0000	0.9805	0.0037	2.936	0.601	0.680	0.777

Table 6. Likelihood ratio tests for restrictions on $\boldsymbol{\beta}$. The table reports the LR statistics and p-values.

	r	$\boldsymbol{\beta} = (1, -1, c)'$		$\boldsymbol{\beta} = (1, -1, 0)'$	
Bank of America	1	3.080	0.079	3.083	0.214
Citigroup	1	26.035	0.000	26.513	0.000
Goldman Sachs	1	0.288	0.592	2.344	0.310
Barclays Bank	1	10.257	0.001	11.564	0.003
Vodafone	1	10.558	0.001	11.202	0.004

5 Simulated Credit Default Swap Prices Data

In order to investigate the size and power of the tests for cointegration between the CDS prices and credit spreads with conditional heteroskedasticity, we simulate the CDS prices data. In the simulations we use the estimated parameters $(\widehat{\boldsymbol{\alpha}}^{(r)}, \widehat{\boldsymbol{\beta}}^{(r)}, \widehat{\boldsymbol{\rho}}^{(r)}, \widehat{\boldsymbol{\Gamma}}_1^{(r)}, \ldots, \widehat{\boldsymbol{\Gamma}}_{k-1}^{(r)}, \widehat{\boldsymbol{\Omega}}^{(r)})$ from the cointegrated VAR models. Constant conditional correlation generalised autoregressive conditional heteroskedasticity, CCC-GARCH$(1,1)$, models are fitted to the residuals from the VAR models and in the simulations of the errors we use the estimated parameters $(\widehat{a}_{01}, \widehat{a}_{11}, \widehat{b}_{11}, \widehat{a}_{02}, \widehat{a}_{22}, \widehat{b}_{22}, \widehat{\rho})$. Table 7 summarises the parameter estimates. We may distill information from the parameter estimates about the persistence in volatility. The sum of the estimated parameters, $\widehat{a}_{11} + \widehat{b}_{11}$ and $\widehat{a}_{22} + \widehat{b}_{22}$, is close to 1 for most companies, which implies very strong persistence in volatility. For example, in the equation for p_t^{CDS} for Goldman Sachs, $\widehat{a}_{11} + \widehat{b}_{11} = 0.999$. The stationarity condition $\lambda(\mathbf{\Gamma_C}) < 1$ is satisfied by all models for the errors. The 4th moment condition $\lambda(\mathbf{\Gamma_{C\otimes C}}) < 1$ is only satisfied by the models for Bank of America and Vodafone.

Table 7. The parameter estimates of the CCC-GARCH$(1,1)$ models fitted to the residuals from the VAR models for the CDS prices and credit spreads. Standard errors are reported in parentheses below the parameter estimates. The stationarity condition is $\lambda(\mathbf{\Gamma_C}) < 1$ and the 4th moment condition is $\lambda(\mathbf{\Gamma_{C\otimes C}}) < 1$.

		Bank of America	Citigroup	Goldman Sachs	Barclays Banks	Vodafone
p_t^{CDS}	a_{01}	0.584	0.639	1.514	0.516	0.392
		(0.317)	(0.823)	(0.898)	(0.310)	(0.665)
	a_{11}	0.084	0.146	0.175	0.116	0.150
		(0.025)	(0.072)	(0.089)	(0.033)	(0.276)
	b_{11}	0.906	0.841	0.824	0.879	0.808
		(0.022)	(0.075)	(0.067)	(0.033)	(0.287)
	$a_{11} + b_{11}$	0.990	0.988	0.999	0.995	0.959
p_t^{CS}	a_{02}	1.777	0.893	39.383	2.698	13.548
		(0.878)	(1.518)	(14.991)	(2.501)	(7.350)
	a_{22}	0.023	0.038	0.352	0.058	0.211
		(0.015)	(0.024)	(0.110)	(0.033)	(0.118)
	b_{22}	0.964	0.957	0.495	0.928	0.643
		(0.012)	(0.013)	(0.124)	(0.046)	(0.155)
	$a_{22} + b_{22}$	0.987	0.995	0.847	0.986	0.854
	ρ	0.038	0.039	0.160	0.035	-0.051
		(0.028)	(0.032)	(0.039)	(0.042)	(0.040)
	$\lambda(\mathbf{\Gamma_C})$	0.990	0.995	0.999	0.995	0.054
	$\lambda(\mathbf{\Gamma_{C\otimes C}})$	0.994	1.017	1.059	1.017	0.955

We simulate 100000 time series of length $T = 804$. The processes are started at the actual initial values. The simulated size and power of the LR tests of cointegration rank are reported in Table 8. For $r_0 = 0$ the table gives the size of the tests, and for $r_0 = 1$ the power of the tests of $r = 0$ and the size of the tests of $r = 1$. The size of the asymptotic $Q_{0,T}$ test is between 7.2% and 15.4%, the size of the bootstrap $Q_{0,T}^{\text{B}}$ test is between 2.5% and 8.7%, and the size of the WB $Q_{0,T}^{\text{WB}}$ test is between 3.9% and 5.0%. From Table 8 we find that the powers of the asymptotic $Q_{0,T}$ test are between 62.4% and 100%, the powers of the bootstrap $Q_{0,T}^{\text{B}}$ test are between 32.8%

and 100%, and the powers of the WB $Q^{\mathrm{WB}}_{0,T}$ test are between 33.2% and 100%. For Citigroup the largest stationary root is 0.904 and the power of all tests is 100%. For Goldman Sachs an interesting result emerges. The largest stationary root is 0.987 and the power of the $Q_{0,T}$ test is 62.4%. The strong persistence in volatility results in a loss of power of the $Q^{\mathrm{B}}_{0,T}$ and $Q^{\mathrm{WB}}_{0,T}$ tests. The power of the former is estimated to be 32.8% and the power of the latter 33.2%. Similar results are obtained for the other companies, as is readily seen in Table 8.

Table 8. Rejection probabilities of a nominal 5% level test for the simulated CDS prices data. The errors are simulated from the CCC-GARCH$(1,1)$ models for the residuals.

$Q_{0,T}$	$Q^{\mathrm{B}}_{0,T}$	$Q^{\mathrm{WB}}_{0,T}$	$Q_{0,T}$	$Q^{\mathrm{B}}_{0,T}$	$Q^{\mathrm{WB}}_{0,T}$	$Q_{1,T}$	$Q^{\mathrm{B}}_{1,T}$	$Q^{\mathrm{WB}}_{1,T}$
	$r_0 = 0$				$r_0 = 1$			
				Bank of America				
0.073	0.027	0.050	0.982	0.859	0.962	0.073	0.040	0.053
				Citigroup				
0.098	0.035	0.039	1.000	1.000	1.000	0.068	0.038	0.048
				Goldman Sachs				
0.154	0.087	0.050	0.624	0.328	0.332	0.118	0.091	0.058
				Barclays Bank				
0.094	0.040	0.046	0.849	0.652	0.727	0.074	0.047	0.048
				Vodafone				
0.072	0.025	0.043	0.816	0.561	0.685	0.067	0.038	0.046

The crucial difference between the simulated CDS prices data with conditional heteroskedasticity and the real CDS prices data is that there is evidence of both unconditional and conditional heteroskedasticity (ARCH) in the real CDS prices data. If there is unconditional heteroskedasticity in the errors, only the WB $Q^{\mathrm{WB}}_{r,T}$ test is valid.

6 Conclusions

In this article we show that there are large stationary roots and strong persistence in volatility in cointegrated systems of CDS prices and credit spreads. Tests of cointegration rank have low power when there are large stationary roots in cointegrated systems. Simulation experiments indicate that asymptotic and bootstrap tests of cointegration rank become unreliable if there is strong persistence in volatility and a condition for the existence of the fourth moment matrix of the errors is not satisfied. This violation does not appear to have a large effect on the size of wild bootstrap tests of cointegration rank. Simulation experiments indicate that wild bootstrap tests of cointegration rank have size close to the nominal level regardless of whether the condition is satisfied or not. This is a useful property of wild bootstrap tests when testing for cointegration between series with strong persistence in volatility. It should be noted that in such cases wild bootstrap tests may have low power to reject the null hypothesis of cointegration rank zero. Obtaining high power for tests of cointegration between CDS prices and credit spreads requires time series of more than 1000 observations, or more than four years of daily observations. Our findings help explain the result typically found in empirical studies that the CDS prices and credit spreads are cointegrated for some but not all companies in the sample.

References

Ahlgren, N. & Antell, J. (2013). The power of bootstrap tests of cointegration rank. *Computational Statistics* 28, 2719–2748.

Blanco, R., Brennan, S. & Marsh, I.W. (2005). An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps. *Journal of Finance* 60, 2255–2281.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics* 72, 498–505.

Cavaliere, G., Rahbek, A. & Taylor, A.M.R. (2012a). Bootstrap determination of co-integration rank in vector autoregressive models. *Econometrica* 80, 1721–1740.

Cavaliere, G., Rahbek, A. & Taylor, A.M.R. (2012b). Bootstrap determination of the co-integration rank in heteroskedastic VAR models. *Econometric Reviews* Forthcoming.

Davidson, R. & MacKinnon, J.G. (2006). The power of bootstrap and asymptotic tests. *Journal of Econometrics* 133, 421–441.

Doornik, J.A. (1998). Approximations to the asymptotic distribution of cointegration tests. *Journal of Economic Surveys* 12, 573–593.

Dötz, N. (2007). Time-Varying Contributions by the Corporate Bond and CDS Markets to Credit Risk Price Discovery. Deutsche Bundesbank, Discussion Paper, Series 2: Banking and Financial Studies, No 8.

Duffie, D. (1999). Credit swap valuation. *Financial Analysts Journal* 55, 73–87.

Forsbäck, A.S. (2012). *En studie om relationen mellan CDS- och obligationspremium under finanskrisen (A Study on the Relation between CDS and Bond Premia during the Financial Crisis, in Swedish)*. Master's thesis, Hanken School of Economics, Helsinki. Unpublished.

He, C. & Teräsvirta, T. (2004). An extended constant conditional correlation GARCH model and its fourth-moment structure. *Econometric Theory* 20, 904–926.

Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Lee, T.H. & Tse, Y. (1996). Cointegration tests with conditional heteroskedasticity. *Journal of Econometrics* 73, 401 – 410.

Nakatani, T. (2013). *ccgarch: An R Package for Modelling Multivariate GARCH Models with Conditional Correlations*. URL http://CRAN.r-project/package=ccgarch. R package version 0.2.0-2.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Swensen, A.R. (2006). Bootstrap algorithms for testing and determining the cointegration rank in VAR models. *Econometrica* 74, 1699–1714.

Swensen, A.R. (2009). Corrigendum to 'Bootstrap algorithms for testing and determining the cointegration rank in VAR models'. *Econometrica* 77, 1703–1704.

Zhu, H. (2006). An empirical comparison of credit spreads between the bond market and the credit default swap market. *Journal of Financial Services Research* 29, 211–235.

MODELING THE EURO–USD EXCHANGE RATE WITH THE GAUSSIAN MIXTURE AUTOREGRESSIVE MODEL[1]

Leena Kalliovirta, Mika Meitz, and Pentti Saikkonen
University of Helsinki

1 Introduction

During the past few decades various nonlinear autoregressive (AR) models have been proposed to model time series data. A comprehensive and up-to-date discussion can be found in Teräsvirta, Tjøstheim & Granger (2010). Perhaps the most popular nonlinear AR models (we confine ourselves to univariate parametric models) are the threshold autoregressive (TAR) models and smooth transition autoregressive (STAR) models (see, e.g., Tong (1990) and Granger & Teräsvirta (1993), respectively). When TAR and STAR models are specified the main interest is focused on finding an adequate description for the conditional expectation (and possibly conditional variance) and not so much on the conditional distribution which in parameter estimation is typically assumed to be Gaussian. In so-called mixture AR models the focus is more on the specification of the entire conditional distribution. In these models the conditional distribution, not only the conditional expectation (and possibly conditional variance), is specified as a convex combination of (typically) Gaussian conditional distributions of linear AR models. Models of this kind were introduced by Le, Martin & Raftery (1996) and further developed by Wong & Li (2000, 2001a,b). Further references include Glasbey (2001), Lanne & Saikkonen (2003), Carvalho & Tanner (2005), Gourieroux & Robert (2006), Dueker, Sola & Spagnolo (2007), and Bec, Rahbek & Shephard (2008). Mixture AR models are related to Markov switching AR models (see, e.g., Hamilton (1994, Ch. 22)) from which they are obtained as special cases with suitable parameter restrictions.

Building on ideas put forth by Glasbey (2001), a recent paper by Kalliovirta, Meitz & Saikkonen (2013) studies a new mixture AR model referred to as the Gaussian mixture autoregressive (GMAR) model. The specific formulation of the GMAR model turns out to have very convenient theoretical implications. To highlight this point, first recall a property that makes the stationary linear Gaussian AR model different from most, if not nearly all, of its nonlinear AR alternatives, namely that the probability structure of the underlying stochastic process is fully known and

[1]We acknowledge financial support from Academy of Finland, Finnish Cultural Foundation, and OP-Pohjola Group Research Foundation.

can be described by Gaussian densities. In nonlinear AR models the situation is typically very different: the conditional distribution is known by construction but what is usually known beyond that is only the existence of a stationary distribution and finiteness of some of its moments. In the GMAR model, stationarity of the underlying stochastic process is a simple consequence of the definition of the model. Moreover, letting p denote the order of the autoregressive part of the model (see Section 2) the stationary distribution of a $(p+1)$–dimensional realization is known to be a mixture of Gaussian distributions with constant mixing weights and known structure for the mean and covariance matrix of the component distributions. From the specification of the GMAR model it also follows that the conditional distribution is a Gaussian mixture with time varying mixing weights. Thus, similarly to the linear Gaussian AR model, and contrary to (at least most) other nonlinear AR models, the structure of stationary marginal distributions of order $p+1$ or smaller is fully known in the GMAR model.

Kalliovirta et al. (2013) apply the GMAR model to a monthly difference between the Euro area and U.S. long-term government bond yields, a series referred to as the interest rate differential. They find that a two-regime GMAR model provides an adequate description of this interest rate differential series, and that forecasts obtained with the GMAR model are superior compared to forecasts from a linear Gaussian AR model and a mixture AR model of Wong & Li (2001a).

In this paper, we provide a more detailed comparison of the GMAR model of Kalliovirta et al. (2013) to other existing mixture (and other nonlinear) AR models, and apply the GMAR model to a monthly Euro-U.S. dollar exchange rate series from January 1999 to January 2012. It turns out that this series is best described by a three-regime GMAR model. The three regimes of the model can be clearly identified in time; one is comprised of the years from 1999 to 2003 and another one of the remaining years except for the year 2008, the early stage of the recent financial crisis.

The rest of the paper proceeds as follows. After discussing general mixture AR models, Section 2 presents the GMAR model of Kalliovirta et al. (2013), followed by a comparison to other mixture AR models proposed in earlier literature. Section 3 deals with issues of specification, estimation, and evaluation of GMAR models. Section 4 presents the empirical example with the Euro-U.S. dollar exchange rate data, and Section 5 concludes.

2 Models

2.1 Mixture autoregressive models

We present two equivalent ways to define mixture autoregressive models. The first one motivates the nomenclature 'mixture autoregressive', whereas the second one clarifies how data from a mixture autoregressive process are generated. To fix notation, let y_t $(t = 1, 2, \ldots)$ be the real-valued time series of interest, and $\mathcal{F}_{t-1}$ the σ–algebra generated by $\{y_{t-j},\ j > 0\}$. According to the first definition, the conditional density function of y_t given its past, $f(\cdot \mid \mathcal{F}_{t-1})$, is assumed to be of the form

$$f(y_t \mid \mathcal{F}_{t-1}) = \sum_{m=1}^{M} \alpha_{m,t} \frac{1}{\sigma_m} \phi\left(\frac{y_t - \mu_{m,t}}{\sigma_m}\right). \tag{1}$$

Here the (positive) mixing weights $\alpha_{m,t}$ are $\mathcal{F}_{t-1}$–measurable and satisfy $\sum_{m=1}^{M} \alpha_{m,t} = 1$ (for all t). Furthermore, $\phi(\cdot)$ denotes the density function of a standard normal random variable, $\mu_{m,t}$ is defined by

$$\mu_{m,t} = \varphi_{m,0} + \sum_{i=1}^{p} \varphi_{m,i} y_{t-i}, \qquad m = 1, \ldots, M, \tag{2}$$

and $\boldsymbol{\vartheta}_m = (\varphi_{m,0}, \boldsymbol{\varphi}_m, \sigma_m^2)$, where $\boldsymbol{\varphi}_m = (\varphi_{m,1}, \ldots, \varphi_{m,p})$ and $\sigma_m^2 > 0$ $(m = 1, \ldots, M)$, contain the unknown parameters introduced in the above equations. By replacing p in (2) with p_m, the autoregressive orders in the component models could be allowed to vary (alternatively, this can be achieved by restricting some of the $\varphi_{m,i}$–coefficients in (2) to be zero). As equation (2) indicates, the definition of the model also requires a specification of the initial values $y_{-p+1}, \ldots, y_0$. Different mixture autoregressive models are obtained by different specifications of the mixing weights $\alpha_{m,t}$. Section 2.3 provides a more detailed discussion of the various specifications proposed in the literature.

To present the model (1)–(2) in an alternative (but equivalent) format, let $P_{t-1}(\cdot)$ signify the conditional probability of the indicated event given $\mathcal{F}_{t-1}$, and let ε_t be a sequence of independent standard normal random variables such that ε_t is independent of $\{y_{t-j},\ j > 0\}$. Furthermore, let $\boldsymbol{s}_t = (s_{t,1}, \ldots, s_{t,M})$ $(t = 1, 2, \ldots)$ be a sequence of (unobserved) M–dimensional random vectors such that, conditional on $\mathcal{F}_{t-1}$, $\boldsymbol{s}_t$ and ε_t are independent. The components of $\boldsymbol{s}_t$ are such that, for each t, exactly one of them takes the value one and others are equal to zero, with conditional probabilities $P_{t-1}(s_{t,m} = 1) = \alpha_{m,t}$, $m = 1, \ldots, M$. Now y_t can be expressed as

$$y_t = \sum_{m=1}^{M} s_{t,m}(\mu_{m,t} + \sigma_m \varepsilon_t) = \sum_{m=1}^{M} s_{t,m}\left(\varphi_{m,0} + \sum_{i=1}^{p} \varphi_{m,i} y_{t-i} + \sigma_m \varepsilon_t\right). \tag{3}$$

This formulation suggests that the mixing weights $\alpha_{m,t}$ can be thought of as (conditional) probabilities that determine which one of the M autoregressive components of the mixture generates the observation y_t.

To gain further insight into mixture autoregressive models, note that the conditional mean and variance of y_t given $\mathcal{F}_{t-1}$ can be expressed as

$$E[y_t \mid \mathcal{F}_{t-1}] = \sum_{m=1}^{M} \alpha_{m,t}\mu_{m,t} = \sum_{m=1}^{M} \alpha_{m,t} \left(\varphi_{m,0} + \sum_{i=1}^{p} \varphi_{m,i} y_{t-i} \right) \tag{4}$$

and

$$Var[y_t \mid \mathcal{F}_{t-1}] = \sum_{m=1}^{M} \alpha_{m,t}\sigma_m^2 + \sum_{m=1}^{M} \alpha_{m,t} \left(\mu_{m,t} - \left(\sum_{m=1}^{M} \alpha_{m,t}\mu_{m,t} \right) \right)^2 . \tag{5}$$

These expressions are straightforward consequences of (1)–(2) or (3) and apply for any specification of the mixing weights $\alpha_{m,t}$. The conditional mean is a weighted average of the conditional means of the M autoregressive components with weights generally depending on the past history of the process. The conditional variance also contains a similar weighted average of the conditional (constant) variances of the M autoregressive components but there is an additional additive term which depends on the variability of the conditional means of the component processes. This additional term makes the conditional variance nonconstant even if the mixing weights are nonrandom and constant over time.

2.2 *The Gaussian Mixture Autoregressive (GMAR) model*

2.2.1 *Definition*

The mixture autoregressive model studied by Kalliovirta et al. (2013) is based on a particular choice of the mixing weights $\alpha_{m,t}$ in (1). In order to define these mixing weights we first use the parameters $\varphi_{m,0}$, $\boldsymbol{\varphi}_m = (\varphi_{m,1}, \ldots, \varphi_{m,p})$, and σ_m (see equation (1) or (3)) and define the M auxiliary Gaussian AR(p) processes

$$\nu_{m,t} = \varphi_{m,0} + \sum_{i=1}^{p} \varphi_{m,i}\nu_{m,t-i} + \sigma_m\varepsilon_t, \qquad m = 1, \ldots, M,$$

where the autoregressive coefficients $\boldsymbol{\varphi}_m$ are assumed to satisfy

$$\varphi_m(z) = 1 - \sum_{i=1}^{p} \varphi_{m,i} z^i \neq 0 \quad \text{for } |z| \leq 1, \qquad m = 1, \ldots, M. \tag{6}$$

This condition implies that the processes $\nu_{m,t}$ are stationary and also that each of the component models in (3) satisfies the usual stationarity condition of the conventional linear AR(p) model. Now, set $\boldsymbol{\nu}_{m,t} = (\nu_{m,t}, \ldots, \nu_{m,t-p+1})$ and $\mathbf{1}_p = (1, \ldots, 1)$ ($p \times 1$), and let $\mu_m \mathbf{1}_p$ and $\boldsymbol{\Gamma}_{m,p}$ signify the mean vector and covariance matrix of $\boldsymbol{\nu}_{m,t}$ ($m = 1, \ldots, M$).[2] Then the random vector $\boldsymbol{\nu}_{m,t}$ follows the p–dimensional multivariate normal distribution with density

$$\begin{aligned} \mathsf{n}_p\left(\boldsymbol{\nu}_{m,t}; \boldsymbol{\vartheta}_m\right) &= (2\pi)^{-p/2} \det(\boldsymbol{\Gamma}_{m,p})^{-1/2} \\ &\quad \times \exp\left\{-\tfrac{1}{2}\left(\boldsymbol{\nu}_{m,t} - \mu_m \mathbf{1}_p\right)' \boldsymbol{\Gamma}_{m,p}^{-1}\left(\boldsymbol{\nu}_{m,t} - \mu_m \mathbf{1}_p\right)\right\}. \end{aligned} \tag{7}$$

Now set $\boldsymbol{y}_{t-1} = (y_{t-1}, \ldots, y_{t-p})$ ($p \times 1$), and define the mixing weights $\alpha_{m,t}$ as

$$\alpha_{m,t} = \frac{\alpha_m \mathsf{n}_p\left(\boldsymbol{y}_{t-1}; \boldsymbol{\vartheta}_m\right)}{\sum_{n=1}^{M} \alpha_n \mathsf{n}_p\left(\boldsymbol{y}_{t-1}; \boldsymbol{\vartheta}_n\right)}, \tag{8}$$

where the $\alpha_m \in (0,1)$, $m = 1, \ldots, M$, are unknown parameters satisfying $\sum_{m=1}^{M} \alpha_m = 1$. (Clearly, the coefficients $\alpha_{m,t}$ are measurable functions of $\boldsymbol{y}_{t-1} = (y_{t-1}, \ldots, y_{t-p})$ and satisfy $\sum_{m=1}^{M} \alpha_{m,t} = 1$ for all t.) Equations (1), (2), and (8) (or (3) and (8)) define the Gaussian Mixture Autoregressive model or the GMAR model abbreviated as GMAR(p, M).[3] We collect the unknown parameters to be estimated in the vector $\boldsymbol{\theta} = (\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_M, \alpha_1, \ldots, \alpha_{M-1})$ ($(M(p+3)-1) \times 1$); the coefficient α_M is not included due to the restriction $\sum_{m=1}^{M} \alpha_m = 1$.

2.2.2 *Theoretical properties*

A major motivation for specifying the mixing weights as in (8) is theoretical attractiveness. A detailed discussion on this point is provided in Kalliovirta et al. (2013) who show that the process $\boldsymbol{y}_t = (y_t, \ldots, y_{t-p+1})$ ($t = 1, 2, \ldots$) with y_t generated by (1), (2), and (8) (or, equivalently, (3) and (8)) is an ergodic Markov chain on $\mathbb{R}^p$ with a stationary distribution characterized by the density

$$f(\boldsymbol{y}; \boldsymbol{\theta}) = \sum_{m=1}^{M} \alpha_m \mathsf{n}_p\left(\boldsymbol{y}; \boldsymbol{\vartheta}_m\right). \tag{9}$$

---

[2]Here $\mu_m = \varphi_{m,0}/\varphi_m(1)$ and each $\boldsymbol{\Gamma}_{m,p}$, $m = 1, \ldots, M$, has the familiar form of being a $p \times p$ symmetric Toeplitz matrix with $\gamma_{m,0} = Cov[\nu_{m,t}, \nu_{m,t}]$ along the main diagonal, and $\gamma_{m,i} = Cov[\nu_{m,t}, \nu_{m,t-i}]$, $i = 1, \ldots, p-1$, on the diagonals above and below the main diagonal (for the dependence of the covariance matrix $\boldsymbol{\Gamma}_{m,p}$ on the parameters $\boldsymbol{\varphi}_m$ and σ_m, see Reinsel (1997, Sec. 2.2.3)).

[3]This definition of the model is a simplifed version of the one introduced in Kalliovirta et al. (2013) where the definition of the mixing weights is based on the q–dimensional vector $\boldsymbol{y}_{t-1} = (y_{t-1}, \ldots, y_{t-q})$ with $q \geq p$. We shall not consider this extension because it is not needed in our empirical application.

Thus, the stationary distribution of $\boldsymbol{y}_t$ is a mixture of M multinormal distributions with constant mixing weights α_m that appear in the time varying mixing weights $\alpha_{m,t}$ defined in (8). This immediately implies that all moments of the stationary distribution exist and are finite. As demonstrated by Kalliovirta et al. (2013) the stationary distribution of the $(p+1)$–dimensional random vector $(y_t, \boldsymbol{y}_{t-1})$ is also a Gaussian mixture with density of the same form as in (9) or, specifically, $\sum_{m=1}^{M} \alpha_m \mathsf{n}_{p+1}\left((y, \boldsymbol{y}); \boldsymbol{\vartheta}_m\right)$ with an explicit form of the density $\mathsf{n}_{p+1}\left((y, \boldsymbol{y}); \boldsymbol{\vartheta}_m\right)$ given in the proof of Theorem 1 of Kalliovirta et al. (2013). Furthermore, the marginal distributions of this Gaussian mixture belong to the same family, as can be seen by integrating the relevant components of $(y, \boldsymbol{y})$ out of the density. It may be worth noting, however, that this does not hold for higher dimensional realizations so that the stationary distribution of $(y_{t+1}, y_t, \boldsymbol{y}_{t-1})$, for example, is not a Gaussian mixture. This fact was already pointed out by Glasbey (2001) who considered a first order version of the same model (i.e., the case $p = 1$) by using a slightly different formulation. Unlike Kalliovirta et al. (2013), Glasbey (2001) did not discuss higher order models explicitly and he did not establish ergodicity, however.

The fact that the stationary distribution of the GMAR model is fully known is not only convenient but also quite exceptional among mixture autoregressive models or other related nonlinear autoregressive models such as TAR models or STAR models. In this respect the GMAR model differs from most, if not nearly all, previous nonlinear autoregressive models (for a few rather simple first order examples, see Tong (2011, Sec. 4.2)). As illustrated in Section 4, a nonparametric estimate of the stationary density of y_t can thus be used (as one tool) to assess the need of a mixture model and the fit of a specified GMAR model. A limitation of the GMAR model is, however, that the components of the mixture are assumed to satisfy the usual stationarity condition of a linear AR(p) model which is not required in all previous models. For instance, Bec et al. (2008) consider a mixture AR model with $M = 2$ and give conditions for its stationarity without any restrictions on the autoregressive parameters of one of the component models.

2.2.3 Interpretation of the mixing weights α_m and $\alpha_{m,t}$

Unless otherwise stated, we are from now on concerned with the stationary version of the GMAR process. According to the preceding discussion, the parameter α_m ($m = 1, \ldots, M$) has an immediate interpretation as the unconditional probability of the random vector $\boldsymbol{y}_t = (y_t, \ldots, y_{t-p+1})$ being generated from a distribution with density $\mathsf{n}_p\left(\boldsymbol{y}; \boldsymbol{\vartheta}_m\right)$, that is, from the mth component of the Gaussian mixture characterized in (9). Consequently, α_m ($m = 1, \ldots, M$) also represents the unconditional probability of the component y_t being generated from a distribution with density $\mathsf{n}_1\left(y; \boldsymbol{\vartheta}_m\right)$ which is the mth component of the (univariate) Gaussian mixture density $\sum_{m=1}^{M} \alpha_m \mathsf{n}_1\left(y; \boldsymbol{\vartheta}_m\right)$ where $\mathsf{n}_1\left(y; \boldsymbol{\vartheta}_m\right)$ is the density function of a normal random variable with mean μ_m and variance $\gamma_{m,0}$.

It can also be shown that α_m represents the unconditional probability of (the scalar) y_t being generated from the mth autoregressive component in (3) whereas $\alpha_{m,t}$ represents the corresponding conditional probability $P_{t-1}\left(s_{t,m} = 1\right) = \alpha_{m,t}$. This conditional probability depends on the (relative) size of the product $\alpha_m \mathsf{n}_p(\boldsymbol{y}_{t-1}; \boldsymbol{\vartheta}_m)$, the numerator of the expression defining $\alpha_{m,t}$ (see (8)). The latter factor of this product, $\mathsf{n}_p(\boldsymbol{y}_{t-1}; \boldsymbol{\vartheta}_m)$, can be interpreted as the likelihood of the mth autoregressive component in (3) based on the observation $\boldsymbol{y}_{t-1}$. Thus, the larger this likelihood is the more likely it is to observe y_t from the mth autoregressive component. However, the product $\alpha_m \mathsf{n}_p(\boldsymbol{y}_{t-1}; \boldsymbol{\vartheta}_m)$ is also affected by the former factor α_m or the weight of $\mathsf{n}_p(\boldsymbol{y}_{t-1}; \boldsymbol{\vartheta}_m)$ in the stationary mixture distribution of $\boldsymbol{y}_{t-1}$ (evaluated at $\boldsymbol{y}_{t-1}$; see (9)). Specifically, even though the likelihood of the mth autoregressive component in (3) is large (small) a small (large) value of α_m attenuates (amplifies) its effect so that the likelihood of observing y_t from the mth autoregressive component can be small (large). This seems intuitively natural because a small (large) weight of $\mathsf{n}_p(\boldsymbol{y}_{t-1}; \boldsymbol{\vartheta}_m)$ in the stationary mixture distribution of $\boldsymbol{y}_{t-1}$ means that observations cannot be generated by the mth autoregressive component too frequently (too infrequently).

2.2.4 Covariance structure

Using the facts that the density of $(y_t, \boldsymbol{y}_{t-1})$ is $\sum_{m=1}^{M} \alpha_m \mathsf{n}_{p+1}\left((y_t, \boldsymbol{y}_{t-1}); \boldsymbol{\vartheta}_m\right)$ and that of y_t is $\sum_{m=1}^{M} \alpha_m \mathsf{n}_1\left(y; \boldsymbol{\vartheta}_m\right)$ we can obtain explicit expressions for the mean, variance, and first p autocovariances of the process y_t. With the notation introduced in equation (7) we can express the mean as

$$\mu \stackrel{def}{=} E\left[y_t\right] = \sum_{m=1}^{M} \alpha_m \mu_m$$

and the variance and first p autocovariances as

$$\gamma_j \stackrel{def}{=} Cov\left[y_t, y_{t-j}\right] = \sum_{m=1}^{M} \alpha_m \gamma_{m,j} + \sum_{m=1}^{M} \alpha_m \left(\mu_m - \mu\right)^2, \quad j = 0, 1, \ldots, p.$$

Using these autocovariances and Yule-Walker equations (see, e.g., Box, Jenkins & Reinsel (2008, p. 59)) one can derive the parameters of the linear AR(p) process that best approximates a GMAR(p, M) process. As higher dimensional stationary distributions are not Gaussian mixtures and appear difficult to handle no simple expressions are available for autocovariances at lags larger than p.

2.3 Discussion of models

In this section, we discuss the GMAR model in relation to other nonlinear autoregressive models introduced in the literature.

2.3.1 Relation to other mixture autoregressive models

If the mixing weights $\alpha_{m,t}$ are assumed constant over time the general mixture autoregressive model (1) reduces to the MAR model studied by Wong & Li (2000). The MAR model, in turn, is a generalization of a model considered by Le et al. (1996). The Logistic MAR (LMAR) model of Wong & Li (2001a) can be obtained from equation (1) by choosing $M = 2$, $\alpha_{2,t} = 1 - \alpha_{1,t}$, and

$$\alpha_{1,t} = \frac{\exp(\beta_0 + \boldsymbol{\beta}'\boldsymbol{y}_{t-1})}{1 + \exp(\beta_0 + \boldsymbol{\beta}'\boldsymbol{y}_{t-1})}, \tag{10}$$

where, unlike before, $\boldsymbol{y}_{t-1}$ denotes the q-dimensional vector $(y_{t-1}, \ldots, y_{t-q})$ with some integer q. Because here β_0 and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)$ are additional unknown parameters, the number of parameters in the LMAR model is expected to be larger than in the two-regime GMAR model (cf. (8)). The case $\boldsymbol{\beta} = 0$ corresponds to the MAR model of Wong & Li (2000).

Related two-regime mixture models with time-varying mixing weights were also considered by Gourieroux & Robert (2006) and Bec et al. (2008). Lanne & Saikkonen (2003) considered a mixture AR model in which multiple regimes are allowed (see also Zeevi, Meir & Adler (2000) and Carvalho & Tanner (2005) in the engineering literature). Lanne & Saikkonen (2003) specify the mixing weights as

$$\alpha_{m,t} = \begin{cases} 1 - \Phi((y_{t-d} - c_1)/\sigma_\eta), & m = 1, \\ \Phi((y_{t-d} - c_{m-1})/\sigma_\eta) - \Phi((y_{t-d} - c_m)/\sigma_\eta), & m = 2, \ldots, M-1, \\ \Phi((y_{t-d} - c_{M-1})/\sigma_\eta), & m = M, \end{cases} \tag{11}$$

where $\Phi(\cdot)$ signifies the cumulative distribution function of a standard normal random variable, $d \in \mathbb{Z}_+$ is a delay parameter, the real constants $c_1 < \cdots < c_{M-1}$ are location parameters, and σ_η is a positive scale parameter. In their model, the probabilities determining which of the M autoregressive components the next observation is generated from depend on the location of y_{t-d} relative to the location parameters $c_1 < \cdots < c_{M-1}$. Thus, when $p = d = 1$ a similarity between the mixing weights in the model of Lanne & Saikkonen (2003) and in the GMAR model is that the value of y_{t-1} gives indication concerning which regime will generate the next observation. However, even in this case the functional forms of the mixing weights and their interpretation are rather different.

An interesting two-regime mixture model with time-varying mixing weights was recently introduced by Dueker et al. (2007) (see also Dueker, Psaradakis, Sola & Spagnolo (2011) for a multivariate extension).[4] In their model, the mixing weights are specified as

$$\alpha_{1,t} = \frac{\Phi\left((c_1 - \varphi_{1,0} - \boldsymbol{\varphi}_1' \boldsymbol{y}_{t-1})/\sigma_1\right)}{\Phi\left((c_1 - \varphi_{1,0} - \boldsymbol{\varphi}_1' \boldsymbol{y}_{t-1})/\sigma_1\right) + \left[1 - \Phi\left((c_1 - \varphi_{2,0} - \boldsymbol{\varphi}_2' \boldsymbol{y}_{t-1})/\sigma_2\right)\right]} \tag{12}$$

and $\alpha_{2,t} = 1 - \alpha_{1,t}$. Here c_1 is interpreted as a location parameter similar to that in the model of Lanne & Saikkonen (2003) whereas σ_1 and σ_2 are positive scale parameters. However, similarly to our model the mixing weights are determined by lagged values of the observed series and the autoregressive parameters of the component models. The probability that the next observation is generated from the first or second regime is determined by the locations of the conditional means of the two autoregressive components from the location parameter c_1 whereas in the GMAR model this probability is determined by the stationary densities of the two component models and their weights in the stationary mixture distribution. The functional form of the mixing weights of Dueker et al. (2007) is also similar to that of the GMAR model except that instead of the Gaussian density function used in the GMAR model, Dueker et al. (2007) have the Gaussian cumulative distribution function.

2.3.2 Relation to Markov switching models

The mixture autoregressive models discussed in this paper are also related to Markov switching AR models (see, e.g., Hamilton (1994, Ch. 22)). The basic form of the Markov switching AR model corresponds to the case where the sequence $\boldsymbol{s}_t$ (see (3)) forms a (time-homogeneous) Markov chain whose transition probabilities correspond to the mixing weights. The special case where the rows of the transition probability matrix are identical yields the MAR model of Wong & Li (2000). In time-inhomogeneous versions of the Markov switching AR model the transition probabilities depend on lagged values of the observed time series and, assuming that the rows of the transition probability matrix are identical, one similarly obtains special cases of mixture AR models with time varying mixing weights (see, e.g.,

[4] According to the authors their model belongs to the family of STAR models and this interpretation is indeed consistent with the initial definition of the model which is based on equations (1)–(4) in Dueker et al. (2007). However, we have chosen to treat the model as a mixture model because the likelihood function used to fit the model to data is determined by conditional density functions that are of the mixture form (1). These (not necessarily Gaussian) conditional density functions are given in equation (7) of Dueker et al. (2007).

Diebold, Lee & Weinbach (1994) and Filardo (1994)). Thus, as noted by Wong & Li (2001a) the Markov chain structure of the sequence $\boldsymbol{s}_t$ makes Markov switching AR models more general than mixture AR models. However, due to this generality, the theoretical properties of (time-inhomogeneous) Markov switching AR models (and, especially, explicit expressions for the stationary distribution) are more difficult to obtain than in the case of the GMAR model.

2.3.3 Relation to threshold and smooth transition AR models

The GMAR model is also related to threshold and smooth transition type nonlinear models. In particular, the conditional mean function $E[y_t \mid \mathcal{F}_{t-1}]$ of the GMAR model (see (4) and (8)) is similar to that of a TAR or a STAR model (see, e.g., Tong (1990) and Teräsvirta (1994)). In a basic two-regime TAR model, whether a threshold variable (a lagged value of y_t) exceeds a certain threshold or not determines which of the two component models describes the generating mechanism of the next observation. The threshold and threshold variable are analogous to the location parameter c_1 and the variable y_{t-d} in the mixing weights used in the two-regime ($M = 2$) mixture model of Lanne & Saikkonen (2003) (see (11)). In a STAR model, one gradually moves from one component model to the other as the threshold (or transition) variable changes its value. In a GMAR model, the mixing weights follow similar smooth patterns. A difference to STAR models is that while the mixing weights of the GMAR model vary smoothly, the next observation is generated from one particular AR component whose choice is governed by these mixing weights. In a STAR model, the generating mechanism of the next observation is described by a convex combination of the two component models. This difference is related to the fact that the conditional distribution of the GMAR model is of a different type than the conditional distribution of the STAR (or TAR) model which is not a mixture distribution. This difference is also reflected in differences between the conditional variances associated with the GMAR model and STAR (or TAR) models (see (5)).

2.3.4 Graphical illustration

Figure 1 illustrates the discussion in the preceding subsections. In the top panels, we plot the mixing weight $\alpha_{1,t}$ of the GMAR model as a function of $y_{t-1} = y$ in the case $M = 2$, $p = 1$, with certain parameter combinations. The bottom left panel shows $\alpha_{1,t}$ in some cases for the LMAR model of Wong & Li (2001a); in the model of Lanne & Saikkonen (2003) $\alpha_{1,t}$ behaves in a comparable way (no picture presented). The two pictures on the left illustrate that the three models can produce

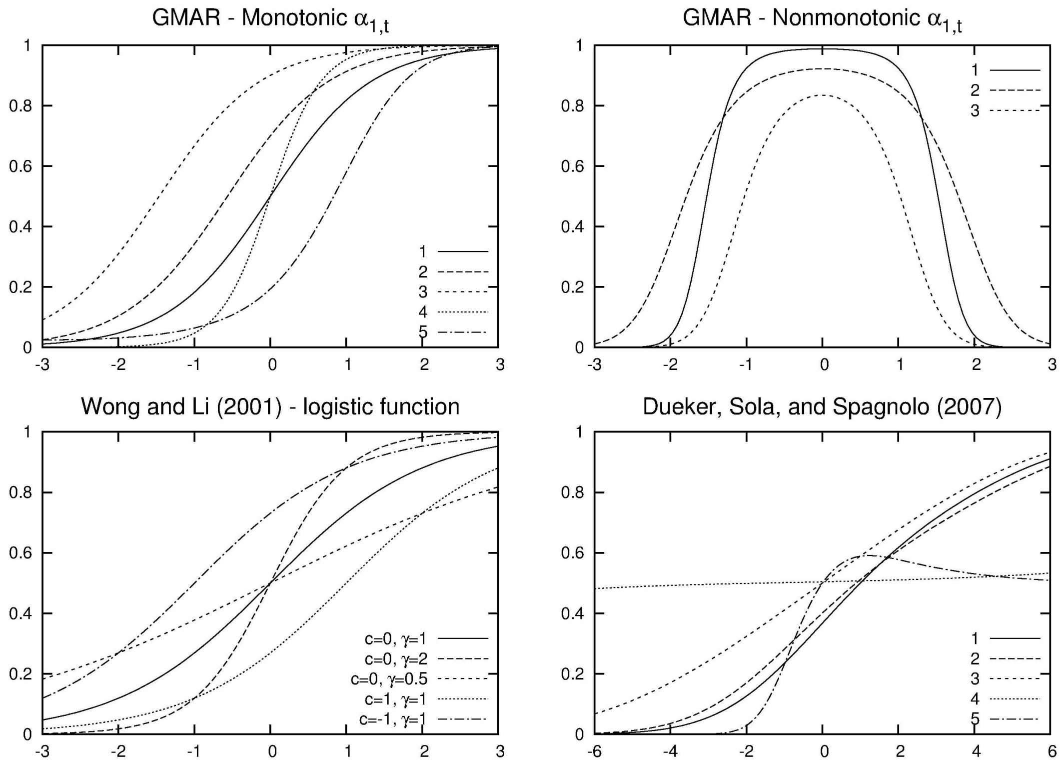


Figure 1. Graphs of the mixing weight $\alpha_{1,t}$ as a function of y_{t-1} for different parameter combinations: the GMAR(1,2) model (top panels), the LMAR model of Wong & Li (2001a) (bottom left panel), and the model of Dueker et al. (2007) (bottom right panel). Details of the parameter combinations used are available in Kalliovirta, Meitz & Saikkonen (2012).

mixing weights of similar monotonically increasing patterns. The figure in the top left panel also illustrates the previously mentioned fact about the mixing weights of the GMAR model that, other things being equal, a decrease in the value of α_m decreases the value of $\alpha_{m,t}$. In the conditional expectation of a basic logistic two-regime STAR model, referred to as the LSTAR1 model in Teräsvirta et al. (2010, Sec. 3.4.1), the transition function, which is the counterpart of the mixing weight $\alpha_{1,t}$, also behaves in a similar monotonically increasing way. Given these observations it is interesting that with suitable parameter values the GMAR model can produce nonmonotonic mixing weights even in the case $M = 2$. The top right panel illustrates this. The models of Wong & Li (2001a) and Lanne & Saikkonen (2003) can produce mixing weights of this form only when $M > 2$. Similarly, in LSTAR models a transition function of this form cannot be obtained with a LSTAR1 model. For that one needs an LSTAR2 model (see Teräsvirta et al. (2010, Sec. 3.4.1)) or some other similar model such as the exponential autoregressive model of Haggan & Ozaki (1981). Thus, once the number of component models is specified the GMAR model appears more flexible in terms of the form of mixing weights than the aforementioned previous mixture models and the same is true when the mixing weights of the GMAR model are compared to the transition functions of LSTAR models.

As far as the mixing weights of the model of Dueker et al. (2007) are concerned they can be nonmonotonic, as illustrated in the bottom right panel of Figure 1. After trying a number of different parameter combinations it seems, however, that (at least in the case $p = 1$) nonmonotonic mixing weights are rather special for this model. The first four (monotonic) graphs in the bottom right panel correspond to parameter configurations in Table 2 of Dueker et al. (2007). The fourth one is interesting in that it produces a nearly constant graph (the graph would be constant if the values of the standard deviations σ_1 and σ_2 were changed to be equal). Finally, note that a common convenience of the GMAR model as well as of the models of Wong & Li (2001a) and Dueker et al. (2007) is that there is no need to choose a threshold variable (typically y_{t-d}) as in the model of Lanne & Saikkonen (2003) or in TAR and STAR models.

3 Model specification, estimation, and evaluation

3.1 Specification

We next discuss some general aspects of building a GMAR model. A natural first step is to consider whether a conventional linear Gaussian AR model provides an adequate description of the data generation process. Thus, one finds an AR(p) model that best describes the autocorrelation structure of the time series, and checks whether residual diagnostics show signs of non-Gaussianity and possibly also of conditional heteroskedasticity. At this point also the graph of the series and a non-parametric estimate of the density function may be useful. The former may indicate the presence of multiple regimes, whereas the latter may show signs of multimodality.

If a linear AR model is found inadequate, specifying a GMAR(p, M) model requires the choice of the number of component models M and the autoregressive order p. A nonparametric estimate of the density function of the observed series may give an indication of how many mixture components are needed. One should, however, be conservative with the choice of M, because if the number of component models is chosen too large then some parameters of the model are not identified. Therefore, a two component model ($M = 2$) is a good first choice. If an adequate two component model is not found, only then should one proceed to a three component model and, if needed, consider even more components. The initial choice of the autoregressive order p can be based on the order chosen for the linear AR model. Again, one should favor parsimonious models, and initially try a smaller p if the order selected for the linear AR model appears large.

3.2 Estimation

After an initial candidate specification (or specifications) is (are) chosen, the parameters of a GMAR model can be estimated by the method of maximum likelihood. As the stationary distribution of the GMAR process is known it is even possible to make use of initial values and construct the exact likelihood function and obtain exact ML estimates, as already discussed by Glasbey (2001) in the first order case. Assuming the observed data is $y_{-p+1}, \ldots, y_0, y_1, \ldots, y_T$ and stationary initial values the log-likelihood function takes the form

$$l_T(\boldsymbol{\theta}) = \log\left(\sum_{m=1}^{M} \alpha_m \mathsf{n}_p(\boldsymbol{y}_0; \boldsymbol{\vartheta}_m)\right) + \sum_{t=1}^{T} \log\left(\sum_{m=1}^{M} \alpha_{m,t}(\boldsymbol{\theta}) \left(2\pi\sigma_m^2\right)^{-1/2} \exp\left(-\frac{(y_t - \mu_{m,t}(\boldsymbol{\vartheta}_m))^2}{2\sigma_m^2}\right)\right), \quad (13)$$

where dependence of the mixing weights $\alpha_{m,t}$ and the conditional expectations $\mu_{m,t}$ of the component models on the parameters is made explicit (see (8) and (2)). Maximizing the log-likelihood function $l_T(\boldsymbol{\theta})$ with respect to the parameter vector $\boldsymbol{\theta}$ yields the ML estimate denoted by $\hat{\boldsymbol{\theta}}$ (a similar notation is used for components of $\hat{\boldsymbol{\theta}}$). Here we have assumed that the initial values in the vector $\boldsymbol{y}_0$ are generated by the stationary distribution. If this assumption seems inappropriate one can condition on initial values and drop the first term on the right hand side of (13). Finally, for reasons of identification, it is necessary to assume that

$$\alpha_1 > \cdots > \alpha_M > 0 \quad \text{and} \quad \boldsymbol{\vartheta}_i = \boldsymbol{\vartheta}_j \text{ only if } 1 \leq i = j \leq M; \quad (14)$$

for further discussion on this issue, see Kalliovirta et al. (2013).

Practical maximization of the likelihood function or its conditional version can be carried out using standard optimization algorithms; in our empirical examples we have used the cmlMT library of Gauss. As usual in nonlinear optimization, good initial values improve the performance of the estimation algorithm. Kalliovirta et al. (2012) discuss possibilities to make use of the stationary distribution of the GMAR process in computing initial estimates.

Kalliovirta et al. (2013) show that the ML estimator is consistent provided the GMAR model is identified, that is, the number of autoregressive components M

is correctly specified and (14) holds. They also discuss the asymptotic normality of the ML estimator and point out that, under appropriate 'high level' conditions similar to those used in Dueker et al. (2007), the ML estimator $\hat{\boldsymbol{\theta}}$ is asymptotically normally distributed with mean vector $\boldsymbol{\theta}$ and covariance matrix the inverse of the Fisher information matrix $E\left[-\partial^2 l_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\right]$ that can be estimated by inverting the observed information matrix $-\partial^2 l_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$. It is worth noting that these results require a correct specification of the number of autoregressive components M. In particular, standard likelihood-based tests are not applicable if the number of component models is chosen too large because then some parameters of the model are not identified. This particularly happens when one tests for the number of component models. For further discussion of this issue, see Dueker et al. (2007, 2011) and the references therein.

3.3 Evaluation

Having estimated a few candidate models, one must check their adequacy and choose the best fitting GMAR(p, M) model. As mentioned above, standard likelihood-based tests cannot be used to test for the number of component models M. Instead of trying to develop proper test procedures for these purposes, Kalliovirta et al. (2013) take a pragmatic approach and employ residual-based diagnostics and information criteria (AIC and BIC) to select a model. In practice, this is often how model selection is done in other nonlinear models as well (cf., e.g., Teräsvirta et al. (2010, Ch. 16); for instance, often the choice of a lag length to be used in a threshold/transition variable is done in a somewhat informal manner). When M is (correctly) chosen, standard likelihood-based inference can be used to choose the autoregressive order p (which can vary from one component model to another).

In mixture models, care is needed when residual-based diagnostics are used to evaluate fitted models. The reason is that residuals with conventional properties are not readily available. This can be seen from the formulation of the GMAR model in equation (3) which shows that, due to the presence of the unobserved variables $s_{t,m}$, an empirical counterpart of the error term ε_t cannot be straightforwardly computed. A more elaborate discussion of this can be found in Kalliovirta (2012). Making use of ideas put forth by Smith (1985), Dunn & Smyth (1996), Palm & Vlaar (1997), and others, Kalliovirta (2012) proposes to use so-called quantile residuals instead of conventional (Pearson) residuals in mixture models (note that quantile residuals have also been called by other names such as normalized residuals and normal forecast transformed residuals).

Quantile residuals are defined by two transformations. Assuming correct specification, the first one (the so-called probability integral transformation) uses the esti-

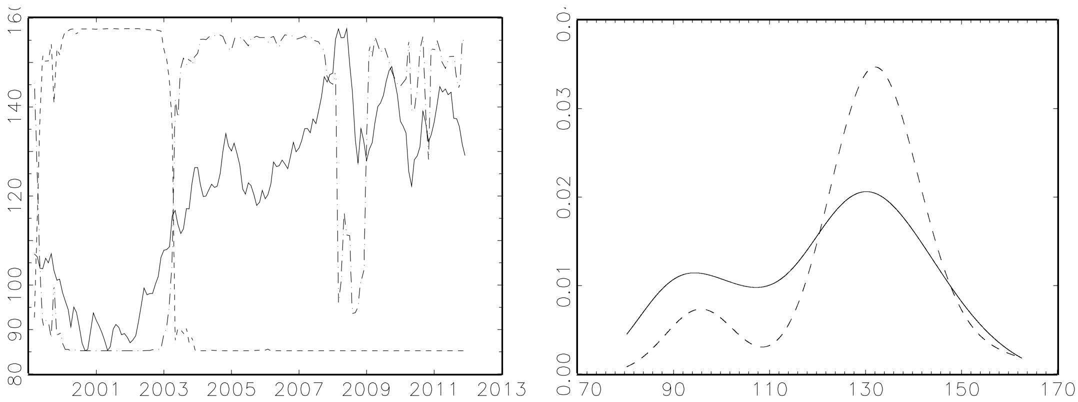


Figure 2. Left panel: Euro and U.S. dollar exchange rate (solid line), and scaled mixing weights based on the estimates of the constrained GMAR(3,3) model in Table 1 (dashed lines, scaled $\hat{\alpha}_{1,t}$ - - and scaled $\hat{\alpha}_{2,t}$ – –). **Right panel:** A kernel density estimate of the observations (solid line) and mixture density implied by the GMAR(3,3) model in Table 1 (dashed line).

mated conditional cumulative distribution function implied by the specified model to transform the observations into approximately independent uniformly distributed random variables. In the second transformation the inverse of the cumulative distribution function of the standard normal distribution is used to get variables that are approximately independent with standard normal distribution. Based on these 'two-stage' quantile residuals Kalliovirta (2012) proposes tests that can be used to check for autocorrelation, conditional heteroskedasticity, and non-normality in quantile residuals. These tests correctly allow for the uncertainty caused by parameter estimation so that, under correct specification, the obtained p–values are asymptotically valid. These are the residual-based diagnostic tests we use in our empirical application along with associated graphical tools to evaluate a fitted model.

4 Empirical example

4.1 A GMAR model for the Euro-U.S. dollar exchange rate

We consider the monthly average Euro-U.S. dollar exchange rate from January 1999 to January 2012, a period of 157 observations that also contains the financial crisis period since 2008 (in an out-of-sample forecasting exercise we also use data till October 2013). This series, multiplied by 100, is depicted in Figure 2 (left panel, solid line).

We began the analysis by fitting several Gaussian AR models to this data. The AIC and BIC information criteria suggested an AR(3) model, which, however, was rejected by residual diagnostics; in particular, the residuals were found conditionally heteroskedastic. Table 1 reports estimation results for this model along with the values of AIC and BIC and outcomes of (quantile) residual-based tests of normality, autocorrelation, and conditional heteroskedasticity (for details of these tests, see Kalliovirta (2012)). Also, the multimodal expression of the kernel density estimate of the original series depicted in Figure 2 (right panel, solid line) points to nonnormality, and hence a potential inadequacy of linear Gaussian AR models.

We fitted several two regime GMAR models to the data, but the conditional heteroskedasticity of quantile residuals was not removed. We then considered three regime GMAR models with different number of lags. The GMAR(2,3) and the GMAR(3,3) models with a common AR structure in all regimes passed the residual diagnostics. We chose the GMAR(3,3) model, because it was suggested both by the AIC and BIC as well as by the likelihood ratio test (p-value 0.003). Furthermore, in our forecasting experiment (see Section 4.1.2) the GMAR(3,3) model turned out to outperform the GMAR(2,3) model (we do not report the forecasts from the latter model).

The estimation results for the constrained GMAR(3,3) model based on conditional likelihood are presented in Table 1 (using exact likelihood gave very similar results). According to diagnostic tests based on quantile residuals (also reported in Table 1), the constrained GMAR(3,3) model provides a good fit to the data. Furthermore, related graphical analyses of the quantile residuals (see Figure 3) give no obvious reason to suspect the adequacy of this model. Thus, unlike the linear AR(3) model, the GMAR(3,3) model seems to provide an adequate description for the exchange rate series, although the linear AR(3) model is preferred by the BIC (see Table 1). The estimated sum of the AR coefficients in our GMAR model is 0.934 which is much less than the corresponding sum (0.984) obtained in the linear AR(3) model. The reduction is presumably related to the differences in the intercept terms of the three AR components which is directly reflected as different means in the three regimes. The estimated error variances of the AR components are also very different and, consequently, the same is true for the variances of the three regimes. This feature is undoubtedly related to the fact that the GMAR model has been able to remove the conditional heteroskedasticity observed in linear modeling. According to the approximate standard errors in Table 1, the estimation accuracy appears quite reasonable except for the parameters α_1 and α_2. A similar finding, presumably due to the shortness of the series, was observed in Kalliovirta et al. (2013).

Table 1. Estimated AR(3) and GMAR(3,3) models (left panel) and means and covariances implied by the GMAR(3,3) model (right panel).

	Estimated Models			Means & Covariances
	AR(3)	GMAR(3,3)		GMAR(3,3)
$\varphi_{1,0}$	1.985 (1.499)	6.295 (2.376)	μ_1	95.64
$\varphi_{2,0}$		8.668 (3.202)	μ_2	131.70
$\varphi_{3,0}$		9.558 (3.445)	μ_3	145.21
φ_1	1.322 (0.080)	1.283 (0.084)	$\gamma_{1,0}$	54.39
φ_2	-0.437 (0.129)	-0.411 (0.133)	$\gamma_{2,0}$	82.32
φ_3	0.099 (0.080)	0.062 (0.083)	$\gamma_{3,0}$	257.09
σ_1^2	8.743 (0.996)	4.942 (1.097)	$\gamma_{m,1}/\gamma_{m,0}$	0.947
σ_2^2		7.480 (1.320)	$\gamma_{m,2}/\gamma_{m,0}$	0.863
σ_3^2		23.360 (11.928)	$\gamma_{m,2}/\gamma_{m,1}$	0.911
α_1		0.135 (0.266)		
α_2		0.739 (0.249)		
$-L_T(\hat{\boldsymbol{\theta}})$	385	377		
AIC	781	776		
BIC	796	809		
N	0.20	0.50		
A_1	0.23	0.26		
A_4	0.02	0.73		
H_1	0.61	0.33		
H_4	0	0.86		

Notes: Left panel: Parameter estimates (with standard errors calculated using the Hessian of the log-likelihood function in parentheses) of the estimated AR and GMAR models. GMAR model is the constrained model ($\varphi_{m,1} = \varphi_1$, $\varphi_{m,2} = \varphi_2$, $\varphi_{m,3} = \varphi_3$, $m = 1, 2, 3$), with estimation based on the conditional likelihood. Rows labelled N, ..., H_4 present p–values of diagnostic tests based on quantile residuals. The test statistic for normality, N, is based on moments of quantile residuals and the test statistics for autocorrelation, A_k, and conditional heteroskedasticity, H_k, are based on the first k autocovariances and squared autocovariances of quantile residuals, respectively. Under correct specification, test statistic N is approximately distributed as χ_2^2 (AR(3)) or χ_3^2 (GMAR) and test statistics A_k and H_k are approximately distributed as χ_k^2 (for details, see Kalliovirta (2012)). A p–value < 0.001 is denoted by 0. **Right panel:** Estimates derived for the expectations μ_m and elements of the covariance matrix $\boldsymbol{\Gamma}_{m,3}$; see Section 2.2.

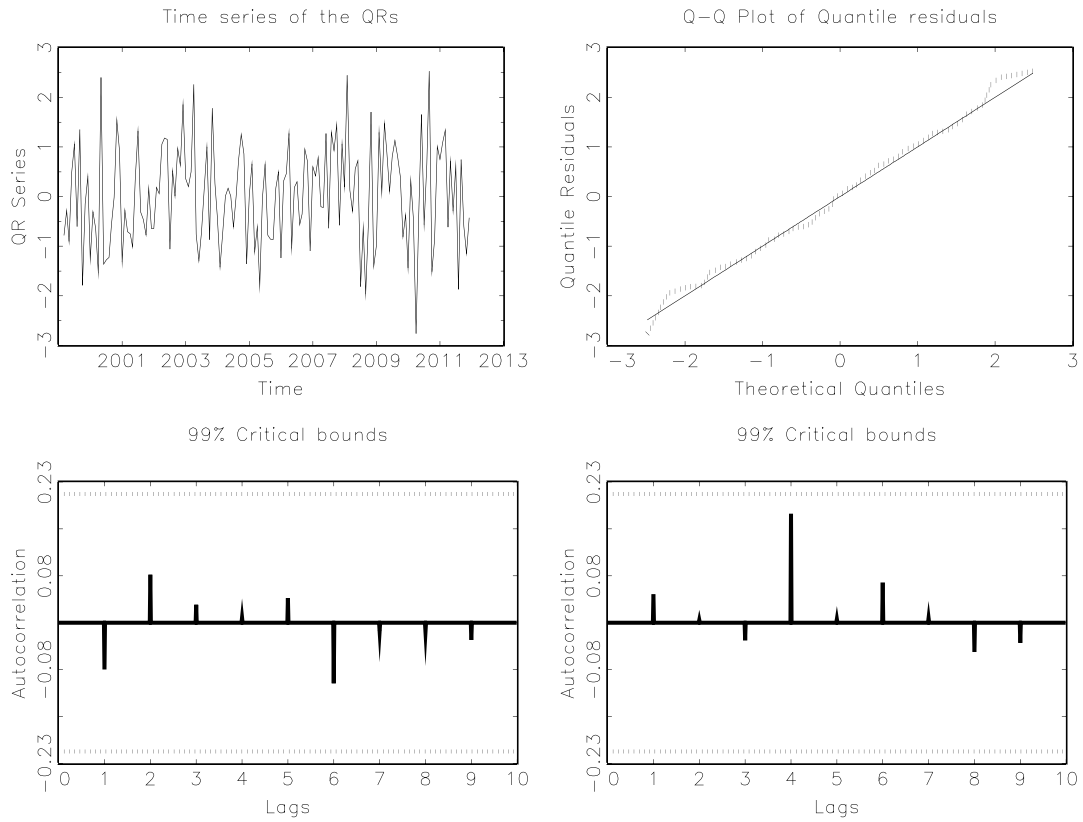


Figure 3. Diagnostics of the restricted GMAR(3,3) model described in Table 1: Time series of quantile residuals (top left panel), QQ-plot of quantile residuals (top right panel), and ten first scaled autocovariances of quantile residuals and squared quantile residuals (bottom left and right panels, respectively). The lines in the bottom panels show approximate 99% critical bounds. For details, see Kalliovirta (2012).

4.2 Interpretations of the GMAR(3,3) model

Based on the conditional ML estimates of the GMAR(3,3) model in Table 1, the top left panel of Figure 4 shows the contour plot of the estimated two-dimensional stationary mixture density $\sum_{m=1}^{3} \hat{\alpha}_m \mathsf{n}_2(\boldsymbol{y}_{t-1}; \hat{\boldsymbol{\vartheta}}_m)$, whereas the corresponding one-dimensional mixture density $\sum_{m=1}^{3} \hat{\alpha}_m \mathsf{n}_1(y_{t-1}; \hat{\boldsymbol{\vartheta}}_m)$ and its three components are depicted in the bottom left panel of the same figure. Of the two distinct peaks of the contour plot, the one centered approximately at $y_{t-1} = y_{t-2} \approx 95$ corresponds to the first regime and the component density $\hat{\alpha}_1 \mathsf{n}_2(\boldsymbol{y}_{t-1}; \hat{\boldsymbol{\vartheta}}_1)$, whereas the other one is comprised of the second and third regimes with component densities $\hat{\alpha}_2 \mathsf{n}_2(\boldsymbol{y}_{t-1}; \hat{\boldsymbol{\vartheta}}_2)$ and $\hat{\alpha}_3 \mathsf{n}_2(\boldsymbol{y}_{t-1}; \hat{\boldsymbol{\vartheta}}_3)$. These one- and two-dimensional densities clearly illustrate the large differences between the shapes of the three component densities already apparent in the estimates of Table 1. When one visually compares the density implied by the model and the nonparametric kernel density estimate

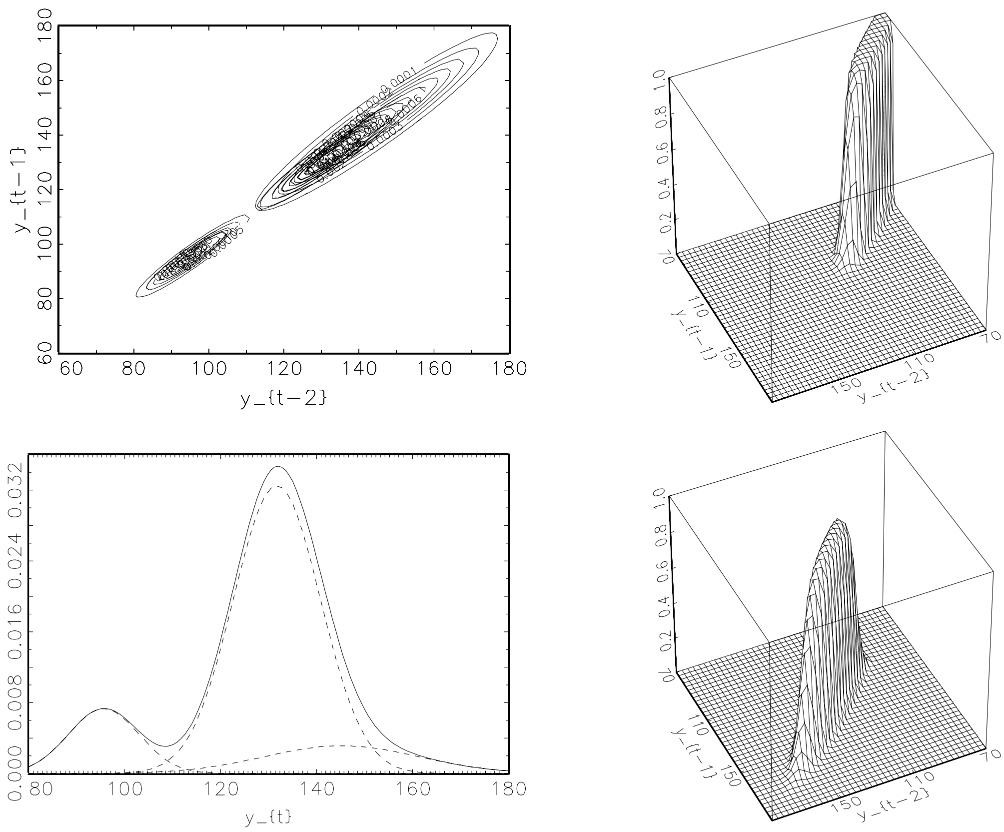


Figure 4. Properties of the restricted GMAR(3,3) model described in Table 1. Contours of the two-dimensional stationary mixture density (top left panel), one-dimensional density function of the stationary mixture distribution (solid) and its three components (dashed) (bottom left panel), probability of regime 1, $\alpha_{1,t}$, as function of the values y_{t-1} and y_{t-2} (top right panel), and probability of regime 2, $\alpha_{2,t}$, as function of the values of y_{t-1} and y_{t-2} (bottom right panel).

of the observations (Figure 2, right panel), there seem to be rather large departures between the two. However, a simulation experiment based on the estimated GMAR(3,3) model indicates that such a difference is observed on average in at least 25 cases out of 100 when the model is correctly specified. Thus, the difference does not indicate misspecification in the model.

Figure 2 (left panel, dashed line) depicts the time series of the estimated mixing weights $\hat{\alpha}_{1,t}$ and $\hat{\alpha}_{2,t}$ scaled so that $\hat{\alpha}_{i,t} = \max y_t$ when $\hat{\alpha}_{i,t} = 1$, and $\hat{\alpha}_{i,t} = \min y_t$ when $\hat{\alpha}_{i,t} = 0$ for $i = 1, 2$. During the period between 1999 and 2003 the first regime (with lowest mean, $\hat{\mu}_1 = 95.64$) is clearly dominating. In 2003 the series switches to regime 2 (with mean $\hat{\mu}_2 = 131.70$), and evolves there for the rest of the time period except for the year 2008 when the series stays in regime 3 (with the highest mean, $\hat{\mu}_3 = 145.21$); after 2009 there are a couple of occasions where the probability of regime 3 increases but there is no change in level. The most noticeable difference between the second and the third regimes is in their estimated variances ($\hat{\gamma}_{2,0} = 82.32$ and $\hat{\gamma}_{3,0} = 257.09$), where the higher variance in the third regime is presumably connected to the hightened uncertainty during the early stages of the recent financial crisis in 2008.

To illustrate how the mixing weights depend on the past observations, Figure 4 (top right and bottom right panels) depicts the estimated time varying mixing weights $\hat{\alpha}_{1,t}$ and $\hat{\alpha}_{2,t}$ as functions of y_{t-1} and y_{t-2}. The functional forms are similar to the estimated density functions $\hat{\alpha}_1 \mathsf{n}_2(\boldsymbol{y}_{t-1}; \hat{\boldsymbol{\vartheta}}_1)$ and $\hat{\alpha}_2 \mathsf{n}_2(\boldsymbol{y}_{t-1}; \hat{\boldsymbol{\vartheta}}_2)$ (not shown). In the top right panel, inside an ellipse roughly corresponding to an ellipse where the estimated density $\hat{\alpha}_1 \mathsf{n}_2(\boldsymbol{y}_{t-1}; \hat{\boldsymbol{\vartheta}}_1)$ has nonnegligible mass, the estimated mixing weight $\hat{\alpha}_{1,t}$ is nearly unity. This corresponds to points such that $y_{t-1}, y_{t-2} \leq 100$. The mixing weight $\hat{\alpha}_{2,t}$ is near unity around points y_{t-1} and y_{t-2} that are inside an ellipse centered approximately at $y_{t-1} = y_{t-2} \approx 125$ (bottom right panel of Figure 4). The more the series exceeds values about 145, the more likely it switches into the highest regime.

4.3 Forecasting experiment

Finally, we compare the estimated AR(3) and GMAR(3,3) models in a small forecasting exercise. The date of forecasting (up until which observations are used) ranges from January 2012 till September 2013, and for each date of forecasting, forecasts are computed for all the subsequent periods up until October 2013. We do not update estimates when the date of forecasting changes so that all forecasts are based on a model whose parameter estimates are given in Table 1. Using initial values known at the date of forecasting, we simulate 500,000 realizations and treat the mean of these realizations as a point forecast, and repeat this for all forecast horizons up until October 2013. This results in a total of 21 one-step forecasts, 20 two-step forecasts, . . . , fourteen 8-step forecasts (as well as a few forecasts for longer horizons which we discard). For each of the forecast horizons 1, 2, . . . , 8, we measure the forecast accuracy by the mean squared prediction error (MSPE) with the mean computed across the 21, . . . , 14 forecasts available.

As expected, the forecast accuracy is best in one-step-ahead prediction and steadily deteriorates when the forecast horizon increases. For the one-step-ahead forecasts, the MSPE of the AR(3) model relative to that of the GMAR(3,3) model is 105%, indicating that the GMAR(3,3) model forecasts this series better than the AR(3) model. The relative MSPE from one-step up to eight-step ahead forecasts is depicted in Figure 5 (left panel). It is seen that, at least during this particular time period, the GMAR(3,3) model is able to forecast the series at step lengths 2 and

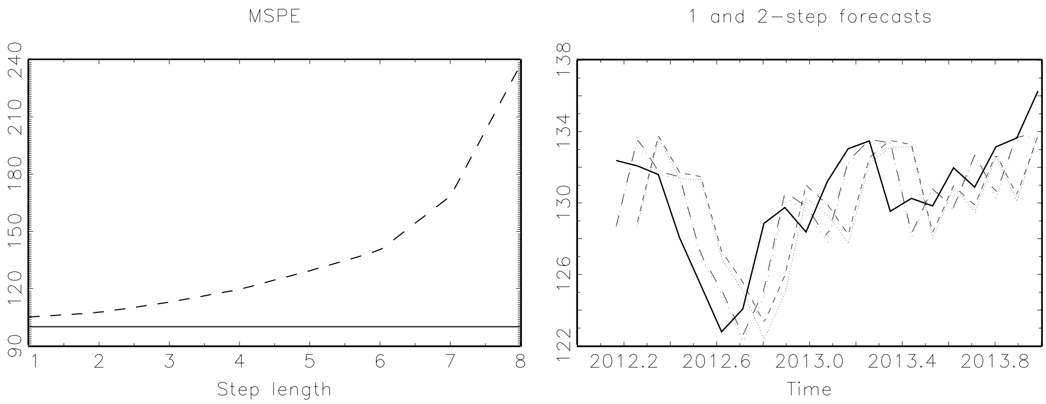


Figure 5. Left panel: Relative forecast accuracies of the two models, measured in relative mean squared prediction error (MSPE): GMAR(3,3) model (solid horizontal line at 100) and AR(3) model (dashed line). **Right panel:** Observed series (solid line) and 1- and 2-step ahead forecasts based on the GMAR(3,3) (dashed lines, 1-step – – and 2-step - -) and AR(3) (dotted lines, 1-step · · · and 2-step ···) models.

larger significantly better than the linear AR(3) model. The right panel of Figure 5 depicts the evolution of the series and the one-step and two-step ahead forecasts of the two models.

5 Conclusion

This paper provides a further study of the Gaussian mixture autoregressive (GMAR) model proposed by Glasbey (2001) and Kalliovirta et al. (2013). The GMAR model can allow for various deviations from linearity and Gaussianity, such as multimodal conditional and marginal distributions as well as shifts in the mean and variance of the series. These features were illustrated by our empirical application to a monthly Euro-U.S. dollar exchange rate series from January 1999 to January 2012. In this application, a three-regime GMAR model turned out to provide an adequate description of the exchange rate series. The three regimes could be clearly identified in time, with one of them corresponding to the early stages of the recent financial crisis of 2008.

References

Bec, F., Rahbek, A. & Shephard, N. (2008). The ACR model: a multivariate dynamic mixture autoregression. *Oxford Bulletin of Economics and Statistics* 70, 583–618.

Box, G.E.P., Jenkins, G.M. & Reinsel, G.C. (2008). *Time Series Analysis: Forecasting and Control*. 4th edition. Hoboken NJ: Wiley.

Carvalho, A.X. & Tanner, M.A. (2005). Mixtures-of-experts of autoregressive time series: asymptotic normality and model specification. *IEEE Transactions on Neural Networks* 16, 39–56.

Diebold, F.X., Lee, J.H. & Weinbach, G.C. (1994). Regime switching with time-varying transition probabilities. In C. Hargreaves (Ed.) *Nonstationary Time Series Analysis and Cointegration*. Oxford: Oxford University Press, 283–302.

Dueker, M.J., Psaradakis, Z., Sola, M. & Spagnolo, F. (2011). Multivariate contemporaneous-threshold autoregressive models. *Journal of Econometrics* 160, 311–325.

Dueker, M.J., Sola, M. & Spagnolo, F. (2007). Contemporaneous threshold autoregressive models: estimation, testing and forecasting. *Journal of Econometrics* 141, 517–547.

Dunn, P.K. & Smyth, G.K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics* 5, 236–244.

Filardo, J.F. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business and Economic Statistics* 12, 299–308.

Glasbey, C.A. (2001). Non-linear autoregressive time series with multivariate Gaussian mixtures as marginal distributions. *Journal of the Royal Statistical Society: Series C* 50, 143–154.

Gourieroux, C. & Robert, C.Y. (2006). Stochastic unit root models. *Econometric Theory* 22, 1052–1090.

Granger, C.W.J. & Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Haggan, V. & Ozaki, T. (1981). Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* 68, 189–196.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Kalliovirta, L. (2012). Misspecification tests based on quantile residuals. *Econometrics Journal* 15, 358–393.

Kalliovirta, L., Meitz, M. & Saikkonen, P. (2012). A Gaussian mixture autoregressive model for univariate time series. HECER Discussion Paper No. 352.

Kalliovirta, L., Meitz, M. & Saikkonen, P. (2013). A Gaussian mixture autoregressive model for univariate time series. Unpublished revision of HECER Discussion Paper No. 352.

Lanne, M. & Saikkonen, P. (2003). Modeling the U.S. short-term interest rate by mixture autoregressive processes. *Journal of Financial Econometrics* 1, 96–125.

Le, N.D., Martin, R.D. & Raftery, A.E. (1996). Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models. *Journal of the American Statistical Association* 91, 1504–1515.

Palm, F.C. & Vlaar, P.J.G. (1997). Simple diagnostic procedures for modelling financial time series. *Allgemeines Statistisches Archiv* 81, 85–101.

Reinsel, G.C. (1997). *Elements of Multivariate Time Series Analysis*. New York: Springer.

Smith, J.Q. (1985). Diagnostic checks of non-standard time series models. *Journal of Forecasting* 4, 283–291.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89, 208–218.

Teräsvirta, T., Tjøstheim, D. & Granger, C.W.J. (2010). *Modelling Nonlinear Economic Time Series*. Oxford: Oxford University Press.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.

Tong, H. (2011). Threshold models in time series analysis – 30 years on. *Statistics and Its Interface* 4, 107–118.

Wong, C.S. & Li, W.K. (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society: Series B* 62, 95–115.

Wong, C.S. & Li, W.K. (2001a). On a logistic mixture autoregressive model. *Biometrika* 88, 833–846.

Wong, C.S. & Li, W.K. (2001b). On a mixture autoregressive conditional heteroscedastic model. *Journal of the American Statistical Association* 96, 982–995.

Zeevi, A., Meir, R. & Adler, R.J. (2000). Non-linear models for time series using mixtures of autoregressive models. technical report, Technion and Stanford University.

THE NUMBER OF SHAREHOLDERS: TIME SERIES MODELLING AND SOME EMPIRICAL RESULTS

Kurt Brännäs
Umeå University

1 Introduction

This short paper models time series of the number of shareholders (owners) in large Finnish and Swedish stocks. The numbers of owners may beyond being interesting in their own right be viewed as an integral part of the time series of trading volume. While market data allows for intra-day studies, the number of owners is typically provided publicly by securities depositories and some companies at much lower frequencies.

The integer-valued autoregressive model of order one (INAR(1)) with constant parameters (McKenzie 1985; Al-Osh & Alzaid 1987) is taken as the basic modelling approach. This model catches the main aspects of the integer-valued, number of owners variable in a relatively simple way. In contrast to conventional time series modelling the adopted approach offers interesting interpretations in terms of the model's parameters. We may obtain measures such as the mean holding time from the parameters, even though sampling individual transactions in the records of securities depositories may be a better alternative (cf. Bøhrens et al., 2006, for a study based on individual transactions). The available low sampling frequency or sparse but equidistant reporting dates brings along the problem of short holdings falling between measurement days. An approach to handling this problem is discussed and used empirically.

Empirically, we focus on the stocks of eight large stock market companies, the four largest (by the end of 2005) in each of Finland and Sweden, in terms of their numbers of shareholders. Some of the companies are traded not only in their home markets, but also elsewhere. The numbers of recorded owners in the Finnish and Swedish depositories also contain registered (but not all) foreign owners. The consequences of these and a few other extensions of the INAR(1) are also analyzed.

2 The Basic Model and Extensions

The number of owners/shareholders of a stock at some time point t is a count data or integer-valued variable y_t taking values $0, 1, 2, \ldots$. The y_t is composed of the number of new shareholders, ε_t, that buy into the stock in the period $(t-1, t]$ and of those owners at $t-1$ that remain owners at time t; these are denoted by $\alpha \circ y_{t-1}$. We write the model as

$$y_t = \alpha \circ y_{t-1} + \varepsilon_t. \tag{1}$$

Here, the symbol $\circ$ indicates a binomial thinning operation and it replaces conventional multiplication. The thinning is defined through $\alpha \circ y = \sum_{i=1}^{y} u_i$, where $\{u_i\}$ is an independent sequence of random $0-1$ variables with $\Pr(u_i = 1) = \alpha$. Importantly, thinning gives integers in the $[0, y]$ range. The α is the probability of remaining an owner to the next time period, i.e. α is a survival probability. Assuming that y and $\{u_i\}$ are independent gives that $E(\alpha \circ y) = E_y\left[E(\alpha \circ y)|y\right] = \alpha E(y)$ and $V(\alpha \circ y) = \alpha^2 V(y) + \alpha(1-\alpha)E(y)$. For the $\{\varepsilon_t\}$ sequence we assume independence and that $E(\varepsilon_t) = \lambda$ and $V(\varepsilon_t) = \sigma^2$. The λ corresponds to the average number of new owners over a time period. For the important special case of the Poisson distribution we have that $\sigma^2 = \lambda$. The model was first introduced by McKenzie (1985) and independently by Al-Osh & Alzaid (1987). It has later been discussed and generalized in a number of studies, cf. the survey of McKenzie (2003).

The basic model has the moment properties

$$\begin{aligned}
E(y_t|\mathcal{F}_{t-1}) &= \alpha y_{t-1} + \lambda \\
E(y_t) &= \lambda/(1-\alpha) \\
V(y_t|\mathcal{F}_{t-1}) &= \alpha(1-\alpha)y_{t-1} + \sigma^2 \\
V(y_t) &= \left[\alpha(1-\alpha)E(y_{t-1}) + \sigma^2\right]/(1-\alpha^2),
\end{aligned} \tag{2}$$

where $\mathcal{F}_{t-1}$ is the available information up through time $t-1$. The expected number of owners is $\lambda/(1-\alpha)$. The probability of remaining an owner over k periods is α^k and the expected duration of ownership (mean holding time at the t-scale) is $MH = 1/(1-\alpha)$. The median duration falls into the interval where the survival function is 0.5. The model also has a conditional heteroskedasticity property as is evident from $V(y_t|\mathcal{F}_{t-1})$.

Given that we write the model as $y_t = \alpha \circ y_{t-1} + \gamma \circ \nu_t$, where ν_t may be viewed as potential entrants, it appears reasonable to consider, i.a., dependence between entry and exit mechanisms. This corresponds to dependence between the thinning operations and can be demonstrated to influence only the second order moment properties in most cases (Brännäs & Hellström, 2001). Hence, considering only first order moment properties is a way of assuring robustness to incorrect assumptions related to the thinning operations. On the other hand, if such assumptions could be justified they would pave the way for even more efficient estimation.

An alternative but more complex model for the number of owners is an integer-valued moving average (INMA) model with a special pattern of dependence between thinning operations, cf. Brännäs & Hall (2001). The INAR model can be viewed as an approximative and simpler dual representation of the INMA model. Obviously, the more parsimoniously parameterized or short lagged INAR(1) model has an empirical advantage with respect to the commonly short time series at hand.

Next we consider model consequences of a number of issues that may arise in this and related areas.

2.1 Low Sampling Frequency

Given the widely spread algorithmic trading at an almost continuous time scale and the availability of data at a monthly or even sparser time scale the question of appropriate time scale for the model in (2) is an important but difficult one. We have chosen to view the model in (2) as generating data on a daily basis. The available data is monthly or even quarterly. To account for a low sampling frequency of $s = 21$ days (for a trading month), 63 days (trading quarter) or even biannual data we consider the following modelling strategy based on temporal aggregation.

By successive substitution over s days from t to $t + s$ we get

$$y_{t+s} = \alpha^s \circ y_t + \sum_{i=1}^{s} \alpha^{s-i} \circ \varepsilon_{t+i}, \tag{3}$$

where equality is in distribution. In this context there are only observations at times that are separated by s.

With $\alpha < 0.9$ the first term in (3) can in practice be disregarded, while for an α closer to one the term remains important, at least, in the monthly or $s = 21$ trading days case.

The conditional expectation of y_{t+s} in (3) conditional on information up through the previous observation time, i.e. t, takes the form

$$E(y_{t+s}|\mathcal{F}_t) = \alpha^s y_t + \lambda \sum_{i=1}^{s} \alpha^{s-i} = \alpha^s y_t + \frac{\lambda(1-\alpha^s)}{1-\alpha}, \tag{4}$$

which indicates that the model is still of the INAR(1) type (cf. Brewer, 1973).

Note that the parameters in the conditional expectation (4) differs from those of (2). In particular, the survival probability is, not surprisingly, smaller at the sparser time scale, and the mean entry is larger. The unconditional mean $E(y_{t+s})$ remains unchanged, while the conditional variance is less dependent on past y than the one in (2), since $\alpha^s(1-\alpha^s) < \alpha(1-\alpha)$.

2.2 *Two Owner Types*

Consider a case of two types of owners with numbers y_{1t} and y_{2t} that are unobservable, while their sum is observed. Both follow INAR(1) models but with different parameters $\alpha_i, \lambda_i, i = 1, 2$. We may consider α_2 to be smaller than α_1 and then to correspond to owners having short holding times. Since $y_{t+s} = y_{1t+s} + y_{2t+s}$ we have that $E(y_{1t+s} + y_{2t+s}|\mathcal{F}_t)$ arises as the sum of expressions of the type in (4), so that we can get, say, the conditional expectation

$$E(y_{t+s}|\mathcal{F}_t) = \alpha_1^s y_t + (\alpha_2^s - \alpha_1^s)y_{2t} + \frac{\lambda_1(1-\alpha_1^s)}{1-\alpha_1} + \frac{\lambda_2(1-\alpha_2^s)}{1-\alpha_2}. \tag{5}$$

When α_2 is small $\alpha_2^s \approx 0$ leaving us with a second term $-\alpha_1^s y_{2t}$ in (5). As we cannot observe y_{2t} we may replace it with its expected value $\lambda_2/(1-\alpha_2)$ to get an approximative expression

$$E(y_{t+s}|\mathcal{F}_t) \approx \alpha_1^s y_t + \frac{\lambda_1(1-\alpha_1^s)}{1-\alpha_1} + \frac{\lambda_2(1-\alpha_1^s)}{1-\alpha_2}. \tag{6}$$

Even if α_1 can be uniquely estimated using (6) the other parameters cannot be separately estimated without further information. Importantly, (6) then suggests that it may be empirically difficult to catch the survival probability α_2 by this modelling approach, at least, with low sampling frequencies for the time series.

2.3 *Two Depositories*

Shareholders in a stock may be registered in either of two depositories, domestically (D) or abroad (A). Over time registered ownership may move between the two depositories according to

$$y_{Dt} = \alpha_D \circ y_{D,t-1} + \beta_A \circ y_{A,t-1} + \varepsilon_{Dt} \tag{7}$$

$$y_{At} = \alpha_A \circ y_{A,t-1} + \beta_D \circ y_{D,t-1} + \varepsilon_{At}, \tag{8}$$

where we expect the β_A and β_D migration probabilities to be small, and much smaller than the α_D and α_A probabilities.

If we only have access to the domestic series $\{y_{Dt}\}$ we may substitute from (8) to get rid of the $y_{A,t-1}$ part in (7). We use $y_{it} = E(y_{it}|\mathcal{F}_{t-1}) + \xi_{it}$, $i = A, D$, where $E(\xi_{it}|\mathcal{F}_{t-1}) = 0$ for both i to get

$$y_{At} = \frac{\beta_D}{1-\alpha_A L} y_{D,t-1} + \frac{\lambda_A}{1-\alpha_A} + \frac{\xi_{At}}{1-\alpha_A L}.$$

Inserting this into the conditional representation of (7) then gives

$$\begin{aligned} E(y_{Dt}|\mathcal{F}_{t-1}) &= \alpha_D y_{D,t-1} + \beta_A\beta_D y_{D,t-2} + \alpha_A\beta_A\beta_D y_{D,t-3} \\ &\quad + \alpha_A^2\beta_A\beta_D y_{D,t-4} \ldots + \lambda_D + \frac{\beta_A\lambda_A}{1-\alpha_A}. \end{aligned} \tag{9}$$

With a small $\beta_A\beta_D$ the autoregression can in practice be expected to be of order one. When this is the case only α_D can be separately estimated by, e.g., a conditional least squares estimator.

Systematically observing only every s observation of a daily series leads to an INAR with the same maximum lag as in (9) (Brewer, 1973). Given the complexity of the general expression it may appear reasonable to use (9) directly and then to interpret the parameters in terms of the s time scale.

2.4 Time Dependence

In practice both λ and α can be expected to vary over time. Liquidity and the share price are examples of variables that can be expected to influence both parameters. Over longer time horizons parameters may also change due to the common use of share expansions in take-overs or due to repurchases of shares.

It is straightforward to introduce time dependence in the parameters of (4) as long as they remain constant within the month for $s = 21$, etc. We may write

$$E(y_{t+s}|\mathcal{F}_t) = \alpha_t^s y_t + \frac{\lambda_t(1-\alpha_t^s)}{1-\alpha_t}. \tag{10}$$

Brännäs (1995) suggested that a logistic distribution function $\alpha_t = 1/[1+\exp(\mathbf{x}_t\boldsymbol{\beta})]$ and an exponential function $\lambda_t = \exp(\mathbf{z}_t\boldsymbol{\gamma})$ provide convenient parameterizations where use is made of exogenously determined variable vectors $\mathbf{x}_t$ and $\mathbf{z}_t$ and the unknown parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Note that the unconditional mean in (2) suggests that incorporating the same variable in $\mathbf{x}_t$ and $\mathbf{z}_t$ will likely lead to strong negative correlation between the associated parameters.

2.5 Unit Root

Given the count data interpretation of the $\{y_t\}$ sequence and the assumption $\varepsilon_t \geq 0$, a unit root or $\alpha \equiv 1$ gives $y_t - y_{t-1} = \varepsilon_t \geq 0$. Hence, a unit root implies that an owner series can remain constant or be growing, but it cannot decrease. Importantly, this corresponds to an infinite holding time. Therefore, a unit root hypothesis may be rejected on logical grounds in most cases. Note that there is some empirical evidence of overestimating α in AR(1) models when it is de facto time-varying rather than time invariant.

3 Estimation

From Al-Osh & Alzaid (1987) and followers several estimators for the INAR(1) model have been studied and compared. Among the simpler ones for the current context of very large numbers of shareholders the Yule-Walker and conditional least squares (CLS) estimators are both simple to use, are consistent estimators, and have been found to perform well in small samples.

The constant parameter model in (4) has for monthly data, $t = 1, \ldots, T$, the one-step-ahead predictor $E(y_t|\mathcal{F}_{t-1}) = \alpha^s y_{t-1} + \lambda(1-\alpha^s)/(1-\alpha)$, and the prediction error is $e_t = y_t - E(y_t|\mathcal{F}_{t-1})$. For the other model specifications the appropriate conditional expectation should be used. The prediction errors have zero means, but conditional variances vary with, e.g., the dependence between and within the binomial thinning operations (cf. Brännäs & Hellström, 2001). For this reason estimation will here not explicitly build on a conditional variance specification as would be required for a conditional weighed least squares estimator. Additionally, we refrain from making distributional assumptions, so that maximum likelihood estimation is precluded.

A linear CLS estimator is based on only a first conditional moment assumption and gives estimates a of α^s and b of $\lambda(1-\alpha^s)/(1-\alpha)$. The underlying $\hat{\alpha}$ and $\hat{\lambda}$ can be obtained from these expressions as $\hat{\alpha} = a^{1/s}$ and $\hat{\lambda} = b(1-a^{1/s})/(1-a)$. Standard errors can be obtained using the Delta method, e.g., for $\hat{\alpha}$ the standard error is $a^{(1/s)-1}V^{1/2}(a)/s$. The CLS estimator should have an accompanying robust covariance matrix for the a and b estimators to account for conditional heteroskedasticity of unspecified form. A nonlinear CLS estimator can also be used to obtain the $\hat{\alpha}$ and $\hat{\lambda}$ estimates directly. Introducing exogenous variables through the α and λ parameters creates no substantially new and difficult obstacles for the nonlinear CLS estimator.

For the Swedish stocks the available time series is initially measured on a biannual scale and later on a quarterly one. The LS criterion function S can be decomposed accordingly, i.e. as $S = S_1 + S_2$, where S_1 is for the biannual data. The prediction errors are then initially based on $s_1 = 126$ and later $s_2 = 63$ days, respectively.

The Yule-Walker estimator of α and λ is a moment estimator based on $\bar{y} = \lambda/(1-\alpha)$ and $r_1 = \alpha^s$, where $\bar{y}$ is the sample average and r_1 is the estimator of the lag one autocorrelation. Hence, this estimator is based on unconditional moment expressions while the CLS estimator is based on the conditional mean expression.

4 Empirical Findings

We present empirical results for the four largest and registered companies in terms of their numbers of owners (December 2005) of the Helsinki (Finland) and Stockholm (Sweden) stock markets. For Finland there are monthly time series 2000:12–2012:10 ($T = 143$) and for Sweden (expressed in months) biannual observations 1999:12–2005:12 followed by quarterly observations 2006:3–2012:9 ($T = 40$).[1] Note that in each case the number of owners corresponds to owners registered domestically. Ownership of American Depository shares may therefore influence the interpretation of $\hat{\lambda}$ (see, e.g., the annual reports 2008 of Nokia and Ericsson).

The Finnish time series are exhibited in Figure 1. The Elisa series contains a few jumps that correspond to takeovers when the Elisa stock was involved. Nokia has an expansion phase up to about the 40th month (2004:3) followed by a long recession before recovering towards the latter part of the series. Note that Nokia to some 90 percent is foreign-owned. The UPM series shows some fluctuation, while for Sampo the predominant impression is one of growth in the number of shareholders.

For each Finnish stock we consider two specifications, one in which the parameters are time invariant and one in which we have a time dependent α_t parameter of the logistic type and λ_t as an exponential function. The time-varying specifications may indicate the role of time variation on, e.g., the MH. Here, α_t is set to depend on the price change in the previous month, i.e. $\Delta p_{t-1} = p_{t-1} - p_{t-2}$ and on the number of owners at the end of previous month (y_{t-1}) (divided by 10000). The λ_t is a function of Δy_{t-1}. For Elisa we also include a dummy variable (d_t) taking value one for the months of larger changes in the stock in the time-varying λ_t.

For Nokia we find that the Yule Walker estimator gives $\hat{\alpha} = 0.9985$ and an implied holding time of about 31 trading months. The CLS estimator gives an even longer MH. The fit is good $(R^2 = 0.97)$ but there is remaining serial correlation of the

[1] TeliaSonera has the first observation at 2000:6.

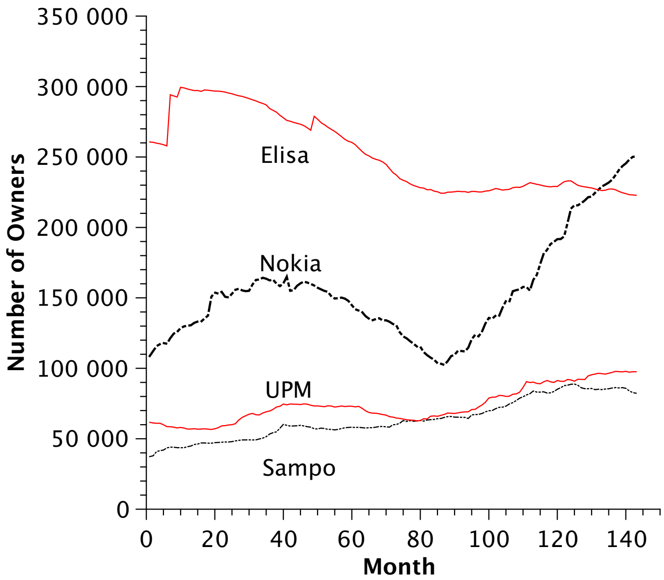


Figure 1. The Finnish shareholder time series vs observation month, 2000:12–2012:10.

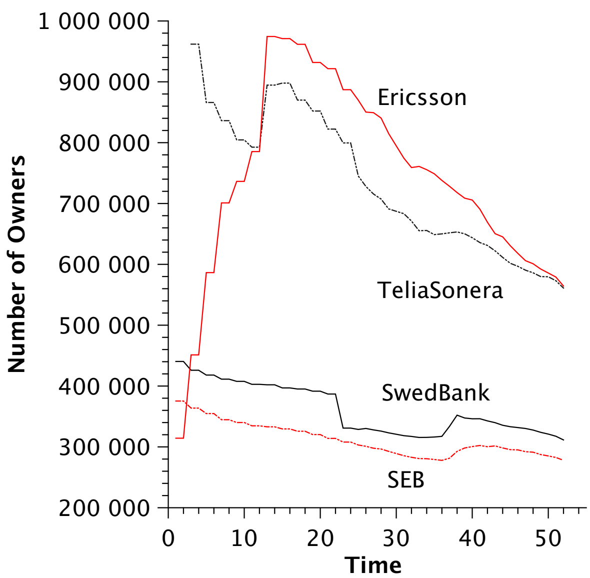


Figure 2. The Swedish shareholder time series vs observation number, 1999:12-2005:12 (biannually), 2006:3-2012:9 (quarterly).

Table 1. Yule-Walker (Finland) and CLS (Sweden) estimates for constant parameter specifications. The mean holding time (MH) is in months.

Stock	$\hat{\alpha}$	$\hat{\lambda}$	T	MH
		Finland		
Elisa	0.9994	3028.1	143	83
Nokia	0.9985	4960.3	143	31
Sampo	0.9990	1274.4	143	50
UPM	0.9991	1390.1	143	53
		Sweden		
Ericsson	0.9982	1559.7	40	26
SEB	0.9992	219.6	40	59
SwedBank	0.9993	208.2	40	68
TeliaSonera	0.9881	8855.4	39	4

Note: The Yule-Walker estimator is based on the mean and lag one autocorrelation.

AR(1) type. Therefore, in an INAR(2) model the lag two parameter not surprisingly comes out with a negative sign. This also speaks against, e.g., a two depository and an INAR(2) interpretation. After some trial and error we give an estimated time dependent parameter model as $\hat{y}_t = y_{t-1}/(1 + \exp(-5.0 + 0.022\Delta p_{t-1} - 0.078y_{t-1}/10000)) + \exp(5.82 + 0.418\Delta y_{t-1})$, where all parameter estimates are significant. A positive price change is estimated to have a reducing effect on α_t, while an increase in the number of owners has an enhancing effect. The implied holding times vary between 1.2 and 4.2 trading years, with an average of about 2.0 years. The longest holding times are noted for the final part of the series, where the share price is lowest. Hence, introducing explanatory variables seems to reduce the mean holding time. Moreover, there is no remaining serial correlation (with the exception of the UPM model) thanks to the Δy_{t-1} included in the λ_t-part.

Table 1 gives the Yule-Walker estimates for all Finnish and CLS estimates for all Swedish stock series. The mean holding times are reported in trading months. In general, the models fit the Finnish series very well, but there are serial correlation problems. The $\hat{\alpha}$ estimates are very close to one, and even closer for the CLS estimator.

Table 2 contains estimates for the time varying α_t and a λ_t model specifications for the Finnish stocks. The price effect appears positive and the lagged y_{t-1} effect comes out as significantly negative. The Elisa specification produces a holding time estimate that appears on the long side, while the other holding times appear more realistic. Average entry is affected positively by an increase in the lagged change in ownership. Only for UPM is there significant remaining serial correlation.

Table 2. Nonlinear CLS estimates for time dependent parameter specifications. The mean holding time (MH) is in years and $y^*_{t-1} = y_{t-1}/10000$.

	α_t			λ_t				
Stock	Const	Δp_{t-1}	y^*_{t-1}	Const	Δy_{t-1}	d_t	MH	LB_5
Elisa	-12.67	0.047	0.143	-22.73	7.17	30.29	37.2	6.2
	(0.01)	(0.009)	(0.000)	(0.03)	(0.09)	(0.03)		
Nokia	-5.00	0.022	-0.078	5.82	0.418	-	2.0	6.8
	(0.36)	(0.011)	(0.007)	(0.21)	(0.14)			
Sampo	-7.73	-0.009	-	3.66	2.80	-	9.0	9.4
	(0.05)	(0.000)		(0.002)	(0.00)			
UPM	-5.45	0.056	-0.146	4.74	1.16	-	2.8	14.2
	(0.01)	(0.024)	(0.000)	(0.01)	(0.22)			

The Swedish stocks are displayed in Figure 2. In contrast to the Finnish stocks there is a negative trend, at least, in the latter parts of the series. Ericsson peaks at about the same time as Nokia. The Swedish series are very short and when estimated individually there is less room to enlarge by using time dependent parameters.

The constant parameter estimation results are given in Table 1. The implied mean holding times range between 0.3 and 5.7 years. The MH estimate of TeliaSonera is surprisingly small and its asymptotic 95 percent confidence interval is 0 - 8.2 months. The downward trend of the series is overall steeper than for any of the other series, which may be the reason for the small $\hat{\alpha}$ estimate. The confidence interval for Ericsson is 19.6 - 32.4 trading months.

5 Concluding Remarks

The only other study from a Nordic country that we are aware of is Bøhrens et al. (2006), whose main interest was in corporate governance issues. They studied Norwegian registered firms and found, using micro-data 1989-1999, that median holding times were in the region 1-2 years across firms, excluding financial ones. Their sample included both small and large firms beyond being based on a different time period. Their results, e.g., indicate that foreign owners have shorter mean holding times and we find shorter ones for Nokia, Ericsson as well as TeliaSonera that have internationally spread ownership. It is obviously possible to use the current modeling approach to studying corporate governance related issues across both time and different stocks.

The empirical results indicate that mean holding can be expected to vary across time. The suggested specifications are, admittedly, quite ad hoc and some more serious thinking is required. Alternatively, some flexible time polynomial could be adopted to catch main features across time. Even so, the simple INAR(1) model provides a very good fit to the data, and there is not much remaining to explain.

References

Al-Osh, M.A. & Alzaid, A.A. (1987). First-order integer valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* 8, 261–275.

Bøhrens, Ø., Priestley, R. & Ødegaard, B.A. (2006). The duration of equity ownership at Oslo Stock Exchange 1989–1999. Research Report. BI Norwegian School of Management.

Brewer, K.R.W. (1973). Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models. *Journal of Econometrics* 1, 133–154.

Brännäs, K. (1995). Explanatory variables in the AR(1) count data model. Umeå Economic Studies 381.

Brännäs, K. & Hall, A. (2001). Estimation in integer-valued moving average models. *Applied Stochastic Models in Business and Industry* 17, 277–291.

Brännäs, K. & Hellström, J. (2001). Generalized integer-valued autoregression. *Econometric Reviews* 20, 425–443.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin* 21, 645–650.

McKenzie, E. (2003). Discrete variate time series, in Shanbhag, D.N. & Rao, C.R. (eds.) *Handbook of Statistics*, Volume 21, 573–606, Elsevier Sciences, Amsterdam.

IV

FINANCE

ASYMMETRIC DYNAMIC LINKAGES BETWEEN RETURNS ON BANKS AND OTHER INDUSTRY PORTFOLIO RETURNS

Kenneth Högholm and Johan Knif
Hanken School of Economics

Gregory Koutmos
Fairfield University

1 Introduction

This study revisits the issue of dynamic linkages between returns on industry portfolios and contributes to the discussion by specifically focusing on the role of the banking industry in the dynamic linkages among industry portfolio returns. Using monthly returns on 48 US industry portfolios, the results show that the dynamic linkages between the returns on the banking industry portfolio and other industries often are asymmetric in two ways. First, there appears to be a one-directional causality relation running from the banking industry to several other industries but seldom the other way around. One-month lagged banking industry returns seem to improve predictability of returns on several industry portfolios. Surprisingly, for many industry portfolios returns on the banking industry portfolio can be regarded as exogenous and Granger cause other industry returns. Second, the cross-autocorrelation is found to be asymmetric in a sense that on average the impact of a one-month lag of the return of the banking portfolio is about twice as high in the lower part of the conditional return distribution than in the upper part. Furthermore, the empirical results indicate that returns on the banking industry portfolio are dynamically connected to two of the classic four asset-pricing risk factors. Returns on the banking industry portfolio Granger cause and lead the returns on the size factor and the momentum factor.

Boudoukh, Richardson, and Whitelaw (1994) documented empirical results that imply that the impact of autocorrelations had been overstated in the existing literature. However, since then a vast amount of literature has documented new information of both autocorrelations and cross-autocorrelations.

As different industries exhibit different exposure to risk factors it could be expected that in some market situations the returns on one industry portfolio could lead or lag the returns on some other industry portfolios with different risk exposure characteristics. However, the source of the empirically observed lead-lag effects is still a subject of debate. The proposed explanations include non-synchronous trading (e.g. Lo and MacKinley (1990a)), time-varying expected returns (e.g. Hameed (1997)), asymmetric information (e.g. Zebedee and Kasch-Haroutounian (2009), imperfect information (Chan (1993)) and slow diffusion of information (Merton (1987) and Lo and MacKinley (1990b)).

Using a static model of multiple stocks where investors have access to limited information, Merton (1987) shows that stocks with a smaller investor base are traded at greater discount due to limited risk sharing. Merton (1987) also suggests that market segmentation and limited participation could be a reason for slowness of investors in one market to absorb information from another market. This argument is often called the gradual-information-diffusion hypothesis. Hou (2007) found that this slow diffusion of information is the leading cause of the lead-lag effect and that it is predominantly an intra-industry phenomenon that is associated with firm size: big firms lead small firms. This explanation is also provided by Ayers and Freeman (2000) and recently thoroughly examined across industries by Cen, Chan, Dasgupta, and Gao (2013). Anderson, Eom, Hahn and Park (2013) find compelling evidence that this partial price adjustment is a major source of the autocorrelation in returns.

Hong, Torous, and Valkanov (2007) also find that the gradual-information-diffusion hypothesis provides a key auxiliary explanation for the lead-lag relations but might not be the only one. Using monthly returns on 34 value-weighted US industry portfolios over the period from 1946 to 2001 they found that 14 industries were able to predict market movements by one month. A few industries such as petroleum, metal, and financial could predict the market up to two months ahead. They also provided remarkably similar empirical evidence for the eight largest non-US equity markets. Their conclusion is that stock markets as a whole might react with a delay to fundamental information contained in industry returns and that information diffuses only gradually across markets.

In a recent paper Laopodis (2013) continues on this issue and empirically investigates the dynamic linkages among industries and the stock market. Using a newer dataset based on monthly returns on seventeen large US industry portfolios and the aggregate stock market over the period from 1957 to 2012, he finds that certain industries provide strong predictive ability both to the aggregate market and many other industries. Examining the dynamic behavior over bull and bear markets separately he finds no consistent patterns of responses.

Chan (1993) and McQueen, Pinegar, and Thorley (1996) find directional asymmetry in the small stock response to large stock movements. International evidence is presented in e.g. Altay (2004) for the German and Turkish markets. Doong, Yang, and Chiang (2005) and Lee, Chen, and Chang (2013) present results for the Asian stock markets.

Recently Bernhardt and Mahani (2007) argued that information asymmetry cannot fully explain asymmetry in lead-lag relations. Additional frictions are necessary to produce asymmetry in cross-autocorrelations of stock returns. Furthermore, the results of Chou, Ho, and Ko (2013) indicate that common risk factors extracted from industry returns contain significant risk premiums and have explanatory power up and above those of size, value and momentum.

Baur, Dimpfl, and Jung (2012) recognize another type of asymmetry of return autocorrelations. They analyze the dependence pattern over a range of quantiles of the conditional return distribution. The results indicate positive dependence on past returns in the lower part of the distribution while the upper quantiles are marked by negative serial dependence. This type of dependence on the outcome of the conditional distribution will make the autocorrelation structure at least partly endogenous and hence empirically time-varying. The time-varying characteristic of the cross-autocorrelation structure is also investigated in Kinnunen (2013).

The special systemic role of the banking industry has been recognized for a long time. The credit channel effect has been thoroughly discussed by Bernanke (1993), Bernanke and Gertler (1995), and Anari, Kolari, and Mason (2004). In a recent paper Hammami and Lindahl (2014) report empirical findings that underscore the relevance of bank credit growth for stock prices. They conclude that bank credit growth is important because it is able to predict business cycle variables and labor income growth.

This study focuses specifically on the role of the banking industry in the dynamic lead-lag relation to other industries. As the banking industry by its liquidity providing nature is closely related to all other industries, it is expected that the banking sector plays a major role in the dynamic interdependencies. On the one hand, the banks are dependent on the performance of other industries. On the other hand, other industries are financially dependent of the performance of the banking sector. The empirical results of this study indicate a dynamic linkage between the returns on the banking portfolio and other industry portfolios that appears to be asymmetric in two ways. First, a one-directional causality relation running from the banking industry to several other industries is found but seldom the other way around. Lagged banking industry returns seem to improve the predictability of returns for several industry portfolios. Surprisingly, for many industry portfolios returns on the banking industry portfolio can be regarded as exogenous and Granger causes other industry returns. Second, in line with Baur et al.

(2012) the results show asymmetry in the autocorrelation structure: positive in the lower part of the conditional return distribution and negative in the upper part. Furthermore, the cross-autocorrelation is found to be asymmetric such that on average the impact of a one-month lag of the return of the banking portfolio is about twice as high in the lower part of the conditional return distribution than in the upper part.

The rest of the study is structured as follows. Section 2 describes the empirical methodology applied. Section 3 presents the data and industry classifications. Section 4 reports the empirical findings and Section 5 summarizes and concludes.

2 Methodology

In order to monitor the role of the banking industry in the dynamic linkages with other industries, we first investigate the individual autocorrelation structure for each industry separately. Traditionally the autocorrelation as a function of the lag-length k is calculated as $\rho(k) = \gamma(k)/\gamma(0)$, where γ is the auto-covariance function, $\gamma(k) = cov(r_t^{ind}, r_{t-k}^{ind})$, and r_t^{ind} is the continuously compounded return on the industry index for time period *t*.

Following the VAR approach applied by Laopodis (2013), we study the Granger causality between the banking industry and the other industries. The basic VAR-model with two lags for this analysis is

$$\begin{aligned} r_t^{banks} &= \beta_{10} + \beta_{11} r_{t-1}^{banks} + \beta_{12} r_{t-2}^{banks} + \beta_{13} r_{t-1}^{ind} + \beta_{14} r_{t-2}^{ind} + \varepsilon_t^{banks} \\ r_t^{ind} &= \beta_{20} + \beta_{21} r_{t-1}^{ind} + \beta_{22} r_{t-2}^{ind} + \beta_{23} r_{t-1}^{banks} + \beta_{24} r_{t-2}^{banks} + \varepsilon_t^{ind} \end{aligned} . \quad (1)$$

The two null hypotheses in the Granger casualty test are that returns on the industry portfolio (*ind*) do not Granger cause returns on the banking industry portfolio (*banks*), $\beta_{13} = \beta_{14} = 0$, and that returns on the banking industry portfolio (*banks*) do not Granger cause returns on the industry portfolio (*ind*), $\beta_{23} = \beta_{24} = 0$. Based on this bivariate VAR-model we also check exogeneity using the Wald block exogeneity test.

As a basic benchmark we estimate a model to monitor the impact of lagged information from the banking industry on other industries. The following model is estimated using OLS:

$$r_t^{ind} = \beta_0 + \beta_1 r_{t-1}^{ind} + \beta_2 r_{t-1}^{banks} + \varepsilon_t^{ind}. \quad (2)$$

We develop the modeling further by estimating the basic model using quantile regression. In line with Baur, Dimpfl, and Jung (2012), we check this model for robustness over the conditional return distribution using the quantile regression approach of Koenker and Bassett (1978). This approach is described in detail in Koenker (2005). The traditional regression approach (OLS) focuses on the conditional mean of the dependent variable and implicitly assumes that this is a good representation for the entire distribution. The quantile regression approach, on the other hand, models the quantiles of the conditional distribution of the dependent variable given conditioning explanatory variables. This enables monitoring and testing of the regression coefficients across different parts of the conditional return distribution. Furthermore, as quantile regression requires weaker distributional assumptions, it provides a more robust method of modeling the conditional return distribution and is, hence, less sensitive to extreme observations. Traditional quantile estimators provide conditional estimates of the lead-lag structure. These estimates are, on the one hand, by definition conditional on the lag specifications in the estimated model. On the other hand, the quantile estimates are also conditional upon the quantile parameter that specifies the weighting across the distribution of the idiosyncratic component.

The quantile regression approach is also applied to empirical finance in Högholm, Knif, and Pynnönen (2011a,b) and Högholm, Knif, Koutmos, and Pynnönen (2011). Using the conditional quantile regression, explicit conditioning variables do not need to be specified as the estimation procedure implicitly accounts for the joint effect of possible conditioning variables by conditioning on the residual distribution. For the quantile regression analysis, equation (2) is rewritten in the form

$$r_t^{ind} = \beta_0(\tau) + \beta_1(\tau) r_{t-1}^{ind} + \beta_2(\tau) r_{t-1}^{banks} + \varepsilon_t^{ind}, \tag{3}$$

where $\beta_i(\tau)$, *i*= 0,1,2, will be functions of the quantile parameter τ. The quantile regression will solve the minimization problem

$$\min_{\beta(\tau)_i, i=0,1,2} \left[\sum_{t: r_t^{ind} \geq \hat{r}_t^{ind}} \tau \left| r_t^{ind} - \hat{r}_t^{ind} \right| + \sum_{t: r_t^{ind} < \hat{r}_t^{ind}} (1-\tau) \left| r_t^{ind} - \hat{r}_t^{ind} \right| \right], \tag{4}$$

where $\hat{r}_t^{ind}$ is the estimated expectation of (3) and τ is the quantile parameter ranging from 0 to 1. In case of $\tau = 1$, the quantile regression will result in a least-absolute-deviation regression for positive residuals. Correspondingly, in case $\tau = 0$, the result is a least-absolute-deviation regression for negative residuals. Setting $\tau = 0.5$ provides a least-absolute-deviation regression at the median. Letting τ vary between 0 and 1, the quantile regression will monitor the regression lead-lag relationship across the entire conditional industry return distribution. As τ defines the weighting pattern over the conditional return distribution for the minimization in (4), the quantile regression will estimate a model that is implicitly non-linear with implicitly time-varying and conditional regression coefficients.

3 Data

The sample data used in the empirical estimation and testing consists of monthly returns on 48 US value-weighted industry portfolios over the period from January 1970 to July 2013. The data is downloaded from Kenneth French data library (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). According to this industry classification structure, the banking industry portfolio contains equity from the sectors listed in Table 1 and the other 47 industries are presented in Appendix A. As seen from this industry classification, all borrowing and lending activities are grouped in the banking industry whereas all financial trading is concentrated to the finance and trading industry, and all insurance activities are grouped in the insurance industry.

Table 1. Banking industry portfolio structure

Kenneth French gives the following description: We assign each NYSE, AMEX, and NASDAQ stock to an industry portfolio at the end of June of year t based on its four-digit SIC code at that time. However, we use not only CRSP, but also Compustat as a source of SIC codes. We use Compustat SIC codes (for the fiscal year ending in calendar year t-1) whenever available. Otherwise, we use CRSP SIC codes (for June of year t).

Banking	
6000-6000 Depository institutions	6010-6019 Federal reserve banks
6020-6020 Commercial banks	6021-6021 National commercial banks
6022-6022 State banks - Fed Res System	6023-6024 State banks - not Fed Res System
6025-6025 National banks - Fed Res System	6026-6026 National banks - not Fed Res System
6027-6027 National banks, not FDIC	6028-6029 Banks
6030-6036 Savings institution	6040-6059 Banks (Other)
6060-6062 Credit unions	6080-6082 Foreign banks
6090-6099 Functions related to deposit banking	6100-6100 Non-depository credit institutions
6110-6111 Federal credit agencies	6112-6113 FNMA
6120-6129 S&Ls	6130-6139 Agricultural credit institutions
6140-6149 Personal credit institutions	6150-6159 Business credit institutions
6160-6169 Mortgage bankers	6170-6179 Finance lessors
6190-6199 Financial services	

Table 2. Descriptive statistics for the return distribution of monthly returns on 48 US industry portfolios over the period from January 1970 to July 2013

	Mean	Max.	Min.	Std. Dev.	Skew-ness	Kurtosis	Jarque-Bera	Prob.
aero	1.24	25.33	-30.28	6.84	-0.38	4.82	84.80	0.00
agric	1.04	28.56	-28.81	6.53	-0.02	4.82	72.05	0.00
autos	0.89	49.45	-36.42	7.16	0.23	8.64	696.60	0.00
banks	0.97	24.94	-27.87	6.25	-0.30	5.10	103.98	0.00
beer	1.13	25.82	-19.74	5.43	-0.04	5.22	107.53	0.00
bldmt	1.02	35.74	-30.86	6.41	-0.04	7.09	364.16	0.00
books	0.91	30.57	-24.70	5.99	0.02	5.14	99.42	0.00
boxes	1.01	20.90	-28.32	5.89	-0.44	4.99	102.91	0.00
bussv	1.06	25.28	-27.56	6.80	-0.17	4.18	32.66	0.00
chems	1.07	22.01	-27.96	5.78	-0.17	5.27	114.41	0.00
chips	0.98	26.80	-31.79	7.79	-0.35	4.46	57.01	0.00
clths	1.08	32.39	-30.86	6.81	-0.06	5.40	125.92	0.00
cnstr	0.94	24.28	-31.21	7.47	-0.13	3.89	18.91	0.00
coal	1.24	46.41	-38.04	10.63	0.32	4.80	79.51	0.00
comps	0.88	24.45	-32.78	7.40	-0.16	4.50	51.38	0.00
drugs	1.08	31.84	-19.10	5.16	0.19	5.76	168.93	0.00
elceq	1.20	23.17	-32.09	6.48	-0.22	4.65	63.60	0.00
fabpr	0.84	30.39	-26.67	7.18	-0.13	4.54	53.22	0.00
fin	1.12	19.46	-26.02	6.46	-0.45	4.22	50.40	0.00
food	1.15	19.34	-17.78	4.59	0.10	5.01	88.88	0.00
fun	1.31	40.68	-32.24	8.05	-0.21	6.02	202.19	0.00
gold	0.95	78.02	-33.63	10.71	0.77	8.14	626.62	0.00
guns	1.27	32.87	-30.08	6.74	-0.18	5.24	111.99	0.00
hlth	0.98	36.41	-41.07	8.47	-0.10	5.77	168.51	0.00
hshld	0.88	18.71	-21.67	4.84	-0.29	4.92	88.09	0.00
insur	1.05	26.60	-26.65	5.70	-0.29	5.11	104.73	0.00
labeq	0.98	22.48	-30.13	7.45	-0.17	4.05	26.51	0.00
mach	1.00	22.98	-31.37	6.52	-0.43	5.38	139.35	0.00
meals	1.06	27.79	-31.27	6.30	-0.55	5.60	174.05	0.00
medeq	0.95	21.10	-20.56	5.42	-0.36	4.33	49.71	0.00
mines	1.00	26.99	-34.85	7.56	-0.42	5.28	128.77	0.00
oil	1.15	24.40	-18.27	5.60	0.05	4.26	34.80	0.00
other	0.50	21.08	-28.64	6.94	-0.50	4.58	76.44	0.00
paper	1.00	24.11	-26.16	5.72	0.10	5.24	110.37	0.00
persv	0.64	24.68	-28.23	6.98	-0.28	4.54	58.60	0.00
rlest	0.54	59.19	-37.10	7.80	0.60	12.17	1864.52	0.00
rtail	1.08	27.08	-29.16	5.68	-0.19	5.03	92.91	0.00
rubbr	1.05	32.09	-30.49	6.16	-0.24	5.96	195.79	0.00
ships	1.07	29.33	-32.22	7.52	-0.01	4.52	50.43	0.00
smoke	1.49	32.46	-24.96	6.36	-0.10	5.55	142.70	0.00
soda	1.15	38.90	-26.28	6.82	0.15	6.78	313.39	0.00
steel	0.77	30.67	-32.52	7.74	-0.25	5.10	101.63	0.00
telcm	0.98	22.12	-15.56	4.84	-0.23	4.22	37.28	0.00
toys	0.82	27.03	-34.51	7.26	-0.24	4.46	51.46	0.00
trans	1.00	19.21	-28.07	6.01	-0.26	4.24	39.41	0.00
txtls	1.03	59.28	-32.63	7.54	0.52	12.35	1929.16	0.00
util	0.93	18.80	-12.65	4.15	-0.15	4.10	28.19	0.00
whlsl	0.98	17.94	-28.67	5.62	-0.35	5.54	150.80	0.00

Table 2 summarizes the descriptive statistics of the return distributions for the 48 individual industry indexes. The mean monthly return on all industry portfolios is positive. The highest monthly mean return of 1.49% is measured for the tobacco-product industry (smoke) and the lowest mean return of 0.50% for the industry group “other”. This “other” industry group contains e.g. sanitary services, steam, air conditioning supplies, and irrigation systems. The standard deviation is highest, 10.71, for the precious metal industry (gold) and lowest, 4.14, for utilities industry (util). The skewness is negative for the majority of the industry return distributions but positive and relatively high for the precious metal industry (gold). All industry portfolios exhibit excess kurtosis and this is especially high for the textiles industry (txtls). The Jarque-Bera statistic does not support the assumption that the unconditional industry return distributions are symmetric with no excess kurtosis.

4 Empirical results

The individual autocorrelation functions for the returns on the 48 industry portfolios are presented in Table 3. For 26 of the 48 industries, the autocorrelation with one lag is positive and significant at the 10% level or lower. This first order autocorrelation is as high as 0.206 for the real estate industry (rlest). Only six industries have a statistically significant autocorrelation of order two. Five of these six second order autocorrelations are negative. There appears to be no autocorrelation for lags 3 or longer. The banking industry (banks) has a first order autocorrelation of 0.108 that is statistically significant at the 5% level.

Table 4 presents the results for the Granger causality tests. For 25 of the 47 industries the null hypothesis that the banking industry does not Granger cause the industry is rejected at the 10% level of significance or lower. On the other hand, there appears to be only two other industries that significantly Granger cause returns on the banking industry portfolio. These two are the bituminous coal industry (coal) and the petroleum and natural gas industry (oil). The VAR-based bivariate exogeneity Wald test indicates that for the 25 industries that are Granger caused by the banking industry the banking industry can be regarded as exogenous. Only the bituminous coal industry (coal) and the petroleum and natural gas industry (oil) can be regarded as exogenous for the banking industry. The petroleum and natural gas industry (oil) did not exhibit any autocorrelation structure in Table 3 and returns on the bituminous coal industry (coal) portfolio only contain an autocorrelation for a two-month lag.

Table 3. Autocorrelation coefficients for monthly returns on 48 US industry portfolios over the period from January 1970 to July 2013

Significant autocorrelations are indicated with bold for significance levels of 10% or lower.

	lag 1	lag2	lag3	lag4	lag5	lag6
aero	**0.135**	-0.037	-0.053	0.026	0.021	-0.053
agric	0.019	-0.023	0.021	-0.045	-0.015	-0.031
autos	**0.118**	-0.046	0.053	0.034	-0.026	-0.138
banks	**0.108**	-0.028	0.005	-0.038	0.046	-0.079
beer	0.026	-0.029	-0.001	-0.011	0.032	0.024
bldmt	**0.090**	**-0.084**	0.008	0.042	-0.017	-0.185
books	**0.192**	-0.001	0.064	0.023	-0.014	-0.057
boxes	0.036	-0.039	0.004	-0.001	0.002	-0.100
bussv	**0.076**	-0.034	0.053	-0.055	-0.011	-0.004
chems	0.030	-0.044	0.037	0.006	-0.011	-0.094
chips	0.054	0.010	0.052	-0.017	-0.011	0.001
clths	**0.174**	-0.016	-0.041	-0.040	0.000	-0.034
cnstr	**0.131**	-0.027	-0.007	0.020	-0.026	-0.079
coal	0.041	**0.097**	-0.010	-0.027	-0.080	-0.029
comps	0.064	0.005	0.055	-0.026	-0.023	0.044
drugs	-0.028	0.018	-0.030	0.012	0.062	-0.018
elceq	0.022	-0.049	0.007	0.025	-0.016	-0.045
fabpr	**0.106**	-0.011	-0.060	-0.008	0.001	-0.114
fin	**0.150**	-0.056	0.009	0.009	0.036	-0.066
food	0.062	-0.004	-0.035	-0.049	0.090	-0.058
fun	**0.170**	-0.023	0.004	0.000	-0.038	-0.137
gold	-0.038	-0.068	0.012	-0.018	-0.031	-0.017
guns	0.043	**-0.112**	-0.037	0.099	0.029	-0.099
hlth	**0.164**	0.039	0.072	0.007	0.025	0.021
hshld	**0.076**	-0.005	-0.013	0.000	0.007	-0.100
insur	**0.117**	**-0.091**	-0.025	0.071	0.065	-0.122
labeq	**0.098**	-0.030	0.009	-0.028	-0.064	-0.017
mach	**0.090**	-0.035	0.020	-0.024	-0.059	-0.110
meals	**0.146**	0.011	-0.010	-0.067	-0.019	-0.112
medeq	0.070	0.008	0.004	0.002	0.042	-0.057
mines	0.048	0.022	0.053	-0.001	-0.014	-0.054
oil	-0.027	-0.042	-0.011	0.037	0.027	-0.051
other	**0.121**	-0.007	-0.022	-0.025	-0.022	-0.003
paper	0.016	-0.066	-0.013	0.023	-0.004	-0.047
persv	**0.108**	-0.030	-0.016	-0.056	0.035	0.011
rlest	**0.206**	-0.025	0.057	0.092	-0.025	-0.110
rtail	**0.144**	-0.048	-0.046	-0.020	-0.004	-0.098
rubbr	**0.105**	-0.012	-0.040	0.000	-0.019	-0.103
ships	0.058	-0.065	-0.012	0.011	-0.032	-0.114
smoke	0.049	0.002	-0.014	-0.018	0.035	-0.032
soda	0.044	-0.076	0.009	-0.022	0.024	0.001
steel	0.032	-0.026	0.009	0.021	-0.018	-0.047
telcm	0.048	-0.027	0.093	0.002	0.096	0.022
toys	**0.084**	-0.052	0.020	-0.027	0.033	-0.096
trans	**0.084**	-0.060	-0.020	0.022	-0.029	-0.049
txtls	**0.158**	**-0.144**	0.044	0.048	-0.097	-0.115
util	0.060	**-0.087**	0.017	0.022	0.133	-0.032
whlsl	**0.143**	-0.007	-0.006	-0.041	-0.047	-0.055

Table 4. Granger causality tests

Bivariate tests with two lags and 523monthly returns over the period from January 1970 to July 2013. Bold indicates statistical significance.

Null hypothesis:	Banking industry do not Granger cause other industry		Other industry do not Granger cause Banking industry	
Other industry	F-Statistic	P-value	F-Statistic	P-value
aero	**2.853**	**0.059**	0.029	0.971
autos	**4.174**	**0.016**	0.095	0.910
beer	1.950	0.143	1.301	0.273
bldmt	**6.581**	**0.002**	0.462	0.630
books	1.346	0.261	2.190	0.113
boxes	2.003	0.136	0.284	0.753
bussv	1.536	0.216	0.718	0.488
chems	1.763	0.173	0.173	0.841
chips	**5.011**	**0.007**	0.074	0.929
clths	0.151	0.860	1.325	0.267
cnstr	**3.692**	**0.026**	0.829	0.437
coal	1.144	0.319	**4.055**	**0.018**
comps	**4.287**	**0.014**	0.252	0.777
drugs	0.916	0.401	1.724	0.179
elceq	**4.093**	**0.017**	0.084	0.920
fabpr	**11.063**	**0.000**	0.320	0.726
fin	0.837	0.434	0.981	0.376
food	0.634	0.531	0.560	0.572
fun	**4.388**	**0.013**	0.794	0.452
gold	0.720	0.487	2.117	0.121
guns	**4.730**	**0.009**	1.090	0.337
htlh	**5.502**	**0.004**	0.077	0.926
hshld	**4.169**	**0.016**	0.220	0.803
insur	0.770	0.464	1.569	0.209
labeq	**3.093**	**0.046**	0.339	0.713
mach	**3.898**	**0.021**	0.114	0.893
meals	2.107	0.123	0.357	0.700
medeq	**2.705**	**0.068**	1.884	0.153
mines	**3.743**	**0.024**	2.076	0.127
oil	1.390	0.250	**2.791**	**0.062**
other	0.623	0.537	1.488	0.227
paper	**6.845**	**0.001**	0.337	0.714
persv	0.977	0.377	0.190	0.827
rlest	**8.943**	**0.000**	0.532	0.588
rtail	0.088	0.915	1.042	0.353
rubbr	**11.678**	**0.000**	0.084	0.920
ships	**3.639**	**0.027**	0.958	0.384
smoke	0.585	0.557	0.872	0.419
soda	1.480	0.229	0.115	0.891
steel	**7.352**	**0.001**	1.368	0.256
telcm	**2.574**	**0.077**	0.050	0.951
toys	**5.292**	**0.005**	0.616	0.540
trans	**2.364**	**0.095**	0.004	0.996
txtls	**5.860**	**0.003**	1.369	0.255
util	0.107	0.899	0.303	0.739
whlsl	2.079	0.126	0.939	0.392

In order to further investigate the impact of the cross-autocorrelation between the banking industry and the other industries, we run an OLS regression of the benchmark model (2). The parameter estimates of the model are presented in Table 5. The results indicate that for many of the industries the observed autocorrelation structure in Table 3 is in fact driven by the cross-autocorrelation with the banking industry. When a one-month lag of the banking industry portfolio return is included in the model, the significance of the autocorrelation is lost for 19 of the 24 industries with significant first order autocorrelation. Only for the printing and publishing industry (books), the apparel industry (clts), entertainment (fun), healthcare industry (hlth), the restaurants, hotels, and motels industry (meals), the business-supply industry (paper), and the retail industry (rtail), the significance of the own autocorrelation remains. The impact of the cross-autocorrelation with the banking industry is positive and varies between 0.084 (for the beer and liquor (beer) and the medical equipment industries (medeq)) and 0.290 (for the real estate industry (rlest)).

Table 6 presents the quantile regression estimates of the autocorrelation coefficient of model (3). For none of the industries is the autocorrelation coefficient statistically significant across the entire conditional distribution (i.e. for all values of the quantile parameter τ). For the majority of the industry portfolios, the significant autocorrelation appears either in the lower or upper part of the distribution. In all significant cases, except for the business supply industry (paper), the autocorrelation is positive in the lower part of the distribution. On the other hand, for many industry portfolios, the autocorrelation is negative in the upper part of the conditional distribution. These empirical results are in line with those of Baur et al. (2012). These results also indicate that on average, using OLS regression, the conditional autocorrelation structure would in many cases not be statistically significant as was shown in Table 5.

The corresponding quantile regression estimates for the cross-autocorrelation coefficient with the banking industry is presented in Table 7. The results are here quite different from the results presented in Table 6 for the estimated autocorrelation coefficients. First, the cross-autocorrelation with the returns on the banking industry portfolio is positive in all significant cases, except for precious metals (gold) and restaurants, hotels and motels (meals) in the extreme upper part of the distribution with τ=0.90. Second, the significance and the values of the cross-autocorrelation with a one-month lag of the banking industry are much higher in the lower part of the conditional distribution (lower values of τ) than in the upper part of the distribution (higher values of τ). The cross-autocorrelation is as high as 0.415 for the real estate industry portfolio (rlest) and 0.400 for both the construction materials (bldmt) and the rubber and plastic products (rubbr) industries.

Table 5. Benchmark OLS regression estimates

$r_t^{ind} = \beta_0 + \beta_1 r_{t-1}^{ind} + \beta_2 r_{t-1}^{banks} + \varepsilon_t^{ind}$ based on 523 monthly returns over the period from January 1970 to July 2013. Statistical significance is indicated in boldface.

	β_0	p-value	β_1	p-value	β_2	p-value	Adj R-sq
aero	**1.067**	**0.000**	0.058	0.293	**0.134**	**0.028**	0.024
agric	**0.968**	**0.001**	-0.029	0.560	**0.104**	**0.049**	0.004
autos	**0.721**	**0.022**	0.020	0.719	**0.174**	**0.008**	0.024
beer	**1.085**	**0.000**	-0.029	0.589	**0.084**	**0.071**	0.003
bldmt	**0.895**	**0.001**	-0.076	0.234	**0.232**	**0.000**	0.028
books	**0.747**	**0.004**	**0.153**	**0.013**	0.054	0.360	0.035
boxes	**0.980**	**0.000**	0.000	0.994	0.055	0.297	0.000
bussv	**0.971**	**0.001**	0.034	0.541	0.073	0.228	0.005
chems	**1.036**	**0.000**	-0.022	0.711	0.071	0.203	0.000
chips	**0.892**	**0.010**	0.004	0.943	0.109	0.103	0.004
clths	**0.892**	**0.003**	**0.154**	**0.009**	0.032	0.625	0.027
cnstr	**0.757**	**0.021**	0.031	0.588	**0.182**	**0.008**	0.027
coal	**1.176**	**0.013**	0.036	0.444	0.023	0.769	-0.002
comps	**0.785**	**0.017**	0.026	0.602	0.092	0.121	0.005
drugs	**1.118**	**0.000**	-0.054	0.315	0.038	0.397	-0.002
elceq	**1.162**	**0.000**	-0.080	0.176	**0.157**	**0.011**	0.009
fabpr	**0.626**	**0.045**	-0.007	0.890	**0.240**	**0.000**	0.038
fin	**0.972**	**0.001**	0.115	0.104	0.047	0.521	0.020
food	**1.089**	**0.000**	0.029	0.610	0.039	0.350	0.002
fun	**1.064**	**0.003**	**0.104**	**0.070**	**0.130**	**0.078**	0.031
gold	**1.027**	**0.032**	-0.034	0.447	-0.051	0.498	-0.002
guns	**1.182**	**0.000**	-0.024	0.630	**0.146**	**0.007**	0.012
hlth	**0.735**	**0.047**	**0.101**	**0.048**	**0.161**	**0.020**	0.033
hshld	**0.802**	**0.000**	-0.030	0.607	**0.124**	**0.006**	0.016
insur	**0.944**	**0.000**	0.099	0.184	0.020	0.769	0.010
labeq	**0.841**	**0.011**	0.046	0.377	**0.114**	**0.066**	0.012
mach	**0.886**	**0.002**	0.024	0.681	**0.107**	**0.073**	0.010
meals	**0.913**	**0.001**	**0.100**	**0.084**	0.070	0.229	0.020
medeq	**0.880**	**0.000**	0.010	0.859	**0.084**	**0.083**	0.007
mines	**0.885**	**0.008**	-0.003	0.951	**0.126**	**0.039**	0.007
oil	**1.185**	**0.000**	-0.062	0.209	0.067	0.128	0.001
other	**0.406**	**0.184**	0.080	0.160	0.072	0.256	0.013
paper	**0.962**	**0.000**	**-0.120**	**0.043**	**0.183**	**0.001**	0.018
persv	**0.567**	**0.062**	0.062	0.264	0.083	0.180	0.012
rlest	**0.252**	**0.450**	0.057	0.298	**0.290**	**0.000**	0.071
rtail	**0.938**	**0.000**	**0.149**	**0.015**	-0.006	0.919	0.017
rubbr	**0.882**	**0.001**	-0.066	0.249	**0.253**	**0.000**	0.044
ships	**0.949**	**0.004**	-0.015	0.782	**0.150**	**0.021**	0.010
smoke	**1.414**	**0.000**	0.029	0.554	0.048	0.332	0.000
soda	**1.096**	**0.000**	0.029	0.586	0.031	0.588	-0.001
steel	**0.627**	**0.066**	-0.057	0.269	**0.205**	**0.001**	0.017
telcm	**0.925**	**0.000**	0.010	0.860	0.050	0.237	0.001
toys	**0.664**	**0.037**	-0.018	0.738	**0.197**	**0.002**	0.022
trans	**0.909**	**0.001**	-0.002	0.971	**0.118**	**0.046**	0.011
txtls	**0.786**	**0.017**	0.041	0.460	**0.221**	**0.001**	0.041
util	**0.885**	**0.000**	0.054	0.298	0.009	0.800	0.000
whlsl	**0.840**	**0.001**	0.076	0.199	0.089	0.097	0.022

Table 6. Quantile regression autocorrelation coefficient $\beta_1(\tau)$ estimates $r_t^{ind} = \beta_0(\tau) + \beta_1(\tau)r_{t-1}^{ind} + \beta_2(\tau)r_{t-1}^{banks} + \varepsilon_t^{ind}$. Based on 523 monthly returns over the period from January 1970 to July 2013. Significant autocorrelations are indicated in boldface for significance levels of 10% or lower.

τ	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
aero	0.136	0.066	0.059	0.037	-0.017	-0.010	-0.012	-0.010	0.121
agric	0.071	0.065	0.021	0.006	0.035	-0.041	**-0.126**	**-0.131**	**-0.206**
autos	0.091	-0.022	0.001	0.032	0.046	0.009	-0.059	-0.124	-0.242
beer	0.093	-0.032	-0.081	-0.080	-0.065	-0.015	-0.023	-0.006	**-0.150**
bldmt	-0.056	0.001	0.014	-0.048	-0.028	-0.083	-0.082	-0.118	**-0.290**
books	**0.160**	**0.237**	**0.164**	**0.143**	**0.138**	0.126	0.107	0.115	0.095
boxes	0.113	0.127	0.055	0.049	0.083	-0.001	-0.042	-0.130	-0.148
bussv	0.083	0.137	**0.145**	0.094	-0.019	-0.032	-0.020	-0.072	-0.145
chems	0.129	0.040	0.003	-0.019	-0.041	**-0.162**	**-0.188**	**-0.225**	-0.193
chips	0.003	**0.114**	0.057	-0.013	-0.017	0.018	-0.028	-0.061	-0.046
clths	**0.296**	**0.282**	**0.197**	**0.132**	0.087	0.028	0.019	-0.036	0.030
cnstr	0.066	0.040	0.001	-0.004	-0.004	-0.003	0.054	0.010	0.177
coal	0.049	-0.008	0.012	-0.063	-0.086	-0.087	-0.010	-0.004	**0.165**
comps	0.133	**0.120**	0.066	-0.044	-0.021	-0.056	-0.009	-0.023	-0.050
drugs	**0.151**	0.021	0.108	0.051	-0.012	-0.097	**-0.140**	**-0.200**	**-0.229**
elceq	-0.035	-0.056	-0.001	0.006	-0.032	-0.087	**-0.219**	**-0.217**	**-0.242**
fabpr	0.011	0.076	0.049	0.091	0.056	-0.019	-0.049	-0.097	-0.092
fin	0.122	**0.242**	**0.216**	0.191	0.089	0.071	0.038	-0.010	-0.049
food	**0.194**	0.118	0.070	0.056	0.065	0.026	-0.027	-0.089	-0.232
fun	0.125	0.038	0.085	0.083	0.115	0.100	0.100	0.082	0.043
gold	0.002	0.018	0.000	-0.058	-0.052	-0.049	-0.025	0.034	-0.014
guns	0.101	0.061	-0.040	-0.007	-0.058	-0.092	-0.101	-0.101	-0.091
hlth	**0.232**	**0.188**	**0.120**	**0.126**	**0.099**	0.070	0.091	0.038	0.024
hshld	**0.250**	0.122	0.033	0.024	-0.023	-0.020	-0.047	**-0.151**	**-0.177**
insur	0.117	0.096	0.090	**0.163**	0.077	0.027	0.005	-0.060	-0.059
labeq	**0.249**	**0.151**	**0.134**	**0.144**	0.044	-0.011	-0.086	**-0.156**	-0.116
mach	0.177	**0.148**	**0.114**	**0.107**	0.060	-0.009	-0.077	**-0.220**	**-0.243**
meals	**0.257**	0.120	0.123	0.113	0.062	0.067	0.027	0.065	**0.143**
medeq	0.130	**0.163**	0.092	0.032	0.017	-0.075	-0.099	**-0.130**	**-0.186**
mines	0.092	0.040	-0.027	-0.022	-0.003	0.023	0.011	-0.033	-0.134
oil	-0.054	-0.040	-0.045	-0.032	-0.045	**-0.126**	**-0.167**	-0.116	-0.123
other	0.113	**0.180**	**0.136**	**0.105**	**0.124**	0.089	0.089	-0.024	**-0.127**
paper	-0.057	**-0.114**	**-0.133**	**-0.111**	**-0.123**	-0.067	-0.101	-0.125	**-0.196**
persv	**0.146**	**0.155**	**0.132**	0.086	0.067	0.035	-0.080	**-0.128**	**-0.234**
rlest	0.094	0.099	0.109	0.124	**0.166**	0.097	0.103	-0.030	-0.104
rtail	0.183	**0.158**	**0.145**	**0.138**	**0.129**	0.048	0.085	0.016	0.178
rubbr	0.022	-0.049	-0.035	-0.014	-0.020	-0.098	**-0.181**	**-0.156**	-0.127
ships	0.107	0.109	0.124	0.087	0.053	0.002	-0.094	-0.138	-0.194
smoke	0.030	-0.005	0.071	-0.015	0.042	0.013	0.034	0.055	0.021
soda	-0.021	-0.070	-0.075	0.018	0.059	0.047	0.061	0.022	0.017
steel	0.106	0.029	-0.057	-0.130	**-0.133**	**-0.161**	**-0.132**	**-0.213**	-0.170
telcm	-0.110	-0.045	-0.036	0.039	0.108	**0.121**	0.071	0.094	0.041
toys	0.000	0.001	0.065	-0.022	0.004	0.006	0.009	-0.040	**-0.127**
trans	0.160	0.045	0.101	0.018	0.048	-0.016	-0.043	**-0.148**	**-0.237**
txtls	**0.166**	0.091	**0.172**	0.076	0.045	-0.035	0.027	-0.073	-0.054
util	-0.043	0.055	0.023	0.052	0.040	0.084	**0.118**	0.122	0.132
whlsl	**0.155**	**0.173**	**0.156**	0.121	0.066	-0.030	-0.017	-0.002	-0.095

Table 7. Quantile cross-autocorrelation coefficients $\beta_2(\tau)$estimates $r_t^{ind} = \beta_0(\tau) + \beta_1(\tau) r_{t-1}^{ind} + \beta_2(\tau) r_{t-1}^{banks} + \varepsilon_t^{ind}$ based on 523 monthly returns over the period from January 1970 to July 2013. Significant cross-autocorrelations are indicated in boldface for significance levels of 10% or lower.

τ	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
aero	**0.229**	**0.213**	**0.192**	**0.145**	**0.163**	0.116	0.091	0.009	-0.092
agric	**0.136**	0.091	**0.138**	**0.142**	**0.118**	**0.129**	**0.119**	0.020	0.000
autos	**0.188**	0.198	0.165	0.167	0.185	0.152	**0.175**	**0.225**	0.178
beer	0.041	**0.123**	**0.148**	**0.111**	0.073	0.046	0.083	0.056	0.082
bldmt	**0.400**	**0.282**	**0.207**	**0.230**	**0.188**	**0.180**	0.127	0.084	**0.213**
books	**0.165**	0.072	0.086	0.079	0.069	0.023	0.026	-0.041	-0.021
boxes	0.098	0.032	0.074	0.047	0.004	0.074	0.072	**0.129**	0.085
bussv	**0.209**	0.148	0.100	0.155	0.169	**0.140**	0.091	0.069	0.009
chems	0.083	**0.155**	0.093	0.071	0.068	**0.154**	**0.171**	**0.163**	-0.041
chips	**0.234**	0.116	0.128	**0.185**	0.139	0.065	0.028	0.079	0.074
clths	0.064	0.070	0.096	0.117	0.127	0.071	0.054	0.107	0.017
cnstr	**0.323**	**0.273**	**0.254**	**0.228**	**0.220**	**0.210**	0.093	0.104	-0.052
coal	**0.353**	0.155	0.045	0.101	0.063	0.020	-0.057	-0.097	-0.169
comps	0.004	0.060	0.061	**0.144**	0.113	0.147	0.029	0.034	0.085
drugs	-0.003	0.067	0.043	0.064	0.076	0.068	0.023	0.030	0.028
elceq	**0.334**	**0.279**	0.117	0.046	0.076	**0.149**	**0.161**	**0.188**	0.132
fabpr	**0.396**	**0.302**	**0.241**	**0.151**	**0.140**	**0.170**	**0.180**	**0.180**	**0.186**
fin	**0.232**	0.036	0.061	0.048	0.078	0.069	0.050	0.022	0.059
food	0.046	0.066	0.054	0.071	0.033	0.035	0.038	0.066	0.063
fun	**0.238**	0.150	0.065	0.040	0.103	0.086	0.030	-0.075	-0.021
gold	0.139	0.024	0.148	0.029	0.075	0.052	0.051	0.022	**-0.210**
guns	**0.218**	**0.170**	**0.190**	**0.155**	**0.145**	**0.136**	0.107	0.090	0.076
hlth	0.119	**0.269**	**0.252**	**0.179**	**0.159**	**0.174**	0.111	0.114	-0.057
hshld	0.118	0.091	0.088	0.066	0.050	**0.096**	0.072	**0.128**	0.110
insur	**0.225**	0.151	0.128	0.042	0.048	0.029	0.021	0.023	-0.039
labeq	0.087	0.131	**0.144**	0.096	0.106	0.088	0.101	0.118	0.075
mach	**0.207**	**0.189**	**0.156**	0.085	0.098	0.111	**0.122**	**0.164**	0.083
meals	0.179	**0.137**	**0.178**	**0.139**	0.090	0.053	0.063	-0.027	**-0.128**
medeq	0.054	0.014	0.047	0.122	0.100	**0.125**	**0.127**	**0.091**	0.091
mines	**0.220**	**0.160**	**0.130**	0.132	**0.160**	0.101	0.077	0.033	0.012
oil	**0.234**	**0.141**	0.080	0.074	0.079	0.071	0.050	-0.022	-0.113
other	**0.241**	0.088	0.068	0.085	0.051	0.033	-0.031	0.045	0.002
paper	**0.331**	**0.322**	**0.274**	**0.199**	**0.165**	**0.123**	0.103	0.050	0.055
persv	**0.205**	**0.198**	**0.137**	0.063	0.101	0.093	0.105	0.158	0.096
rlest	**0.415**	**0.353**	**0.258**	**0.201**	**0.186**	**0.172**	**0.137**	**0.274**	**0.283**
rtail	0.057	0.066	0.040	-0.017	-0.025	0.030	-0.010	-0.015	-0.103
rubbr	**0.400**	**0.329**	**0.262**	**0.219**	**0.182**	**0.188**	**0.235**	**0.216**	0.147
ships	**0.259**	**0.262**	0.117	0.115	**0.141**	0.099	0.095	0.090	-0.039
smoke	0.101	0.048	0.010	0.055	0.016	0.019	-0.012	-0.013	0.074
soda	0.159	**0.156**	0.067	0.028	0.029	0.024	-0.007	0.003	-0.110
steel	**0.298**	**0.282**	**0.259**	**0.233**	**0.203**	**0.207**	0.113	**0.142**	0.049
telcm	**0.220**	**0.192**	**0.151**	0.061	0.012	-0.015	-0.016	-0.055	-0.093
toys	**0.301**	**0.303**	**0.183**	**0.200**	**0.154**	0.132	0.091	0.072	0.036
trans	**0.137**	**0.266**	0.110	0.132	0.054	0.056	0.013	0.085	0.127
txtls	**0.236**	**0.257**	**0.191**	**0.167**	**0.166**	**0.210**	0.105	0.148	0.099
util	**0.188**	0.065	0.043	0.019	0.005	-0.035	-0.070	-0.095	-0.095
whlsl	**0.227**	**0.167**	**0.124**	**0.156**	**0.146**	0.091	0.068	0.004	0.081

The average adjusted R-squares of the quantile regression model (3) are presented in Figure 1 over different values of the quantile parameter τ. The adjusted R-squares are on average about twice as high (0.04) in the extreme lower part of the distribution, with τ =0.10, than in the middle and upper part of the distribution (below 0.02) with τ=0.40 to 0.80.

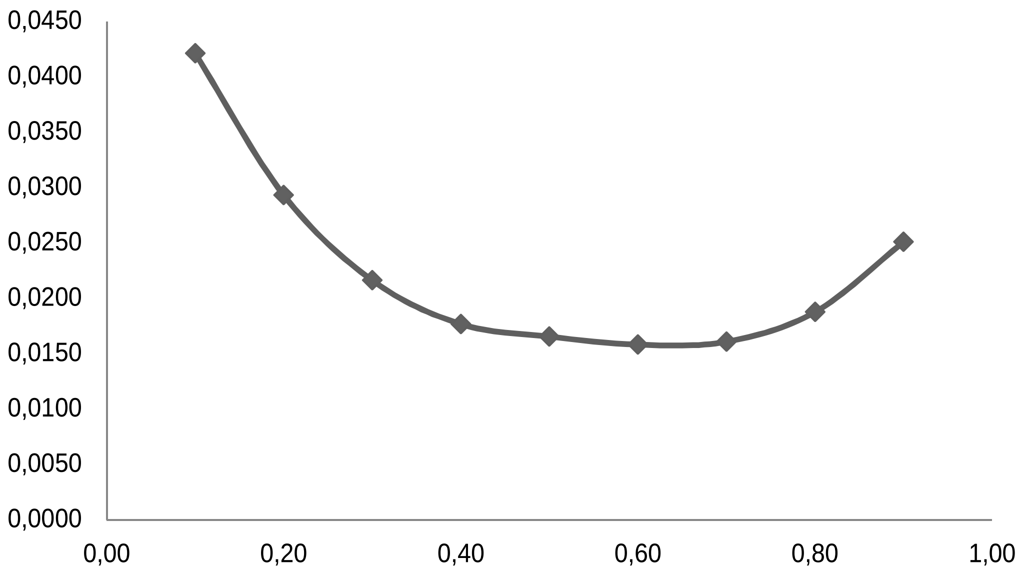


Figure 1. Average adjusted R-squares over the 47 industry portfolios from the quantile regressions $\boldsymbol{r_t^{ind}} = \boldsymbol{\beta_0(\tau)} + \boldsymbol{\beta_1}(\boldsymbol{\tau})\boldsymbol{r_{t-1}^{ind}} + \boldsymbol{\beta_2}(\boldsymbol{\tau})\boldsymbol{r_{t-1}^{banks}} + \boldsymbol{\varepsilon_t^{ind}}$ Based on 523 monthly returns over the period from January 1970 to July 2013.

The results shown in Table 7 and Figure 1 indicate that industries are especially dependent on the performance of the banking industry when the conditional returns on the industry portfolios are low. This result may be seen as an indication of an increased dependence on liquidity provided by banks in situations of poorer than expected realized returns on industry portfolios.

As a final empirical step the dynamic linkages between returns on the banking industry portfolio and four standard asset-pricing risk factors are monitored for the time series sample data at hand. First, a Granger causality test is conducted and the results are presented in Table 8.

Table 8. Granger causality tests

Bivariate tests using 523 monthly returns over the period from January 1970 to July 2013 and two lags used for the tests. Significant statistics are indicated in boldface for significance levels of 10% or lower.

Null hypothesis:	Banking industry do not Granger cause risk factors		Risk factor do not Granger cause Banking industry	
Risk factor	F-Statistic	P-value	F-Statistic	P-value
market	2.091	0.125	0.163	0.849
size	**16.698**	**0.000**	1.115	0.329
value	1.305	0.272	0.278	0.758
momentum	**7.110**	**0.001**	0.709	0.493

From the results of Table 8 it is evident that the returns on the banking industry portfolio are dynamically connected to the size and the momentum factors. The banking industry portfolio leads and Granger causes size and momentum but none of the four risk factors seem to lead or Granger cause the returns on the banking industry portfolio. More detailed empirical evidence on this leading relation of the banking industry portfolio over the four risk factors is presented in Table 9. This table presents the results of the quantile regression of the model:

$$r_t^{factor} = \beta_0(\tau) + \beta_1(\tau) r_{t-1}^{factor} + \beta_2(\tau) r_{t-1}^{banks} + \varepsilon_t^{ind}.$$

From Table 9 it is clear that especially the value factor is significantly positively auto-correlated across the entire conditional return distribution. The influence of the banking portfolio is only evident in the extreme lower part of the conditional distribution for the value factor. For the size factor, on the other hand, the dynamic influence of the returns on the banking portfolio is stable and statistically significant across the entire conditional return distribution of the long-short size portfolio. The momentum factor has no significant autocorrelation but is especially in the extreme tails driven by the cross-autocorrelation with the returns on the banking industry portfolio. Finally, although the returns on several of the 47 individual industry portfolios were affected by the cross-autocorrelation with the banking portfolio, on a general value-weighted market level this relation is not fully clear. The autocorrelation as well as the cross-autocorrelation with the banking portfolio is high and significant in the extreme lower part of the distribution. However, none of these correlations stays statistically significant across the entire conditional distribution of the market risk factor.

Table 9. Quantile regression cross- and autocorrelation coefficient estimates of the model $r_t^{factor} = \beta_0(\tau) + \beta_1(\tau) r_{t-1}^{factor} + \beta_2(\tau) r_{t-1}^{banks} + \varepsilon_t^{ind}$. Based on 523 monthly returns over the period January 1970 to July 2013 and two lags used for the tests. Significant correlations are indicated in boldface for significance levels of 10% or lower.

τ	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
Autocorrelation coefficient									
market	**0.15**	0.07	0.08	0.02	-0.06	**-0.18**	-0.16	-0.16	-0.17
size	0.07	**0.17**	**0.13**	**0.12**	0.07	0.05	-0.00	-0.04	-0.05
value	**0.14**	**0.20**	**0.20**	**0.18**	**0.17**	**0.20**	**0.23**	**0.23**	**0.20**
momentum	-0.05	0.01	0.03	0.06	0.08	0.09	0.09	0.06	0.03
Cross- autocorrelation coefficient									
market	**0.17**	**0.18**	0.09	**0.15**	0.12	**0.18**	**0.13**	0.10	0.06
size	**0.10**	**0.12**	**0.13**	**0.12**	**0.11**	**0.09**	**0.14**	**0.11**	**0.11**
value	**0.06**	**0.05**	0.05	0.02	0.02	0.04	0.03	0.00	0.02
momentum	**-0.11**	-0.01	0.01	-0.00	-0.05	**-0.06**	**-0.07**	-0.04	**-0.11**

5 Summary and conclusion

The purpose of this empirical study was to revisit the issue of dynamic linkages between equity returns on different industry portfolios and contribute to the discussion by specifically focusing on the role of the banking industry portfolio. Using monthly returns on 48 US industry portfolios, the results show that the dynamic linkages between the returns on the banking portfolio and other industries often are asymmetric in two ways. First, there appears to be a one-directional causality relation running from the banking industry to several other industries but seldom the other way around. Lagged returns on the banking industry portfolio seem to be able to predict returns on several other industry portfolios. Surprisingly, for many industry portfolio returns the banking industry portfolio can be regarded as exogenous and Granger cause several other industry portfolio returns. Second, the cross-autocorrelation is found to be asymmetric such that on average the impact of a one-month lag of the return of the banking portfolio is about twice as high in the lower part of the conditional return distribution than in the upper part. It is also found that returns on the banking industry portfolio seem to be dynamically connected to two of the four classic asset-pricing risk factors. Returns on the banking industry portfolio Granger cause and lead the returns on the size and the momentum portfolios. The relation to the market and value risk factor is not fully clear.

References

Altay, E. (2004). Cross-autocorrelation between small and large cap portfolios in the German and Turkish stock markets. *Journal of Financial Management and Analysis* 17, 77–92.

Anari, A., Kolari, J. & Mason, J. (2004). Further evidence of nonmonetary effects on the U.S. economy during the Great Depression. *Journal of Money, Credit and Banking* 28, 2331–2351.

Anderson, R., Eom, K., Hahn, S. & Park, J.-H. (2013). Autocorrelation and partial price adjustment. *Journal of Empirical Finance* 24, 78–93.

Ayers, B. & Freeman, R. (2000). Why do large firms' prices anticipate earnings earlier than small firms' prices? *Contemporary Accounting Research* 17, 191–212.

Baur, D., Dimpfl, T. & Jung, R. (2012). Stock return autocorrelation revisited: A quantile regression approach. *Journal of Empirical Finance* 19, 254–265.

Bernanke, B. (1993). Credit in the macroeconomy. *Federal Reserve Bank of New York Quarterly Review*. 257–276.

Bernanke, B. & Gertler, M. (1995). Inside the black box: The credit channel of monetary policy transmission. *NBER Working Papers* 5146, National Bureau of Economic Research, Inc.

Bernhardt, D. & Mahani, R. (2007). Asymmetric information and stock return cross-autocorrelations. *Economic Letters* 96, 14–22.

Boudoukh, J., Richardson, M. & Whitelaw, R. (1994). A tale of three schools: Insights on autocorrelations of short-horizon stock returns. *The Review of Financial Studies* 7, 539–573.

Cen, L., Chen, K., Dasgupta, S. & Gao, N. (2013). When the tail wags the dog: Industry leaders, limited attention and spurious cross-industry information diffusion. Working Paper available at http://ssrn.com/abstract=964106.

Chan, K. (1993). Imperfect information and cross-autocorrelation among stock prices. *Journal of Finance* XLVIII, 1211–1230.

Chou, P.-H., Ho, P.-H. & Ko, K.-C. (2013). Do industries matter in explaining stock returns and asset-pricing anomalies? *Journal of Banking & Finance* 36, 355–370.

Doong, S.-H., Yang, S.-Y. & Chiang, T. (2005). Response asymmetries in Asian stock markets. *Review of Pacific Basin Financial Markets and Policies 8*, 637–657.

Hameed, A. (1997). Time-varying factors and cross-autocorrelations in short-horizon stock returns. *Journal of Financial Research* 20, 435–458.

Hammami Y. & Lindahl, A. (2014). An intertemporal capital asset pricing model with bank credit growth as a state variable. *Journal of Banking and Finance* 39, 14–28.

Hong, H., Torous, W. & Valkanov, R. (2007). Do industries lead stock markets? *Journal of Financial Economics* 83, 367–396.

Hou, K. (2007). Industry information diffusion and the lead-lag effect in stock returns. *Review of Financial Studies* 20, 1113–1138.

Högholm, K., Knif, J. & Pynnönen, S. (2011a). Cross-distributional robustness of weekday effects: Evidence from European equity index returns. *European Journal of Finance* 17, 377–390.

Högholm, K., Knif, J. & Pynnönen, S. (2011b). Fund performance robustness: An evaluation using European large-cap equity funds. *Frontiers in Finance and Economics* 8, 1–26.

Högholm, K., Knif, J., Koutmos, G. & Pynnönen, S. (2011). Distributional asymmetry of loadings on market co-moments. *Journal of International Financial Markets, Institutions & Money* 21, 851–866.

Kinnunen, J. (2013). Dynamic cross-autocorrelation in stock returns. In *Risk-return trade-off and autocorrelation*. PhD dissertation. Acta Universitatis Lappeenrantaensis 511. Lappeenranta University of Technology.

Koenker, R. (2005). *Quantile Regression*. New York: Cambridge University Press.

Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica* 4, 41–55.

Laopodis, N. (2013). Dynamic linkages among industries and the stock market. Working Paper presented at the Eastern Finance Association annual meeting in St. Pete Beach, FL April 2013.

Lee, C.-C., Chen, M.-P. & Chang, C.-H. (2013). Dynamic relationships between industry returns and stock market returns. *North American Journal of Economics and Finance* 26, 119–144.

Lo, A. & MacKinley, C. (1990a). An econometric analysis of nonsynchronous trading. *Journal of Econometrics* 45, 181–211.

Lo, A. & MacKinley, C. (1990b). When are contrarian profits due to stock market overreaction? *Review of Financial Studies* 3, 175–205.

McQueen, G., Pinegar, M. & Thorley, S. (1996). Delayed reaction to good news and the cross-autocorrelation of portfolio returns. *Journal of Finance* LI, 889–919.

Merton, R. (1987). A simple model of capital market equilibrium with incomplete information. *Journal of Finance* XLII, 483–510.

Zebedee A. & Kasch-Haroutounian, M. (2009). A closer look at co-movements among stock returns. *Journal of Economics and Business* 61, 279–294.

Appendix A

List of the included industry portfolios with abbreviations used in the text

Aircraft (aero)
Agriculture (agric)
Automobiles and Trucks (autos)
Banking (banks)
Beer & Liquor (beer)
Construction Materials (bldmt)
Printing and Publishing (books)
Shipping Containers (boxes)
Business Services (bussv)
Chemicals (chems)
Electronic Equipment (chips)
Apparel (clths)
Construction (cnstr)
Coal (coal)
Computers (comps)
Pharmaceutical Products (drugs)
Electrical Equipment (elceq)
Fabricated Products (fabpr)
Financial Trading (fin)
Food Products (food)
Entertainment (fun)
Precious Metals (gold)
Defense (guns)
Healthcare (hlth)
Consumer Goods (hshld)
Insurance (insur)
Measuring and Control Equipment (labeq)
Machinery (Mach)
Restaurants, Hotels, Motels (meals)
Medical Equipment (medeq)
Non-Metallic and Industrial Metal Mining (mines)
Petroleum and Natural Gas (oil)
Almost Nothing (other)
Business Supplies (paper)
Personal Services (persv)
Real Estate (rlest)
Retail (rtail)
Rubber and Plastic Products (rubbr)
Shipbuilding, Railroad Equipment (ships)
Tobacco Products (smoke)
Candy & Soda (soda)
Steel Works Etc (steel)
Communication (telcm)
Recreation (toys)
Transportation (trans)
Textiles (txtls)
Utilities (util)
Wholesale (whlsl)

Source: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/changes_ind.html

FISHER, WICKSELL, TAYLOR AND THE NEGATIVE INTEREST RATE ELASTICITY OF INFLATION RATES

Ali Anari and James Kolari
Texas A&M University

1 Introduction

This paper shows that real and nominal interest rates have a negative impact on inflation rates and that this negative impact is foundational to the Fisher puzzle and Taylor principle. A model of joint determination of interest rates and inflation rates is proposed wherein an important part of the positive relationship between nominal interest rates and inflation rates due to the Fisher effect is offset by the negative Wicksell price effect of real or nominal interest rates on inflation. Hence, researchers using ex-post interest rate and inflation rate data for estimating the ex-ante Fisher effect have not been able to obtain the full Fisher effect, known as the *Fisher puzzle*. We also show that the greater than unity inflation coefficient in the Taylor rule, known as *Taylor principle*, is due to the negative Wicksell price effect. Empirical evidence is provided using US interest and inflation rate series for the period 1960 to 2012.

Irving Fisher's (1930) theory of interest rates states that there is a positive one-to-one relationship between nominal interest rates and expected inflation rates and that causality runs from inflation rates to interest rates (e.g., see Peek 1982, Carmichael & Stebbing 1983, Rose 1988, Mishkin 1992, 1995, Pelaez 1995, Peng 1995, Phillips 1998, Sun & Phillips 2004, Yoon 2010, and others). By contrast, a closely-related interest rate theory by Wicksell (1898) posits a negative association between real interest rates and inflation rates with causality running from interest rates to inflation rates. The Wicksell price effect plays an important role in modern monetary policy. The Federal Reserve has utilized the federal funds rate to help control inflation in the U.S since 1981.

The dual existence of both Fisher and Wicksell processes implies that interest and inflation rates are jointly determined by the two effects. Beginning with the Wicksell relation, we show that the negative relationship between real interest rates and inflation rates can be derived by solving the Phillips curve and investment-saving (IS) curve. The resultant equation of the trade-off between interest rates and inflation rates together with the Fisher equation form a model of joint inflation and interest rate determination. Solving these equations results in reduced equations that provide theoretical insights into the Fisher puzzle and Taylor principle. The *Fisher puzzle* refers to less than unity estimates of the inflation coefficient in empirical studies of the Fisher equation. We hypothesize that one possible explanation for the Fisher puzzle is that an important part of the positive relationship between interest rates and inflation rates due to the Fisher effect is offset by the negative Wicksell price effect. In other words, inflation rates have been reduced by the negative impact of real interest rates on prices. We further show that the Wicksell effect, or negative impact of interest rates on inflation rates, is foundational to the *Taylor principle* also. According to this principle, monetary policy makers should raise nominal interest rates more than one percent to decrease the inflation rate by one percent. This greater than unity negative effect of interest rates on inflation rates can be attributed to the partially offsetting Fisher effect.

Given that interest and inflation rates are jointly determined by full processes of Fisher and Wicksell effects, we proceed to conduct empirical tests of these effects. To circumvent a previously unaddressed empirical problem of the appropriate interest rate maturity in Fisher-Wickell relations, we construct a macro-interest rate series that takes into account aggregate debt outstanding and aggregate interest paid by the household, business, and the government sectors. Employing different inflation rate measures in combination with macro-interest rates, short-run Fisher (Wicksell) coefficients estimates range from 0.089 to 0.116 (-0.086 to -0.103). Depending on the measure of inflation rates, long-run Fisher coefficient estimates are between 1.009 and 1.344 implying tax rates of 1 to 25 percent, and long-run Wicksell coefficient estimates are in the range -1.268 to -1.443. These long-run estimates strongly support *both* Fisher and Wicksell hypotheses. Evidence from impulse response functions for macro-interest and inflation rates over a ten-year forecast horizon further support positive (negative) long-run Fisher (Wicksell) effects. The time paths of the impulse responses show that the initial response of interest rates to inflation rates, or Fisher effect, is immediate and strong in the first year of the forecast horizon, such that it dominates the Wicksell price effect in the short-run. Thereafter in the second year of the forecast horizon the Wicksell effect becomes more salient, a finding consistent with those of other researchers (e.g., see Friedman 1961). Lastly, variance decomposition analyses reveal that interest (inflation) rate shocks can explain large proportions of the forecast error variance of inflation (interest) rates over multi-year horizons. An important implication of the dual Fisher-Wicksell processes is that, as proposed under the Taylor rule, a greater than one percent increase in interest rates is required to decrease inflation rates by one percent, which can be attributed to the partially offsetting Fisher effect. We conclude from empirical evidence that the Wicksell effect helps to explain the Fisher puzzle and Taylor principle.

The next section provides background discussion. Section 3 derives a model of the negative impact of interest rates on inflation rates, which is shown to be foundational to the Fisher puzzle and the Taylor principle. Section 4 presents empirical results of the model of joint determination of inflation and interest rates based on U.S. data series, including the estimation of short- and long-run Fisher and Wicksell coefficients and other related findings. Section 5 concludes.

2 Background Discussion

According to Fisher (1930), the nominal interest rate is the sum of real interest rate and the expected inflation rate, which implies that the estimated coefficient relating expected inflation rates to interest rates is unity. Empirically testing his theory using annual data for the U.S., Fisher obtained the highest correlation coefficient of 0.86 between long-term interest rates and inflation rates when inflation was lagged over 20 years in a distributed lag model. Using interest rates and inflation rates data for the U.K., he obtained a correlation coefficient of 0.98 when a lag order of 28 years was used. Fisher concluded that it takes years for the effects of inflation to be incorporated in interest rates.

Unfortunately, two well-known empirical problems arise in tests of the Fisher relation. One problem is that the estimated Fisher coefficients tend to be considerably less than unity.[1] Known as the *Fisher puzzle*, researchers have attempted to apply almost all newly-developed econometric methods, including ordinary least squares (OLS), rational expectations, and more recently cointegration analysis, in efforts to resolve this problem. Using cointegration methods, Peng (1995) found a long-run Fisher effect between expected inflation rates and interest rates for the U.S, U.K., France, Germany, and France, but weaker Fisher effects for Germany and Japan attributable to anti-inflationary policies of monetary authorities in those countries. By contrast, Sun & Phillips (2004) reported little support

[1] For example, based on a variety of different least squares models, Tanzi (1980) estimated Fisher coefficients between 0.3 and 0.7 for 6- and 12-month Treasury bills, respectively. Mishkin (1992) estimated the Fisher coefficient to be 0.7 using 3-month ex post inflation and nominal interest rates, which is similar to Söderlind’s (1998) results for 1-year rates. Using Livingston survey data on price expectations, Gibson (1972) reported estimates as high as 0.9 for 6- and 12-month bills but, like Pyle (1972), other estimates were well below unity. See Rose (1986) and Cooray(2002) for a survey of Fisher coefficient estimates.

for the presence of a cointegrating relation among the Fisher variables and therefore the long-run Fisher hypothesis.

Upon reviewing the extensive literature on the Fisher effect, Cooray (2002) concluded that, despite the fact that empirical studies using U.S. data have generally found a positive relation between interest rates and inflation rates, the evidence did not support the hypothesized one-to-one relationship. Furthermore, evidence for other countries was mixed with respect to the Fisher effect but generally less convincing than U.S. findings.

A wide variety of theories have been proposed to explain the Fisher puzzle.[2] Crowder and Wohar (1997) summarized four possible explanations for low Fisher coefficient estimates: (1) Tobin's (1965, 1969) argument that an increase in expected inflation causes investors to shift from nominal to real assets, (2) Tanzi's (1980) fiscal illusion of investors who fail to take into account the tax burden associated with real interest rates, (3) Evans & Lewis' (1995) peso problem explanation wherein investors underestimate the probability of high inflation, and (4) various econometric problems of small samples, inappropriate estimators, data errors, and misspecified equations (see Lahiri 1976, Tanzi 1980, Rose 1988, Mishkin 1992, Crowder & Hoffman 1996, and Ng & Perron 1997). Other work by Levi & Makin (1979) takes into account changes in real rates associated with changes in anticipated inflation and obtains results in the period 1950–1970 that tend to support the Fisher hypothesis. However, results for the period 1947–1975 indicated that inflation had a less than unity impact on nominal interest rates (see also Levi & Makin 1978). Sun & Phillips (2004) posited that ex-post data can be viewed as noisy observations of the ex-ante variables, such that univariate long memory estimates based on ex-post data tend to underestimate the persistence of ex-ante variables. In line with this logic, Evans & Lewis (1995) argued that investors rationally anticipate infrequent shifts in inflation that induces small-sample biases in long-run Fisher estimates. Additionally, studies by Bierens (2000), Lanne (2006), and Christopoulos & León-Ledesma (2007) attributed low Fisher coefficients to nonlinear common trends between interest rates and inflation rates. Also, Barsky (1987) argued that changes in the stochastic process of inflation, as opposed to changes in the structural relation between nominal rates and expected inflation, can affect Fisher coefficients over time. Unfortunately, according to Cooray (2002), evidence presented with respect to different arguments seeking to explain the failure of the Fisher effect has not been consistent.

[2] See also Crowder & Wohar (1997) for further discussion and references.

Adding to the Fisher puzzle, Darby (1975) and Feldstein (1976) reformulated the Fisher equation and hypothesized an even larger Fisher coefficient greater than one due to the tax factor $1/(1-T)$, where T is a tax rate on interest rate income (see also Carrington & Crouch 1987). Studies by Peek (1982) and Crowder & Wohar (1997) have provided empirical support for the notion that bonds exposed to taxes have higher Fisher coefficients than those not exposed to taxes. For example, based on a variety of different regression methods and a sample period from January 1950 to December 1995, Crowder & Wohar obtained estimated Fisher coefficients ranging from 0.89 to 1.49 for Treasury bills exposed to taxes but a much lower range of 0.51 to 0.90 for municipal bonds. However, studies by Nielson (1981) and Gandolfi (1982) incorporated the impact of capital gain taxation in the Darby-Feldstein version of the Fisher equation and found a Fisher coefficient of more than one but not as high as hypothesized.

Another complicating issue is that empirical evidence does not support a short-run Fisher relation. Using inflation and interest rates data for Australia, Mishkin & Simon (1995) found evidence for the Fisher effect in the long run but not the short run. Of course, the absence of a short-run Fisher effect is troubling, as bond investors should seek compensation for expected inflation in both the short and long runs.

We hypothesize that an important error of omission in Fisher studies is the Wicksell price effect, which suggests that there is a negative relationship between inflation rates and interest rates. He believed that prices in the economy change in response to movements in an unobservable natural interest rate[3] (i.e., the marginal productivity of capital) relative to the bank lending rate (i.e. cost of capital), both of which are real rates.[4] According to his "cumulative process" model of business cycles, if the natural rate is greater (less) than the lending rate, then inflation (deflation) will occur as business firms expand (contract) their operations through increased (decreased) borrowing. In this respect his ideas were ahead of the times, as central banks around the world today practice neo-Wicksellian monetary policies of price stability, hereafter the Wicksell price effect.[5] Clinton (2006) ob-

---

[3] Laubach & Williams (2003) attempted to empirically estimate Wicksell's natural rate of interest but found the resultant estimates were imprecise. See also Orphanides & Williams (2002).

[4] Sargent (1969) has noted that Keynes (1930) adopted the Wicksellian view that prices respond to interest rate movements.

[5] We use the term *Wicksell price effect* to differentiate from the *Wicksell effect* used in another context. Excellent overviews of the Wicksell price effect are provided in Graboyes & Humphrey (1990) and Humphrey (1997). Woodford (2003) is attributed for reviving Wicksell's notion of using interest rates to control prices. See also Hoover (2006) for a review of Woodford's

serves that Wicksell's now famous book *Interest and Prices* did not get translated into English until 10 years after his death in 1936. Despite the slow dissemination of his ideas, he is recognized by some economics historians as one of the founding fathers of modern macroeconomics (Blaug 1986).

In the aftermath of abandoning the gold standard in the Great Depression, central banks of many industrialized countries relied on the quantitative theory of money to control inflation. Alternatively, influenced by Wicksell's interest rate theory of price determination, the central bank of Sweden (Sveriges Riksbank) adopted his program of price stabilization (Berg & Jonung 1999 and Woodford 2003). This monetary experiment succeeded to control inflation in Sweden but implementing monetary policy by means of a central bank's policy interest rate was not conducted by other central banks until half a century later when in 1981 the U.S. Federal Reserve began an important monetary policy change. In response to inflation at that time, the Fed employed a policy of raising interest rates to control inflation. Due to the Wicksell price effect, the negative impact of real interest rates helped to reduce inflation (Blanchard 1984).

A variety of related explanations for the negative relationship between interest and inflation rates have been proposed over the past 50 years. According to Mundell (1963), Tobin (1965), and Fischer (1979), higher expected inflation reduces real returns to capital due to increasing the capital stock. Darby (1975) and Feldstein (1976) have argued that higher expected inflation reduces after-tax real returns due to the taxing of nominal capital gains. Stulz (1986) attributed the negative relationship to the inverse association between inflation and output as well as uncertainty about monetary policy. Kandel, Ofer & Sarig (1996) reported a negative relationship between ex-ante real interest and inflation rates and concluded that this finding contradicted the Fisher hypothesis (of independence between real interest and inflation rates[6]) but was consistent with the theories of Mundell, Tobin, Darby, Feldstein, and Stulz. Unifying this body of work, we believe that the negative relationship between real interest rates and inflation is due to the Wicksell price effect and that his cumulative process model of inflation more generally encompasses these explanations.

Wicksellian approach as well as Formaini (2004) for a discussion of Wicksell's impact on modern monetary policy.

[6] See also Mishkin (1981), who rejected the hypothesis that the real rate is constant in the periods 1931-1952 and 1953-1979. Adjustment for taxes lent further support for this finding. He found that a negative real-rate/inflation relationship was present in these sample periods, which he contrasted with Fama's (1975) non-rejection of the Fisher hypothesis in favor of a constant real rate during a period that lacked variation in the real rate data .

3 Theoretical Foundations of the Fisher Puzzle and the Taylor Principle

Consider a macroeconomic model represented by the following expressions**:**

$$x_t = \alpha_0 + \alpha_1(i_t - \pi_t) \tag{1}$$

$$\pi_t = \beta_0 + \beta_1 x_t \tag{2}$$

$$i_t = \lambda_0 + \lambda_1 \pi_t^*, \tag{3}$$

where equations (1) to (3) are the IS (investment-saving) curve, Phillips curve, and Fisher equation, respectively.[7] In IS equation (1), x is the deviation of actual output from potential output or unemployment rate, i is the nominal interest rate, π is the inflation rate, t is the time period, and $\alpha_1 > 0$. In Phillips curve (2), $\beta_1 < 0$. Lastly, in Fisher equation (3), π_t^* is expected inflation in period *t*, with expectations formed at the beginning of period *t*, and λ_1 is the Fisher coefficient equal to one, which can be more than one in the tax-adjusted Darby-Feldstein version. The positive relation between expected inflation and interest rates is known as the *Fisher effect*.

Empirical studies of the Fisher effect commonly re-specify equation (3) as:

$$i_t = \lambda_0 + \lambda_1 \pi_t + e_t, \tag{4}$$

Where π_t is the actual inflation rate in period *t*, and e_t is inflation forecast error.

Combining equations (1) and (2) gives the following relationship:

$$\pi_t = \varphi_0 + \varphi_1(i_t - \pi_t), \tag{5}$$

where $\phi_0 = \beta_0 + \alpha_0 \beta_1$, and $\phi_1 = \alpha_1 \beta_1 < 0$. Equation (5) shows that there is a negative relationship between real interest rates and inflation rates, which we refer to as the *Wicksell price effect*.[8]

---

[7] For recent studies of the IS curve, see Kara & Nelson (2004), Hafer & Jones (2008), and Stracca (2010). For studies of Phillips curve, see Gordon (1990) and Gali, Gertler & Lopez-Salido (2005).

[8] Fama (1975) developed a model for estimating the short-run relationship between inflation rates as the dependent variable and interest rates as the independent variable. He found a positive,

The joint existence of both Fisher and Wicksell processes implies that the observed time series of interest rates and inflation rates used for estimating the Fisher coefficient are ex-post data generated from the interactions of these two processes. Consequently, researchers using ex-post interest rate and inflation rate data series for estimating the ex-ante Fisher effect have not been able to obtain the full Fisher effect, i.e., the Fisher puzzle. Beginning with Fisher himself, as summarized in the previous section, more than seven decades of research work has been spent on explaining or attempting to resolve the estimation of Fisher coefficients less than unity.

Ignoring forecast error, equations (4) and (5) can be solved and re-written as follows:

$$i_t = \left(\frac{\lambda_0 + \lambda_0 \phi_0}{1 - \lambda_1 \phi_1} \right) + \left(\frac{-\lambda_1 \phi_1}{1 - \lambda_1 \phi_1} \right) \pi_t \tag{6.1}$$

$$\pi_t = \frac{\varphi_0 + \lambda_0 \varphi_1}{1 - \varphi_1 (\lambda_1 - 1)}. \tag{6.2}$$

Equation (6.1) is an ex-post Fisher equation after the full effects of the negative impact of interest rates on inflation rates are incorporated in inflation and interest rates. In this equation $\frac{-\lambda_1 \phi_1}{1 - \lambda_1 \phi_1}$ is the coefficient of ex-post π_t. The numerator $-\lambda_1 \phi_1$ of this coefficient is positive, and the denominator $1 - \lambda_1 \phi_1$ is more than one given Wicksell coefficient $\phi_1 < 0$. Using ex-post interest rate and inflation rate time series data, Fisher and subsequent researchers have regressed nominal interest rates on inflation rates and obtained ex post $\frac{-\lambda_1 \phi_1}{1 - \lambda_1 \phi_1} < 1$ as the Fisher coefficient. However, we hypothesize that the true ex-ante Fisher coefficient is given by λ_1. Because ϕ_1 is a fraction less than zero resulting in $(1 - \lambda_1 \phi_1) > 1$, it is obvious that the negative Wicksell coefficient of real interest rates ϕ_1 in equation (5) reduces the ex post Fisher coefficient. It also seriously biases the intercept estimate coinciding with the estimated ex-post real rate.

Equation (6.2) shows that lower inflation rates are associated with values of λ_1 more than unity. Given that $\phi_1 < 0$, when λ_1 is larger (smaller) than one, then the

---

significant short-run relationship between inflation and interest rates (see also Kugler (1982)). However, when longer lags are introduced, as shown in forthcoming empirical findings in the next section, the Wicksell effect dominates the Fisher effect.

denominator of equation (6.2) is more (less) than one leading to lower (higher) inflation rates. If equation (4) is to be employed as a monetary policy rate (e.g., targeting the federal funds rate), then the size of λ_1 should be more than unity in order to reduce the inflation rate, as suggested by the Taylor principle.[9] Equation (6.2) shows that, given ϕ_1 is negative, the higher the magnitude of ϕ_1 the lower is equilibrium inflation.

Solving for π in equation (5) yields:

$$\pi_t = \frac{\phi_0}{1+\phi_1} + \frac{\phi_1}{1+\phi_1} i_t \tag{7}$$

or

$$\pi_t = \eta_0 + \eta_1 i_t . \tag{8}$$

Equations (7) and (8) indicate that there is also a negative relationship between nominal interest rates and inflation rates. Combining these equations, it is clear that the relationship between the Wicksell coefficient when the real interest rate is used (ϕ_1) and when nominal interest rate is used (η_1) is as follows:

$$\varphi_1 = \frac{\eta_1}{1-\eta_1} . \tag{9}$$

Equation (9) shows that, when real interest rates are used, the Wicksell coefficient is negative and less than one as long as the estimated Wicksell coefficient is negative when nominal interest rates are used.

Equations (4) to (7) assume that the Fisher and Wicksell processes are fully realized in a single period. In the real world, the realization of these two processes

---

[9] The Taylor rule was first proposed by John B. Taylor (1993) as well as Dale W. Henderson & Warwick McKibbin (1993). Its general form is: $i_t = \bar{i} + \alpha_\pi (\pi_t - \pi_t^*) + \alpha_y (y_t - \bar{y}_t)$, where i_t is the central bank's short-term policy rate, $\bar{i}$ is the long-run policy rate, π_t is the inflation rate, π_t^* is the target inflation rate of central bank, and $(y_t - \bar{y}_t)$ is the output gap (viz., aggregate output minus potential output). The Taylor principle says that central banks can stabilize the macroeconomy by adjusting the short-term policy rate more than one percent for one percent inflation rate, setting $\alpha_\pi > 1$. See Davig & Leeper (2007) for a model of the Taylor rule with endogenous monetary regimes.

takes time and is distributed over several periods. Thus, the dynamic Fisher and Wicksell process can be represented as:

$$i_t = \lambda_0 + \sum_{j=0}^{n} \lambda_{1j} \pi_{t-j} + \sum_{j=1}^{n} \lambda_{2j} i_{t-j} \quad (10.1)$$

$$\pi_t = \phi_0 + \sum_{j=0}^{n} \phi_{1j} i_{t-j} + \sum_{j=1}^{n} \phi_{2j} \pi_{t-j}, \quad (10.2)$$

where $\sum_{j=0}^{n} \lambda_{1j}$ and $\sum_{j=0}^{n} \phi_{1j}$ capture the short-run Fisher effect and Wicksell price effect, respectively, whereas their long-run effects are given as:

$$\Lambda = \sum_{j=0}^{n} \lambda_{1j} / (1 - \sum_{j=1}^{n} \lambda_{2j}) \quad (11.1)$$

$$\Theta = \sum_{j=0}^{n} \phi_{1j} / (1 - \sum_{j=1}^{n} \varphi_{2j}) \quad (11.2)$$

provided $\sum_{j=1}^{n} \lambda_{2j} < 1$ and $\sum_{j=1}^{n} \varphi_{2j} < 1$.

4 Empirical Analysis

An unaddressed empirical problem in the Fisher and Wicksell relations is that it is not clear what maturity of government interest rates is appropriate. A wide variety of government maturities are available to investors. Because all interest rates are affected by inflation and vice versa, we circumvent this problem by constructing a macro-interest rate, which is defined as aggregate interest paid by the nonfinancial sector of the U.S. economy divided by the aggregate debt outstanding of this sector in the U.S. economy. Aggregate interest paid is the sum of monetary interest paid by the household, business, and the government (federal, state, and local) sectors. This data series is available from the U.S. Bureau of Economic Analysis on an annual basis from 1960 to 2012 (Table 7.1 of the BEA). Total debt outstanding is from series Z1 of the Federal Reserve's Flows of Funds Account.[10] Our inflation rate series are the annual growth rate of gross domestic product (GDP) deflator, consumer expenditure price (CE) deflator, and consumer price index (CPI) for all urban consumers, all items.

---

[10] The website for the U.S. Bureau of Economic Analysis is: http://www.bea.gov. The website for the Flow of Funds Accounts is: http://www.federalreserve.gov/releases/z1/.

Table 1. Descriptive Statistics for Macro-Interest Rates and Selected Interest Rates 1960-2012. Macro denotes an average macro-interest rate defined as aggregate interest paid divided by the aggregate debt outstanding in the U.S. economy, Aaa and Baa are Moody's seasoned Aaa and Baa corporate bond yields, respectively, GS10 is the 10-year Treasury constant maturity rate, and federal funds is the interbank loan rate. Correlation coefficients between the macro-interest rates and different interest rate series are shown in the last column. Sources: Federal Reserve and Bureau of Economic Analysis.

Interest Rates	Mean	Median	Max.	Min.	Standard Deviation	Correlation Coefficient
Macro	8.10	8.14	12.44	4.69	2.15	1.00
Aaa	7.47	7.26	14.17	3.67	2.50	0.93
Baa	8.50	8.05	16.11	4.83	2.76	0.90
GS10	6.54	6.21	13.91	1.80	2.67	0.89
Federal funds	5.55	5.11	16.38	0.10	3.45	0.77

Table 1 gives descriptive statistics for the macro-interest rate as well selected interest rate series available from 1960 to 2012. As shown in Figure 1, among different interest rates, our macro-interest rate has the largest Pearson correlation coefficient (0.93) with Moody's seasoned Aaa corporate bond yield.

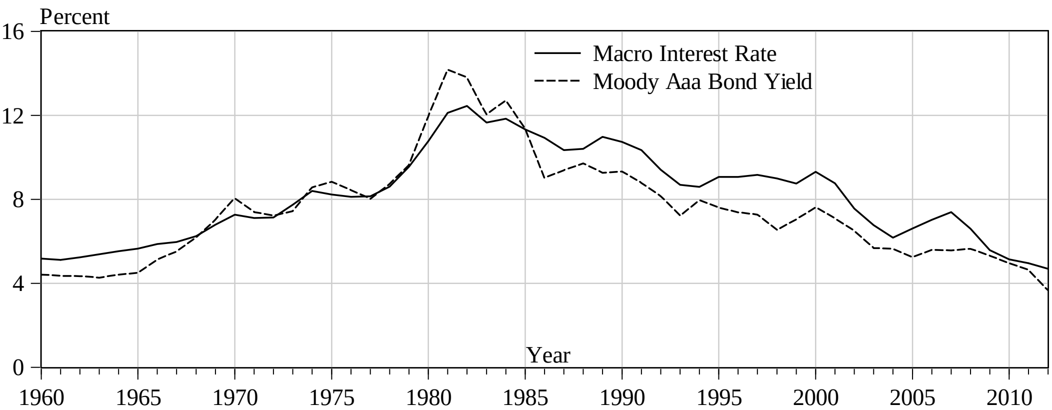


Figure 1. Macro-Interest Rates and Moody's Seasoned Aaa Corporate Bond Yields 1960–2011. Sources: Federal Reserve and Bureau of Economic Analysis.

Figures 2.1 and 2.2 provide scatter diagrams together with nearest neighborhood fits for the Fisher and Wicksell effects using macro-interest rates and GDP inflation rates. Figure 2.1 shows higher nominal macro-interest rates are associated with higher inflation rates. Figure 2.2 illustrates that lower inflation rates are associated with higher real interest rates (i.e., nominal rates minus inflation rates). Similar results were obtained using CE and CPI inflation rates.

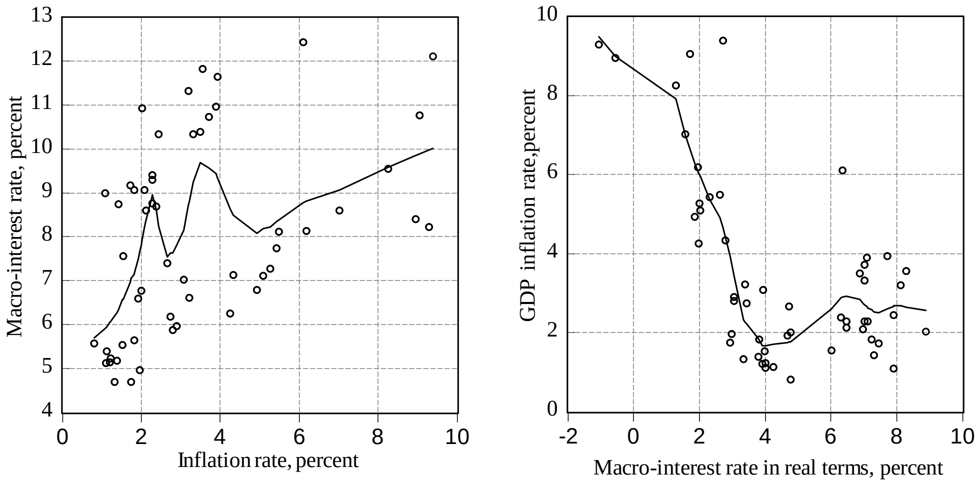


Figure 2. The Fisher Effect and Wicksell Price Effect. Nearest neighborhood fits showing in Figure 2.1 the positive Fisher effect of GDP inflation rates on nominal macro-interest rates and in Figure 2.2 the negative Wicksell effect of macro-interest rates in real terms on inflation rates.

The bivariate system of equations (10.1-2) is a structural model of the relationships between inflation and interest rates. The reduced form of the system in more compact notation can be specified as a vector autoregression (VAR) model:

$$Y_t = C + \sum_{k=1}^{n} A_k Y_{t-k} + BX_t + u_t, \tag{11}$$

where $Y = (i \quad \pi)$ is a bivariate vector of interest and inflation rates, C is a 2x1 vector of constants, A_k are 2x2 matrices of coefficients to be estimated, X_t is a vector of exogenous variables, B is a vector of coefficients of exogenous variables, and u_t represents unexpected movements in i and π. It is assumed that $E(u_t) = 0$, $E(u_t u_v') = 0$ when $v \neq t$. The order of VAR is be determined based on an optimal lag length criterion and should be long enough to avoid the residual serial correlation problem. The VAR method affords estimation of impulse response functions as well as forecast variance decompositions. Following Gordon (1977), to take into account the imposition and lifting of the Nixon wage-price controls in 1971, we augmented our VAR model with 0-1 dummy variables *NIXON* and *NIXOFF*, respectively.

Table 2. Order of Lag Selection Criteria in the Bivariate Vector Autoregression Models of the Fisher and Wicksell Effects Between GDP Inflation Rates and Macro-Interest Rates. An asterisk denotes the suggested criterion at the 5% significance level. LR is the sequential modified LR test, FPE is the final prediction error, and AIC, SC, and HQ are the Akaike information criterion, Schwarz information criterion, and Hannan-Quinn information criterion, respectively.

Lag Order	LR	FPE	AIC	SC	HQ
1	235.455	0.165	3.875	4.243*	4.017
2	11.767	0.150	3.773	4.288	3.972
3	12.783*	0.131*	3.637*	4.300	3.892*
4	0.944	0.149	3.763	4.573	4.075
5	2.155	0.166	3.859	4.816	4.228

To determine the lag order in the VAR system, we examined the sensitivity of the results to lag structure (Hafer & Sheehan (1991)). Table 2 presents the results from five alternative lag selection criteria: the sequential modified Langrage ratio (LR) test, final prediction error test (FPE), Akaike (1973) information criterion (AIC), Schwarz (1978) information criterion, and Hannan-Quinn (1979) information criterion. The test statistics at the five percent level suggest a lag order of three according to the LR, FPE, AIC, and HQ criteria but lag order of one with SC. Similar results were obtained based on CE and CPI inflation rates. Accordingly, we specified a VAR with a lag order of three.

Table 3 presents estimated VAR models of the Fisher and Wicksell effect in bivariate models of the relationship between macro-interest rates and the three inflation rate measures using ordinary least squares (OLS). The first two columns of Table 3 show the estimated VAR model for macro-interest rates and GDP inflation rates. In the Fisher effect equation shown in the first column, the estimated coefficient of the first lag of inflation rate (π_{t-1}) is positive (0.163) and statistically significant at the five percent significance level. The sum of the estimated coefficients of lags on inflation rates in this equation is positive suggesting a positive short-run relationship between nominal macro-interest rates and inflation rates in line with the Fisher effect. Turning to Wicksell price effect equation shown in the second column of Table 3, the estimated coefficient of the second lag on macro-interest rates (i_{t-2}) is negative (-1.375) and statistically significant at the one percent significance level. The sum of the estimated coefficients of lags on macro-interest rates in this equation is negative suggesting a negative short-run relationship between inflation rates and nominal macro-interest rates per the Wicksell price effect.

Table 3. Estimated Bivariate Vector Autoregression Models of the Fisher and Wicksell Effects. Variables are defined as follows: i = macro-interest rates, π = inflation rate, and dummy variables *NIXON* and *NIXOFF* represent the imposition and lifting of the Nixon wage-price controls, respectively. Figures in parentheses are *t*-statistics, where $**$, $*$, and + denote statistical significance at the 1, 5, and 10 percent significance levels, respectively. Adjusted *R*-squared values (R^2) and Durbin-Watson statistics (*DW*) are reported also.

	GDP inflation Rate		CE Inflation Rate		CPI Inflation Rate	
	Fisher	Wicksell	Fisher	Wicksell	Fisher	Wicksell
Regressand	i_t	π_t	i_t	π_t	i_t	π_t
Intercept	0.324	0.799+	0.370	0.999	0.380	1.098
	(1.39)	(1.68)	(1.56)	(1.62)	(1.55)	(1.36)
π_{t-1}	0.163^{*}	0.962^{**}	0.143^{*}	0.747^{**}	0.087^{+}	0.922^{**}
	(2.16)	(6.22)	(2.30)	(4.60)	(1.71)	(5.44)
π_{t-2}	-0.152	-0.234	-0.095	-0.117	-0.028	-0.298
	(-1.48)	(-1.11)	(-1.27)	(-0.60)	(-0.43)	(-1.41)
π_{t-3}	0.105^{+}	0.212	0.062	0.296^{*}	0.029	0.295^{*}
	(1.66)	(1.64)	(1.26)	(2.29)	(0.69)	(2.15)
i_{t-1}	1.421^{**}	0.433	1.426^{**}	0.759^{+}	1.430^{**}	0.746
	(9.05)	(1.35)	(9.22)	(1.88)	(8.57)	(1.36)
i_{t-2}	-0.738^{**}	-1.375^{**}	-0.767^{**}	-1.991^{**}	-0.792^{**}	-2.187^{*}
	(-2.90)	(-2.64)	(-3.10)	(-3.07)	(-2.96)	(-2.48)
i_{t-3}	0.230	0.857^{**}	0.252^{+}	1.128^{**}	-0.274^{+}	-1.337^{*}
	(1.53)	(2.79)	(1.72)	(2.94)	(-1.74)	(-2.56)
NIXON	-0.283	0.123	-0.207	-0.019	-0.181	-0.456
	(-0.89)	(0.19)	(-0.65)	(-0.02)	(-0.55)	(-0.42)
NIXOFF	-0.559	2.177^{**}	-0.600	2.515^{*}	-0.402	2.055^{+}
	(-1.56)	(2.97)	(-1.59)	(2.55)	(-1.09)	(1.69)
R^2	0.96	0.86	0.96	0.79	0.96	0.73
DW	2.09	1.65	2.10	1.77	2.08	1.87

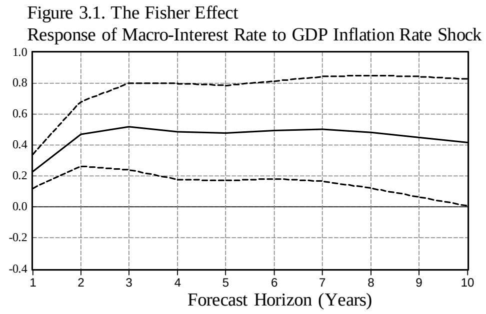
Moreover, the negative value and the statistical significance of the second lag of nominal macro- interest rates on inflation rates shows that the full impact of interest rates on inflation rates takes time as confirmed by previous researchers (e.g., see Fisher (1930), Friedman (1961) and others).

Columns 3 and 4 of Table 3 present the estimated results for the VAR model using CE inflation rates, and column 5 and 6 show the results with CPI inflation rates. The similarity of the results to those for GDP inflation rates demonstrates the robustness of the model to alternative measures of inflation rates in the U.S. economy.

Table 4 presents estimated short-run and long-run Fisher and Wicksell effects from the VAR models. The short-run Fisher effect – namely, the sum of the estimated coefficients of lagged inflation rates in the Fisher equation ($\sum_{j=0}^{n} \lambda_{1j}$) -- is positive with magnitudes ranging from 0.089 to 0.116 with respect to different inflation rate measures. The magnitudes of the long-run Fisher effect defined as $\sum_{j=0}^{n} \lambda_{1j} /(1-\sum_{j=1}^{n} \lambda_{2j})$ lie between 1.009 and 1.344 implying tax rates of 1 to 25 percent. The short-run Wicksell price effect – specifically, the sum of the estimated coefficients of lagged interest rates in the Wicksell equation ($\sum_{j=0}^{n} \phi_{1j}$) -- is negative with magnitudes ranging from -0.086 to -0.103. And, the long-run Wicksell price effect defined as $\sum_{j=0}^{n} \phi_{1j} /(1-\sum_{j=1}^{n} \varphi_{2j})$ lies between -1.268 and -1.443. Due to controversies concerning whether CPI inflation rate underestimates (via excluding volatile food and energy prices) or overestimates inflation rates, the results based on GDP inflation rates may be more reliable. In general, we interpret these results to suggest that short- and long-run negative relationships between interest rates and inflation rates help to explain the Fisher puzzle and the Taylor principle.

Table 4. Short-run and Long-run Fisher and Wicksell Effects

Coefficients	GDP Inflation Rate	CE Inflation Rate	CPI Inflation Rate
Short-run Fisher coefficient, $\sum_{j=0}^{n} \lambda_{1j}$	0.116	0.110	0.089
Short-run Wicksell coefficient, $\sum_{j=0}^{n} \phi_{1j}$	-0.086	-0.103	-0.103
Long-run Fisher effect, $\sum_{j=0}^{n} \lambda_{1j} /(1-\sum_{j=1}^{n} \lambda_{2j})$	1.344	1.229	1.009
Long-run Wicksell effect, $\sum_{j=0}^{n} \phi_{1j} /(1-\sum_{j=1}^{n} \varphi_{2j})$	-1.443	-1.374	-1.268



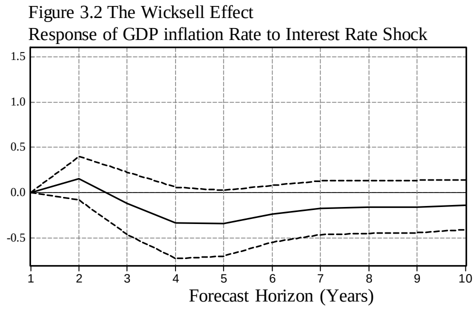


Figure 3. Graphs of Cholesky Impulse Response Functions. Figure 3.1 shows responses of macro-interest rates to a one-standard deviation shock in GDP inflation rates. Figure 3.2 shows responses of GDP inflation rates to a one-standard deviation shock in macro-interest rates. Plus and minus two standard deviation confidence bands are illustrated with dotted lines.

The Fisher and Wicksell effects not only exist between expected values of interest and inflation rates but between unpredicted movements in these two rates also. Figure 3.1 shows the impulse response functions for macro-interest rates and GDP inflation rates from the VAR model for a ten-year forecast horizon based on a Cholesky decomposition of the covariance matrix. Figure 3.1 shows that the response of macro-interest rates to a positive shock in inflation. The time path of the response makes clear that the Fisher effect is long lasting over a period of five-to-ten years. Furthermore, Figure 3.2 shows that the response of inflation rates to a shock in interest rates is initially positive but then becomes negative in the second year of the forecast horizon and thereafter for several years.

Table 5 presents the forecast error variance of macro-interest rates attributable to GDP inflation rates as well as forecast error variance (FEV) of inflation rates attributable to interest rates. Because the results from Cholesky decomposition of the covariance matrix depend on the ordering of the variables, we estimated variance decompositions based on alternative orderings of macro-interest rates and inflation rates. Panel A of Table 5 shows that interest rate shocks can explain more than 37 percent of the FEV of inflation rates in the second year of the forecast horizon and more than 26 percent after the second year. Panel B shows that inflation rates can explain more than 29 percent of the FEV of macro-interest rates in the first year of the forecast horizon increasing to more than 70 percent after 10 years.

Table 5. Forecast Error Variance Decomposition of Macro-Interest Rates and GDP Inflation Rates in the Bivariate VAR Model of Fisher and Wicksell Effects. Variables are defined as follows: i = macro-interest rates, and π = GDP inflation rates.

A. Percentage of Squared Prediction Error of Inflation Rates Explained by Macro-Interest Rates

Forecast Horizon	1	2	3	4	5	6	7	8	9	10
Ordering: i_t, π_t	29.24	37.78	33.34	30.88	29.46	27.99	27.01	26.51	26.35	26.30
Ordering: π_t, i_t	0.00	1.44	1.94	6.89	11.21	12.89	13.60	14.32	15.07	15.64

B. Percentage of Squared Prediction Error of Macro-Interest Rates Explained by Inflation Rates

Forecast Horizon	1	2	3	4	5	6	7	8	9	10
Ordering: i_t, π_t	0.00	2.17	3.83	6.10	9.24	12.99	16.78	20.28	23.40	26.16
Ordering: π_t, i_t	29.24	41.39	46.57	51.12	55.67	60.07	63.92	67.00	69.36	71.15

The VAR model can be augmented further by including more variables, such as the federal funds rate, inflation rate of imported goods and services, etc. However, these analyses are beyond the scope of the present work.

We also conducted cointegration tests to evaluate the possibility of employing an error-correction VAR model but found little empirical evidence to support this method. In the long-run, the Fisher and Wicksell effects are expected to offset each other leading to equilibrium inflation and interest rates, otherwise there would not be lower and upper bounds on the two time series variables.

5 Conclusion

This paper has argued that the long-standing Fisher puzzle of a less than unity estimated relationship between nominal interest rates and inflation rates can be explained in large part by the Wicksell price effect. That is, the negative Wicksell relation between real or nominal interest rates and inflation rates offsets the positive Fisherian inflation effect on nominal interest rates to a considerable degree. In line with this reasoning, we proposed a novel model of the joint determination of inflation and interest rates based on combined Fisher and Wicksell effects and, subsequently, applied the model to U.S. data. To empirically test our model, we circumvent a previously unaddressed empirical problem of the appropriate interest rate maturity in Fisher-Wickell relations by constructing a macro-interest rate series that takes into account aggregate debt outstanding and aggregate interest paid by the household, business, and the government sectors.

Empirical results documented a highly significant and negative relationship between interest rates and inflation rates. Both short- and long-run Fisher and Wicksell effects were estimated in addition to the time paths of these respective processes. Depending on the measure of inflation rates, short-run Fisher coefficients were estimated in the range of 0.089 to 0.116 compared to a short-run Wicksell coefficients of -0.086 to -0.103. Long-run Fisher coefficients were estimated in the range of 1.009 and 1.344 implying tax rates of 1 to 25 percent, whereas the long-run Wicksell coefficient was estimated in the range -1.268 and -1.443. Impulse response functions for macro-interest rates and inflation rates from the VAR model for a ten-year forecast horizon revealed: (1) inflation rate shocks have positive, Fisherian long-run effects on macro-interest rates over this horizon, and (2) macro-interest rate shocks have negative, Wicksellian long-run effects on inflation rates over many years. Further variance decomposition analyses indicated that interest rate shocks can explain more than 37 percent of the forecast error variance of inflation rates after two years and that inflation rate shocks can explain more than 29 percent of the forecast error variance of macro-interest rates in the first year of forecast horizon and over 70 percent after 10 years.

An important implication of the large negative impact of interest rates on inflation rates is that, as proposed under the Taylor rule, a greater than one percent increase in interest rates is required to decrease inflation rates by one percent. Further implications of our findings relevant to the study of the relationships between interest and inflation rates include the following: (1) single equation approaches for estimating the Fisher equation should be replaced by a system approach that takes into account both Fisher and Wicksell effects; (2) the model should be made dynamic by including lags of both interest and inflation rates that are long enough to incorporate both opposing effects; and (3) the impact of all interest rates in an economy on inflation should be utilized in Fisher-Wicksell analyses in view of the fact that interest rates of different maturities and risks can potentially impact inflation.

We conclude that empirical evidence supports our earlier theoretical propositions that the Wicksell effect plays a crucial role in explaining the Fisher puzzle and Taylor principle. Future research is recommended to extend our framework to decision-making methods employed by central banks to control inflation. For example, our model could be employed to forecast inflation and thereby assist monetary authorities. Also, we hope that researchers will find our model useful to various financial and macroeconomic applications.

Acknowledgements: The authors gratefully acknowledge financial support from the Center for International Business Studies and Real Estate Center, Mays Business School, Texas A&M University. We have benefited from comments by seminar participants at the annual conferences of the Eastern Finance Association (2010) in New York, NY and the Western Economic Association International (2011) in San Diego, CA. Helpful comments have been provided by Will Armstrong, Johan Knif, Wei Liu, Heather Mitchell, and Margarita Sapozhnikov.

References

Akaike, H. (1973). Information theory and the extension of the maximum likelihood principle. In Proceedings of the *2nd International Symposium on Information Theory* (B.N. Petrov & F. Caski, eds.). Budapest, 206–281.

Barsky, R. (1987). The Fisher hypothesis and the forecastibility and persistence of inflation, *Journal of Monetary Economics* 19, 3–24.

Berg, C. & Jonung, L. (1999). Pioneering price level targeting: The Swedish experience 1931-37, *Journal of Monetary Economics* 43, 525–551.

Bierens, H.J. (2000). Nonparametric nonlinear cotrending analysis, with an application to interest and inflation in the U.S. *Journal of Business and Economic Statistics* 18, 323–337.

Blanchard, O. & Fischer, S. (1989). *Lectures on Macroeconomics*. Cambridge, MA: MIT Press, Chapter 4.

Blaug, M. (1986). *Great Economists Before Keynes: An Introduction to the Lives and Works of 100 Economists of the Past*. Brighton, UK: Harvester Wheatsheaf.

Carmichael, J. & Stebbing, P.W. (1983). Fisher's paradox and the theory of interest, *American Economic Review* 63, 619–630.

Carrington, S. & Crouch, R. (1987). A theorem on interest rate differentials, risk and anticipated inflation, *Applied Economics* 19, 1675–1683.

Choudhry, A. (1997). Cointegration analysis of the inverted Fisher effect: Evidence from Belgium, France and Germany. *Applied Economics Letters* 4, 257–260.

Christopoulos, D. & Leon-Ledesma, M. (2007). A long-run nonlinear approach to the Fisher effect, *Journal of Money, Credit and Banking* 39, 543–559.

Clinton, K. (2006). Wicksell at the Bank of Canada. Working Papers 1087, Queen's University, Department of Economics.

Cooray, A. (2002). The Fisher effect: A review of the literature, Research papers 0206. Macquarie University, Department of Economics.

Crowder, W.J. & Wohar, M. (1997). Are tax effects important in the long-run Fisher relationship? Evidence from the municipal bond market, *Journal of Finance* 54, 307–317.

Crowder, W.J & Hoffman, D.L. (1996). The long-run relationship between nominal interest rates and inflation: The Fisher equation revisited. *Journal of Money, Credit, and Banking* 28, 102–118.

Darby, M.R. (1975). The financial and tax effects of monetary policy on interest rates. *Economic Inquiry* 13, 266–276.

Davig, T. & Leeper, E.M. (2007). Generalizing the Taylor principle, *American Economic Review* 97, 607–635.

Evans, M. & Lewis, K. (1995). Do expected shifts in inflation affect estimates of the long-run Fisher relation? *Journal of Finance* 50, 225–253.

Fama, E.F. (1975). Short-term interest rate as predictors of inflation. *American Economic Review* 65, 269–282.

Feldstein, M. (1976). Inflation, income taxes, and the rate of interest: A theoretical analysis. *American Economic Review* 66, 809–820.

Fischer, S. (1979), Anticipations and the nonneutrality of money, *Journal of Political Economy*, 87, 225-252.

Fisher, I. (1930). *The Theory of Interest*. New York, NY: Macmillan.

Formaini, R.L. (2004). Knut Wicksell: The birth of modern monetary policy. *Economic Insights* 9, Federal Reserve Bank of Dallas.

Friedman, M. (1961). The lag in effect of monetary policy. *Journal of Political Economy* 69, 447–466.

Gali, J., Gertler, M. & Lopez-Salido, D. (2005). Robustness of the estimates of the hybrid new Keynesian Phillips curve. *NBER Working Paper* 11788.

Gandolfi, A.E. (1982). Inflation, taxation and interest rates. *Journal of Finance* 37,797–807.

Gibson, W.E. (1972). Interest rates and inflationary expectations: New evidence. *American Economic Review* 62, 854–865.

Gordon, R.J. (1977). Can the inflation of the 1970s be explained? *Brookings Papers on Economic Activity* 1, 253–276.

Gordon, R.J. (1990). The Phillips curve now and then. *NBER Working Paper* 3393.

Graboyes, R.F. & Humphrey, T.M. (1990). Wicksell's monetary framework and dynamic stability. Working Paper 90–7. Federal Reserve Bank of Richmond.

Hafer, R. & Sheehan, R. (1991). Policy inference using VAR models. *Economic Inquiry* 29, 44–52.

Hafer, R.W. & Jones, G. (2008). Dynamic IS curves with and without money: An international comparison. *Journal of International Money and Finance* 27, 609–616.

Hannan, E.J. & Quinn, B.G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society B* 41, 190–195.

Henderson, D.W. & McKibbin, W. (1993). A Comparison of some basic monetary policy regime for open economies: Implications of different degrees of instrument adjustment and wage persistence. *Carnegie-Rochester Conference Series on Public Policy* 39, 221–318.

Hoover, K.D. (2006). A neoWicksellian in a new classical world: The methodology of Michael Woodford's Interest and Prices. *Journal of the History of Economic Thought* 28, 143–149.

Humphrey, T.M. (1997). Fisher and Wicksell on the quantity theory. *Economic Quarterly* 83, Federal Reserve Bank of Richmond, 71–91.

Kandel, S., Ofer, A. & Sarig, O. (1996), Real Interest and Inflation: An Ex-Ante Empirical Analysis. *Journal of Finance* 51: 1, 205–225.

Kara, A. & Nelson, E. (2004). International evidence on the stability of the optimizing IS equation. *Oxford Bulletin of Economics & Statistics* 66, 687–712.

Keynes, J.M. (1930). *A Treatise on Money Vol. I: The Pure Theory of Money; Vol. II: The Applied Theory of Money*. London, UK: Macmillan.

Kugler, P. (1982). The dynamic relationship between interest rates and Inflation: An empirical investigation. *Empirical Economics* 7, 125–137.

Lahiri, K. (1976). Inflationary expectations: Their formation and interest rate effects. *American Economic Review* 66, 124–131.

Lanne, M. (2006). Nonlinear dynamics of interest and inflation. *Journal of Applied Econometrics* 21, 1157–1168.

Laubach, T. & Williams, J.C. (2003). Measuring the natural rate of interest. *Review of Economics and Statistics* 85, 1063–1070.

Levi, M.D. & Makin, J.H. (1978). Anticipated inflation and interest rates: Further interpretation of findings on tbe Fisber equation. *American Economic Review* 68, 801–812.

Levi, M.D. & Makin, J.H. (1979). Fisher, Phillips, Friedman and the measured impact of inflation on interest. *Journal of Finance* 34, 35–52.

Mishkin, F.S. (1981). The real interest rate: An empirical investigation. *NBER Working Paper* 622.

Mishkin, F.S. (1992). Is the Fisher effect for real? *Journal of Monetary Economics* 30, 195–215.

Mishkin, F.S. (1995). Nonstationarity of regressors and test on the real rate behavior. *Journal of Business and Economics Statistics* 13, 47–51.

Mishkin, F.S. & Simon, J. (1995). An empirical examination of the Fisher effect in Australia. *Economic Record* 71, 217–229.

Mundell, R.A. (1963). Inflation and real interest. *Journal of Political Economy* 71, 280–283.

Ng, S. & Perron, P. (1997). Estimation and inference in nearly unbalanced nearly cointegrated systems. *Journal of Econometrics* 79, 53-81.

Nielsen, N. C. (1981), Inflation and taxation: Nominal and real rates of return, *Journal of Monetary Economics* 7, 261–270.

Orphanides, A. & Williams, J.C. (2002). Imperfect knowledge, inflation expectations and monetary policy. *Finance and Economics Discussion Series* 2002-27. Washington, DC: Board of Governors of the Federal Reserve System.

Peek, J. (1982). Interest rates, income taxes and anticipated inflation. *American Economic Review* 72, 980–991.

Pelaez, R.F. (1995). The Fisher effect: Reprise. *Journal of Macroeconomics* 17, 333–346.

Peng, W. (1995). The Fisher hypothesis and inflation persistence evidence from five major industrial countries. *IMF Working Paper* 95/118. Washington, DC: International Monetary Fund.

Phillips P.C.B. (1998). Econometric analysis of Fisher’s equation. Cowles Foundation Discussion Paper 1180. Yale University.

Pyle, D. (1972). Observed price expectations and interest rates. *Review of Economics and Statistics* 54, 275–280.

Rose, L. (1986). A respecified tax-adjusted Fisher relation. *Economic Inquiry* 24, 319–331.

Rose, A.K. (1988). Is the real interest rate stable? *Journal of Finance* 43, 1095–1112.

Sargent, T.J. (1969). Commodity price expectations and the interest rate. *Quarterly Journal of Economics* 83, 127–140.

Schwarz, G.E. (1978). Estimating the dimension of a model. *Annals of Statistics* 6, 461–464.

Soderlind, P. (1998). Nominal interest rates as indicators of inflations expectations, *Scandinavian Journal of Economics* 100, 457–472.

Stracca, L. (2010). Is the new Keynesian IS curve structural? Working Paper 1236. European Central Bank.

Stulz, R.M. (1986). Interest rates and monetary policy uncertainty. *Journal of Monetary Economics* 17, 331–347.

Sun, Y. & Phillips, P.C.B. (2004). Understanding the Fisher equation. *Journal of Applied Econometrics* 19, 869–886.

Tanzi, V. (1980). Inflationary expectations, economic activity, taxes and interest rates. *American Economic Review* 70, 12–21.

Taylor, J.B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy* 39, 195–214.

Tobin, J. (1965). Money and economic growth. *Econometrica* 33, 671–684.

Tobin, J. (1969). A general equilibrium approach to monetary theory. *Journal of Money, Credit, and Banking* 1, 15–29.

Wicksell, K. (1898), *Interest and Prices: A Study of the Causes of Regulating the Value of Money*, Translated by R. F. Kahn, 1936, London, UK: MacMillan.

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, NJ: Princeton University Press.

Yoon, G. (2010). Does nonlinearity help resolve the Fisher effect puzzle? *Applied Economics Letters* 17, 823–828.

HEDGE FUNDS: MARKET TIMING AND THE DYNAMICS OF SYSTEMATIC RISK

Johan Knif
Hanken School of Economics
Dimitrios Koutmos
Worcester Polytechnic Institute
Gregory Koutmos
Fairfield University

1 Introduction

This study investigates the stochastic properties of the systematic (market) risk of several hedge fund index returns and the market timing abilities of hedge fund managers. The empirical evidence shows that the systematic risk of all hedge fund index returns is highly variable over time, implying that reported alpha returns are unreliable. In almost all cases volatility is asymmetric and the range of estimated betas is rather large. The degree of persistence is also very high. Both the simple regression and the vector EGARCH model show no evidence of successful market timing.

The growth in money under management by hedge funds has been spectacular during the last fifteen years. As of the second quarter of 2012, the amount stood at 1.7 trillion US dollars according to Barclays' Alternative Investment Databases. There are many reasons for this growth. Hedge funds have a great deal of flexibility given that they are largely unregulated pools of money managed by professional managers. Such flexibility allows managers to exploit perceived market inefficiencies more easily than traditional buy-and-hold mutual fund managers. More specifically, hedge funds can use leverage, take short positions and use derivative securities in their hedging, speculative and arbitrage activities. Another measure of the importance of hedge funds is their trading activity which according to Stulz (2007) accounts for approximately half the trading in the New York and London stock exchanges.

The performance of hedge fund managers and the compensation structure has been a subject of considerable research in the academic world as well as a subject of intense debate in the investment world. As discussed in Stulz (2007), typically the total fee consists of the standard management fee plus an incentive fee. The first, ranges between 1%-2% and the latter is calculated as a percentage of the profits that can range between 15%-25%. There are many variations and many stipulations about the particular incentive fee structure but in all instances the higher fees charged by hedge fund managers are based on the perceived ability of these managers to earn "alpha returns", i.e., returns that cannot be explained (replicated) by common risk factors. Ibbotson, Chen & Zhou (2011) find that during the period 1995-2009, hedge funds were able to add a significant amount of alpha to a typical portfolio consisting of stocks bonds and cash. On average, the alpha return was 3% whereas beta exposure contributed 4.7% during the same period.

There is no general agreement as to what constitutes "alpha returns" or equivalently what is the appropriate set of risk factors to be used when estimating alpha. In most cases the statistical framework used is based on the Capital Asset Pricing Model whereby excess returns on the hedge fund are regressed against excess returns on a suitable proxy for the market portfolio. The intercept of the regression is the alpha return component. The underlying assumption is that the model parameters, i.e., the market exposure parameter, the so-called beta, and the intercept are constant. A growing body of literature finds that returns on speculative assets are non-stationary, in their second moments and co-moments. If the betas are time varying it is possible that returns due to market co-movements are mistakenly interpreted as alpha returns. This is a very important issue since management fees can only be justified on the basis of alpha returns.

There is also a small but growing body of literature dealing with the market timing abilities of hedge fund managers and the degree to which these managers can earn alpha returns, i.e., returns unrelated to general market movements. The assumption however almost always is that the relationship of hedge fund returns to markets movements, the so-called beta parameter, is constant over time. There has been plenty of evidence however that the returns of speculative assets in general do not follow stationary distributions. For hedge fund returns in particular Brooks & Kat (2002) report that hedge fund index returns exhibit high kurtosis which has been linked to non-stationarity of higher moments. Similar findings are reported by Fuss, Kaiser & Adams (2006) and Kat & Lu (2002). Such findings cast doubt on the validity of traditional measures of investment performance such as the well-known and extensively used Sharpe Ratio (see Sharpe 1994).

Regarding the market timing ability of hedge fund managers, the evidence reported up to now is mixed. For example, Chen & Liang (2007), based on a sample of 221 funds self-identified as market timers for the 1994–2005 period, find that the evidence supports timing ability, especially in bear and volatile markets. Similar findings are reported by Xin Li & Shawky (2013) for Long/Short type hedge funds. On the other had Cave, Hubner & Sougne (2012) report that depending on the particular style, there have been positive, negative and mixed market timers during the financial crisis of 2008. Fung, Xu & Yau (2002) find that global hedge fund managers do not show positive market timing ability but instead demonstrate superior security selection ability.

The mixed evidence is, to some extent due to different time horizons and sample sizes employed in different studies. Almost all of these studies however employ statistical methods that assume stationary distributions. Thus, exposures to risk factors are assumed constant over the sample period. Given the evidence against stationarity however, such an assumption is not warranted and it may lead to erroneous conclusions.

The purpose of this paper is to test the market timing abilities of hedge fund managers and the possibility that beta coefficients are time-varying. More specifically, this study attempts to provide answers to the following questions: a) Is the systematic risk (beta) of hedge funds with a variety of investment styles time varying?; b) Is the systematic risk higher during market downturns (i.e., asymmetric)?; c) Is time variation and/or asymmetry related to the particular investment style?: d) What is the degree of persistence and predictability in systematic risk?

2 Data and Methodology

That data used in this study are weekly returns on hedge funds with the following investment styles: Convertible Arbitrage (CONV), Dedicated Short Bias (DEDS), Event Driven (DRIV), Emerging Markets (EMGM), Equity Market Neutral (NEUT), Fixed Income Arbitrage (FIAR), Global Macro (GLMA), Long-Short Equity (LSEQ), Managed Futures (FUTR), and Multiple Strategy (MSTR).

The data cover the period 9/12/2005 till 3/12/2012 for total of 340 weekly observations. The return series data are obtained from the TASS Hedge Funds Data Base which produces indexes of investment performance of several hedge fund classes.

We use a bivariate EGARCH model described by the following set of equations:

$$R_{i,t} = \alpha_i + \beta_{i,t} R_{m,t} + \theta_i R_{i,t-1} + \varepsilon_{i,t} \quad (1)$$

$$R_{m,t} = \alpha_m + \varepsilon_{m,t} \ , \quad (2)$$

where $R_{i,t}$ and $R_{m,t}$ are the weekly excess returns on the individual security and the market portfolio respectively; $\beta_{i,t}$ is the time-varying beta; α_i and α_m are constants and; $\varepsilon_{i,t}$ and $\varepsilon_{m,t}$ are innovations or, error terms for the individual security and the market respectively.

The elements of the variance/covariance matrix of the two error terms follow a bivariate EGARCH model described by the following set of equations (see Koutmos and Booth 1995):

$$\sigma^2[\varepsilon_{i,t}] = exp\{a_{i,0} + a_{i,1}(|z_{i,t-1}| - E|z_{i,t-1}| + c_i z_{i,t-1}) + b_i ln(\sigma^2[\varepsilon_{i,t-1}])\} \quad (3)$$

$$\sigma^2[\varepsilon_{m,t}] = exp\{a_{m,0} + a_{m,1}(|z_{m,t-1}| - E|z_{m,t-1}| + c_m z_{m,t-1}) + b_m ln(\sigma^2[\varepsilon_{m,t-1}])\} \quad (4)$$

$$\sigma_{i,m,t} = (\rho_{i,m} + \gamma_i R_{t-1}^{-})(\sigma^2[\varepsilon_{i,t}]\sigma^2[\varepsilon_{m,t}])^{1/2}, \quad (5)$$

where $ln(.)$ are natural logarithms, $z_{i,t} = \varepsilon_{i,t}/\sigma[\varepsilon_{i,t}]$ and $z_{m,t} = \varepsilon_{m,t}/\sigma[\varepsilon_{m,t}]$ are normalized innovations; $\sigma_{i,m,t}$ and $\rho_{i,m}$ are the conditional covariance and the conditional correlation coefficient; and $a_{i,0}$, $a_{i,1}$, c_i, b_i, $a_{m,0}$, $a_{m,1}$, c_m, b_m, ρ_{im}, γ_i are fixed parameters to be estimated. R_{t-1}^{-} is an indicator variable taking the value of 1 if the lagged market return is negative and 0 otherwise. The purpose is to test for asymmetries in the conditional correlation. If for example γ_i is positive and significant then the correlation with the market will be $\rho_{i,m} + \gamma_i$ if the lagged market return is negative. The beta of the individual hedge fund is given by

$$\beta_{i,t} = (\sigma_{i,m,t}/\sigma^2 [\varepsilon_{m,t}]) \quad (6)$$

Assuming that the returns of each hedge fund and the market index are jointly conditionally normally distributed, we can estimate the fixed parameters of the model described by (1)-(5) by maximizing the sample log-likelihood function. The latter can be written as

$$L(\boldsymbol{\Theta}) = -Tln(2\pi) - (1/2)\boldsymbol{\Sigma}(ln|\boldsymbol{H}_t| + \boldsymbol{E}_t\boldsymbol{H}_t^{-1}\boldsymbol{E}_t'), \quad (7)$$

where T is the number of observations, $\boldsymbol{\Theta}$ is the parameter vector to be estimated, $\boldsymbol{E}_t = [\varepsilon_{i,t}, \varepsilon_{m,t}]$ is the 1x2 vector of innovations at time *t*, $\boldsymbol{H}_t = Cov_{t-1}(\boldsymbol{E}_t)$, where the diagonal elements of $\boldsymbol{H}_t$ are given by (3) and (4) and the cross diagonal elements are given by (5). Because of nonlinearities in the log-likelihood function numerical maximization techniques are used to obtain parameter estimates. The particular algorithm used is based on Berndt et al. (1974).

3 Main Empirical Findings

The descriptive statistics reported on Table A1 in the Appendix suggest that all returns are negatively skewed and highly leptokurtic. The latter is mostly due to time variation in the variance of the returns. Unconditional normality is rejected in all instances on the basis of the Jarque-Bera statistic. Using the Sharpe ratio, it can be seen that the majority of hedge fund styles have underperformed the market index, the only exceptions being the Event Driven (DRIV) and the Managed Futures (FUTR).

Table A2 in the Appendix reports pairwise correlations among hedge fund styles and the market index. The estimated values do not follow a particular pattern. Dedicated Short Bias (DEDS) has a significant negative correlation with the market (-0.544) and Managed Futures (FUTR) also has a negative correlation with the market (-0.133). The rest of the hedge fund indices have positive correlations ranging from a high of 0.652, for Long-Short Equity (LSEQ) to low of 0.098 for the Equity Market Neutral (NEUT). Overall, the correlations with the market are rather low, suggesting that hedge funds focus on alpha strategies with low correlations with the market.

Table 1 reports the results of a market model where we allow for an asymmetric exposure to the market and for autocorrelation in the returns. The form of the regression equation is as follows:

$$R_{i,t} = \alpha_i + \beta_i R_{m,t} + \delta_i R^{+}_{m,t} + \theta_i R_{i,t-1} + \varepsilon_{i,t}, \tag{8}$$

where $R^{+}_{m,t} = R_{m,t}$ when the market return is positive and zero otherwise. The objective is to capture any asymmetric exposure to the market which would be an indication of market timing. If for example δ_i is positive and significant then we have evidence of successful market timing. The results show that most hedge funds have positive market exposure with the exception of Dedicated Short Bias (DEDS) and Managed Futures (FUTR). Surprisingly, the asymmetry parameter is negative and statistically significant across all hedge fund indices with the exception of Dedicated Short Bias and Managed Futures. This implies that that exposure to the market is low when the market return is positive and high when the market return is negative. If there was any effort at timing the market the results are exactly opposite of what we might have expected. This is contrary to the findings of Li & Shawky (2013) and Chen & Liang (2007). This may be due to the fact that this study is using more recent data that include the financial crisis of 2008 or, to the different methodologies. Even though the evidence on market timing is negative, there is support for the notion that hedge fund managers are doing a good job on the security selection front. Most alpha parameters (α_i) are positive and statistically significant. This is in agreement with the findings of Fung, Xu & Yau (2002).

Table 1. OLS regression results

$$R_{i,t} = \alpha_i + \beta_i R_{m,t} + \delta_i R^{+}_{m,t} + \theta_i R_{i,t-1} + \varepsilon_{i,t}$$

	α_i	β_i	δ_i	θ_i	$Q^2(5)$	$Q_{i,m}(5)$
CONV	0.176 (2.52)**	0.170 (6.19)**	-1.76 (-3.74)**	0.455 (9.89)**	41.89**	66.63**
DEDS	-0.063 (-0.45)	-0.353 (-6.66)**	-0.074 (-0.81)	0.016 (0.35)	10.30**	14.78*
EMGM	0.120 (1.28)	0.294 (8.03)**	-0.124 (-1.97)**	0.263 (5.60)**	12.06**	9.86**
NEUT	0.121 (1.36)	0.135 (3.91)**	-0.211 (-3.58)**	0.033 (0.61)	31.27**	12.11**
DRIV	0.244 (3.83)**	0.165 (6.68)**	-0.215 (-5.10)**	0.184 (3.63)**	15.69**	17.73**
FIAR	0.191 (2.45)**	0.218 (6.94)**	-0.277 (-5.13)**	0.238 (4.61)**	21.59**	72.69**
GLMA	0.120 (1.51)	0.176 (5.77)**	-0.130 (-2.48)**	0.131 (2.56)**	20.14**	19.13**
LSEQ	0.099 (1.74)*	0.277 (12.31)**	-0.123 (-3.19)**	0.262 (6.59)**	17.55**	23.31**
FUTR	0.137 (1.34)	-0.027 (-0.69)	-0.063 (-0.94)	0.092 (1.69)*	20.64*	23.50**
MSTR	0.115 (2.13)**	0.133 (6.34)**	-0.122 (-3.35)**	0.296 (5.88)**	17.11*	16.62**
PWA	0.134 (3.34)**	0.123 (7.76)**	-0.146 (-5.38)**	0.307 (6.36)**	16.96*	15.52**
Equally-Weighted Average	0.133 (3.26)**	0.116 (7.11)**	-0.152 (-5.48)**	0.295 (5.97)**	11.75*	13.72**

Notes: (**) and (*) indicate statistical significance at the 5% and the 10% levels respectively. Numbers in parentheses are t-statistics. $Q^2(5)$ and $Q_{i,m}(5)$ are the Ljung-Box statistics for the squared residuals and the cross product of the residuals testing for serial correlation up to 5 lags.
MKT=Market Index; CONV=Convertible Arbitrage; DEDS=Dedicated Short Bias; DRIV=Event Driven; EMGM=Emerging Markets; FIAR=Fixed Income Arbitrage; GLMA=Global Macro; LSEQ=Long-Short Equity; FUTR=Managed Futures; MSTR=Multiple Strategy; NEUT=Market Neutral; PWA=Price Weighted Index.

Table 2 reports the results from the Vector EGARCH model. The alphas cease to be significant indicating that the earlier finding may have been spurious. The parameters describing the variance covariance matrix are highly significant confirming time variation in second moments and cross-moments.

Table 2. Vector EGARCH results

$$R_{i,t} = \alpha_i + \beta_{i,t}R_{m,t} + \theta_i R_{i,t-1} + \varepsilon_{i,t}, \qquad R_{m,t} = \alpha_m + \varepsilon_{m,t}$$

$$\sigma^2[\varepsilon_{i,t}] = exp\{a_{i,0} + a_{i,1}(|z_{i,t-1}| - E|z_{i,t-1}| + c_i z_{i,t-1}) + b_i ln(\sigma^2[\varepsilon_{i,t-1}])\}$$

$$\sigma^2[\varepsilon_{m,t}] = exp\{a_{m,0} + a_{m,1}(|z_{m,t-1}| - E|z_{m,t-1}| + c_m z_{m,t-1}) + b_m ln(\sigma^2[\varepsilon_{m,t-1}])\}$$

$$\sigma_{i,m,t} = (\rho_{i,m} + \gamma_i R_{t-1}^{-})(\sigma^2[\varepsilon_{i,t}]\sigma^2[\varepsilon_{m,t}])^{1/2}$$

Fund type	$a_{i,0}$	$\beta_{i,t}$ average, minimum maximum	θ_i	$a_{i,0}$	$a_{i,1}$	c_i	b_i	$\rho_{i,m}$	γ_i	$Q^2(5)$ $Q_{i,m}(5)$
CONV	0.030 (0.96)	0.030 0.008 0.177	0.405 (8.89)**	-0.052 (-3.85)**	0.317 (9.23)**	-0.153 (-0.29)	0.998 (57.09)**	0.097 (2.27)**	0.025 (0.45)	5.93 6.57
DEDS	-0.294 (-0.03)	-0.339 -2.394 -0.087	0.016 (0.36)	1.308 (10.29)**	0.554 (11.8)**	-0.450 (-5.67)**	0.360 (6.33)**	-0.485 (-11.35)**	0.111 (2.53)**	8.96 1.61
EMGM	0.024 (0.35)	0.160 0.050 0.350	0.286 (5.83)**	0.416 (3.89)**	0.196 (1.88)*	-0.962 (-1.60)	0.962 (15.96)**	0.364 (7.80)**	-0.106 (-1.82)*	8.10 5.12
NEUT	-0.065 (-1.81)*	0.030 0.007 0.142	0.032 (0.61)	-0.101 (-1.35)	0.214 (17.8)**	-0.356 (-6.11)**	0.906 (34.42)**	0.060 (1.27)	0.042 (0.06)	2.56 0.12
DRIV	0.056 (1.43)	0.054 0.014 0.173	0.174 (3.03)**	-0.015 (-1.56)	0.196 (4.49)**	0.067 (0.72)	0.906 (59.51)**	0.120 (2.10)**	0.116 (1.67)*	0.33 5.83
FIAR	-0.085 (-5.92)**	0.028 0.000 0.171	0.256 (4.10)**	-0.003 (-1.28)	0.015 (5.59)**	-0.096 (-4.41)**	0.966 (26.20)**	0.019 (0.33)	0.142 (2.21)*	4.94 3.31
GLMA	0.032 (0.66)	0.053 0.020 0.145	0.132 (2.56)**	0.007 (0.62)	0.250 (4.39)**	-0.012 (-0.98)	0.998 (37.50)**	0.130 (2.46)**	0.002 (0.03)	8.32 7.11
LSEQ	0.002 (0.052)	0.120 0.060 0.205	0.260 (6.53)**	-0.048 (-3.22)**	0.040 (1.98)**	-0.275 (-1.66)	0.947 (44.41)**	0.355 (7.66)**	0.127 (2.49)**	5.82 4.08
FUTR	0.208 (2.94)**	0.013 -0.000 0.051	0.062 (1.48)	0.027 (1.23)	0.070 (2.23)**	-0.413 (-1.27)	0.935 (34.71)**	0.040 (0.878)	-0.040 (-0.77)	7.26 3.38
MSTR	0.055 (3.32)**	0.041 0.011 0.156	0.346 (5.84)**	-0.020 (-3.21)	0.127 (7.20)**	-0.319 (-4.14)**	0.962 (38.92)**	0.143 (2.96)**	0.087 (1.44)	4.73 5.07
PWA	0.005 (0.18)	0.035 0.016 0.107	0.285 (5.06)**	-0.228 (-2.83)**	0.188 (5.38)**	-0.077 (-0.76)	0.924 (15.44)**	0.164 (3.33)**	0.008 (0.14)	1.69 6.95
Equally-Weighted Average	0.001 (0.05)	0.022 0.011 0.108	0.295 (5.97)**	-0.250 (-3.24)**	0.215 (5.44)**	-0.061 (-0.63)	0.913 (15.05)**	0.107 (1.97)**	-0.002 (-0.04)	0.28 2.81

Notes: (**) and (*) indicate statistical significance at the 5% and the 10% levels respectively. Numbers in parentheses are t-statistics. $Q^2(5)$and $Q_{i,m}(5)$ are the Ljung-Box statistics for the squared residuals and the cross product of the residuals testing for serial correlation up to 5 lags.

MKT=Market Index; CONV=Convertible Arbitrage; DEDS=Dedicated Short Bias; DRIV=Event Driven; EMGM=Emerging Markets; FIAR=Fixed Income Arbitrage; GLMA=Global Macro; LSEQ=Long-Short Equity; FUTR=Managed Futures; MSTR=Multiple Strategy; NEUT=Market Neutral; PWA=Price Weighted Index.

Table 3. Beta dynamics

$$\boldsymbol{\beta}_{i,t} = \boldsymbol{g}_{i,0} + \boldsymbol{g}_{i,1}\boldsymbol{\beta}_{i,t-1} + \boldsymbol{g}_{i,2}\boldsymbol{R}_{m,t} + \boldsymbol{g}_{i,3}\boldsymbol{R}^{+}_{m,t} + \boldsymbol{v}_{i,t}$$

Fund type	$g_{i,0}$	$g_{i,1}$	$g_{i,2}$ x 10^2	$g_{i,3}$ x 10^4
CONV	0.006 (4.91)**	0.790 (23.36)**	-0.061 (-0.15)	0.097 (0.18)
DEDS	-0.308 (-11.43)**	0.123 (2.28)**	0.875 (1.26)	0.911 (0.76)
EMGM	0.094 (10.91)**	0.502 (11.17)**	0.564 (4.01)**	-1.354 (-5.52)**
NEUT	0.003 (2.79)**	0.865 (31.30)**	-0.033 (-0.87)	0.061 (0.93)
DRIV	0.022 (7.43)**	0.566 (12.50)**	-0.041 (-0.06)	0.069 (0.61)
FIAR	0.018 (7.32)**	0.149 (2.80)**	-0.202 (-2.55)**	0.438 (3.32)**
GLMA	0.007 (4.93)**	0.893 (37.36)**	0.897 (37.36)**	-0.144 (3.32)**
LSEQ	0.046 (8.13)**	0.620 (14.32)**	0.038 (0.55)	-0.069 (-0.57)
FUTR	0.015 (11.38)**	0.139 (2.63)**	0.123 (3.15)**	-0.320 (-4.75)**
MSTR	0.005 (3.298)**	0.846 (29.08)**	-0.0365 (-0.82)	0.100 (1.30)
PWA	0.011 (7.37)**	0.725 (19.68)**	0.023 (1.19)	-0.089 (-2.66)**
Equally-Weighted Average	0.0100 (8.75)**	0.596 (13.71)**	0.078 (0.44)	-0.063 (-2.04)**

Notes: (**) and (*) indicate statistical significance at the 5% and the 10% levels respectively. Numbers in parentheses are t-statistics. MKT=Market Index; CONV=Convertible Arbitrage; DEDS=Dedicated Short Bias; DRIV=Event Driven; EMGM=Emerging Markets; FIAR=Fixed Income Arbitrage; GLMA=Global Macro; LSEQ=Long-Short Equity; FUTR=Managed Futures; MSTR=Multiple Strategy; NEUT=Market Neutral; PWA=Price Weighted Index.

The parameter c_i is negative and significant suggesting that volatility is asymmetric in the sense that it is higher following negative returns than it is following positive returns. The estimated beta coefficients are time varying and the range of values is typically very large. This shows that models assuming constant market exposure are misspecified. The time-series behavior of the estimated betas is explored using the following regression:

$$\beta_{i,t} = g_{i,0} + g_{i,1}\beta_{i,t-1} + g_{i,2}R_{m,t} + g_{i,3}R^{+}_{m,t} + v_{i,t}, \qquad (9)$$

where $g_{i,1}$ captures persistence, $g_{i,2}$, market timing and $g_{i,3}$ asymmetric timing. The results presented in Table 3 show that in all instances there is a high degree of persistence over time in the estimated betas. There is some evidence of successful market timing but in most cases that is more than offset by the asymmetric parameter on market timing which implies timing of the wrong type.

Overall, the evidence suggests that hedge funds do not exhibit superior skill in timing the market. There is however, evidence of security selection skill.

4 Conclusion

The evidence in this paper shows that the betas of several hedge fund style indexes are highly variable over time. This in turn renders the estimated alpha returns unreliable. In almost all cases volatility is asymmetric and the range of estimated betas is rather large. The degree of persistence is also very high. Both, the simple regression and the vector EGARCH model show no evidence of successful market timing, though there is positive evidence of superior security selection.

References

Brent, E.K., Hall, H.B. Hall, R.E. & Hausman, J.A. (1974). Estimation and inference in nonlinear models, *Annals of Economic and Social Measurement* 4, 653-666.

Brooks, C. & Kat, H.M. (2002). The statistical properties of hedge fund returns and their implications for investors. *The Journal of Alternative Investments*, Fall 2002, 26–44.

Cave A., Hubner G. & Sougne, D. (2012). The market timing skills of hedge funds during the financial crisis. *Managerial Finance* 38, 4–26.

Chen Y. & Liang B. (2007). Do market timing hedge funds time the market? *Journal of Financial and Quantitative Analysis* 42: 4, 827–856.

Fung H.G., Xu, X.E. & Yau, J. (2002). Global hedge funds: Risk, return and market timing. *Financial Analysts Journal* 58, 19–31.

Fuss, R., Kaiser, D.G. & Adams, Z. (2007). Value at risk, GARCH modeling and the forecasting of hedge fund return volatility. *Journal of Derivatives & Hedge Funds* 13, 2–25.

Ibbotson, R.G., Chen, P. & Zhu, K.X. (2011). The ABCs of hedge funds: Alphas, betas, and costs. *Financial Analysts Journal* 67, 15–25.

Kat, H.M. & Lu, S. (2002). An excursion into the statistical properties of hedge fund returns. Working Paper. the University of Reading.

Koutmos, G. & Booth, G. (1995). Asymmetric volatility transmission in international stock markets. *Journal of International Money and Finance* 14, 747–762.

Sharpe, W. (1994). The sharpe ratio. *The Journal of Portfolio Management*, Fall 1994, 49–59.

Stulz, R. (2007). Hedge funds: Past, present, and future. *Journal of Economic Perspectives* 21, 175–194.

Xin, L. & Shawky, A. (2013). The market timing skills of long/short equity hedge fund managers. Working Paper. Department of Economics at Albany.

Appendix

Table A1. Summary statistics

Fund type	MKT	CONV	DEDS	DRIV	EMGM	FIAR	FUTR	GLMA	LSEQ	MSTR
Mean	0.130	0.038	-0.131	0.067	0.065	-0.081	0.111	0.032	0.026	0.031
Median	0.272	0.129	-0.183	0.169	0.269	0.066	0.208	0.072	0.190	0.131
Maximum	13.037	6.029	15.636	4.096	4.588	9.462	3.744	6.956	2.908	5.564
Minimum	-18.390	-7.263	-10.598	-4.100	-6.332	-9.18	-8.231	-8.509	-7.002	-7.106
Std. Dev.	3.043	1.119	2.163	0.913	1.482	1.154	1.358	1.119	1.067	0.796
Sharpe Ratio	0.030	-0.001	-0.079	0.031	0.018	-0.105	0.053	-0.006	-0.012	-0.001
Skewness	-0.555	-2.032	1.097	-0.643	-1.025	-2.260	-1.130	-1.343	-1.877	-1.506
Kurtosis	8.575	20.734	12.403	5.995	5.707	41.090	8.079	19.440	11.353	30.441
Jarque-Bera	456.4	4675.9	1317.0	150.1	162.9	20781.7	436.5	3919.3	1184.7	10764.7
Probability	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sum	43.914	12.724	-44.362	22.689	22.246	-27.606	37.776	10.945	8.957	10.349
Sum Sq. Dev.	3131.1	423.4	1580.8	281.8	742.0	449.8	623.4	423.2	384.9	214.1
Observations	339	339	339	339	339	339	339	339	339	339

Notes: MKT=Market Index; CONV=Convertible Arbitrage; DEDS=Dedicated Short Bias; DRIV=Event Driven; EMGM=Emerging Markets; FIAI
GLMA=Global Macro; LSEQ=Long-Short Equity; FUTR=Managed Futures; MSTR=Multiple Strategy; NEUT=Market Neutral; PWA=Price Weig

Table A2. Cross-correlations

Fund type	MKT	CONV	DEDS	DRIV	EMGM	FIAR	FUTR	GLMA	LSEQ	MSTR	NEUT	PWA
MKT	1.000	0.278	-0.544	0.235	0.501	0.265	-0.133	0.315	0.652	0.296	0.098	0.309
CONV	0.278	1.000	0.039	0.471	0.465	0.529	-0.077	0.384	0.422	0.641	0.191	0.673
DEDS	-0.544	0.039	1.000	-0.060	-0.300	-0.024	-0.104	-0.099	-0.338	-0.019	0.072	0.061
DRIV	0.235	0.471	-0.060	1.000	0.410	0.298	0.156	0.257	0.475	0.494	0.248	0.677
EMGM	0.501	0.465	-0.300	0.410	1.000	0.364	0.133	0.438	0.658	0.466	0.231	0.735
FIAR	0.265	0.529	-0.024	0.298	0.364	1.000	-0.142	0.340	0.371	0.416	0.148	0.528
FUTR	-0.133	-0.077	-0.104	0.156	0.133	-0.142	1.000	0.194	0.046	0.001	0.050	0.325
GLMA	0.315	0.384	-0.100	0.257	0.438	0.340	0.194	1.000	0.374	0.427	0.099	0.617
LSEQ	0.652	0.422	-0.338	0.475	0.658	0.371	0.046	0.374	1.000	0.356	0.173	0.637
MSTR	0.296	0.641	-0.019	0.494	0.466	0.416	0.001	0.427	0.356	1.000	0.349	0.689
NEUT	0.098	0.191	0.072	0.248	0.231	0.148	0.050	0.099	0.173	0.349	1.000	0.457
PWA	0.309	0.673	0.061	0.677	0.735	0.528	0.325	0.617	0.637	0.689	0.457	1.000

Notes: MKT=Market Index; CONV=Convertible Arbitrage; DEDS=Dedicated Short Bias; DRIV=Event Driven; EMGM=Emerging Markets; FIAR=Fixed Income Arbitrage
GLMA=Global Macro; LSEQ=Long-Short Equity; FUTR=Managed Futures; MSTR=Multiple Strategy; NEUT=Market Neutral; PWA=Price Weighted Index.

A MODERN ECONOMETRIC ANALYSIS OF AN ANCIENT EXCHANGE RATE MARKET

Richard T. Baillie, G. Geoffrey Booth
Michigan State University

Sanders S. Chang
University of Dayton

1 Introduction

McCloskey (1976) cogently argues that history provides a plethora of varied economic facts and circumstances to test plausible propositions that may lead to better economic theory and policy, and, hence, better economists. We embrace this argument and suggest that an understanding of economics may lead to a better interpretation of historical events and their sociological implications, a suggestion supported by Granovetter (1985), Greif and Laitin (2004) and Greif (2006) and others. Despite the apparent reasonableness of both these assertions, the amount of empirical work using modern econometric methods to analyze pre-modern data is relatively sparse. We believe that this discrepancy is not due to the lack of interest but rather to the lack of data, which may be a result of information not being collected, irregularly collected, or carefully recorded but misplaced or destroyed.

A notable exception to the above is the local exchange rate dataset housed in the Archivo di Stato di Firenze (State Archive of Florence) and made available by Bernocchi (1974). This dataset contains the daily observations of the gross and net exchange rates between the florin (a gold coin) and the denaro (a silver/copper coin).[1] The exchange rates were set at the end of the day using a rubric devised by the Arte del Cambio, Florence's banking guild. This protocol incorporated the

[1] First minted in 1252, the florin was the first pure gold coin to appear in significant quantities in Western Europe following of the Dark Ages. In the period covered by the dataset there was no significant change in its metal content or design. The original petty coins were the denaro and the quattrino, with four denari equaling one quattrino by government edict. The petty coins consisted of an alloy of silver (less than 50%) and copper (a mixture called billion). Because of gradual debasement over time (e.g., an average 0.8% annually from 1252 to 1500 (Spufford, 1988, p. 291)), the denaro fell into disuse for everyday transactions, although it remained a unit of account for record keeping purposes.

economic beliefs of the guild's members concerning the value of the rates for the next day and was used for all member transactions that occurred prior to the next setting. Thus, the exchange rates could and did vary greatly over time, although oftentimes there were extended periods in which they did not change. The data were initially recorded by various guild members (their names were often written on the original documents) and cover the 44 year period beginning January 4, 1389 and ending February 11, 1432 for a total of 10,741 daily observations.

To our knowledge, at the present time only two studies have explored these data using the modern notion that these exchange rates, like the prices of most financial assets, are a result of their market's infrastructure and its method of impounding relevant pricing information.[2] Specifically, Booth & Gurun (2010) document that the florin-denaro exchange rate bid-ask spread is determined, at least in part, by the exchange rate's previous implied volatility and, following the current market microstructure paradigm, this volatility is positively associated with the amount and diversity of information known by the guild members. In a follow-up study, Booth & Chang (2013) show that the way in which the exchange rate is determined results in an evolutionarily stable Nash equilibrium with a trajectory consistent with the manner in which information most likely became embedded in the Florentine financial market.

Both papers, however, limit their statistical analysis to only what is necessary to support their hypotheses. The purpose of this paper, therefore, is to more thoroughly investigate the time series properties of the Florentine local exchange rate data and the related explanatory abilities of the associated models. In particular, we focus our attention on the behavior of the commission series, which is the difference between the gross and net exchange rates.[3] First, we conduct a series of diagnostic tests to assess the orders of integration and cointegration that are suggested by the data. We then estimate a variety of models that seek to capture various features of the commission, including possible nonlinear dynamics and long memory in the series. Finally, we compare the forecasting performance of various models, both in sample and out of sample.

We find that the commission series, which is defined as half the bid minus the ask rate, is characterized by slow hyperbolic decay in its autocorrelation function that is consistent with a long memory process rather than either a short memory (i.e., exponential decay) or unit root process. Thus, improvements in modeling the se-

2 For a review of the general literature, see Madhavan (2000) and Bias, Glosten & Spatt (2005). For exchange rate applications see Lyons (2001) and Dominguez & Panthaki (2006).

3 As we discuss in fn. 9, the commission is also equal to one-half the bid-ask spread.

ries are obtained over the standard, short memory autoregressive moving average (ARMA) models when also taking into account fractional integration (ARFIMA) in conjunction with a generalized autoregressive conditional heteroskedasticity (GARCH) model for error variance process. In terms of forecasting, we find that the ARMA and ARFIMA classes of models perform comparably well, but accounting for nonlinearities in the commission series leads to increased forecast accuracy in terms of root means square error (RMSE).

The remainder of this paper is divided into four sections. Section 2 provides the Florentine economic and cultural background during the Italian Renaissance with particular attention to the span of time defined by our data. Section 3 discusses the way in which the exchange rate was determined and reported. Section 4 presents the various statistical models that we use and the relevant estimation results, along with statistics assessing the relative explanatory and forecasting power of the estimated models. The final section contains concluding remarks, including suggestions for future research.

2 Economic and Cultural Environment

Located on the banks of the River Arno, in what is now northern upper-central modern Italy, Florence, a city-state, became an independent republic in 1115 following the death of Countess Matilda of Tuscany and remained so (at least in name) until 1532 when it was conquered and turned into a monarchy by the then-reigning pope.[4] This 417 year period witnessed numerous regional wars and foreign invasions, political intrigues and conspiracies, devastating bouts of disease and pestilence, and religious upheavals as most (in)famously evinced in the 1490s by Girolamo Savonarola and his "Bonfire of the Vanities".[5] Concomitant with these events, Florence also made major contributions to the artistic and cultural rebirth of humanity based on early Greek and Roman scholarship and art that historians have dubbed the Renaissance. Examples of these cultural contributions are legion with Dante Alighieri's (1265–1321) Divine Comedy, Leonardo da

[4] At the time of her independence Florence was over a thousand years old. According to Osmond (2000), her name is attributed to Julius Caesar, who named the city Fliorentia in honor of Consul Fliorinus.

[5] In an effort to direct the Florentines away from what he considered anti-Christian behavior, especially the desire for worldly goods, Savonarola, a Dominican friar who at the time had religious and political control of Florence, ordered that each household provide objects of value to the government as a sacrifice. These were collected and burned in an effort to appease a supposedly angry Christian God. As a postscript, shortly thereafter Savonarola was forcefully removed from his position and, ironically, publicly executed by burning at the stake.

Vinci's (1452–1519) The Last Supper and Vitruvian Man, and Niccolo Machiavelli's (1469–1527) The Prince serving as enduring and well known examples of literature, art, science and political theory, respectively.

These intellectual accomplishments were made possible by an economy based on international trade, steeped in entrepreneurship and bolstered by a strong banking system. Padgett (2010), among many others, points out that the Florentines invented many of the modern business practices that are still in use today, including such innovations as the partnership system, limited liability, and double-entry accounting. Commercial mathematics was taught in "reckoning" schools under the vigilance of a "reckoning master," and the city-state was the home to more than a dozen of these schools (Swetz 1987). According to Van Egmond (1976: 229) their curricula typically covered the topics found in modern secondary education. Instruction in the Florentine monetary system and its relationships to its foreign counterparts was also taught. After all, as pointed out by Cippola (1956), the florin and the Venetian ducat were the "global" currencies of their day and not only were they often exchanged for one another but also for pounds sterling (London), livre tournois (Paris), besans (Rhodes and Tunis) and carlins (Naples), among many others. Thus, needed banking skills included not only straightforward arithmetic and algebra but also more arcane (by modern standards) operations such as vigesimal and duodecimal divisions (Van Egmond 1976: 128). As a result, Florence became known as the birthplace of financial capitalism and a caretaker of what McCloskey (2006) calls "bourgeoisie virtues".

The economy was structured around 21 guilds with seven being classified as major or the most important. Membership in at least one guild was required to be an active voting citizen of the city state and those few who failed to join were shunned. The Arte del Cambio belonged to this elite group along with three guilds involved in the buying, selling and manufacturing of foreign, woolen and silk cloth. Staley (1906: 172) suggests that the financial needs of the latter three guilds was influential in the establishment and growth of Florence's banking guild. Once started, however, the banking industry took on a life of its own. Not only did some of the banks support Florentine businesses but also they became bankers to the pope and designed new financial instruments to promote trade and loans throughout Europe, the Levant and the Maghrib. Staley (1906: 174) points out that banks and the members of these three guilds often competed with each other in both the goods and the financial services markets

The Arte del Cambio was a self-regulating body. Not only was it a business organization; but also it was a secular social unit. Pagett (2001) posits that during the period of our analysis banks were organized as an extended family, which included blood relatives, in-laws, and amiciza (friends). Obtaining membership required passing a rigorous exam that included assessing the applicant's character and paying a substantial initiation fee. Guild membership, then, indicated to the external community that the individual was honest, financially sound, and committed to obeying all of the guild's rules. The rules pertained to local banking activities with international activities governed the relevant convention. The penalties for disobedience could be very harsh by today's standards, with the use of the rack and the strappado (both being disfiguring acts of torture) not being unknown.[6]

Banks entered and exited the market on a regular basis and the life spans of their individual businesses varied greatly. According to de Roover (1963: 16), 57 banks were in operation in Florence in the middle of the 14th century, and this number increased to 71 by the century's end. By the middle of the 15th century, however, the number of banks had been halved and remained relatively stable over the next 50 years. There were two kinds of banks. One type, the banco grosso (large bank), engaged in interregional transactions and often, according to de Roover (1968), profitably speculated on exchange rates, an assertion confirmed by Booth (2009). They also engaged in money-changing on the local level. The other type, banco a minuto (small bank), focused primarily on local money-changing and related activities.

Banks that engaged in money-changing ranged in size from a one-person operation to one having several partners served by a clerical staff of a half dozen or more. These money-changers were typically open for business daily except for Sunday with exceptions made for holidays, inclement weather and epidemics. Money-changing typically occurred in four different plazas in which people often congregated for shopping or public events. Transactions took place either in one of the plazas or the bank's own building on the plaza's edge. The money-changers were easy to spot as they sat behind a bench (banco) covered by a green cloth on which were placed stacks of coins of various values and origins as well as an account book. They acted as dealers and were always ready to physically exchange coins, make transfer entries in their account books, and accept and record deposits. Transactions that involved more than one money-changer were periodically netted out by the money-changers involved. Money-changers were familiar with

[6] In Dante's epic poem, *The Divine Comedy* (Alighieri, 1300), which employs many references to Florence and its local environs, individuals committing financial misdeeds were not treated kindly. Usurers were relegated to the seventh circle of hell (Canto XVII) while counterfeiters resided for eternity in the ninth circle (Canto XXX). In the allegory the counterfeiter is immolated by the local populace for his deeds.

the quality of the coins and were able to tell whether their physical condition was worse than that associated with normal wear and tear. They could also detect counterfeit coins and coins that had been tampered with. Detailed written records were kept and archived in locked strongboxes along with coins held in reserve, and in the case of a dispute these records were used by the guild manager to assist in its resolution.

There was typically an active demand for the money-changer’s services. By statute the members of some guilds transacted their business and kept their accounts in florins, while members of other guilds reckoned in petty coins. Wholesale prices were usually stated in gold and retail prices in silver. Interregional transactions, however, were generally conducted in gold. Although wages were usually paid in silver regardless of the type of business, Spufford (1988: 335) points out that ordinary people used both gold and silver coins. Thus, money-changing was a necessary part of business and everyday life.

The number of coins in circulation was not regulated by the Florentine government, but city leaders influenced the amount indirectly by deciding the silver content of the petty coins. Thus, if the monetary value of the small change was higher than its commodity value, bullion was minted into coin at the city’s mint. If the value was lower, coins were melted into bullion. This behavior follows Sargent and Velde’s (2002: 37) model that postulates that for two coins to circulate freely after accounting for costs of minting, the exchange rate between the two coins cannot be greater (less) than the exchange rate value that aligned the melting (minting) point of one coin with the minting (melting) point of the other.

3 Exchange Rate Determination[7]

At the end of each trading day, the new official denaro-florin exchange rate and the trading commission for the next trading day were established and announced in the four markets located in the city. These values were used by all money-changers, who acted as a single economic unit, for all transactions throughout the next trading day. Relying on an archival document dated May 21, 1492, which reports a discussion held by the Guardians of the Republican Mint, Bernocchi (1974) reports that the official rate was determined in a two-step process. First, the money-changers individually and simultaneously submitted to the guild man-

[7] This section is largely based on C.C. Graig’s Italian to English translation of the relevant passages of Bernocchi (1974) and Booth & Chang (2013).

ager their best individual estimates of the next day's "true" price of the florin in terms of denari. Second, the guild manager then set the price equal to the arithmetic average of the submitted estimates.[8] This average rate was referred to as the gross exchange rate. Because it was not mandatory that a money-changer submit a price, the number of estimates submitted and, hence, included in the calculation of the average could range from one to all the guild members. Bernocchi (1974) asserts that this protocol had been in effect since at least the 14th century.

At the same time the guild recorded the gross rate it also recorded the net rate, with the difference between the gross and net rates being the commission. Bernocchi (1974) does not precisely describe how the commission on net rate is determined. However, the way that the gross and net rates are recorded in the State Archive suggests that the money changers likely thought in terms of the rates, whether they be bid, ask or the midpoint between the two, so that the commission was the residual in their calculation.[9] Regardless, a money-changer who submitted an estimate undoubtedly had a commission value in mind. This commission undoubtedly reflected not only profits but costs attributed to adverse information (i.e., the cost of exchanging money with a counterparty more informed than the collective level of knowledge exhibited by the guild members), processing orders, and maintaining an inventory of coins.[10] In our empirical exercise it does not matter whether the money-changer supplied this information to the guild as an estimate of a commission that when combined with the gross rate resulted in the net rate or the net rate itself.

The portion of profits that accrued to an individual member was then determined by his market share, which in turn may have depended on factors such as the money-changer's place of business, customer relationships, and other attributes related to non-price competition. In the context of Vickery's (1961) taxonomy, the guild acted as an "exclusive marketing agency" that attempted to determine

---

8 This rate setting protocol is similar to several pricing schemes that were recently or are currently being used, e.g., the British Bankers Association determination of London interbank offer rate (LIBOR) and the London Bullion Market Association's setting of the gold forward offer rate (GOFO).

9 The ask rate is the price at which the money-changer sells one florin in terms of denari and the bid rate is the price at which he buys one florin. The mid-point rate, then, is often thought to be the true price.

10 The commission was thought by the Roman Catholic Church and the economic thinkers of the day to be compensation for performing a service that included a risk factor and a charge could be levied to compensate for the bearing of risk. Thus, charging a commission was not usury as long as the transaction was truly speculative. Houkes (2004) provides numerous references in this regard.

the equilibrium exchange rate by aggregating the supply and demand information provided by its members.

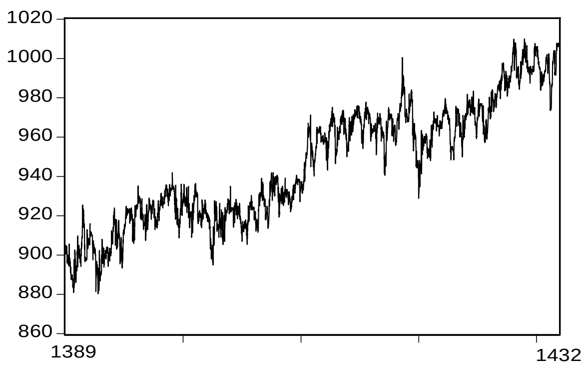
4 Statistical Models and Results

Plots of the daily gross exchange rates, the exchange rate returns, and the commissions are displayed in Figure 1. The exchange rates, which are expressed denari per florin, are actual data and the returns are continuous being the first difference in the natural logarithm of the exchange rate series. The commission is the difference between the gross and net exchange rates and is equal to one-half the bid-ask spread. A review of this figure indicates a strong upward trend in the exchange rate indicating that the aforementioned debasement of the denaro was present during our sample period.[11] The returns appear trendless, aperiodically clustered, and are sometimes subject to abrupt movements. The commission seems to slightly increase over time and is characterized by frequent positive spikes. Its most common value is two denari per florin and this value occurs almost 25 percent of the time. On occasion, however, commissions exceeded 12 denari.

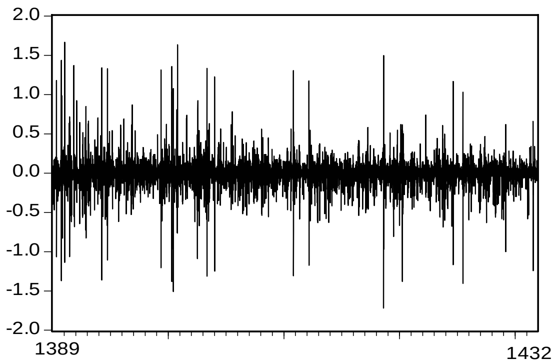
Since the aim of this paper is to further explore and understand the properties of the commission in the ancient denaro-florin market, which may give a better insight into the economics of how the commission was set by the money-changers guild, in what follows we discuss and present results from various diagnostic tests, modeling approaches, and forecasting exercises related to the commission series.

---

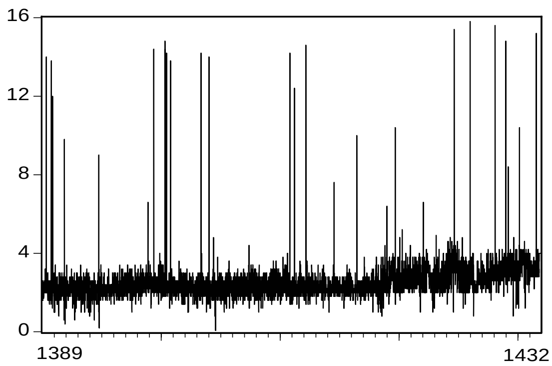
[11] The quotations were recorded in soldi and denari with one soldo equaling 12 denari. The tick size was one-fifth of a denaro. Prior to February 1418, the exchange rate supplied to the guild was expressed in quattrini. See fn 1. Subsequently, and most likely because of debasement, the money-changers submitted their exchange rate quotes estimates in grossi, coins that contained more silver that the quattrini.



(a) Gross Exchange Rates



(b) Gross Returns



(c) Commissions

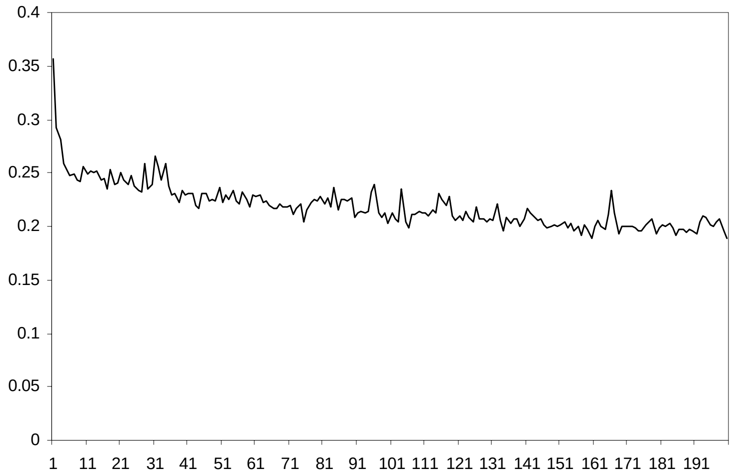
Figure 1. Daily denaro-florin (a) gross exchange rates, (b) gross returns and (c) commissions, 1389–1432. Note: Gross returns are defined as the first difference of the logarithm of the daily denaro-florin gross exchange rates. Daily commissions, expressed in denari per florin, are the difference between the gross and net exchange rates and are equivalent to one-half the bid-ask spread. Source: Bernocchi (1974).

4.1 Diagnostic Tests

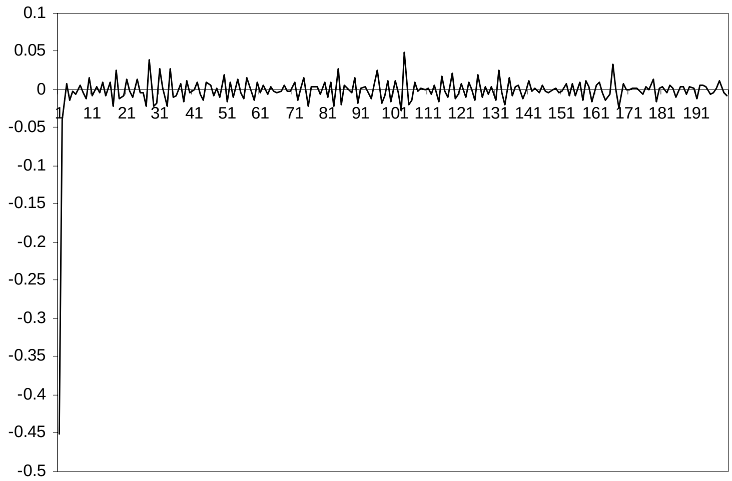
Figure 2 shows the autocorrelation function (ACF) of the denaro-florin commission. Visual inspection of this figure shows that the ACF exhibits very slow hyperbolic decay. This indicates that the commission is likely not an $I(1)$ unit root process (i.e., integrated to the first order), but instead is more likely a long memory $I(d)$ process, where d is the fractional differencing parameter (i.e., fractional integration) that can either be stationary or nonstationary. Figure 2 also shows the ACF for the first-differenced commission. Notably, first differencing induces a large and highly significant negative first-order lag of -0.45, with all other lags being insignificant, which is a classic sign of over-differencing. This result thus furthers our suspicion that the commission is likely a fractionally integrated process.

To explore these issues related to the order of integration further, we perform various diagnostic tests on the commission series. The Augmented Dickey-Fuller (ADF) test soundly rejects the null hypothesis of a unit root (p-value = 0.0000), as does the Phillips-Perron test (p-value = 0.0001). Thus, there does not appear to be a unit root in the commission series. At the same time, however, there is evidence that the series is nonstationary. The Kwiatkawski-Phillips-Schmidt-Shin (KPSS) test squarely rejects the null of stationarity (KPSS LM-statistic = 8.338). The results of the preceding tests point in the direction of fractional integration and nonstationarity. Since the commission is the difference between the gross and net rates, these results also imply that these two rates are fractionally cointegrated (i.e., a linear combination of the gross and net rates results in a process that is $I(d)$, where $0.5 < d < 0.5$ corresponds to stationary, while $0.5 < d < 1$ corresponds to nonstationary).[12] We discuss further the estimation of d below, but begin with establishing some alternative baseline models of the commission with which to make comparisons.

[12] Indeed, a Johansson test of cointegration rejects the null hypothesis of no cointegration (both the Trace and Maximum Eigenvalue tests indicate the presence of at least one cointegrating equation at the 0.05 level of significance). However, the Johansen test is based on an $I(1)$ vector autoregression (VAR) where some linear combinations are $I(0)$. In our case, the linear combination appears $I(d)$. It is common in finance for transformations of $I(1)$ variables to appear to be $I(d)$, e.g., spreads on interest rates, forward premiums, etc.



(a) ACF of the commission series



(b) ACF of the first-differenced commission

Figure 2. Correlograms of (a) the commissions and (b) the first-differenced commissions. The plots show the value of the autocorrelation function (ACF) for lags $k = 1,...,200$, with the ACF values on the y-axis and k lags on the x-axis.

4.1 Model Selection and Estimation

4.2.1 Baseline Models

As a first pass at modeling the commission series, we simply (and naively) employ autoregressive (AR(p)) and moving average models (MA(q)) models to derive baseline estimations, where p and q are lag orders. Given the evidence of a high degree of persistence in the commission, and recognizing that a normal trading week consisted of six trading days, we employ $p = q = 6$ lags for each model. Letting the demeaned commission at time t be denoted by y_t, the estimated AR(6) model is as follows:

$$\begin{aligned}\hat{y}_t = &\underset{(0.006)}{0.0005} + \underset{(0.033)}{0.224}\, y_{t-1} + \underset{(0.027)}{0.111}\, y_{t-2} + \underset{(0.017)}{0.103}\, y_{t-3} + \underset{(0.022)}{0.074}\, y_{t-4} + \underset{(0.014)}{0.077}\, y_{t-5} \\ &+ \underset{(0.014)}{0.085}\, y_{t-6} \\ &T = 10{,}465,\ F = 438.96\ (0.000),\end{aligned}$$

with Newey-West standard errors, which are robust to autocorrelation and heteroskedasticity (HAC), reported in parentheses below their corresponding parameter estimates. T denotes the effective sample size, while F represents the F-statistic used for testing the joint significance of all estimated coefficients (besides intercept), with its corresponding p-value in parentheses. The MA(6) estimation results are:

$$\begin{aligned}\hat{y}_t = &\underset{(0.016)}{0.0002} + \underset{(0.031)}{0.247}\, \varepsilon_{t-1} + \underset{(0.024)}{0.175}\, \varepsilon_{t-2} + \underset{(0.025)}{0.164}\, \varepsilon_{t-3} + \underset{(0.019)}{0.127}\, \varepsilon_{t-4} + \underset{(0.017)}{0.109}\, \varepsilon_{t-5} \\ &+ \underset{(0.014)}{0.084}\, \varepsilon_{t-6} + \varepsilon_t \\ &T = 10{,}471,\ F = 358.11\ (0.000),\end{aligned}$$

where the lagged ε_t terms are the residual moving average terms. For both models, all lags are highly statistically significant.

The above results from our naive baseline specifications are largely expected given the high degree of persistence in the commission. Indeed, increasing the number of lags to $p = q = 12$ lags in each model yields very similar results, with all lags highly significant (results not reported but available upon request). However, even with such a large number of lags, there still remains a large degree of autocorrelation in the residuals. The Ljung-Box Q-statistic for the residuals at 20 lags ($Q(20)$) for both models is highly significant, indicating the presence of remaining autocorrelation in the residuals (the Q-statistic tests the null hypothesis of no autocorrelation up to and including lag k and is Chi-square distributed with

k degrees of freedom). This suggests that simply adding more lags does not adequately account for the persistence of the commission (and violates the principle of parsimony). Below, we refine the estimation approach beyond the naive baseline specifications.

4.2.2 Autoregressive Moving Average (ARMA) Specification

We next select an appropriate model for the commission with a combination of autoregressive and moving average components in an ARMA(p,q) model with general form $\phi(L)y_t = \theta(L)\varepsilon_t$, where $\phi(L) = (1-\phi_1 L - ... - \phi_p L^p)$ is the autoregressive polynomial and $\theta(L) = (1-\theta_1 L - ... - \theta_q L^q)$ is the moving average polynomial with lag operator L. Following the well-known Box-Jenkins approach, our model selection criteria is based on the balancing of three aims: (1) achieving statistical significance in the estimated parameters, (2) removing residual serial correlation, and (3) parsimony.

As a first pass, we estimate an ARMA(1,1) model and find the following results:

$$\hat{y}_t = \underset{(0.0003)}{0.0002} + \underset{(0.001)}{0.998}\, y_{t-1} - \underset{(0.006)}{0.956}\, \varepsilon_{t-1} + \varepsilon_{t-1}$$

$$T = 10{,}470,\ F = 1{,}633.56\ (0.000),$$

which suggests the presence of a unit root in both the AR and MA polynomial. Further evidence of this comes from the ARMA(2,2) model:

$$\hat{y}_t = \underset{(0.375)}{0.2301} + \underset{(0.176)}{1.466}\, y_{t-1} - \underset{(0.176)}{0.466}\, y_{t-2} - \underset{(0.192)}{1.312}\, \varepsilon_{t-1} + \underset{(0.188)}{0.326}\, \varepsilon_{t-2} + \varepsilon_t$$

$$T = 10{,}469,\ F = 880.67\ (0.000),$$

where there appears to be an exact unit root in the AR polynomial of $\phi(L) = (1+1.466L - 0.466L^2)$ so that $\phi(1) = 1$ and a near unit root in the MA polynomial, consistent with the ARMA(1,1) results above. Thus the ARMA specifications appear to reduce to an AR(1) with unit root. Such results point further in the direction that the commission series is (1) fractionally integrated (or, more precisely, that the gross and net rates are fractionally cointegrated so that subtracting one from the other results in over-differencing, which is known to cause a unit root in a MA polynomial); and (2) nonstationary (which is suggested by the unit root in the AR polynomial). Thus, the ARMA class of models for short memory stationary processes is not well suited for the purposes of modeling the denaro-florin commission, for which evidence thus far suggests is a long memory and possibly nonstationary process.

4.2.3 Autoregressive Fractionally Integrated Moving Average (ARFIMA) models

Given the likely presence of long memory dynamics in the commission series as discussed above, we next apply the ARFIMA(*p*,*d*,*q*) model, which is specified as follows for the demeaned commission y_t:

$$\phi(L)(1-L)^d y_t = \theta(L)\varepsilon_t, \quad t = 1,...,T ,$$

where, again, *d* is real and represents the order of fractional integration, or long memory, and $(1-L)^d$ is the fractional difference operator defined by the binomial expansion:

$$(1-L)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-L)^j .$$

For $-1/2 < d < 1/2$, the process is stationary and invertible, and the ACF exhibits hyperbolic decay, while for $1/2 \le d \le 1$, the process does not have finite variance but has finite cumulative impulse response weights (i.e., Wold decomposition, or infinite order moving average MA(∞) representation). Hence, the process is mean reverting for $d < 1$. With the assumption that $\varepsilon_t \sim NID(0, \sigma^2)$, estimation is conducted using the maximum likelihood (MLE) estimator (see Baillie (1996) for further discussion of ARFIMA theory and estimation).

In the spirit of Box-Jenkins, we find that the ARFIMA(2,*d*,1) model best balances the aims of minimizing residual serial correlation, achieving parameter significance, and parsimony. For the ARFIMA(2,*d*,1) model, the estimated long memory parameter (standard error) is $\hat{d} = 0.535$ (0.057); while $\hat{\phi}_1 = 0.510$ (0.069), $\hat{\phi}_2 = 0.074$ (0.040), and $\hat{\theta}_1 = 0.889$ (0.024). The estimated long memory parameter appears to be slightly in the nonstationary region, along the lines of our suspicions above. Also, the problem of unit roots in the AR and MA polynomials that occurred in the ARMA specifications above seems to be ameliorated, suggesting that upon fractional differencing, the remaining short memory component is well behaved (i.e., stationary and not over-differenced, and thus well estimated by ARMA). Lastly, in contrast to the baseline AR(*p*) and MA(*q*) models, the Ljung-Box statistic *Q*(20) is now insignificant for the ARFIMA(2,*d*,1) model, indicating that serial correlation in the residuals has been removed at 20 lags after fractional integration is taken into account (whereas it was still present to a very large degree in the purely short-memory baseline models above).

4.2.4 Volatility Models

Visual inspection of Figure 1 suggests that the commission exhibits time-varying volatility and volatility clustering, with small (large) changes in the commission tending to be followed by similarly small (large) movements. Upon examining the correlogram of the squared residuals of the ARFIMA(2,*d*,1) model estimated above and conducting a Lagrange Multiplier (LM) tests for autoregressive conditional heteroskedasticity (ARCH) effects on the residuals, there is evidence of remaining conditional volatility (these are not reported but available upon request). The decay in the correlogram of the squared residuals indicates normal, exponential decay so that there does not appear to be further long memory in the volatility process.

To explore the volatility dynamics of the commission further, we employ the Generalized ARCH, or GARCH(p,q) class of models of Bollerslev (1986) in which the error variance process is specified as:

$$\varepsilon_t = v_t \sqrt{h_t},$$

$$v_t \sim i.i.d.\,(0,1),$$

$$h_t = c + \alpha(L)\varepsilon_{t-1}^2 + \beta(L)h_{t-1},$$

where the equation for h_t represents the conditional heteroskedasticity process, which depends on p lagged squared innovations ε_{t-1}^2 and q values of the process itself, h_{t-1}.

Using MLE, the estimated coefficients (standard errors) of the ARFIMA(2,*d*,1)-GARCH(1,1) model are as follows: $\hat{d} = 0.551$ (0.050), $\hat{\phi}_1 = 0.5836$ (0.046), $\hat{\phi}_2 = 0.096$ (0.023), and $\hat{\theta}_1 = 0.912$ (0.014), with estimated volatility equation:

$$\underset{}{\hat{h}_t} = \underset{(0.059)}{0.335} + \underset{(0.025)}{0.070}\,\varepsilon_{t-1}^2 + \underset{(0.089)}{0.169}\,h_{t-1}$$

Thus, all estimated parameters in both the mean and variance equation are significant at conventional levels, and residual serial correlation is further reduced when comparing Ljung-Box Q-statistics across the models (results not reported but available upon request). Thus, it appears that an ARFIMA(2,*d*,1)-GARCH(1,1) is most adequate for modeling the denaro-florin commission series.

4.3 Forecasting Performance of Various Alternative Models

In this section, we compare the forecasting performance of the various models. In addition to the ARMA and ARFIMA class of models described above, we employ several other specifications to broaden the comparison. First, we discuss the various models that are considered for forecasting purposes. Then we present the results of various forecasting exercises.

4.3.1 Alternative Forecasting Models

4.3.1.1 ARFIMA Forecasts

For the ARFIMA-based forecasts, we use the estimated ARFIMA(2,*d*,1) model to compute in-sample RMSE, but use a fractional white noise ARFIMA(0,*d*,0) model to compute out-of-sample forecasts. As discussed in Baillie (1996), there are complications involved when using ARFIMA models to make predictions. Generally, forecasts are obtained from the infinite order autoregressive AR(∞) representation of the general ARFIMA(*p*,*d*,*q*) model, which is written as follows:

$$y_t = \sum_{j=1}^{\infty} \pi_j(d) + \varepsilon_t ,$$

where

$$\pi_j(L) = (1-L)^d \phi(l)\theta(L)^{-1} .$$

Predictions are then made using a truncated version of this representation at k lags. The form that these autoregressive parameters actually take are not readily available and involve complex combinations when there are several AR and MA terms, and there are other unresolved issues regarding the truncation lag because there is no finite state-space representation for ARFIMA models (Baillie, 1996).

Thus, for simplicity, we use instead the fractional white noise process specified as:

$$(1-L)^d y_t = \varepsilon_t, \quad t = 1,...,T ,$$

where the j-th autoregressive coefficient in the AR(∞) representation is well known and given by:

$$\pi_j(d) = \{(-d)(1-d)(2-d)...(j-1-d)\} / j!.$$

The fractional white noise model has an estimated fractional difference parameter (standard error) of $\hat{d} = 0.216$ (0.021). For this specification, $Q(20) = 62.233$ (so, as would be expected without ARMA terms for the short-memory component of the series, serial correlation is still present in the residuals). For our forecasting exercises, we use $j = 25$ lags.

4.3.1.2 Forecasts with Volatility Effects

In many economic applications, volatility effects in the variance equation influence the dynamics of the mean equation. For example, when financial market volatility is high, asset returns tend to be negative, a phenomenon known as the leverage effect. To examine whether a similar form of dynamics is present in the denaro-florin commission series and whether this may be useful for forecasting, we estimate a GARCH-in-mean model:

$$\begin{aligned}
\hat{y}_t &= \underset{(0.146)}{0.2339} + \underset{(0.186)}{0.392} \log(\hat{h}_t) + \underset{(0.032)}{0.281}\, \varepsilon_{t-1} + \varepsilon_t \\
\hat{h}_t &= \underset{(0.065)}{0.2530} + \underset{(0.133)}{0.307}\, \varepsilon_{t-1}^2 + \underset{(0.085)}{0.344}\, h_{t-1} \\
& T = 10{,}471,\ F = 152.71\ (0.000),
\end{aligned}$$

where $\log(\hat{h}_t)$ is the natural logarithm of the lagged generalized heteroskedasticity term. The positive GARCH-in-mean parameter suggests that when volatility in the commission increases, commissions themselves tend to rise as well. This result is consistent with Booth & Gurun (2008).

4.3.1.3 Nonlinearities and Seasonal Effects

We next examine whether there are any nonlinearities, including seasonal effects, in the dynamics of the commission series and whether these may be useful for forecasting purposes. First, we examine whether there are monthly effects by creating dummy variables for the months February through December. We then regress the commission on an intercept, the monthly dummies, the lagged commission, and the lagged commission interacted with each monthly dummy. The presence of a month effect would then manifest itself in either a significant shift in the intercept or slope of the fitted regression line, or both. After various specifications, the only significant monthly effect that we are able to detect is an interaction in the month of February:

$$\hat{y}_t = \underset{(0.010)}{0.0005} + \underset{(0.033)}{0.345}\, y_{t-1} + \underset{(0.058)}{0.307}\, feb_{t-1} * y_{t-1}$$

$$T = 10{,}470, = 0.130, F = 784.87\ (0.000),$$

where *feb* is a dummy variable that equals unity if the observation falls in February, and zero otherwise. The February effect is quite strong and indicates that the autoregressive term has twice the impact in February than it does in other months.

Recall that Figure 1 reveals that the commission series is punctuated by large and frequent spikes. Moreover, these spikes appear to all be in the positive direction, a result of the commission always being positive by design. Such dynamics suggest a form of nonlinearity that may be captured by the Threshold Autoregression (TAR) model of Tong (2010), which is a regime-switching model that switches discretely between two AR(*p*) processes. Specifically, letting I_t denote an indicator variable (threshold) that takes on the value of unity if the first difference of the gross rate $\Delta y_t > 0$, and zero otherwise, the first-order TAR(1) estimation results are:

$$\hat{y}_t = (\underset{(0.012)}{0.3085} + \underset{(0.022)}{0.901}\, y_{t-1}) * I_t + (\underset{(0.011)}{-0.3045} + \underset{(0.037)}{0.238}\, y_{t-1})(1 - I_t)$$

$$T = 10{,}470, F = 1{,}342.2\ (0.000).$$

In this model all coefficients (including intercepts) are significant and jointly significant at conventional levels and, furthermore, the corresponding coefficients in each regime are also significantly different from each other (results from Wald tests are not reported but available upon request). These results provide evidence for the presence of discrete regime switching in the commission series. Notably, for regimes characterized by declines in the commission, the AR process has a negative intercept with moderate autoregressive persistence. In contrast, in states characterized by a rising commission, the model intercept is positive and there is very strong persistence.

4.3.2 Comparison of Forecasting Results

The first column of Table 1 lists the 10 models that are used in the forecasting exercises and for which comparisons will be drawn. For each model, we compute the in-sample root mean-square error (RMSE) and the out-of-sample RMSE. Out-of-sample forecasts are computed by estimating the model using 8,500 observations (i.e., the training set), and then using the estimated model to forecast the remaining 1,971 observations. These forecasts are of two types: (1) static forecasts, which consist of a series of one-step-ahead forecasts using the actual lagged value once it is realized, and (2) dynamic forecasts, which use forecasted values to form new forecasts (i.e., n-step ahead forecasts, n = 1,...,1,971).

Table 1. Comparison of Forecasting Results.

Model	In-sample RMSE	Out-of-sample RMSE (Static)	Out-of-sample RMSE (Dynamic)
AR(6)	0.703	0.845	1.159
AR(12)	0.694	0.815	1.158
MA(6)	0.717	0.893	1.160
MA(12)	0.707	0.862	1.160
ARMA(2,2)	0.681	0.801	0.989
MA(1)-GARCH-in-mean(1,1)	0.754	0.990	1.130
TAR(1)	0.638	0.798	1.060
AR(1) with month effect	0.734	0.943	1.161
ARFIMA(2,*d*,1)	0.681		
ARFIMA(0,*d*,0)	0.687	0.854	1.062

The results of the forecasting comparison are reported in the next three columns of Table 1. The benchmark AR(6) model has an in-sample RMSE of 0.703. Adding autoregressive lags does not substantially improve the in-sample fit, nor does using an MA model. The GARCH-in-mean specification performs worse, as does the AR(1) with a month effect. The models that notably increase the in-sample goodness of fit are the ARMA(2,2), TAR(1), and ARFIMA models. Of these, in turn, the TAR(1) model noticeably has the lowest RMSE of 0.638. Thus, it appears that regime-switching provides the best in-sample predictions.

For static out-of-sample forecasts, the TAR(1) model again provides the best performance with an RMSE of 0.798, but now only negligibly so when compared to the ARMA(2,2) model (RMSE = 0.801). Notably, without the presence of further ARMA terms, the ability of the ARFIMA(*0,d,0*) model to make static one-step ahead forecasts cannot beat these two models, although it performs closely to the benchmark AR(6) model and outperforms the MA(6) and MA(12) models, as well as the GARCH-in-mean and month-effect models. Lastly, the model that gives the best dynamic out-of-sample forecasts is the ARMA(2,2) model, which is followed very closely by the TAR(1) and ARFIMA(*0,d,0*) models (the last two have very similar performance).

Thus, it would appear that for purely forecasting purposes, the TAR(1) model does the best all-around job. However, from a modeling perspective, the TAR(1) model is less than adequate as it does not remove the residual serial correlation at any lag, and specifying a higher order TAR model does not resolve this problem (results not reported but available upon request). In addition, for reasons discussed above, while the ARMA(2,2) provides good out of sample forecasts, it suffers from misspecification (i.e., there are unit roots in the lagged AR and MA polynomials). Thus, on balance, it appears that the ARFIMA class models are the most appropriate choice all around – they provide good forecasts and are the most sound from a modeling standpoint.

5 Concluding Remarks

The aim of this paper has been to further explore and understand the properties of the commission in the ancient florin market, which may give a better insight on how the commission was set by the money-changers guild. First, we conducted a series of diagnostic tests to assess the orders of integration and cointegration and found that the commission series is best characterized as a long memory process. We then estimated a variety of models and found that the ARFIMA(2,*d*,1)-GARCH(2,2) model provides the best fit using the Box-Jenkins approach. However, there are regime-switching effects that are worth future examination. Finally, while the TAR(1) and ARMA(2,2) models yields the best out-of-sample forecasting results, these are not entirely sound models from an estimation standpoint. On the other hand, the ARFIMA-type models are theoretically sound and deliver nearly comparable forecasting results.

Further work along these lines, however, can and should be done in computing forecasts for higher-order ARFIMA(*p*,*d*,*q*) models and considering aspects of long memory in conjunction with regime switching. Directions for this work can be found, for instance, in Ballie & Kapetanios (2008), Baillie & Morana (2010) and Gross-Klussman & Hautsch (2011). Moreover, in addition to finding models that provide more accurate forecasts of the Florentine commission, efforts should be directed toward understanding the economics underlying these statistical processes. Important examples of such work include LeBaron & Yamamoto (2007, 2008) and Feng, et al. (2012), among others. Moreover, the economics should be consistent with the notions espoused by Granovetter (1985), Grief & Laitin (2004) and Grief (2006). We save these tasks for future research and note that this research will not only provide insights into the internal workings of an ancient financial market but also into the behavior of modern ones.

Acknowledgements: We thank Cynthia C. Craig for her expert Italian to English translation of excerpts from Bernocchi (1974), William R. Lane for pointing out its existence, James R. Booth for his research assistance in the very early stages of this project, and Elizabeth B. Booth and Ryan Timmer for their able editorial assistance. All errors are our own.

References

Alighieri, D. (circa 1300). *The Divine Comedy – Inferno.* Trns. Henry Wadsworth Longfellow, 1909. www.online-literature.com/dante/inferno.

Baillie, R.T. (1996). Long memory processes and fractional integration in economics. *Journal of Econometrics* 73, 5–59.

Baillie, R.T., & Kapatenios, G. (2008). Nonlinear models for strongly dependent processes with financial applications. *Journal of Econometrics* 147, 60–71.

Baillie, R.T. & Morana, C. (2010). Modeling long memory and structural breaks in conditional variances: An adaptive FIGARCH approach. *Journal of Economic Dynamics and Control* 33, 1577–1592.

Bernocchi, M. (1974). *Le Monete della Repubblica Florentina*, Vol. 4, Documentazione Arte e Archeologia, Studi e Document 11. Firenze, Italy: L.S. Olschki.

Biais, B., Glosten, L.R. & Spatt, C. (2005). Market microstructure: a survey of microfoundations, empirical results, and policy implications. *Journal of Financial Markets* 8, 217–264.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Booth, G.G. (2009). Foreign exchange profits in two early Renaissance money markets. *Journal of European Economic History* 38, 123–144.

Booth, G.G. & Chang, S.S. (2013). Domestic exchange rate determination in Renaissance Florence. Eli Broad Graduate School of Management, Michigan State University.Working Paper. East Lansing, Michigan, U.S.

Booth, G.G. & Gurun, U.G. (2008). Volatility clustering and the bid-ask spread: exchange rate behavior in early Renaissance Florence. *Journal of Empirical Finance* 15, 133–144.

Cipolla, C.M., (1956). *Money, Prices, and Civilization in the Mediterranean World.* London, U.K.: Princeton University Press.

de Roover, R. (1963). *The Rise and Decline of the Medici Banks.* Cambridge, MA: Harvard University Press.

de Roover, R. (1968). *The Bruges Money Market around 1400.* Brussels, Belgium: Paleis der Academien.

Dominguez, K.M.E. & Panthaki, F. (2006). What defines “news” in foreign exchange markets? *Journal of International Money and Finance* 25:1, 168–198.

Feng, L., Li. B., Podobnik, B., Preis, T. & Stanley, H.E. (2012). Linking agent-based and stochastic models of financial markets. *Proceedings of the National Academy of Sciences* (*PNAS*) 109:22, 8388–8393.

Friedman, D. (1991). Evolutionary games in economics. *Econometrica* 58, 637–666.

Glosten, L.R., Jaganathan, R. & Runkle, D.E. (1993). On the relationship between the expected value and volatility of nominal excess returns on stocks. *Journal of Finance* 48:5, 1779–1801.

Granovetter, M.S. (1985). Economic action, social structure, and embededness. *American Journal of Sociology* 91, 481–510.

Greif, A. (2006). *Institutions and the Path to the Modern Economy*. New York, NY: Cambridge University Press.

Greif, A. & Laitin, D.D. (2004). A theory of endogenous institutional change. *American Political Science Review* 98:4, 633–652.

Gross-Klussmann, A. & Hautsch, N. (2011). Predicting bid-ask spreads using long memory autoregressive conditional Poisson models. SFB Discussion Paper. Humbolt University, Berlin, Germany.

Houkes, J.M. (2004). *An Annotated Bibliography on the History of Usury and Interest from the Earliest Times through the Eighteenth Century*. Lampeter, Ceredigion Wales, UK: Edwin Mellen Press.

LeBaron, B. & Yamamoto, R. (2007). Long-memory in an order-driven market. *Physica A* 383, 85–89.

LeBaron, B. & Yamamoto, R. (2008). The impact of imitation on long memory in an order-driven market. *Eastern Economics Journal* 34, 504–517.

Lyons, R.K. (2001). *The Microstructure Approach to Exchange Rates*. Cambridge, MA: MIT Press.

Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets* 3, 205–258.

McCloskey, D.N. (1976). Does the past have useful economics? *Journal of Economic Literature* 14, 434–461.

McCloskey, D.N. (2006). *The Bourgeois Virtues*. Chicago, IL: University of Chicago Press.

Osmond, P.J. (2000). Catiline in Fiesole and Florence: The afterlife of a Roman conspirator. *International Journal of the Classical Tradition* 7:1, 3–38.

Padgett, J.F. (2001). Organizational genesis, identity, and control: The transformation of banking in Renaissance Florence. In J.E. Rauch & A. Casella (Eds). *Networks and Markets.* New York, NY: Russell Sage Foundation. 211–257.

Padgett, J.F. (2010). Open elite? Social mobility, marriage, and family in Florence, 1282–1494. *Renaissance Quarterly* 63, 357–411.

Sargent, T.J. & Velde, F.R. (2002). *The Big Problem of Small Change.* Princeton, NJ: Princeton University Press.

Spufford, P. (1988). *Money and Its Use in Medieval Europe.* Cambridge, U.K.: Cambridge University Press.

Staley, E. (1906). *Guilds of Florence.* London, U.K.: Methuen & Co.

Swetz, F.J. (1987). *Capitalism and Arithmetic: The New Math of the 15th Century, Including the Full Text of the Treviso Arithmetic.* Trans. by D.E. Smith. LaSalle, IL: Open Court Publishing.

Tong, H. (2010). *Threshold Models in Time Series Analysis – 30 Years On*. Research Report, Serial No. 471, Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong, China.

Van Egmond, W. (1976). *The Commercial Revolution and the Beginning of Western Mathematics in Renaissance Florence*, 1300–1500. Indiana University dissertation. Ann Arbor, MI: University Microfilm International.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance* 16, 8–37.

A NOTE ON THE CALCULATION OF THE RISK-FREE RATE FOR TESTS OF ASSET PRICING MODELS AND EVENT STUDIES

Mika Vaihekoski
Turku School of Economics, University of Turku

1 Introduction

The risk-free rate has a special role in finance theory. It is one of the key parameters in most asset pricing models and companies' cost of capital calculations. It is the reference point for risky asset returns, and as such, asset pricing models are typically tested using excess returns. In event studies, it is used when abnormal returns are measured using the excess market model.

To calculate excess returns, one needs to have time series for the risk-free rate of return. In asset pricing tests, this return is often taken to be the risk-free rate of return over a period of one month. For this reason, researchers in the USA have typically used Ibbotson Associates' or CRSP's monthly risk-free rate series, calculated originally from the U.S. Treasury bills closest to the one month investment period (see e.g. CRSP 2006). In other countries, however, researchers usually do not have access to similar standardized data series. Treasury bills are sometimes not traded, or their price series are not readily available. Therefore many studies on other markets have used local interbank money market rates (e.g. Euribor rates) or some other international money market rates (e.g. Eurodollar rates) as an approximation for the risk-free rate.[1]

However, even when money market rates are available, there are several practical issues that need to be solved before one has appropriate time series for the risk-free rate of return that can be used in asset pricing studies.[2] First, the time to ma-

---

1 An example of a research paper where the risk free rate has been calculated from the local interbank rates is e.g. Nummelin and Vaihekoski (2002). In many papers studying international asset pricing models, authors have calculated risk-free rates from Eurodollar rates (see e.g. Carrieri, Errunza, and Majerbi 2006; Chaieb and Errunza 2007) or from other Eurocurrency rates (see e.g. Dumas, Harvey, and Ruiz 2003). Several studies have augmented local interbank rates with Euribor rates after the adoption of the Euro (see e.g. Vaihekoski 2009).

2 In addition to the issues raised here, researchers and practitioners have to decide whether long term rates are better suited for the task at hand (see Damodaran 2008, for discussion).

turity of the traded Treasury bills or certificates of deposits do not necessarily match exactly the length of the return period of researcher's interest. This can be the case if, for example, the instruments are not issued every day. Thus it is unlikely to have instruments that give the researcher always directly the desired risk-free rate of return for the period that she is interested in.

Second, money market rates are often quoted as annualized simple rates (yields) which cannot be used as such in empirical research. Additional problems arise when a researcher studies daily (e.g., while conducting an event study) or weekly excess returns, but the shortest available rates are for one month. As a result, one often has to select a method to give the best estimate for the risk-free rate from the available rate series.

Researchers have used several ways to estimate the risk-free rate of return even though only a few papers actually report which method was used in the calculation. This is understandable as the risk-free rate has only minor effects on the results of asset pricing tests. This may have led researchers to think that the calculation is a trivial exercise even though this is typically not the case. In fact, some of the methods used by researchers are producing small, yet unnecessary bias in their risk-free return series. In any case, it is important that researchers acknowledge the approach taken to calculate the proxy for the risk-free rate.

The aim of this paper is to provide a compact comparison of the different methods of estimating the risk-free rates from the available money market rates. Furthermore, the main issues that one faces when calculating the risk-free rate series for tests of asset pricing models are discussed. Finally, the role of the risk-free rate in event studies is briefly discussed. The rest of this paper is organized as follows. The next section presents two alternative approaches to calculate the risk-free rate. The advantages and disadvantages of both approaches are discussed. The last section concludes.

2 Risk-free rate of return

2.1 Definition

There are typically two different main uses for the risk-free rates of return in empirical research. First, researchers and practitioners alike use the risk-free rate to measure the long-term required rate of return for an investment or a company (e.g. using the weighted average cost of capital) which typically requires some kind of knowledge about the equity market premium. Second, the risk-free rate is used in tests of the asset pricing models or in event studies. Here, the latter use for the risk-free rate is the main interest. The main difference between these two alternatives is the length of the period over which the risk-free rate is calculated. The first use emphasizes the long-term perspective whereas the latter one typically favors shorter ones. As a result, they typically use different time series to estimate the risk-free rate.

In order to estimate the risk-free rate of return, we should first agree on what we mean by the risk-free rate of return. First, risk-free rate should measure the return on an asset whose return is completely without risk or uncertainty over the period of investigation. Mathematically the rate of return is a constant, and as a result its variance is zero as well as its correlation with other random variables. Second, the risk-free rate has to be known at the time of the analysis for the period ahead of time. In other words, the risk-free rate is a forward looking measures and known *a priori*, not after the fact. Investing in a risk-free asset, whose rate of return is not known until the maturity, is not truly risk-free. In addition, sometimes the definition is even extended so that the return has to be risk-free in real terms, but typically researchers have settled for the risk-free rate in nominal terms. This is out of practical convenience as there may not be assets that provide a risk free return in real terms or their time series is not readily available for the whole sample period of interest.

The definition above has a number of practical implications. First, in an international setting, one has to make sure that the return is measured in the same numeraire currency as the asset returns are measured. For example, studying returns from US investors' point of view, one should choose the risk-free rate in the United States. Similarly, researcher studying the same problem from the point of view of German investors should choose the German risk-free rate.

Second, the risk-free rate does not have to be constant over multiple periods. To test asset pricing models, one typically uses short-term data and thus the risk-free rate is estimated to be risk-free over the chosen horizon (e.g., one month). As a result, the risk-free rate varies from one period to another and the time-series shows variability.

Third, if one is interested to estimate e.g. the long-term equity premium or required rate of return for an investment, one typically uses risk-free rates calculated from bonds with longer maturity rather than money market securities. As a result, the bond can be risk-free over its maturity (save the reinvestment risk), but during its lifespan, its price can vary i.e. the risk-free asset is practically risk-free over the long period, but not necessarily over the short-term period.

2.2 Calculation

To calculate excess returns for assets, researchers need time series for the risk-free rates of return. Typically they are not readily available and in many cases, one cannot observe them directly. Sometimes it is even hard to find a truly risk-free asset (e.g. in studies dealing with historical time periods prior to traded money market instruments). This raises a number of practical issues and one has to make some choices in estimating the risk-free rates. Typically, one has to settle for some kind of a proxy. Such a proxy should be based on real market trading information, and the time series should be readily available. If there is no information available from the market, one has to settle to some other proxy that best measures the risk-free rate of return attainable for professional investors. For periods long enough in the history, this can mean even using negotiated rates (e.g., rates set by the Central Bank). Moreover, the selected proxy should measure the risk-free rate over the period in question. Finally, out of practical convenience, the trades required to attain risk-free rates should be fairly simple and achievable in real life. This rules out approaches based on complex derivate positions.

Sometimes a researcher has to consider trade-offs between these criteria in constructing the series. First, there might be several different instruments traded in the money market that could be used as a proxy for a money market instrument. A commonly accepted proxy for the risk-free asset is the government issued short-term money market instrument i.e. T-bill in the USA. Researchers, however, cannot always observe the T-bill rates for the numeraire country under analysis (because they are not issued, not traded, or the price series is otherwise not available), and they have to use some other instruments. As a result, researchers have used interbank money market rates (e.g. Euribor-rates in the EU). This is justifiable since banks' certificates of deposits traded on the money markets are virtually risk-free since governments usually provide at least some kind of implicit guarantees against bank bankruptcies. However, the assumption of risk-free came to a halt during the financial crisis in late 2007 when the market clearly priced a risk premium into the Euribor rates without collateral over the Eurepo rates with collateral (c.f. Figure 1).[3]

[3] Similarly, the observed manipulation of the Libor-rates has raised some doubt over their validity as well.

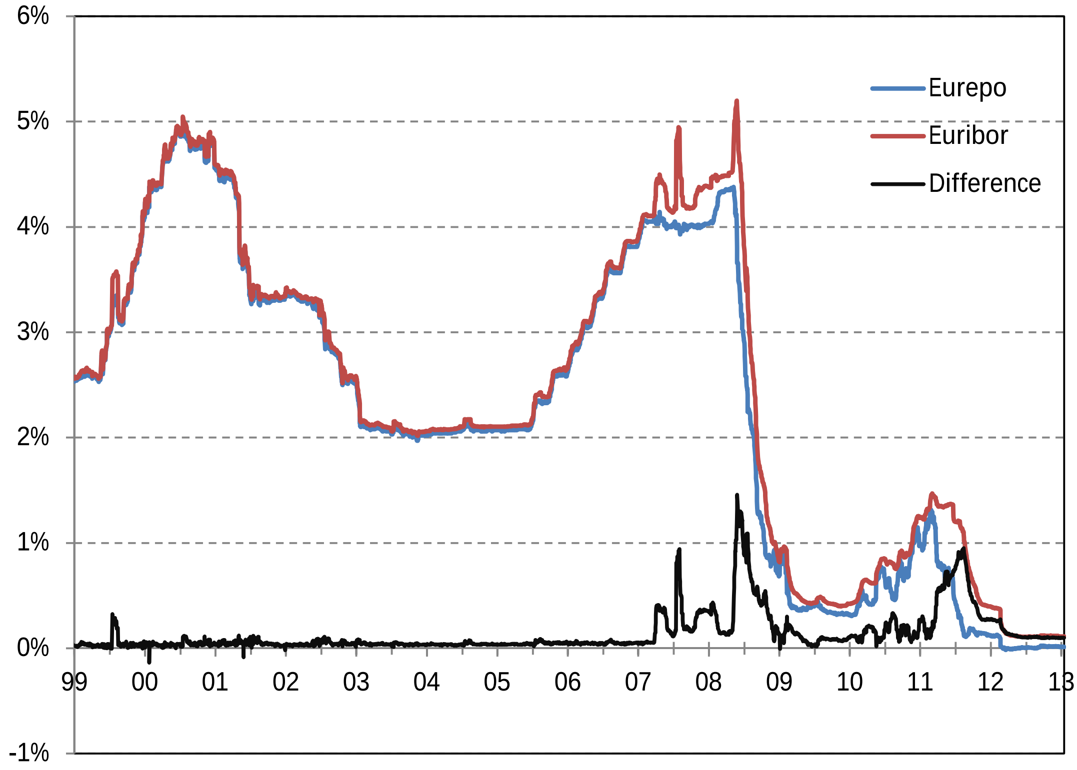


Figure 1. Eurepo vs. Euribor rates. One month daily rates are from 1999 to May 2013. Their difference is also shown.

Second, researcher typically observes money market rates only for one month, two months etc. However, if the researcher wants to calculate risk-free rates for e.g. one day or one week, one has to calculate approximate returns from the rates available for longer maturities.[4] Third, even if the researcher studies monthly (or quarterly) returns, the money market rates may not be suitable without adjustment. Namely, money market rates are typically quoted as simple rates per annum. Moreover, the maturity of the instruments do not necessarily precisely match that of the researcher's stock (or other asset) data. Namely, T-bills or CD instruments closest to one month could have maturities less or more than one month.

Furthermore, the time series for the risk-free rate has to be available "one observation earlier" than the risky asset return series, as risk-free rates are always forward looking measures and known *a priori*, whereas realized asset returns are

---

[4] Note that researchers, who have access to monthly (e.g. from CRSP), but not daily risk free rates, have a similar problem.

always measured *ex post*.[5] Finally, researcher should take into account the day counting conventions and calendar time difference between two stock market observations. For example, for monthly returns, the difference can vary between 28 and 31 days in most cases.[6]

To solve the problem of converting annual yields into shorter periods, researchers should take the market rate for the period that best matches the length of the investment period as the starting point. Typically in asset pricing tests with monthly data, it is the one month money market rate. Given one month money market rates per annum, researchers have used several methods to calculate rates of return for shorter periods. These methods can be categorized into two different categories, namely *the interest compounding approach* and *the price difference approach*.[7] The first approach can generally be considered inferior (or plain wrong) to the second one, but it can give the same results as the second approach if adjusted to match the same assumptions.

The interest compounding approach uses common interest on interest equations to match longer period rate to shorter period rate by equating the sought risk-free rate on a shorter period to the interest rate on longer period usually on an interest on interest basis. Given a one month money market rate, the transformation equation that can be used to get the *d*-day percentage risk-free rate of return, R_t^d, can be written as follows:

$$\begin{aligned} 1 + R_t^{1m} &= (1 + R_t^d)^m \Rightarrow \\ R_t^d &= (1 + R_t^{1m})^{1/m} - 1 \end{aligned} \tag{1}$$

and similarly for the continuously compounded risk-free rate of return, r_t^d, as

[5] For example, if a researcher has monthly stock returns for January, he or she needs the risk free rate of return for January known before the period began. In practice, the risk free rate is calculated using money market information from the last day of the previous year. Note also, that one should be careful with the notation depending whether one uses *t* to denote time or period. Researchers using the time notation, use risk free rates from $r_{ft\text{-}1}$ to calculate excess returns for asset *i*'s return r_{it}, (i.e. E[r_{it} - $r_{ft\text{-}1}$]) whereas those using the period notation, calculate returns in excess of r_{ft}.

[6] Occasionally, the length of the month under investigation can vary even more in cases where the stock market has not been open at the end of the periods of interest due to a weekend or holiday.

[7] If one is aiming to estimate risk-free rates for a period (say, a week) in a situation, where there are rates available for shorter (e.g. overnight) and longer (e.g. a month) periods, one could also utilize some kind of extrapolation methods. They are not discussed here.

$$\begin{aligned} 1+R_t^{1m} &= \exp(m \cdot r_t^d) \Rightarrow \\ r_t^d &= m^{-1}\ln(1+R_t^{1m}) \end{aligned} \tag{2}$$

where R_t^{1m} is the observed one month money market rate (quoted in per annum as typically is the case), d is the length of the period in days over which the risk-free rate of return is calculated, and m is the number of periods with length d days in a year.

This approach, however, has at least three problems. First, we have to select the proper m for the equation. Prior research has used several choices for m. The simplest choice in the case of percentage returns is to use $m = \log_a(1+R_t^{1m})/\log_a(1 + d \cdot R_t^{1m}/rdpy)$, where $rdpy$ is the real number of days per year, and a is the base of one's choice. This somewhat complex looking formulation suggests merely a linear relationship between R_t^{1m} and R_t^d as the risk-free rate of return for d days is simply the money market rate divided by ($rdpy/d$). The method is clearly not correct, as it does not take into account the interest on interest effect.

A better alternative, and probably the one most commonly used, is to use $m = (rdpy/d)$. However, this approach still produces a small bias even though it takes the interest on interest effect into account. Namely, it forgets the fact that the money market rates are quoted per annum in simple interest rate form. In effect, this approach makes the implicit assumption that the one month money market rate is also the annual rate of return.

If the researcher accepts a slight error in his or her risk-free rates and wants to choose the method above, he or she should also be aware of additional problems which are not always properly addressed. Namely, he or she has to choose the number of days in a year. For example, if d is one day, m is sometimes chosen to be 360, 365, the real number of days in a year (365/366), or even the number of trading days in a year (approximately 250). Clearly the best alternative is to use the real number of days as the interest rate is paid for the passing of time, but given historical interest rate series, the calculation required one to backtrack years with leap days, which causes additional work. Finally, the researcher could also take into account the length of the investment period. Namely, if the risk-free rate is calculated over the weekend, we need to adjust d accordingly. Similarly, if one just wants to convert simple annual rates (based on one month market rates) to usable risk-free rates, *d* has to be chosen to match the real number of days in a month.[8]

A slight variation of the method above to calculate daily (or weekly) returns is to take e.g. monthly risk-free rates from a research database (say, the CRSP database), and calculate returns for *d* days by dividing monthly rates by the number of *d* days in a month. Even though this method is easily applicable, it produces even more erroneous values for the risk-free rates as it completely ignores changes in the interest rates. As a result, the risk-free rate basically stays the same within a month, and if not carefully calculated, even the effect of weekends might be ignored.[9] In addition, this method can produce a slight bias into the risk-free rate series, if the market rates show some kind of anomalous month-end behavior. Moreover, this approach produces unwarranted autocorrelation in the (excess) return series. Furthermore, it can be difficult to use if one is not studying end-to-end monthly returns. Finally, using the same risk-free rate for a number of periods in asset pricing tests is likely to affect the estimates of alpha.

The price difference approach takes a slightly different route for calculating, say, daily or weekly risk-free returns. It takes the shortest available money market rate (e.g., one month rate), assumes that the interest rate curve is flat for periods shorter than the available rate, and that the risk-free rate of return is known at the beginning of the period and it stays the same for the whole period (e.g. one day or one week) as it should for the risk-free asset. If this is the case, we can calculate the price of the money market instrument at the beginning and at the end of the investment period with the same rate of return i.e. as if the rate of return had not changed. As a result, the only compensation for the investors is due to the passing of time, as it should be for the holding period.[10]

Using this approach, the risk-free holding period rate of return is now simply the percentage (or ln) difference in the prices of the money market instrument at time t-1 and t. Now the percentage risk-free rate of return for a period of d days can be written as follows:

---

8 Sometimes researchers have chosen to ignore the difference in the length of months and set *d* to 30.4 (c.f. CRSP 2006).

9 Of course, it is not always the case that researchers want to treat weekends differently from other weekdays.

10 This is essentially the method used in Ibbotson Associates' yearly publication ("Stocks, Bonds, Bills, and Inflation Yearbook") that provides monthly (end-to-end) risk free returns for the US market since 1925.

$$R_t^d = \frac{P_{t+d}^{1m}}{P_t^{1m}} - 1, \quad (3a)$$

which can be written as follows for most non-US markets

$$R_t^d = \frac{R_t^{1m} d}{dpy + R_t^{1m}(dtm - d)}, \quad (3a)$$

where R_t^{1m} is the observed money market one month rate (per annum), *dtm* and *dpy* are days to maturity and days per year, respectively. Note that they are related to the pricing of the money market instrument and to the day counting convention used in the market. For example, for the Euribor rates (as well as e.g. money market rates in the USA and in the Eurocurrency rates in London save GBP), the used day counting convention is currently real/360, so *dpy* is 360 and *dtm* is the actual number of days to maturity. In case of one month Euribor, it is the number of days between today and the same calendar day in the next month, unless today is the last day of the month.[11] The parameter *d* is the length of the period over which the risk-free rate of return is calculated. For daily returns, *d* is typically one (from Monday to Thursday) or three (from Friday to Monday), but it can vary if non-business day holidays are taken into account.[12] For weekly returns from Wednesday to Wednesday, *d* is typically seven or occasionally slightly less or more depending on the holidays occurring in the middle of the week.

In the USA, the market convention for quoting the prices for T-bills differs slightly from most other countries. T-bills are priced deducting the discount ($D = R \times dtm / dpy$) from the face value (i.e., $P = 100 - D$), where the discount

[11] Note that the day counting conventions differ from market to market and can be quite complex since there are typically a number of exceptions to the main rule. For example, in the Euribor market, one month Euribor is typically priced to the next month with the same calendar day i.e. Euribor one month quoted February 15th matures March 15th with the actual running time of 32 days. If the same day does not exist in the next month (as is the case e.g. for the last two or three days of January), one month rates are quoted to the last day of the next month (e.g. February 28th or 29th in the case of a leap year). However, month-end Euribor rates are always quoted until the end of the next month, i.e. *one month quote for 28 February has* the maturity date of 31 March, not 28 March. Furthermore, if the maturity date falls on a non-business day, the maturity date is moved to the next business day except in the case of the month-end quotes. In these cases the maturity date is brought back to the last business day of that month (see Euribor 2006)

[12] Note that the number of non-business days can differ for the money market and the asset return data under investigation. Typically, when testing asset pricing models using stock data, researchers should try to match the trading days given in the equity data since the stock market is typically more often closed than the money market.

yield (R) is quoted as an annual rate (see e.g. Zipf 2003). As a result, equation (3a) should be written as follows (see Appendix for details)

$$
\begin{aligned}
R_t^d &= \frac{100 - D_{t+d}}{100 - D_t} - 1 = \frac{D_t - D_{t+d}}{100 - D_t} \\
&= \frac{R_t^{1m} \cdot d}{dpy \cdot 100 - R_t^{1m} \cdot dtm}
\end{aligned}
\tag{3b}
$$

where the market convention is to use real/360 for *dpy*/*dtm* at the time of the quote.

A similar equation can be derived for the continuously compounded risk-free rate of return (here shown for non-US-markets).

$$
\begin{aligned}
r_t^d &= \ln \frac{P_{t+d}^{1m}}{P_t^{1m}} \\
&= \ln\left(\frac{dpy + R_t^{1m} dtm}{dpy + R_t^{1m}(dtm - d)} \right)
\end{aligned}
\tag{4a}
$$

and for US T-bills (see Appendix for details)

$$
r_t^d = \ln\left[\frac{100dpy - R_t^{1m}(dtm - d)}{100dpy - R_t^{1m} dtm} \right]. \tag{4b}
$$

This second approach has the drawback that we need to know the day counting convention used in the market to be precise in the analysis. Moreover, if the convention has changed during the period under investigation, we need to change the equation accordingly. For example, studying excess asset returns in Europe over a period which dates back to times before the introduction of Euribor rates in before 1999, the researcher has to augment Euribor rates that with the local money market rates (e.g. Fibor rates) which might have had a different day counting convention.

The interest compounding and price difference approaches can be made to give the same results if m is set to match the assumptions made in the second approach. For example, in a case of the percentage returns, m in equation (1) should be set (non-US-case) to

$$
m_t = \frac{\ln(1 + R_t^{1m})}{\ln\left(\frac{R_t^{1m} d}{dpy + R_t^{1m}(dtm - d)} + 1 \right)}, \tag{5}
$$

which obviously is not the most intuitive equation to work with.

We can study the error one makes when using daily risk-free returns and confining to equation (1) with *m* set to 365 (*d* equals one) which disregards the effect caused by the weekends (method 1a) or *m* set to 365/*d* which takes the true length of the holding period into account (method 1b). Panel A in Table 1 compares the result against the equation (3) using one month Euribor rates and its day counting conventions. We have also included a comparison against less than perfect use of equation (3) where the researcher has applied fixed 30/360 day counting convention for all observations. Turn of the month has been chosen for demonstration purposes to highlight the differences between the methods.

We can see that the difference (error) is not large, especially if the weekends are properly taken into account also in the first method or if one alternatively decides to treat weekends as one trading day and equal to all other days. The difference between the methods probably does not make a big difference in the asset pricing tests, but at least one should aware of the method used to calculate the risk-free rate. We can also see that the casual use of equation (3) gives fairly accurate returns – the bias against the benchmark model is considerably less than for the compounding method.

We can also perform a similar analysis for the monthly risk-free rates. Panel B in Table 1 shows the results for month-end Euribor rates. In the analysis one month interest rate (per annum) is set to 4.50 per cent to reflect typical value during the last twenty years. Method 1 uses the compounding approach either by assuming fixed length for the months regardless of real length of the holding period (method 1a). Method 1b takes into account different lengths of the months. Method 2 uses the price difference formula, which transforms the quoted interest rates properly into risk-free rates usable in asset pricing studies.[13]

The results again confirm a small difference between the methods. Performing sensitivity analysis (not reported) reveals that the difference is sensitive to the level on interest rates. The higher the interest rate level the larger the difference (realistic interest rates levels give differences for method 1b against 2b that vary between one and five per cent).

---

[13] Note that the risk-free rate for April is assumed to be based on a holding period from 31.3. to 28.4., since 30.4. is assumed to be a Sunday to illustrate the effect of weekends.

Table 1. Analysis of different time aggregation approaches.

One month Euribor rate is used to calculate percentage daily and monthly risk-free (holding period) rates of return using three different methods. Method 1a uses an approximation (equation (1) with m = 365 or 12 in Panels A and B, respectively) and disregards the length of the holding period (i.e. in Panel A the effects of the weekends is disregarded), whereas method 1b takes them into account (m = 365/d), where d is the length of the holding period in days. Method 2 uses equation (3a) with the exception that 2a assumes fixed 30/360 day counting convention. Diff (%) column shows the percentage difference in proxies from methods 1a, 1b, and 2a against the method 2b.

	1 month	Method	Method	Method	Method	Diff (%) against Method 2b		
Day	Euribor	1a	1b	2a	2b	Method 1a	Method 1b	Method 2a
Panel A: Daily returns								
Friday 27.1.	4.500 %	0.01206 %	0.03618 %	0.03736 %	0.03737 %	-67.7270 %	-3.1695 %	-0.0125 %
Monday 30.1.	4.500 %	0.01206 %	0.01206 %	0.01245 %	0.01246 %	-3.1811 %	-3.1811 %	-0.0125 %
Tuesday 31.1.	4.500 %	0.01206 %	0.01206 %	0.01245 %	0.01246 %	-3.1932 %	-3.1932 %	-0.0249 %
Wednesday 1.2.	4.500 %	0.01206 %	0.01206 %	0.01245 %	0.01246 %	-3.1932 %	-3.1932 %	-0.0249 %
Thursday 2.2.	4.600 %	0.01232 %	0.01232 %	0.01273 %	0.01273 %	-3.2078 %	-3.2078 %	0.0000 %
Friday 3.2.	4.450 %	0.01193 %	0.03579 %	0.03695 %	0.03696 %	-67.7245 %	-3.1620 %	-0.0246 %
					Average:	-24.7045 %	-3.1845 %	-0.0166 %
Panel B: Monthly returns								
Tuesday 31.1.	4.500 %	0.36748 %	0.33823 %	0.34874 %	0.35000 %	4.9946 %	-3.3616 %	-0.3612 %
Tuesday 28.2.	4.500 %	0.36748 %	0.37454 %	0.38610 %	0.38750 %	-5.1662 %	-3.3441 %	-0.3612 %
Friday 31.3.	4.500 %	0.36748 %	0.36244 %	0.37365 %	0.37509 %	-2.0296 %	-3.3741 %	-0.3861 %
Friday 28.4.	4.500 %	0.36748 %	0.37454 %	0.38610 %	0.38750 %	-5.1662 %	-3.3441 %	-0.3612 %
					Average:	-1.8419 %	-3.3559 %	-0.3674 %

The price difference approach has also some additional benefits. First, it is robust to the choice of the money market rate, i.e., one could use one day (overnight) rate or even use three month money market rates to calculate rates for *d* days, if required.[14] Second, the risk-free rate is always positive as long as the observed market rate is positive. Third, equations (3) and (4) solve the second problem mentioned at the beginning of this section. Namely, we are often not able to use published money market rates as such even if they match our data frequency (e.g., we are studying monthly excess returns and we have access to monthly money market rates) partly due to market pricing conventions and non-matching business days in some rare instances, but mostly due to the simple interest compounding convention used in the market. Finally, the price difference approach is also very flexible since we can choose to calculate our monthly returns from any calendar day forward (say mid-month), even though the month-end-to-end is the commonly used convention in asset pricing research.

2.3 Risk-free rate and event studies

In event studies, the use of risk free rate series occurs less often and it depends on what kind of model the researcher has chosen for the normal return of the asset when calculating the abnormal returns. Kolari and Pynnönen (2010) list the most commonly used models.

The market model is probably the most common approach in the event studies. It has the benefit that the risk-free rates of return are not needed in the estimation. In a multi-country setting (e.g., Europe in pre-Euro-era) acquiring risk-free rates data for each country could be burdensome and time consuming. Moreover, even if one is working in a single-country setting, the market model approach avoids the need to calculate daily risk-free series which is beneficial if they are not readily available.

However, if one assumes an asset pricing model (e.g., the CAPM) for the normal returns, ignoring the risk-free rate can produce a slight bias in the event study estimation if not appropriately taken care of as pointed out by Binder (1998). Namely, the CAPM implies the following form for asset *i*'s alpha if estimated using the market model: $\alpha_{MM,i} = \alpha_{J,i} + r_f(1 - \beta_i)$, where r_f is the mean risk-free

---

14 For example, we want to use Black-Scholes pricing formula throughout time to maturity, T (say, three months i.e. 90 days) and we need risk free rates for periods T, T-1, T-2, …, 1 days. Typically, we have only a limited number of rates available to us (say, rates for one, two, and three month).

rate during the estimation sample and $\alpha_{J,i}$ is Jensen's alpha (risk-adjusted excess return) from the excess return market model. The standard event study methodology uses the alpha and beta from the market model to estimate abnormal returns during the event window. Now, given that the CAPM applies, Jensen's alpha should be zero and left out when calculating the abnormal returns during the event window.[15]

Obviously one can alternatively chooses to estimate the excess market model using returns in excess of the risk-free rate returns or utilize some other asset pricing model for the excess returns (c.f., e.g., OLS model, Fama-French model and FF industry model in Kolari and Pynnönen 2010). If this is the case, one requires risk-free rates over a period of one day. The approach presented earlier can be easily applied to estimate the daily returns from T-bills or other money market instruments.

3 Conclusion

This paper has briefly reviewed two commonly used approaches to calculate proxies for the percentage and continuously compounded holding period risk-free rates of return from the observed money market rates for empirical tests of asset pricing models and event studies. The price difference approach can be shown to have several beneficial properties against the more commonly used compounding method. Their difference is typically fairly small, and it does not have any major effect on the empirical financial research, but one should acknowledge the approach taken to calculate the risk-free rates. In addition, the issues that one needs to consider when one calculates risk-free rates were discussed. Special care was put on analyzing situations where one is forced to estimate risk-free rates for periods shorter than the maturity of the shortest available market rate.

---

[15] Depending on one's assumptions, different approaches have been used. For example, Kolari and Pynnönen (2010) include alpha from the market model when calculating abnormal returns, whereas De Jong (2007) does not. If the CAPM is not the correct pricing model or it only assumed to be a proxy for the true model, Jensen's (and thus the market model's) alpha can be taken to capture the average pricing error due to choice of the wrong pricing models, and thus it could be applied when calculating the abnormal returns during the event window.

References

Binder, J. J. (1998). The event study methodology since 1969. *Review of Quantitative Finance and Accounting* 11, 111–137.

Carrieri, F., Errunza, V. & Majerbi, B. (2006). Local risk factors in emerging markets: Are they separately priced? *Journal of Empirical Finance* 13, 444–461.

Chaieb, I. & Errunza, V. (2007). International asset pricing under segmentation and PPP deviations. *Journal of Financial Economic* 86, 543–578.

CRSP (2006). *CRSP Monthly US Treasury Guide*. Center for Research in Security Prices.

Damodaran, A. (2008). *What is Riskfree Rate? A Search for the Basic Building Block.* Unpublished, available at SSRN: http://ssrn.com/abstract=1317436.

De Jong, F. (2007). Event studies methodology. Unpublished manuscript, Tilburg University.

Dumas, B., Harvey, C. R. & Ruiz, P. (2003). Are correlations of stock returns justified by subsequent changes in national outputs? *Journal of International Money and Finance* 22, 777–811.

Euribor (2006). *Technical Features*. E-document available at www.euribor.org. Accessed December 5th, 2006.

Kolari, J. W. & Pynnönen, S. (2010). Event study testing with cross-sectional correlation of abnormal returns. *Review of Financial Studies* 23, 3996–4025.

Nummelin, K. & Vaihekoski, M. (2002). International capital markets and Finnish stock returns. *European Journal of Finance* 8, 322–343.

Vaihekoski, M. (2009). Pricing of liquidity risk: Empirical evidence from Finland. *Applied Financial Economics* 19, 1547–1557.

Zipf, R. (2003). *Fixed Income Mathematics*. USA: Academic Press.

Appendix

Derivation of equation (3b):

$$
\begin{aligned}
R_t^d &= \frac{100 - D_{t+d}}{100 - D_t} - 1 = \frac{D_t - D_{t+d}}{100 - D_t} \\
&= \left(R_t^{1m} \frac{dtm}{dpy} - R_t^{1m} \frac{dtm - d}{dpy} \right) \cdot \frac{1}{100 - D_t} \\
&= \frac{R_t^{1m} \cdot d}{dpy} \cdot \frac{1}{100 - D_t} = \frac{R_t^{1m} \cdot d}{dpy} \cdot \frac{1}{100 - R_t^{1m} \dfrac{dtm}{dpy}} \\
&= \frac{R_t^{1m} \cdot d}{dpy} \cdot \frac{dpy}{dpy \cdot 100 - R_t^{1m} \cdot dtm} \\
&= \frac{R_t^{1m} \cdot d}{dpy \cdot 100 - R_t^{1m} \cdot dtm}
\end{aligned}
$$

Derivation of equation (4b):

$$
\begin{aligned}
r_t^d &= \ln \frac{P_{t+d}^{1m}}{P_t^{1m}} \\
&= \ln \frac{100 - D_{t+d}}{100 - D_t} = \ln \frac{100 - R_t^{1m} \dfrac{dtm - d}{dpy}}{100 - R_t^{1m} \dfrac{dtm}{dpy}} \\
&= \ln \left[(100 - R_t^{1m} \frac{dtm - d}{dpy}) \cdot \frac{1}{100 - R_t^{1m} \dfrac{dtm}{dpy}} \right] \\
&= \ln \left[\frac{100dpy - R_t^{1m} (dtm - d)}{100dpy - R_t^{1m} dtm} \right]
\end{aligned}
$$

SHORT-TERM VALUE CREATION FOR THE BIDDER: EVIDENCE FROM FINLAND

Kenneth Högholm, Johan Knif and Tobias Romar
Hanken School of Economics

1 Introduction

Most empirical analyses of the takeover market report that bidding firm's shareholders earn small and insignificant returns, while target firm's shareholders earn large and significant returns in takeover transactions.[1] The reported returns are skewed in favor of the target firm also after adjusting for differences in firm size. Bradley, Desai & Kim (1988) reported a 90/10 split of the value-weighted takeover gain in their US sample, while Högfeldt & Högholm (2000) observed a similar distribution in a sample of Swedish takeovers.

Several different hypotheses have emerged trying to explain these empirical findings. One line of thought suggests that takeovers occur because of management incentives or mistakes, i.e., takeovers occur because management of bidding firms wishes to grow (e.g., Morck, Shleifer & Vishny 1990), or they overestimate the value of the target firm (e.g., Roll 1986). Alternatively, the management of the bidding firm takes advantage of a temporary overvaluation of the firm, hence, taking advantage of a window of opportunity to make an acquisition (e.g., Jensen 2004). The main core in these hypotheses is that a takeover does not create any additional value, but is more of a redistribution of wealth from shareholders in the bidding firm to shareholders in the target firm.

Other hypotheses suggest that takeovers do create additional value, but that the target firm's shareholders for some reason obtain a larger share of the gain. Grossman & Hart's (1980) free riding problem is one explanation. In the extreme case the free riding hypothesis suggests that target shareholders capture the entire gain. Alternatively, e.g., competition among bidders (Fishman 1980), takeover defense measures (Harris 1990), or target ownership structure (Högfeldt & Hög-

---

[1] For an overview of the wealth effects of takeover transactions, see, e.g., Jensen & Ruback (1983), Jarrell, Brickley & Netter (1988), Agrawal & Jaffe (2000), Bruner (2004), and Martynova & Renneboog (2008a).

holm 2000) may also explain the skewed distribution in favor of the target firm's shareholders.

The purpose of this study is to investigate the short-term abnormal return to the bidding firm's shareholders in takeover transactions in Finland during the time period from January 2000 to December 2009. We estimate the abnormal return around 249 individual takeover announcements and investigate determinants of the abnormal returns. Our results show that the takeover announcement on average yields a positive abnormal return to the bidding firm's shareholders. The announcement effect on the announcement day is 1.1 % and statistically significant. Both pre-event and post-event abnormal returns are statistically insignificant, although there is sign of negative revaluation in the post-event period. Among the takeover characteristics, we document a significant impact on the bidder's abnormal return on the announcement day for a) *small deals* yielding a higher abnormal return, b) *cross-border deals* giving a smaller abnormal return, and c) *diversification deals* giving a higher abnormal return to the bidder's shareholders.

The reminder of the paper is organized as follows. Section 2 summarizes the literature review on the motivations for takeovers and the determinants of the share price reaction to the takeover announcement. Section 3 describes the methodology and the data, while the empirical results are presented in Section 4. Section 5 concludes the study.

2 Motives for takeovers

Three major takeover motives have been advanced in the literature implying gains for both the bidder and the target, or negative return to the bidder. These are the synergy motive, the agency motive and the hubris hypothesis.

2.1 The synergy motive

The synergy motive assumes that managers maximize shareholders' wealth and would engage in takeover activities only if it results in gains to the shareholders. Among the synergy motives, the first set of motives is consistent with the assumption that additional value is created by takeovers. The second set of motives cast doubt on whether any additional value is created by takeovers, or if the resulting gains to shareholders' is at the expense of other stakeholders (e.g., employees, customers, suppliers, tax payers).

According to the *inefficient management* motive, more efficient firms will acquire less efficient firms and realize gains by improving their efficiency; this implies excess managerial capabilities in the acquiring firm (e.g., Bradley, Desai & Kim 1988). The *operating synergy* motive postulates, e.g., economics of scale and that takeovers help achieve levels of activities at which they can be obtained (e.g., Williamson 1975). The *financial synergy* motive hypothesizes complementaries between merging firms, not in managerial capabilities, but in the availability of investment opportunities and internal cash flows. A merged firm will have lower cost of capital due to lower cost of internal funds as well as possible risk reduction, savings in flotation costs, and improvements in capital allocation (e.g., Levy & Sarnat 1970, Galai & Masulis 1976, Prescott & Visscher 1980).

The theory of *strategic alignment* to changing environments motivates takeovers to take place as a response to environmental changes. External acquisitions of needed capabilities allow firms to adapt more quickly to changes, than by developing capabilities internally (e.g., Summer 1980). The *undervaluation* theory states that takeovers occur when the market value of the target firm for some reason does not reflect its true or potential value, or its value in the hands of an alternative management. Firms can acquire assets for expansion more cheaply by buying the stock of existing firms than by buying or building the assets, when the target's stock price is below the replacement cost of its assets (e.g., Chappell & Cheng 1984). In line with the undervaluation hypothesis, the *information or signaling* theory attempts to explain why target shares seem to be permanently revalued upward in a takeover whether or not it is successful. The information hypothesis states that the takeover sends a signal to the market that the target shares are undervalued, or alternatively, the offer signals information to target management which inspires them to implement a more efficient strategy on their own (e.g., Dodd & Ruback 1977, Bradley, Desai & Kim 1988).

All of the above presented motives suggest that additional value is created in takeovers. The remaining three motives argue that the gains accruing to target and bidder shareholders are merely wealth redistribution from other stakeholders in the respective firms. The *market power* hypothesis states that shareholder's wealth increases at the expense of customers (or suppliers), due to increased concentration leading to collusion and monopoly effects (e.g., Eckbo 1992). Redistribution of wealth is also the case if takeovers are motivated by *tax considerations*. In this case, shareholders gain at the expense of tax payers (e.g., Auerbach & Reishaus 1987). Finally, according to the *redistribution* hypothesis, shareholders’ gain can also accrue from bondholders due to unexpectedly increased leverage (e.g., Dennis & McConnel 1986), or from employees, who are deprived of their benefits (e.g., Shleifer & Summers 1988).

According to the synergy motives, there should always be a positive gain in takeovers for all shareholders, stemming from efficiency improvements or from other stakeholders. Therefore, it follows that the measured gain to both target and bidder shareholders is expected to be positive. The division of the gain between target firm and acquiring firm shareholders may, though, not be equally distributed, but may be skewed in favor of the target due to a number of reasons. Grossman & Hart (1980) argue that due to potential free riding by the target firm's atomistic shareholders, the smallest tender offer price the shareholders will accept is the full improvement value after a successful takeover by the bidder. Hence, the extreme case of the free riding problem suggests that the target captures the entire gain, and consequently, there is no incentive to make takeover bids at all. Fishman (1988), among others, offers bidder competition as one reason for a larger target share of the takeover gain. On the other hand, Harris (1990) argues that takeover defense measures, taken by the target firm's management, force the bidder to pay out a large share of the gain to target shareholders. Another reason for a larger target share of the gain is an upward-sloping supply curve as a result of heterogeneity in beliefs and differences in tax status, as suggested by, e.g., Stulz, Walking and Song (1990). Finally, one line of thought suggests that if the target has some bargaining power, mainly because it can resist the bidder, target shareholders may be able to extract a larger fraction of the takeover gain in an explicit or implicit negotiation with the bidder (e.g., Israel 1992, Högfeldt & Högholm 2000).

2.2 The agency motive

According to the agency theory (Jensen & Meckling 1976, Jensen 1986) it has been suggested that some takeovers are primarily motivated by the self-interest of the acquirer management. Several reasons have been advanced to explain this divergence. Among them are diversification of management's personal portfolio (Amihud & Lev 1981), use of free cash flow to increase the size of the firm (Jensen 1986), and acquiring assets that increase the firm's dependence on the management (Shleifer & Vishny 1989). The basic idea in most of these explanations is that acquisitions result in an extraction of value from the acquirer shareholders by acquirer management. For example, specialist management acquires firms in their own line of business so that the success of the combined entity will depend even more on their specific skills. The management can exploit this dependency to increase perquisite consumption or defeat rivals who are better in running some of the operations of the firm. Such management actions result in agency costs that reduce the total value of the combined firm available to shareholders.

The important aspect of the above argument is that the target firm has been identified by the acquirer management as one that is most suited to increase its own welfare. Therefore, target shareholders, realizing their value to the acquirer management, will attempt to obtain some of this value. To the extent that target shareholders have some bargaining power, they will succeed in doing so, and the value they obtain will increase with the amount that the acquirer management can appropriate. Therefore, the more severe the agency problem, the higher is the target's gain. Since greater appropriation by acquirer management also results in lower (or a negative) total gain, the observed gain to acquirer shareholder's should be small (compared to target shareholder's gain) or negative.

2.3 The hubris hypothesis

Roll (1986) hypothesizes that managers commit errors of over optimism in evaluating takeover opportunities due to excessive pride or hubris. Suppose that the bidder management is equally likely to overestimate as underestimate the synergy. If there is no synergy gain, the mean of the valuations will be the current market price. When the valuation turns out to be below the current market price, no offer is made. Consequently, an offer is made only if the valuation exceeds the current market price. Hence, the takeover premium is a random error, a mistake by the bidder. Even if gains exist but are small, some valuations will be below the current market price. No bids are made in these cases and, therefore, fewer negative errors will be observed than positive errors. At least part of the average observed takeover premium could still be the result of valuation errors and hubris.

The hubris hypothesis assumes market efficiency. Stock prices reflect all information; redeployment of productive resources cannot bring gains, and management cannot be improved through reshuffling or combinations across firms. Roll (1986) claims that the hubris hypothesis thus serves as a benchmark for comparison and is the null hypothesis against which other hypotheses should be compared. Further, the hypothesis does not require conscious pursuit of self-interest by managers. Managers may have good intentions, but can make mistakes in judgment.

Since the takeover gain, according to the hubris hypothesis, is presumed to be close to zero, the payment to target shareholders represents a transfer between the target and the acquirer. It follows that the higher the target gain, the lower is the bidder gain, and that the total gain is close to zero (e.g., Berkovitch & Narayanan 1993, Malmendier & Tate 2005).

2.4 The value creation for bidders and its determinants

All of the above presented motives suggest that target shareholders experience a gain in takeovers. On the basis of the presented motives, however, the effect for the acquirer firm's shareholders is not clear. This is also evident in the presented empirical results across different stock markets, where some find positive, some negative and some insignificant bidder returns. The takeover literature has also shown that the characteristics of the deal will affect takeover returns, and, hence, the gain to the shareholders.[2]

Kane (2000) and Moeller, Schlingemann & Stulz (2004) argue that large transactions result in value creation for the shareholders. On the other hand, e.g., Al-Sharkas (2003) shows a negative correlation between bidder abnormal return and relative size. Likewise, Bradley & Sundaram (2004) show that the announcement effect is more negative with increased target size. As argued by Hansen (1987), a possible revaluation loss will be larger for the bidder the larger the target company. Hence, there are contrasted results regarding the *size effect*, but we expect to find a negative relationship between the size of the target and the abnormal return to the bidder on the announcement day.

Cross-border takeovers may open up an opportunity for the bidder to exploit market imperfections and to expand their business into new, international markets (e.g., Moeller & Schlingemann 2005). Since these effects are unavailable in domestic takeovers, one may expect a higher wealth effect in cross-border deals. Martynova & Renneboog (2008b) argue that takeover gains may be caused by improvements in governance of the bidder and the target firm. However, if there are large institutional differences in the bidder's and the target's countries, there may also be difficulties in the post-takeover process to utilize the perceived synergies. If the market anticipates such difficulties it may discount the expected gain. Conn, Cosh, Geust & Hughes (2005) and Moeller & Schlingemann (2005) present evidence consistent with this hypothesis. Mixed results are documented with respect to cross-border acquisitions. However, we expect to find a more positive announcement effect to a cross-border deal.

Conglomerate takeovers (diversification) may create operational and financial synergies, which may lower the financial risk, and, hence, the probability that the company goes bankrupt. This may also lower the cost of debt for the company (Agrawal, Jaffe & Mandelker 1992). Diversification is also associated with a

---

[2] For an overview of the empirical evidence, see, e.g., Jensen & Ruback (1983), Agrawal and Jaffe (2000), Bruner (2004), Martynova & Renneboog (2008a).

number of disadvantages stemming from the agency problem between managers and shareholders (e.g., Doukas, Holmen & Travos 2002, Schafstein & Stein 2000), which may lead to lower takeover returns to bidders engaging in conglomerate takeovers. Mixed results have been documented for the value creation to the bidder's shareholders in conglomerate acquisitions. We expect, however, a more positive announcement response to a focused takeover compared to an announcement of a conglomerate takeover.

The *legal status* of the target company may also affect the takeover gain. A takeover involving a privately held target company may result in a higher return to the bidder's shareholders than a corresponding transaction involving a publicly traded target company (e.g., Moeller, Schlingemann & Stulz 2004, Faccio, McConnell & Stolin 2006). One reason can be a required illiquidity premium, another that the bidder may have a better negotiation power buying a private company compared to launching a public tender offer. Also the probability of the bid to succeed is higher in a private transaction. However, buying a private firm can also be considered being more risky, since there is less information available about the target firm. We expect a more positive announcement effect in a takeover transaction involving a privately held target company.

Martynova & Renneboog (2009) suggest that the *method of payment* may affect the short-term value effect of a takeover announcement. All cash bids are expected to generate a higher return to the shareholders than all-equity bids. The explanation is that asymmetric information implies that the bidder uses shares as a mean of payment when the share is overvalued, and uses cash when it is undervalued (Myers & Majluf 1984). Several studies have confirmed that the market reaction to announcements of equity offerings is significantly negative (e.g., Moeller, Schlingemann & Stulz 2004, Moeller & Schlingemann 2005, Martynova & Renneboog 2011). In line with previous studies we expect a more positive announcement effect when the bid is an all-cash offering.[3]

---

[3] Other proposed takeover characteristics that may affect the short-term value creation to the bidder are, e.g., *value vs. growth* (Lakonishok, Shleifer & Vishny 1994, Sudarsanam & Mahate 2003); *friendly vs. hostile* (Goergen & Renneboog 2004); *tender offers vs. mergers* (Rau & Vermaelen 1998); *target ownership structure* (Högfeldt & Högholm 2000, Martynova & Renneboog 2008a); *bidder toehold* (Stulz, Walking & Song 1990, Hamza 2011); *investor protection* (La Porta, Lopez-de-Silanes, Shleifer & Vishny 2002, Goergen, Martynova & Renneboog 2005, Martynova & Renneboog 2008b); *partial acquisitions* (La Porta, Lopez-de-Silanes, Shleifer & Vishny 2002); *takeover waves* (Martynova & Renneboog 2011).

3 Methodology and data

We study the short-term announcement effect to the bidder's shareholders and investigate several factors that may affect the stock market reaction to the takeover announcement. We measure the announcement effect as the sum of the daily average abnormal returns[4] (CAAR) over different windows around the announcement day[5], with a total event window of 41 days, 20 days prior to and 20 days after the event day. We also study alternative event windows before and after the announcement day to capture any effect of a price run-up before the event, or a possible value readjustment after the announcement day. The daily abnormal return (AR) is calculated as the difference between the actual return and the expected return. The expected return is calculated using the market adjusted model, the market model, and the market model with adjusted beta (the estimated beta adjusted for mean reversion (Blume 1979)). We use the value-weighted OMXHelsinki cap[6] as a proxy for the market portfolio. To estimate the parameters in the market model we use a window of 241 days, starting 300 days and ending 60 days prior to the event day.

To further study the market reaction to the takeover announcement, we regress the bidder's short-term abnormal return on several explanatory factors with respect to the characteristics of the acquisition. The key characteristics we use are the size of the transaction, the origin of the target company, the legal status of the target company, the strategic scope of the transaction (focus or diversification), and the mean of payments.

We study a sample of takeovers made by Finnish stock market listed companies during the time period from January 2000 to December 2009. The information about the acquisition is collected from the Thomson ONE Banker Database and corresponding stock exchange releases. There are a total of 1337 acquisitions during the ten year time period where the bidder is publicly traded. We restrict the sample to acquisitions where the bidder acquire a majority stake in the target (more than 50% ownership), leaving us with 855 observations. We also eliminate transactions that may be considered too small to yield an observable stock market

4 The returns are continuously compounded returns.

5 The event day (announcement day) is defined as the day when the information of the takeover was announced for the first time (or the day after if the announcement occurred after the closing of the trading day).

6 We use the restricted version of the market index, which restricts the weight of any individual company to a maximum of ten percent in the index. This is due to the large weight of some companies in the unrestricted market index, e.g., the weight of Nokia was about 60 percent in the index in year 2000.

reaction to the announcement. We set the transaction value limit to 10 million USD, leaving us with a final sample of 249 transactions.

We collect information about the characteristics of the acquisition from the Thomson ONE Banker database, from stock exchange releases and from companies' homepages. In several acquisitions, when the target firm is privately held, there is a lack of reliable information mostly regarding the transaction value and the term of payments. Some bidders disclose all the details about the bid, but since this is not mandatory for small transactions of privately held targets, there are bidders that do not disclose all details regarding the characteristics of the bid.

Table 1 presents the total sample of 249 takeovers by deal characteristics over the time period January 2000 – December 2009. The takeover activity was largest in year 2000 (49 transactions) and in the years preceding the financial crises (2005-2007) with a total of 105 transactions. The mean transaction value is 253 million USD (median 55 million USD). The relative number of cross-border bids is almost 65 percent, with year 2008 as an extreme year with 23 cross-border bids out of a total of 24 bids during that year. This is consistent with the number of cross-border bids (60 percent) for a sample of 53 Finnish bidders during the time period 1993-2001 (Martynova & Renneboog 2011). A large part of the bids are for privately held target firms (82 percent). Diversification was the dominant takeover strategy for the bidders in our sample. About 62 percent of the bids were for a target company operating in a different industry than the bidder. The terms of payment are undisclosed in 106 transactions (almost 43 percent). Of the bids for which the payment method is disclosed, the majority is cash bids (31 percent). Of the remaining bids, 34 are all equity bids (14 percent), while 22 bids (9 percent) are a mix of cash and equity.

Table 1. Descriptive statistics of takeovers in Finland

The table provides information about the characteristics of 249 takeovers with a Finnish acquirer that took place during the time period Jan 2009-Dec 2009. The table provides information about the distribution of takeovers over the sample period, and the distribution of the takeovers partitioned over different characteristics of the acquisition. *Transaction value* is the value of the bid in million USD, *cross-border bid* is a bid where the target company is of foreign origin, *public target* is a target company that is publicly traded, *diversification* is when the target operates in a different industry than the bidder according to their industry classification (SIC-code), and finally, deals classified base on the *terms of payment* (cash, equity, mixed bids and deals with undisclosed terms).

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total period	%
Total sample	49	19	23	13	15	33	34	28	24	11	249	100
Total sample (%)	19.7	7.6	9.2	5.2	6.0	13.3	13.7	11.2	9.6	4.4		
Transaction value (mean)	317	467	168	67	165	208	115	400	255	303	253	
Transaction value (median)	74	54	67	62	37	56	41	60	61	42	55	
Cross-border bid	31	12	14	8	9	20	20	17	23	7	161	64.7
Domestic bid	18	7	9	5	6	13	14	11	1	4	88	35.3
Public target	11	4	5	2	1	5	4	7	3	3	45	18.1
Private target	38	15	18	11	14	28	30	21	21	8	204	81.9
Diversification	28	9	15	8	12	17	20	21	17	7	154	61.8
Industry focus	21	10	8	5	3	16	14	7	7	4	95	38.2
Cash bid	10	4	6	3	2	15	10	11	11	5	77	30.9
Equity bid	13	5	4	3	2	2	0	1	1	3	34	13.7
Mixed bid	6	2	1	0	3	2	4	2	1	1	22	8.8
Undisclosed payment	10	8	12	7	8	14	20	14	11	2	106	42.6

4 Results

In this section we first present the results from the univariate analysis of the bidder's cumulative abnormal return in takeover transactions in Finland during the time period from January 2000 to December 2009. Secondly, we analyze the determinants of the abnormal return to the takeover announcement.

Table 2 shows that the announcement of a takeover bid on average yields a statistically significant positive abnormal return to the bidder's shareholders on the announcement day (T = 0). The average abnormal return is about 1.1 percent irrespective of how the normal return is measured; using the market adjusted model, the market model or the market model with adjusted beta. For longer event windows, up to 11 days centered around the announcement day, the cumulative average abnormal return (CAAR) increases to about 2 percent (statistically significant). This result is comparable to the results (CAAR of 2.16 percent for an 11 day event window) reported in Martynova & Renneboog (2011) for a sample of Finnish bidders during the time period 1993–2001. However, looking at the longest event windows (21 days and 41 days, respectively) we see that the CAARs using the market model and the market model with adjusted beta becomes insignificant, though still positive. This seems to be mainly driven by a share price readjustment during the time period 11 to 20 days after the announcement.

We do not document any evidence of a price run-up in the pre-event period (20 days to 1 day before the announcement). None of the pre-event windows exhibits a significant CAAR, irrespective of how we measure the normal return. We do, however, see indications of a delayed market reaction to the announcement with a positive CAAR of 0.5 percent in the event window one day to three days after the announcement. For the longer post-event windows the CAARs are not statistically significant, although there seems to be some price adjustment in the longest event window (for the market model and the market model with adjusted beta), as is evident also from Figure 1.

Overall, the results show that short-term bidder returns are positive and statistically significant. The returns accumulate mostly during a seven day event window centered on the event day, with the majority of the market reaction to the bid occurring at the announcement day. Hence, the announcement is on average a value creating event for the bidder's shareholders, indicating that the main motive for

Table 2. Cumulative average abnormal returns (CAARs) for the bidding firms

The table reports the cumulative average abnormal returns for the bidding firm in 249 takeover transactions in Finland during the time period Jan 2000-Dec 2009 over different event windows. The abnormal returns are calculated using three different approaches; the market adjusted model, the market model and the market model with adjusted beta. The market return is based on the value-weighted total return index OMX Helsinki cap. The market model parameters are estimated over a period of 241 days starting 300 days before the announcement. Statistical significance are denoted by */**/*** (10%/5%/1%).

	Market adjusted model		Market model		Market model (adjusted beta)	
Event window	(%)	(p-value)	(%)	(p-value)	(%)	(p-value)
[-20, +20]	1.908**	0.023	0.548	0.555	0.858	0.357
[-10, +10]	1.704**	0.012	1.323**	0.050	1.318*	0.051
[-5, +5]	2.114***	0.000	1.839***	0.000	1.873***	0.000
[-3, +3]	2.085***	0.000	1.891***	0.000	1.917***	0.000
[-1, +1]	1.497***	0.000	1.382***	0.000	1.411***	0.000
[T=0]	1.094***	0.000	1.177***	0.000	1.110***	0.000
[-20, -1]	0.307	0.632	-0.149	0.806	-0.054	0.931
[-10, -1]	0.092	0.841	0.079	0.858	0.041	0.926
[-5, -1]	0.447	0.240	0.390	0.280	0.407	0.264
[-3, -1]	0.324	0.315	0.257	0.404	0.293	0.349
[+1, +20]	0.508	0.408	-0.480	0.451	-0.199	0.757
[+1, +10]	0.518	0.310	0.067	0.888	0.167	0.730
[+1, +5]	0.574	0.144	0.271	0.463	0.356	0.343
[+1, +3]	0.667**	0.023	0.457	0.101	0.514*	0.066

the transaction is to create value to the shareholders.[7] This is also evident from Table 3, in which we report the number of announcements that yields a positive and a negative market reaction, respectively. Out of the 249 announcements, 156 had a positive abnormal return on the announcement day. However, we cannot rule out that some of the acquisitions are driven by hubris or agency motives, since 96 announcements yielded a negative abnormal announcement day return.

---

[7] In addition, the cumulative average abnormal return for the sample of publicly traded target firms was 32.84 percent.

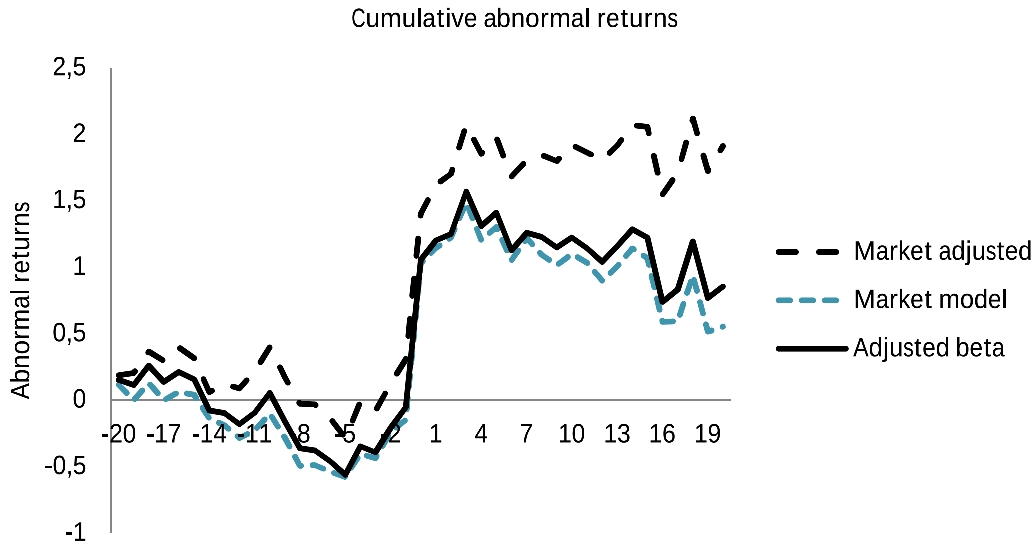


Figure 1. Cumulative average abnormal returns (CAARs) for the bidding firms

The figure shows the cumulative average abnormal returns for the bidding firm in 249 takeover transactions in Finland during the time period Jan 2000-Dec 2009 during an event window of 41 days. The abnormal returns are calculated using three different approaches; the market adjusted model, the market model and the market model with adjusted beta. The market return is based on the value-weighted total return index OMX Helsinki cap. The market model parameters are estimated over a period of 241 days starting 300 days before the announcement.

Our results are robust to the choice of estimation model of the benchmark returns and to the length of the event window. In the remainder of the paper we report the results on the announcement day using the market model returns as the benchmark returns. Using a longer event window and/or the two alternative estimation models does not materially change the results.[8]

Table 4 reports the market reaction to takeover announcements by characteristics of the deal. Both large and small transactions yield a statistically significant abnormal return on the announcement day. However, large transactions, defined as transactions larger than the median transaction value, seem to yield a lower positive announcement reaction than small transactions (0.69 versus 1.66 percent). The difference is statistically significant (prob-value 0.097). During the pre- and the post-event period we do not find any significant CAARs, neither is there any significant difference between the two groups.

---

[8] The results for alternative event windows and using the market adjusted model or the market model with adjusted beta are available upon request.

Table 3. Number of positive and negative cumulative abnormal returns (CARs)

The table reports the number of events with positive and negative cumulative abnormal returns over different event windows for bidding firms in 249 takeover transactions in Finland during the time period Jan 2000-Dec 2009. The abnormal returns are calculated using the market adjusted model, where the market return is based on the value-weighted total return index OMX Helsinki cap.

	Pre-event [-20,-1]		Event day [T=0]		Post-event [+1,+20]	
	Nobs	%	Nobs	%	Nobs	%
Positive	111	44.6	156	62.7	123	49.4
Negative	138	55.4	93	37.3	126	50.6
Total	249	100	249	100	249	100

Most of the bids made by Finnish bidders are for a foreign target company. Overall, the bidder experience a positive announcement effect for both cross-border and domestic bids, but the announcement effect is significantly lower for bidders engaging in cross-border transactions (0.60 versus 2.22 percent). No significant differences can be found during the pre- and the post-event period.

The announcement of an acquisition of a private target yields a statistically significant abnormal return of 1.36 percent, whereas the announcement of an acquisition of a public target yields a small positive (insignificant) return of 0.36 percent. There seems to be a small price run-up during the pre-event period for bidder's acquiring a public firm. The return is, though, not statistically significant.

Most of the acquisitions are diversification takeovers. In contrast to our expectations, the announcement of a diversification takeover yields a statistically significant higher return than an announcement of a related takeover (1.69 versus 0.35 percent). There are small, and insignificant, differences between the two types in the pre- and post –event period.

Finally, in Table 4, we also show the difference in the returns for cash bids versus non-cash bids. Contrary to our expectations there is no difference in the announcement of a bid based upon the terms of payment. However, looking at the entire event window the total CAAR for cash bids is about 2.7 percent, while the corresponding CAAR for non-cash bids is about -0.4 percent. The difference is, however, not statistically significant. The results may, though, be affected by the fact that in almost 43 percent of the bids the terms of payment is not disclosed, hence, there may be quite a large number of acquisitions that are misclassified.

Table 4. Cumulative average abnormal returns (CAARs) for the bidding firms by different characteristics of the acquisition

The table reports the cumulative average abnormal returns for the bidding firm in 249 takeover transactions in Finland during the time period Jan 2000-Dec 2009 over different event windows and for different characteristics of the acquisition. The abnormal returns are calculated using the market adjusted model, where the market return is based on the value-weighted total return index OMX Helsinki cap. The acquisition characteristics are the size of the bid (large transactions > 55 million USD), the origin of the target company (cross-border or domestic), the bid being for a publicly traded or privately held company, the bid being for a company within the same industry or for an unrelated company (based upon the SIC-code), and the terms of payment; cash or non-cash bid. Statistical significance are denoted by */**/*** (10%/5%/1%).

	Pre-event [-20,-1]		Event day [T=0]		Post-event [+1,+20]		
	(%)	(p-value)	(%)	(p-value)	(%)	(p-value)	Nobs
Total sample	-0.149	0.806	1.177***	0.000	-0.480	0.451	249
Large transactions	-0.561	0.455	0.689*	0.076	0.094	0.895	124
Small transactions	0.261	0.785	1.661***	0.000	-1.050	0.322	125
Diff	-0.822	0.498	-0.972*	0.097	1.144	0.370	
Cross-border bid	-0.751	0.288	0.604*	0.068	-0.450	0.581	161
Domestic bid	0.954	0.397	2.225***	0.000	-0.537	0.602	55
Diff	-1.705	0.200	-1.621**	0.013	0.087	0.947	
Public target	0.788	0.617	0.363	0.682	-1.382	0.198	45
Private target	-0.355	0.588	1.357***	0.000	-0.282	0.705	204
Diff	1.143	0.503	-0.994	0.290	-1.100	0.397	
Diversification	-0.127	0.867	1.690***	0.000	-0.470	0.565	154
Industry focus	-0.184	0.856	0.346	0.363	-0.498	0.630	95
Diff	0.057	0.964	1.344**	0.016	0.028	0.983	
Cash bid	1.091	0.219	0.917**	0.014	0.695	0.594	77
Non-cash bid	-0.703	0.369	1.294***	0.001	-1.007	0.161	172
Diff	1.794	0.129	-0.377	0.481	1.702	0.253	

The univariate tests show that there is a difference in the market response to an announcement based upon the characteristics of the bid. In Table 5 we report the results from an OLS-regression of the market reaction using the bid characteristics as explanatory variables. We also add a control variable to the regression, capturing the effect of the sixth takeover wave. For example, Jensen (2004) and Moeller et al. (2005) argue that there is a positive correlation between the sentiment on the stock market and takeover activity, and that bidders in times of high stock market valuation tend to bid more aggressively and, hence, increase the bid premium. As a consequence, the gain that accrues to the bidder's shareholders decreases. To control for this potential effect, we define the sixth takeover wave as the time period between June 2003 and December 2007 (Alexandridis, Mavrovitis & Travlos 2012), and include a dummy variable taking the value 1 if the takeover is announced during that time period. A total of 112 announcements were recorded during the sixth takeover wave.

In the analysis of the market response on the announcement day we see that most of the results from the regression analysis are consistent with the findings in the univariate analysis. Specifically, we see that there is a significant negative relationship between the bidder's abnormal return and the size of the deal, indicating that the market expect that the bidding firm may face large post-acquisition integration costs which will reduce the takeover synergy (Martynova & Renneboog 2011). Lower bidder announcement returns are observed for cross-border acquisitions, relative to domestic acquisitions. The results are consistent with findings reported in Conn et al. (2005) and in Moeller & Schlingemann (2005), indicating that the bidding firm may have difficulties in the post-takeover process to utilize the perceived synergies.

The market perceives diversification announcements to be good news, rewarding the bid with a higher positive abnormal return than a corresponding announcement of a focused acquisition. Hence, the investors consider the positive effect of risk reduction being larger than the negative effect of the agency problem. There are some indications of a lower abnormal return when the target company is publicly listed, and a higher abnormal return when the acquisition is paid for in cash, but these effects are not statistically significant. The control variable for the sixth takeover wave is also insignificant.

We also investigate the period prior to and after the acquisition. As the pre-event period we define the event window as [-20, -1], and the post-event window as [+1, +20]. Hence, we analyze the CAARs of the respective event window using the deal characteristics as the explanatory variables. As reported in Table 5, and consistent with the univariate analysis, we do not find any significant effects in the pre- or in the post-acquisition period.

Table 5. Determinants of the cumulative abnormal returns (CARs)

The table reports the results of the OLS regression of the cumulative abnormal return for the bidders in 249 takeover transactions in Finland during the time period Jan 2000-Dec 2009. *Transaction value* is the value of the bid in million USD, *cross-border bid* is a dummy variable taking the value 1 when the target company is of foreign origin, *diversification* is a dummy variable taking the value 1 when the target operates in a different industry than the bidder according to their industry classification (SIC-code), *public target* is a dummy variable taking the value 1 when the target company is publicly traded, *cash bid* is a dummy variable taking the value 1 when the acquisition is paid for in cash, and sixth wave is a dummy variable taking the value 1 if the acquisition takes place during the time period June 2003 to December 2007. All regressions contain White's heteroskedastic-consistent standard errors. Statistical significance are denoted by */**/*** (10%/5%/1%).

	Pre-event [-20,-1]		Event day [T=0]		Post-event [+1,+20]	
	Coeff	(p-value)	Coeff	(p-value)	Coeff	(p-value)
Intercept	0.879	0.545	1.584**	0.021	-1.206	0.432
Transaction value	-0.001	0.407	-0.001**	0.021	0.000	0.731
Cross-border	-1.949	0.131	-1.534**	0.012	-0.053	0.969
Diversification	0.159	0.899	1.392**	0.019	0.159	0.905
Public target	0.748	0.652	-0.778	0.318	-1.448	0.409
Cash bid	2.077	0.130	0.182	0.777	1.999	0.168
Sixth wave	-1.033	0.405	0.094	0.872	0.852	0.516
Nobs	249		249		249	
Adjusted R^2	-0.001		0.056		-0.012	
F-value	0.943	0.465	3.455***	0.003	0.514	0.798

5 Conclusions

In this study, we analyze the short-term market reactions to takeover announcements in a sample of 249 acquisitions made by stock market listed Finnish companies during the time period from January 2000 to December 2009. We document, on average, a significant positive stock market reaction to the announcement. The announcement effect is statistically significant yielding an average abnormal return of 1.1 percent on the announcement day. This result is consistent with the assumption that most of the acquisitions are motivated by synergy. Neither the pre-event nor the post-event abnormal returns are statistically significant, although there is sign of a negative price revaluation in the post-event period.

We also investigate the relationship between the market reaction to the announcement and deal characteristics. We document a significant, negative relationship between deal size and the abnormal return on the announcement day. The market reaction is more favorable to an acquisition of a domestic target company, indicating that the market believes that the acquirer may face substantial post-acquisition integration costs in cross-border transactions. We also document that an acquisition motivated by diversification yield a higher abnormal return to the bidder shareholders than an acquisition of a target firm within the same industry. Hence, the decrease in the financial risk seems to be more important than a potential increase in the agency costs. We do not find any significant relationship between the announcement effect and legal status of the target, the terms of payment or the sixth takeover wave.

References

Agrawal, A. & Jaffe, J. (2000). The post-merger performance puzzle. *Advances in Mergers and Acquisitions* 1, 7–41.

Agrawal, A., Jaffe, J. & Mandelker, G. (1992). The post-merger performance of acquiring firms: a re-examination of an anomaly. *Journal of Finance* 47, 1605–1621.

Alexandridis, G., Mavrovitis, C. & Travlos, N. (2012). How have M&As changed? Evidence from the sixth merger wave. *European Journal of Finance* 18, 663–688.

Al-Sharkas, A. (2003). Shareholder wealth effect in bank mergers: new evidence during the period 1980–2000. In *2003 Proceedings from AFFI*, Paris.

Amihud, Y. & Lev, B. (1981). Risk reduction as a managerial motive for conglomerate mergers. *Bell Journal of Economics* 12, 605–617.

Auerbach, A. & Reishaus, D. (1987). The effects of taxation on the merger decision. National Bureau of Economic Research, Working Paper.

Berkovitch, E. & Narayanan, M. (1993). Motives for takeovers: an empirical investigation. *Journal of Financial and Quantitative Analysis* 28, 347–362.

Blume, M. (1979). Betas and their regression tendencies: some further evidence. *Journal of Finance* 34, 265–267.

Bradley, M., Desai, A. & Kim, H.E. (1988). Synergistic gains from corporate acquisitions and their division between the stockholders of target and acquiring firms. *Journal of Financial Economics* 21, 3–40.

Bradley, M & Sundaram, A. (2004). Do acquisitions drive performance or does performance drive acquisitions? *SSRN Working Paper*.

Bruner, R. (2004). Does M&A pay? *Applied Mergers and Acquisitions* 1, 11–15.

Chappell, H. & Cheng, D. (1984). Firms' acquisition decisions and Tobin's Q ratio. *Journal of Economics and Business* 2, 29–42.

Conn, R., Cosh, A., Guest, P. & Hughes, A. (2005). The impact on UK acquiriers of domestic, cross-border, public and private acquisitions. *Journal of Business Finance and Accounting* 32, 815–870.

Dennis, D. & McConnell, J. (1986). Corporate mergers and security returns. *Journal of Financial Economics* 16, 143–187.

Dodd, P. & Ruback, R. (1977). Tender offers and shareholders returns: an empirical analysis. *Journal of Financial Economics* 5, 351–374.

Doukas, J., Holmen, M. & Travlos, N. (2002). Diversification, ownership, and control of Swedish corporations. *European Financial Management* 8, 281–314.

Eckbo, E. (1992). Mergers and the value of antitrust dererrence. *Journal of Finance* 47, 1005–1029.

Faccio, M., McDonnell, J. & Stolin, D. (2006). Returns to acquirers of listed and unlisted targets. *Journal of Financial and Quantitative Analysis* 41, 197–220.

Fishman, M. (1988). A theory of preemptive takeover bidding. *Rand Journal of Economics* 19, 88–101.

Galai, D. & Masulis, R. (1976). The option pricing model and the risk factor of stock. *Journal of Financial Economics* 4, 53–82.

Goergen, M., Martynova, M. & Renneboog, L. (2005). Corporate governance convergence: evidence from takeover regulation reforms. *Oxford Review of Economic Policy* 21, 243–268.

Goergen, M. & Renneboog, L. (2004). Shareholders wealth effects of European domestic and cross-border bids. *European Financial Management* 10, 9–45.

Grossman, S. & Hart, O. (1980). Takeover bids, the free rider problem, and the theory of the corporation. *Bell Journal of Economics* 11, 42–64.

Hamza, T. (2011). Determinants of short-term value creation for the bidder: evidence from France. *Journal of Management & Governance* 15, 157–186.

Hansen, R. (1987). A theory for the choice of exchange medium in mergers and acquisitions. *Journal of Business* 60, 75–95.

Harris, E. (1990). Anti-takeover measures, golden parachutes, and target firm shareholder welfare. *Rand Journal of Economics* 21, 614–625.

Högfeldt, P. & Högholm, K. (2000). A law and finance theory of strategic blocking and preemptive bidding in takeovers. *Journal of Corporate Finance* 6, 403–425.

Israel, R. (1992). Capital and ownership structures, and the market for corporate control. *The Review of Financial Studies* 5, 181–198.

Jarrell, G., Brickley, J. & Netter, J. (1988). The market for corporate control: the empirical evidence since 1980. *Journal of Economic Perspectives* 2, 49–68.

Jensen, M. (1986). Agency cost of free cash flow, corporate finance and takeovers. *American Economic Review* 76, 323–329.

Jensen, M. (2004). The agency cost of overvalued equity and the current state of corporate finance. *European Financial Management* 10, 549–565.

Jensen, M. & Meckling, W. (1976). Theory of the firm: managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3, 305–360.

Jensen, M. & Ruback, R. (1983). The market for corporate control: the scientific evidence. *Journal of Financial Economics* 11, 5–50.

Kane, E. (2000). Incentives for banking megamergers: what motives might regulators infer from event-study evidence? *Journal of Money, Credit and Banking* 32, 671–701.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. & Vishny, R. (2002). Investor protection and corporate valuation. *Journal of Finance* 57, 1147–1170.

Lakonishok, J., Shleifer, A. & Vishny, R. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance* 49, 1541-1578.

Levy, H. & Sarnat, M. (1970). Diversification, portfolio analysis and the uneasy case for conglomerate mergers. *Journal of Finance* 25, 795–802.

Malmendier, U. & Tate, G. (2005). Does overconfidence affect corporate investment? CEO overconfidence measures revisited. *European Financial Management* 11, 649–659.

Martynova, M. & Renneboog, L. (2008a). A century of corporate takeovers: what have we learned and where do we stand? *Journal of Banking and Finance* 32, 2148-2177.

Martynova, M. & Renneboog, L. (2008b). Spillover of corporate governance standards in cross-border mergers and acquisitions. *Journal of Corporate Finance* 14, 200–223.

Martynova, M. & Renneboog, L. (2009). What determines the financing decision in corporate takeovers: cost of capital, agency problems or the means of payment? *Journal of Corporate Finance* 15, 290–315.

Martynova, M. & Renneboog, L. (2011). The performance of the European market for corporate control: evidence from the fifth takeover wave. *European Financial Management* 17, 208–259.

Moeller, S. & Schlingemann, F. (2005). Global diversification and bidder gains: a comparison between cross-border and domestic acquisitions. *Journal of Banking and Finance* 29, 533–564.

Moeller, S., Schlingemann, F. & Stulz, R. (2004). Firm size and the gain from acquisitions. *Journal of Financial Economics* 73, 201–228.

Moeller, S., Schlingemann, F. & Stulz, R. (2005). Wealth destruction on a massive scale? A study of acquiring-firm returns in the recent merger wave. *Journal of Finance* 60, 757–782.

Morck, R., Shleifer, A. & Vishny, R.W. (1990). Do managerial objectives drive bad acquisitions? *Journal of Finance* 45, 31–48.

Myers, S. & Majluf, N. (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics* 13, 187–221.

Prescott, E. & Visscher, M. (1980). Organization capital. *Journal of Political Economy* 88, 446–461.

Rau, R. & Vermaelen, T. (1998). Glamour, value and post-acquisition performance of acquiring firms. *Journal of Financial Economics* 49, 223-253.

Roll, R. (1986). The hubris hypothesis of corporate takeovers. *The Journal of Business* 59, 197–216.

Scharfstein, D. & Stein, J. (2000). The dark side of internal capital markets: divisional rent-seeking and inefficient investment. *Journal of Finance* 55, 2537–2564.

Shleifer, A. & Summers, L. (1988). Breach of trust in hostile takeovers. In *Corporate takeovers: causes and consequences,* University of Chicago Press, Chicago.

Shleifer, A. & Vichny, R. (1989). Managerial entrenchment: the case of managerial-specific investments. *Journal of Financial Economics* 25, 123–139.

Stulz, R., Walking, R. & Song, M. (1990). The distribution of target ownership and the division of gains in successful takeovers. *Journal of Finance* 45, 817–833.

Sudarsanam, S. & Mahate, A. (2003). Glamour acquirers, method of payment and post-acquisition performance: the UK evidence. *Journal of Business Finance and Accounting* 30, 299–341.

Summer, C. (1980). *Strategic Behavior in Business and Government*. Boston: Little, Brown and Company.

Williamson, O. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications.* New York: Free Press.

PRICE CLUSTERING OF STOCK INDEX DERIVATIVES: THE CASE OF THINLY TRADED FINNISH MARKETS IN THE LATE 1990S

Jussi Nikkinen and Sami Vähämaa
University of Vaasa

1 Introduction

Instead of being uniformly distributed across all possible prices, the transaction prices of financial assets appear to cluster on a small subset of price grids. Empirical evidence of price clustering in stock markets is documented, for example, in Harris (1991), Christie & Schultz (1994a, 1994b), Godek (1996), Grossman, Miller, Fischel, Cone & Ross (1997), Booth, Kallunki, Lin & Martikainen (2000), Anh, Cai & Cheung (2005), and Ohta (2006), in gold markets by Ball, Torous & Tschoegl (1985), and in foreign exchange markets by Grossman et al. (1997). Furthermore, recent studies have shown that also the transaction prices of financial derivatives exhibit considerable clustering (see e.g., Chueh, 2000; Chung & Chiang, 2006; ap Gwilym & Alibo, 2003; ap Gwilym, Clare & Thomas, 1998; ap Gwilym, McManus & Thomas, 2005; Schwartz, van Ness & van Ness, 2004).[1]

Using data on FTSE 100 index futures and options, ap Gwilym et al. (1998) find evidence of extreme price clustering in derivatives markets. Similar findings have subsequently been documented by Chueh (2002) for the Nikkei 225 futures markets and in Schwartz et al. (2004) for the S&P 500 futures markets. ap Gwilym & Alibo (2003) examine the impact of the trading mechanism on price clustering, and find that the price clustering of the FTSE 100 index futures declined dramatically after the introduction of electronic trading. The impact of tick size reduction on the price clustering of Long Gilt futures contracts is investigated in ap Gwilym, McManus & Thomas (2005). They show that the reduced tick size led to an increase in the clustering of Long Gilt futures. Price clustering in open-outcry and electronic trading settings of Dow Jones IA, S&P 500 and Nasdaq-100 index

[1] Several theories have been presented to explain the observed price clustering. These include, as reviewed by Chung & Chiang (2006), haziness and bounded rationality, attraction hypothesis of Goodhart & Curcio (1991), negotiation hypothesis of Harris (1991), price resolution hypothesis of Ball et al. (1985), the implicit collusion of dealers as suggested by Christie & Schultz (1994), and human bias.

futures is compared in Chung & Chiang (2006). Although they document pervasive price clustering in both trading settings, their findings indicate more significant clustering in the prices of open-outcry traded derivatives.

In this paper, the price clustering of financial derivatives is approached from a novel perspective. Unlike previous studies, this paper examines the price clustering of stock index derivatives in less-developed and thinly traded markets. By using a unique intraday transaction data set on stock index futures and options from the Finnish markets in the late 1990s, we are able to address the impact of thin derivatives markets' special characteristics on price clustering. In thinly traded markets, the negotiation costs associated with the waiting time for a trade to occur are likely to be high. Thus, given the attraction theory proposed by Goodhart & Curcio (1991) and examined, for example, in Cooney, van Ness & van Ness (2003), the use of certain price quotations to attract trading may be hypothesized to be higher in thinly traded than in highly liquid markets. Furthermore, in thinly traded derivatives markets, market makers have an important role as liquidity providers. Since market makers have an obvious incentive to avoid odd tick quotes and thereby increase the bid-ask spread (see e.g., Christie & Schultz, 1994; Dutta & Madhavan, 1997; Grosssman et al., 1997), it may be expected that trades in which market makers acts as the second party are clustered to a higher degree.

Besides the above primary hypotheses, this paper also addresses two additional phenomena in thin market context. First, it is examined how the introduction of the euro in 1999 and the following redefinition of minimum price variation affect the price clustering in thinly traded derivatives markets. This analysis is closely related to previous studies that investigate the effects of decimalization (see e.g., ap Gwilym et al., 2005). Second, this paper examines whether thin trading affects the intraday patterns of price clustering. ap Gwilym et al. (1998), for instance, have previously documented a U-shaped clustering pattern in the FTSE 100 index futures and options markets. Similarly, Chueh (2002) reports an intraday clustering pattern for the Nikkei 225 futures markets.

A number of insights emerge from this paper. First, although electronic trading is expected to reduce the degree of clustering (see e.g., Chung & Chiang, 2006), our empirical findings suggest that both futures and option prices exhibit extreme price clustering in thinly traded derivatives markets with electronic trading settings. Second, the results show, as hypothesized, that the involvement of market makers increases the degree of price clustering. Furthermore, it is found that trading volume increases price clustering, while the price level and the volatility of the underlying index are found to be negatively related to clustering. Thus, the findings in a thin derivatives market support the negotiation hypothesis of Harris (1991).[2] Fourth, the results show that after the introduction of the euro and the redefinition of minimum price variations in terms of euros, the degree of price clustering increases significantly. This finding is consistent with the price resolution hypothesis of Ball et al. (1985).[3] Finally, in contrast to previous studies for liquid open-outcry markets, we find no evidence of intraday clustering patterns in thinly traded derivatives markets.

The remainder of the paper is organized as follows. The second section describes the institutional environment, the market structure and the dataset on Finnish stock index futures and options. The third section presents the research methodology and reports the empirical findings on price clustering in thinly traded markets. Finally, the fourth section provides concluding remarks.

2 Data

The data consist of intraday transaction prices of FOX (Finnish Options Index) index futures and options from the Helsinki Securities and Derivatives Exchange (HEX), currently known as NASDAQ OMX Helsinki. The underlying FOX index is a value-weighted index of the 25 most actively traded stocks on the HEX.[4] The sample period used in the analysis extends from January 4, 1998 to December 31, 1999. During this period, the only available intraday FOX index derivatives dataset was recorded. This dataset includes all trades on the FOX index futures and options contracts with the contract identification codes, information on the time to the nearest second, contract type, expiration date, price and trading volume. Descriptive statistics for the FOX futures and options trades are presented in Table 1. There are 11,681 futures trades in 1998 and 9,224 futures trades in 1999. The corresponding numbers of option trades are 8,554 in 1998 and 6,538 in 1999.

[2] According to the negotiation hypothesis, traders use a subset of available prices to simplify their negations, and thereby lower their cost of negotiation. The degree of price clustering is determined by the tradeoff between lower negotiation costs and lost gains-from-trade. Negotiation costs are lower if traders use a certain subset of ticks, for example only full index points rather than both full and half index points.

[3] According to the price resolution hypothesis, the observed clustering is determined by the optimal degree of price accuracy, i.e. the desired price resolution. Thus, clustering is inversely related to the level of precision concerning the true price of asset, which is determined by the amount of information available in the price discovery process.

[4] The FOX index is currently known as the OMXH25 index. Futures and options on the OMXH25 index are now traded at Eurex.

The FOX options and futures contracts expire six times a year in February, April, June, August, October, and December. The derivatives expire on the fourth Thursday of the expiration month. The time to maturity of option contracts is normally up to four months, so that there are typically two expiration months simultaneously available for trading. Similarly to the FTSE 100 futures contracts (see Gwilym et al. 1998), the trading in the FOX futures and options is concentrated on the front month up to the expiration weeks. The mean (median) time to maturity of futures is 33.2 (30) and 39.4 (38) calendar days in 1998 and 1999, respectively, and the mean (median) time to maturity of options is 31.3 (24) and 40.0 (37) days in 1998 and 1999, respectively.

Continuous trading with the FOX derivatives starts at 10:30 a.m. and ends at 5:30 p.m. local time (GMT +2 hours). Trading is organized through the members of the exchange. In 1998 and 1999, there were 30 and 26 exchange members respectively. Members operate either as dealers or market makers. Market making of the stock index derivatives is continuous from 10:45 a.m. to 5:30 p.m. Trading of the FOX derivatives is organized electronically via a computerized trading system similar to those now used e.g. in Eurex, Sydney futures exchange, and Swedish options exchange. The trading system consists of a limit order book in which every order is displayed individually. The orders are put into the book in price and time priority. In 1998, lots are multiples of five, i.e. lot sizes from 5 to 9995 in multiples of five can be traded via the computerized trading system. In 1999, along with the introduction of euro, the lot unit became one, and the lot sizes range from 1 to 9999. Smaller trades are possible to execute via telephone trading system by using quotes provided by market makers.

The Finnish currency, markka, was replaced by the euro in stock and stock derivatives trading at the end of the year 1998 (EUR 1 = FIM 5.94573). In addition, the value of an index point, FIM 100 in 1998, was changed to EUR 10 from the beginning of year 1999. Table 1 shows that in 1999, the mean trade size has increased both in the case of futures and options. The mean trade size for futures is 16.9 contracts and 24.3 contracts in 1998 and 1999, respectively. The corresponding figures for options are 26.3 and 31.7 contracts.

Table 1. Descriptive statistics of futures and options trades.

Maturity is expressed in calendar days. The Finnish currency, markka was replaced by the euro in equity and equity derivatives trading at the end of the year 1998 (EUR 1 = FIM 5.94573). The price and trading volume figures are expressed in markkas in 1998 and in euros in 1999. The value of index point is FIM 100 in 1998 and EUR 10 in 1999. To convert the FIM price figures comparable with the EUR price figures, apply the conversion factor 10 x FIM price / 5.94573.

	Futures		Options	
	1998	1999	1998	1999
No of trades	11 681	8 889	8 554	6 392
Maturity				
Mean	33,2	38,2	31,3	39,3
Median	30	37	24	36
Std Dev	21,2	25,9	24,4	29,5
Minimum	1	1	1	1
Maximum	122	122	122	122
Trade size				
Mean	16,9	24,7	26,3	31,9
Median	10	10	10	10
Std Dev	30,7	80,8	67	95,5
Minimum	1	1	1	1
Maximum	770	5045	1300	1500
Price	(FIM)	(EUR)	(FIM)	(EUR)
Mean	1 390,70	1 889,60	26,9	32,8
Median	1 369,00	1 835,00	20	25
Std Dev	178,7	377	28,4	46,6
Minimum	999	0,1	0,1	0,04
Maximum	1 740,00	3 310,00	449	1 230,00
Trading volume	(FIM)	(EUR)	(FIM)	(EUR)
Mean	23 439,60	46 737,80	804,5	1 377,50
Median	13 560,00	18 097,50	180	217,9
Std Dev	42 366,70	139 292,30	3 254,20	7 480,00
Minimum	1 000,00	1	0,3	0,4
Maximum	1 257 410	7 274 890	81 270	217 125

Table 1 also reports the mean transaction prices expressed in FIM in 1998 and in euros in 1999. The 1998 figures are comparable with the 1999 figures after applying the conversion factor $10 \cdot \text{FIM price} / 5.94573$. Table 1 shows that the mean futures transaction price measured in euros has actually decreased by 19 percent. The corresponding decrease in the case of options is 27 percent. As a result of the increased sizes of the trades, the total trading volume of futures contracts has, however, increased by 17 percent, while that of the options has remain relatively unchanged.

Besides the change of the currency unit used in trading, also another major change took place on the FOX derivatives market between the years 1998 and 1999. For futures contracts, regardless of the price, the minimum variation (tick size) is 0.25 markkas in 1998. Options contracts, in contrast, are divided into three price categories with different minimum price variation to equalize the relative tick size as shown in Table 2. This procedure is referred to as a proportional tick size, and it is used, for example, in Hong Kong, Singapore and Tokyo stock exchanges. In 1998, the tick size is 0.1 markkas for options under FIM 10, 0.25 markkas for options equal to or above 10 FIM but under FIM 30 and 0.5 markkas for options above FIM 30. In 1999, along with the introduction of the euro in trading, the proportional tick approach was abandoned. As a result, a common tick size of 1 cent was adopted for the FOX index futures and options from the beginning of 1999.

Table 2. Tick sizes.

	1998		1999	
Price category	Index Points	EUR / contract	Index Points	EUR / 10 contracts
Price < FIM 10 Price < EUR 1.68	0,10	1,68	0,01	1,00
FIM 10 ≤ Price < FIM 30 EUR 1.68 ≤ Price < EUR 5.05	0,25	4,2	0,01	1,00
Price ≥ FIM 30 Price ≥ EUR 5.05	0,50	8,41	0,01	1,00

3 Empirical findings

3.1 Price clustering in thinly traded derivatives markets

Table 3 reports preliminary statistics for the FOX index futures and option trades occurring at odd and even ticks. A tick is defined to be an even tick if the last digit of the price occurs at “0” index point and an odd tick otherwise. The Pearson chi-square test is used to test whether the observed frequencies equal the expected frequencies. Moreover, we use z-statistics to test whether the proportions of even and odd transactions are equal. Table 3 shows that the degree of price clustering in the FOX index derivatives market is high, as all the relative frequencies are statistically highly significant. In 1998, 91.4 percent (10,677) of all futures trades (11,681) occurred at even ticks, while 97.8 percent (9,022) of futures trades (9,224) occurred at the even ticks in the following year. This indicates that after the introduction of euro and the redefinition of minimum price variation in terms of euros, the degree of price clustering increases.

Table 3. Number and percentage of FOX futures and option trades occurring at odd and even ticks.

Even tick indicates that the last digit of the price occurs at 0 index point. Odd tick indicates that the last digit of the price is other than 0. Chi-square is the Pearson chi-square statistic to test whether the observed frequencies equal the expected frequencies. z-statistic is for a test of equality of the proportions of even and odd ticks.

	Futures		Options	
	1998	1999	1998	1999
No of trades	11 681	8 889	8 554	6 392
Even ticks	10 677	8 691	6 425	6 014
	91,40 %	97,80 %	75,10 %	94,10 %
Odd ticks	1 004	198	2 129	378
	8,60 %	2,20 %	24,90 %	5,90 %
Chi-square	8 010	8 115	2 158	4 969
p-value	0,000	0,000	0,000	0,000
z-statistic	89,47	90,68	46,42	70,59
p-value	0,000	0,000	0,000	0,000

The degree of price clustering in the thinly traded Finnish markets seems to be well in line with the extreme price clustering (98.4%) of the FTSE 100 futures trades reported in ap Gwilym et al. (1998). In contrast, the price clustering of FOX futures contracts is much higher than that of the Nikkei 225 futures contracts (59.6%) reported by Chueh (2000). For the FOX options, the degree of price clustering is somewhat lower than that for the futures contracts. In 1998, 71.5 percent (6,425) of the option trades (8,554) occurred at even ticks, whereas the corresponding figure in 1999 is 93.9 percent (6,140). After the price refinement of 1999 clustering appears to be close to the proportion (95.9%) documented by ap Gwilym, Clare & Thomas (1998) in the FTSE 100 stock index options market. The results reported in Table 3 are robust across sub-samples.

Table 4. Number and percentage of FOX futures and option trades occurring at odd and even ticks.

Even tick indicates that the last digit of the price occurs at 0 index point. The occurrence figure is defined as the number of trades occurring at even ticks to the total number of trades during the time interval. Median is 100% in all the cases. Probabilities of the sign test to test the hypothesis that the occurrence figure equals to 50% are less than 0.0001 in all the cases. Chi-square refers to the Kruskal-Wallis chi-square statistic to test whether the relative frequencies across the trading hours are equal.

	Futures		Options	
	1998	1999	1998	1999
10:30 – 11:29	93,10 %	97,80 %	84,60 %	95,20 %
11:30 – 12:29	91,40 %	97,50 %	77,10 %	95,20 %
12:30 – 13:29	91,70 %	96,70 %	80,40 %	95,10 %
13:30 – 14:29	90,10 %	98,00 %	76,60 %	93,90 %
14:30 – 15:29	91,80 %	98,20 %	76,10 %	92,20 %
15:30 – 16:29	91,30 %	97,80 %	77,60 %	94,90 %
16:30 – 17:30	91,20 %	97,50 %	77,00 %	91,90 %
Chi-square	17,9	3,8	32,9	20,3
p -value	0,006	0,695	0,000	0,002

Table 4 presents the intraday distribution of even ticks. The occurrence figure is defined as the number of trades occurring at even ticks to the total number of trades during a one hour time interval. The Kruskal-Wallis test is applied to test whether the relative frequencies across the trading hours are equal. Although some variation seems to exist, no systematic intraday patterns can be observed from Table 4. This is in contrast to ap Gwilym et al. (1998), who document a U-shaped intraday pattern for even-tick trades of the FTSE 100 futures contracts. ap Gwilym et al. (1998) postulate that clustering is affected by the type of the market. In open outcry settings, the use of odd ticks may be cumbersome during busy periods, which in turn may cause the U-shaped clustering pattern. As we find no evidence of intraday clustering patterns in electronic settings, the results may be interpreted to support hypothesis of ap Gwilym et al. (1998).

As a next step, this paper focuses on the negotiation hypothesis of Harris (1991) in thinly traded markets. According to this hypothesis, the probability of odd-ticks should increase with trade size. Table 5 reports the mean trade sizes for FOX futures and options at odd and even ticks. Even / Odd is the average size of trades occurring at even ticks divided by the average size of the trades occurring at odd ticks. *F*-test and Wilcoxon sign rank test are used to test the equality of the trade sizes. As can be noted from the table, odd ticks are more likely to occur in large trades. Thus, the results support the negotiation hypothesis of Harris (1991) and are also consistent with the findings reported in ap Gwilym et al. (1998).

Table 5. Number and percentage of FOX futures and option trades occurring at odd and even ticks.

Even tick indicates that the last digit of the price occurs at 0 index point. Odd tick indicates that the last digit of the price is other than 0. Even / Odd is the average size of trades occurring at even ticks divided by the average size of the trades occurring at odd ticks. F-test and Wilcoxon sign rank test are used to test the equality of the trade sizes.

	Futures		Options	
	1998	1999	1998	1999
No of trades	11 681	8 889	8 554	6 392
Even ticks	16	23,7	25	31,1
Odd ticks	26,5	67,4	30,1	44,3
Even / Odd	0,6	0,35	0,83	0,7
F-statistic	109,24	56,91	9,03	7,9
p-value	0,000	0,000	0,003	0,005
Wilcoxon *z*	4,28	3,47	5,56	–2.40
p-value	0,000	0,001	0,000	0,016

3.2 Role of market makers

In 1998, market makers acted as a second party in all odd lot trades. This feature enables us to examine whether the trades with a market maker acting as a second party are clustered to a higher degree than other trades. We compare the clustering of odd lot trades to that of the smallest round lot trades (trade size equal to five) separately for futures and options both in 1998 and in 1999. Given that odd lot trade in 1998 always implies market maker involvement, it is hypothesized that odd lot trades are more clustered than round lot trades in 1998. In contrast, no differences between the clustering of odd lot and round lot trades in 1999 are expected to be found.

Table 6 reports the results regarding the impact of market maker involvement on price clustering. Panel A shows the number of odd lot trades and round lot trades with sizes equal to five in 1998. There were 2,498 odd lot futures trades and 1,270 option trades, and 2,232 and 1,722 round lot futures and option trades, respectively. Similarly, for the year 1999, Panel A reports the number of trades which are not multiples of five corresponding to the odd lot trades in 1998, and trades with sizes are equal to five. These figures are 2,658 and 1,188 for futures trades and 1,418 and 769 for options. The number of trades that are not multiples of five (i.e., odd lots in 1998), has slightly increased, while the number of the trades with sizes equal to five (round lots in 1998) has decreased.

Panel A of Table 6 also presents the proportions of odd lot and round lot trades occurring at even tick. The table shows that in 1998, 91.4 percent of the odd lot futures trades and 91.5 percent of the round lot futures trades occur at even tick. The difference is -0.1 percentage points and statistically insignificant. This finding is inconsistent with the hypothesis of market maker involvement. However, it should be noted that market makers may also be substantially involved in round lot futures trades. In 1999, the difference between the groups of futures trades is -0.7 percentage points and not statistically significant, as expected.

Turning the focus onto the option trades, Panel A of Table 6 shows evidence of significant differences between the price clustering of odd lot and round lot trades. In 1998, 81.1 percent of the odd lot trades and 75.1 percent of the round lot trades occur at even tick. The difference of 6 percentage points is statistically significant with a *p*-value less than 0.001. In contrast, and as hypothesized, there is no statistically significant difference between the groups in 1999. These findings provide support for the hypothesis that market maker involvement increases the degree of price clustering.

Table 6. Number and percentage of FOX futures and option trades occurring at odd and even ticks.

Even tick indicates that the last digit of the price occurs at 0 index point. Odd tick indicates that the last digit of the price is other than 0. Even / Odd is the average size of trades occurring at even ticks divided by the average size of the trades occurring at odd ticks. *z*-statistic is for a test of equality of the proportions of even and odd ticks. Panel B reports the estimation results of a probit model in which clustering depends on trading volume, price, volatility, time-to-maturity and market maker involvement. *LR* is the likelihood ratio test statistic and *LRI* is the likelihood ratio index. Pseudo R^2 is the McKelvey and Zavoina R^2.

Panel A. The occurrence of even ticks in round and odd lot trades

	Futures		Options	
	1998	1999	1998	1999
No of trades	4 730	3 846	2 992	2 187
Odd lots	2 498	2 658	1 270	1 418
Round lots	2 232	1 188	1 722	769
Odd lots	91,40 %	97,30 %	81,10 %	94,40 %
Round lots	91,50 %	98,00 %	75,10 %	92,80 %
Difference	-0,1	-0,7	6,0	1,6
z-statistic	-0,12	-1,36	3,96	1,44
p-value	0,902	0,173	0,000	0,151

Panel B. The relationship between odd ticks and market maker involvement in option trades

Variable	Sign	Estimate	*t*-stat.	*p*-value
Constant		-0,298	-4,93	0,000
Volume	+	2,043	1,83	0,067
Price	–	-3,077	-8,54	0,000
Volatility	–	-0,769	-4,1	0,000
Expiration	–	-0,186	-1,28	0,202
Market maker	–	-0,137	-2,55	0,011
		Total	Even	Odd
No of trades		2 992	77,70 %	22,30 %
LR-statistic			135	0,000
LRI			0,04	
Pseudo-R^2			0,13	

Since the degree of price clustering may be expected to depend also on several other factors besides market maker involvement, we examine the robustness of our findings in the following model setting:

$$\Pr(ODD_i = 1) = \Phi(\mathbf{x}_i'\boldsymbol{\beta}) \ , \ \ i = 1, 2, ..., n \ , \tag{1}$$

where ODD_i is a dummy variable indicating that the last digit of the price is other than "0", $\mathbf{x}_i' = [1 \ \ Volume_i \ \ Price_i \ \ Volatility_i \ \ Expiration_i \ \ Market\ maker_i]$, $\boldsymbol{\beta}$ is a (6×1) parameter vector and $\Phi(\cdot)$ is a cumulative standard normal distribution function. The variables of the $(n \times 6)$ design matrix, $\mathbf{X}' = [\mathbf{x}_1 \ \ \mathbf{x}_2 \ \ ... \ \ \mathbf{x}_n]$, are defined as follows: $Volume_i$ is the number of derivative contracts in trade *i* divided by the mean trading volume (number of derivative contracts) in 1998, $Price_i$ is the price in trade *i* divided by the mean price in 1998, $Volatility_i$ is the volatility of the underlying stock index estimated from the daily price observations applying the Parkinson (1980) extreme value estimator, $Expiration_i$ is a dummy variable that equals one if trade *i* occurs one day prior to the expiration week, and $Market\ maker_i$ is a dummy variable that indicates whether the second party of trade *i* is a market maker.

The model is estimated by maximizing the following likelihood function:

$$\ln \ell = \sum_{i=1}^{T} \{ ODD_i \log[\Phi(\mathbf{x}_i'\boldsymbol{\beta})] + (1 - ODD_i) \log[1 - \Phi(\mathbf{x}_i'\boldsymbol{\beta})] \} \ . \tag{2}$$

The maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ is consistent, asymptotically efficient and asymptotically normally distributed (see e.g. Amemiya 1985). Based on the negotiation hypothesis, the coefficient estimate of $Volume_i$ is expected to be positive, whereas the estimates for $Price_i$ and $Volatility_i$ are expected to be negative. Moreover, it is hypothesized that the coefficient estimates of $Expiration_i$ and $Market\ maker_i$ are negative.

Panel B of Table 6 reports the estimates on the relation between odd ticks and market maker involvement with trading volume, trade price, volatility variable, and expiration dummy as control variables. The table also reports the likelihood ratio test statistic (*LR*), the McFadden likelihood ratio index (*LRI*), and the McKelvey and Zavoina pseudo-R^2 that is considered to be close to what the OLS R^2 would be using the underlying latent index implicit in the model. As can be noted from the table, 77.7 percent of the 2,992 analyzed trades occur at even ticks and 22.3 percent at odd ticks. The likelihood ratio statistic is 135 with a *p*-value smaller than 0.001, thereby indicating that the coefficient estimates in the multivariate analysis are significant together. The likelihood ratio index is 0.04, while the pseudo-R^2 of the model is 0.13.

All the coefficient estimates reported in Table 6 have their expected signs. Of the control variables, the coefficients of $Price_i$ and $Volatility_i$ are negative and statistically significant at the 0.1 percent level. The coefficient estimate of $Volume_i$ is positive with a p-value of 0.067. This relatively high p-value is likely to be caused by the fact that there is not much variation in the trading volumes of the trades included in the regression. The coefficient estimate of $Expiration_i$ is negative albeit statistically insignificant. The primary issue of interest in Table 6, however, is the sign and the significance of the coefficient estimate of $Market\ maker_i$. As can be seen from the table, the coefficient of the $Market\ maker_i$ dummy is negative and has a p-value of 0.011. This suggests that odd ticks are less evident in trades in which a market maker is the second party than in round lot trades. In general, these findings provide further support for the hypothesis according to which market maker involvement increases the degree of price clustering.

3.3 Tick size and price clustering

To test the hypothesis according to which there exists a negative relationship between tick size and the degree of price clustering, we next examine clustering simultaneously across price categories and across two time periods. In the case of futures contracts, the model given by Equation (1) is re-estimated with the following definition of the design matrix:

$$\mathbf{x}_i' = \begin{bmatrix} 1 & Y98_i & Volume_i & Price_i & Volatility_i & Expiration_i \end{bmatrix},$$

where $Y98_i$ is a dummy variable that equals one if futures trade i occurs in 1998 and zero if the trade occurs in 1999, $Volume_i$ is the number of futures contracts in trade i divided by the mean trading volume (number of derivative contracts) of the corresponding year, and $Price_i$ is the price in trade i divided by the mean price of the corresponding year. The other variables are defined as previously. $\boldsymbol{\beta}$ is a (6×1) vector of parameters and it is estimated by maximizing the likelihood function given by Equation (2).

For options, the model is estimated with the following definition of the design matrix:

$$\mathbf{x}_i' = \begin{bmatrix} 1 & D10_i & D25_i & D50_i & Volume_i & Price_i & Volatility_i & Expiration_i & Market\,maker_i \end{bmatrix}$$

where $D10_i$, $D25_i$, and $D50_i$ are dummy variables that indicate if option trade i belongs to the lowest price category, to the middle price category, or to high price category, respectively. The remaining variables are defined as previously. $\boldsymbol{\beta}$ is a (9×1) vector of parameters. Based on the hypothesis presented in the introduction, the coefficient estimate of the dummy variable $Y98_i$ and those of the dummy variables $D10_i$, $D25_i$, and $D50_i$ are expected to be positive. Moreover, the coefficient estimate of $Volume_i$ is expected to be positive, while the coefficient estimates of $Price_i$, $Volatility_i$, $Expiration_i$, and $Market\ maker_i$ are expected to be negative.

Table 7 presents the results on the impact of 'decimalization', trading volume, price level, volatility, expiration and market maker involvement on price clustering of thinly traded stock index derivatives. The table shows that 5.8 percent of the 20,570 futures trades occur at odd ticks. For options, the corresponding proportion is 16.8 percent of the 14,946 trades. In case of futures trades, the likelihood ratio statistic is 516.8 with a *p*-value smaller than 0.001, thereby suggesting that the coefficient estimates are significant together. The value of the likelihood ratio index is 0.06 and the pseudo-R^2 is 0.10.

Table 7. Number and percentage of FOX futures and option trades occurring at odd and even ticks.

The table reports the estimation results of a probit model in which clustering depends on tick size dummy variable(s), trading volume, price, volatility, time-to-maturity and market maker involvement. *Y98* is a dummy variable that equals one if futures trade i occurs in 1998 and zero if the trade occurs in 1999. *D10*, *D25*, and *D50* are dummy variables that indicate if option trade *i* belongs to the lowest price category, to the middle price category, or to high price category, respectively. LR is the likelihood ratio test statistic and LRI is the likelihood ratio index. Pseudo R^2 is the McKelvey and Zavoina R^2.

Variable		Sign	Futures Estimate	Futures *t*-stat.	Futures *p*-value	Options Estimate	Options *t*-stat.	Options *p*-value
Constant			-1,881	-17,26	0,000	-1,423	-39,53	0,000
	D*10*	+				1,342	32,45	0,000
Y*98*	D*25*	+	0,669	19,52	0,000	0,880	25,97	0,000
	D*50*	+				0,567	13,16	0,000
Volume		+	3,219	8,86	0,000	1,871	4,28	0,000
Price		–	-5,724	-0,57	0,567	-1,748	-1,09	0,277
Volatility		–	-0,597	-5,3	0,000	-0,646	-6,16	0,000
Expiration		–	0,085	0,99	0,321	-0,333	-4,44	0,000
Market		–				-0,167	-3,70	0,000
			Total	Even	Odd	Total	Even	Odd
No of trades			20 570	94,20 %	5,80 %	14 946	83,20 %	16,80 %
LR-statistic				516,8	0,000		1 514,90	0,000
LRI				0,06			0,11	
Pseudo-R^2				0,10			0,20	

The coefficient estimate of the $Y98_i$ dummy is positive and statistically significant at the 0.1 percent level. This indicates, as hypothesized, that the frequency of futures trades occurring at odd ticks is significantly higher in 1998 than in 1999. The coefficient of $Volume_i$ is also positive and statistically significant at the 0.1 percent level, thus providing support for the negotiation hypothesis. The larger futures trades are more likely to occur at odd ticks than the smaller ones. The coefficient estimate of the trade price has the expected sign, but surprisingly it is not statistically significant. In contrast, the coefficient for volatility is negative, as expected, and statistically significant at the 0.1 percent level. Consistent with Harris (1991), this suggests that high volatility decreases the propensity to trade futures at odd ticks.

In the case of option trades, the signs of all the coefficient estimates are as expected. The likelihood ratio statistic for the option regression is 1514.9 and the corresponding p-value is smaller than 0.001. The coefficient estimates of the $D10_i$, $D25_i$, and $D50_i$ dummy variables are positive and statistically significant at the 0.1 percent level. This suggests that the frequency of option trades occurring at odd ticks is significantly higher in 1998 than in 1999. This finding is consistent with the results for the futures trades discussed above. Furthermore, as can be seen from Table 7, the coefficient estimate of $D10_i$ is larger than those of $D25_i$ and $D50_i$ and the coefficient estimate of $D25_i$ is larger than that of $D50_i$. This indicates that the propensity to trade at odd ticks is highest in the low price category and lowest in the high price category.

The results based on the entire sample are consistent with the results obtained with the sub-sample in Table 7. Moreover, the findings regarding price clustering in options markets are very similar to the findings in futures markets. The results on the relation between clustering and trading volume, price level and volatility are also consistent with the existing literature.

4 Conclusions

This paper examines the price clustering of stock index derivatives in less-developed and thinly traded markets. By using a unique historical transaction data set on stock index futures and options from the Finnish markets in the late 1990s, this paper addresses the impact of thin derivatives markets' special characteristics on price clustering. In thinly traded markets, the negotiation costs associated with the waiting time for a trade to occur are likely to be high. Thus, given the attraction theory, the use of certain price quotations to attract trading may be hypothesized to be higher in thinly traded than on highly liquid markets. Furthermore, in thinly traded derivatives markets, market makers have an important role as liquidity providers. Since market makers have an obvious incentive to avoid odd tick quotes and thereby increase the bid-ask spread, it may be a priori expected that trades in which market makers acts as the second party are clustered to a higher degree. By focusing on the price clustering of thinly traded Finnish stock index derivatives, this paper provides new insights into the clustering and price formation in derivatives markets.

This paper also addresses two additional phenomena in thin market context. First, it is examined how the introduction of the euro in 1999 and the following redefinition of minimum price variation affect the price clustering in thinly traded derivatives markets. This analysis is closely related to previous studies that investigate the effects of decimalization. Second, this paper examines whether thin trading affects the intraday patterns of price clustering.

A number of insights emerge from this paper. First, the results show that both futures and option prices exhibit extreme price clustering in thinly traded derivatives markets with electronic trading settings. Second, the empirical findings demonstrate, as hypothesized, that the involvement of market makers increases price clustering. Furthermore, it is found that trading volume increases price clustering, while the price level and the volatility of the underlying asset are found to be negatively related to clustering. Thus, the findings in a thin derivatives market support the negotiation hypothesis of Harris (1991). Fourth, the results show that after the introduction of the euro and the redefinition of minimum price variations in terms of euros, the degree of price clustering increases significantly. This finding is consistent with the price resolution hypothesis of Ball et al. (1985). Finally, in contrast to previous studies for liquid open-outcry markets, there is no evidence of intraday clustering patterns in thinly traded derivatives markets.

References

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, MA, USA: Harvard University Press.

Anh, H.-J., Cai, J. & Cheung, Y.L. (2005). Price clustering on the limit-order book: Evidence from the Stock Exchange of Hong Kong. *Journal of Financial Markets* 8, 421-451.

ap Gwilym, O., Clare, A. & Thomas, S. (1998). Extreme price clustering in the London equity index futures and options markets. *Journal of Banking and Finance* 22, 1193–1206.

ap Gwilym, O. & Alibo, E. (2003). Decreased price clustering in FTSE100 futures contracts following a transfer from floor to electronic trading. *Journal of Futures Markets* 23, 647–659.

ap Gwilym, O., McManus, I. & Thomas, S. (2005). Fractional versus decimal pricing: evidence from the UK long gilt futures market. *Journal of Futures Markets* 25, 419–442.

Ball, C.A., Torous, W.A. & Tschoegl, A.E. (1985). The degree of price resolution: the case of gold market. *Journal of Futures Markets* 5, 29-43.

Booth, G., Kallunki, J.-P., Lin, J.C. & Martikainen, T. (2000). Internalization and stock price clustering: Finnish evidence. *Journal of International Money and Finance* 19, 737–751.

Brown, S., P. Laux & B. Schachter (1991). On the existence of an optimal tick size. *Review of Futures Markets* 10, 50–72.

Brown, P., Chua, A. & Mitchell, J. (2002). The influence of cultural factors on price clustering: evidence from Asia-Pacific stock markets. *Pacific-Basin Finance Journal* 10, 308–322.

Christie, W.G. & Schultz, P.H. (1994). Why do NASDAQ market makers avoid odd-eight quotes? *Journal of Finance* 49, 1813–1840.

Cooney, J., van Ness, B. & van Ness, R. (2003). Do investors avoid odd-eighth prices? Evidence from NYSE limit orders. *Journal of Banking and Finance* 27, 719–748.

Chueh, H. (2000). Price clustering in the Nikkei 225 stock index futures contract on the SIMEX: and intraday empirical analysis. *Review of Pacific Basin Financial Markets and Policies* 3, 519–533.

Chung, H. & Chiang, S. (2006). Price clustering in E-mini and floor-traded index futures. *Journal of Futures Markets* 26, 269–295.

Dutta, P. & Madhavan, A. (1997). Competition and collusion in dealer markets. *Journal of Finance* 52, 245–276.

Goodhart, C. & Curcio, R. (1991). The clustering of bid-ask prices and the spread in the foreign exchange market. London School of Economics, Financial Markets Group. Discussion Paper No. 110.

Godek, P. (1996). Why Nasdaq market makers avoid odd-eight quotes. *Journal of Financial Economics* 41, 465–474.

Grossman, S.J., Miller, M.H., Fischel, D.R., Cone, K.R. & Ross, D.J. (1997). Clustering and competition in asset markets. *Journal of Law and Economics* 40, 23–60.

Harris, L (1991). Stock price clustering and discreteness. *Review of Financial Studies* 4, 389–415.

Huang, R.D. & Stoll, H.R. (2001). Tick size, bid-ask spreads, and market structure. *Journal of Financial and Quantitative Analysis* 36, 503–522.

Kandel E. & Marx, L.M. (1999). Odd-eight avoidance as a defense against SOES bandits. *Journal of Financial Economics* 51, 85–102.

Mitchell, J. (2001). Clustering and psychological barriers. The importance of numbers. *Journal of Futures Markets* 21, 395–428.

Ohta, W. (2006). An analysis of intraday patterns in price clustering on the Tokyo Stock Exchange. *Journal of Banking and Finance* 30, 1023–1039.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business* 53, 61–65.

Schwartz, A. Van Ness, B.F. & Van Ness, R.A. (2004). Clustering in the futures markets: evidence from S&P 500 futures contracts. *Journal of Futures Markets* 24, 413–428.

FOREIGN DIRECT INVESTMENT OWNERSHIP MODE STRATEGY AND THE INFLUENCES OF PILLARS OF ECONOMIC FREEDOM

Ahmad Arslan
University of Vaasa
Syed Mujahid Hussain
Hanken School of Economics
Jorma Larimo
University of Vaasa

1 Introduction

The political and economic maps have changed due to the process of globalization in the last two decades and now the countries are more dependent on each other due to increased flow of goods and services across borders (Dunning 1993; Dunning & Lundan 2008; Krugman 2008). Increase in the flows of foreign direct investment (FDI) is also a clear indicator of increased interdependence among economies globally (UNCTAD 2009, 2011). When multinational corporations (MNCs) enter new international markets using FDI mode; one of their key decisions relate to the choice between formation of wholly owned subsidiaries (WOSs) or joint ventures (JVs) with one or more local/third country partners (e.g. Delios & Beamish 1999; Padmanabhan & Cho 1999; Chang & Rosenzweig 2001; Jung et al. 2008). This decision is referred as FDI ownership mode strategy in the international business (IB) literature (e.g. Arslan & Larimo 2010). Many previous IB studies addressed the ownership mode strategy of MNCs by studying the WOSs vs. JVs strategy using a variety of theories notably transaction cost economics, resource based view, OLI paradigm and institutional theories (e.g. Anderson & Gatignon 1986; Luo 2001; Xu et al. 2004; Dunning 2001, 2004; Brouthers & Hennart 2007; Jung et al. 2008; Slangen & van Tulder 2009; Arslan 2012; Arslan & Larimo 2012).

MNCs entering foreign markets through FDI mode encounter an important issue that each host country represents a unique institutional environment, which can influence their strategies substantially (e.g. Peng 2002, 2003; Hitt et al. 2006; Estrin et al. 2009; Arslan 2011). Therefore, MNCs are confronted with different aspects of institutional environments of the host countries that may differ significantly from their home countries (e.g. North 1990; Delios & Henisz 2003; Dikova et al. 2010; Arslan 2011). An important characteristic of institutional environments of the host countries concerns the differing levels of economic freedom in the host economies (Friedman 1982; Krugman 1991; Gwartney et al. 2008). Famous economists like Friedman (1982), North (1990, 2005), North et al. (2009), and Krugman (1991, 2008) have stressed the importance of economic freedom in the country for economic development as well as increased FDI inflows and activities by foreign MNCs in those economies. Economic freedom in a country can be defined by its key ingredients which are personal choice, voluntary exchange coordinated by markets, freedom to enter and compete in markets, and protection of persons and their property from aggression by others" (Gwartney et al. 2008: 3).

Previous IB studies addressing the market entry strategies of firms have addressed institutional environment and institutional differences across countries by mostly concentrating on institutional development (e.g. Brouthers 2002; Dikova & van Witteloostuijn 2007), and institutional distance (e.g. Xu & Shenkar 2002; Xu et al. 2004; Gaur & Lu 2007; Estrin et al. 2009; Arslan & Larimo 2010, 2011). However, the factors concerning economic freedom have generally been rather scarcely researched as being key determinants of FDI entry mode strategies of MNCs. Arslan (2011), Arslan & Larimo (2012), and Demirbag et al. (2011) used some aspects of economic freedom from economic freedom of the world annual reports (Gwartney et al. 2008, 2009) in their studies on different aspects of MNC's strategies. However, so far no previous IB study at least to our knowledge has analyzed the impacts of different dimensions of economic freedom together on FDI and market entry strategies of MNCs in their international markets.

Our study attempts to address and hypothesize FDI market entry strategies of the MNCs based on different aspects of economic freedom including size of government, legal structure and security of property rights, access to sound money, freedom to trade internationally and regulation of credit, business and labor. This study contributes to IB and market entry studies as it is one of the pioneering efforts to analyze impacts of different aspects of economic freedom in the host country on the FDI ownership entry strategy of MNCs. It has been mentioned earlier that IB studies lack a comprehensive analysis of impacts of different dimensions of economic freedom on FDI ownership strategy of MNCs. Our study aims to perform in-depth analysis of the FDI ownership mode strategy of the investing MNCs in response to different pillars of economic freedom. Therefore, this paper aims to fill the existing research gap as well as enhance the understanding of this important strategy of MNCs from these unique dimensions. The empirical context of this study is based on FDIs made in emerging economies of Asia by the MNCs from a small, open and highly internationalized Nordic economy (Finland). The home country of investing MNCs in our study represents a country with relatively high level of economic freedom, while most of the host economies are still classified as emerging economies and are in transition to development of market economy. We believe that this empirical context along with theoretical advancement of IB research adds further value to our paper making it interesting for both academics as well as practitioners.

Our paper starts with a theoretical discussion and hypothesis development, followed by a discussion about the sample and the empirical analysis. The paper concludes with discussion about study findings, managerial implications as well the directions for future research.

2 Theoretical Discussion and Study Hypotheses

Economic freedom in different societies has been referred as one of the most important pre-requisite of economic activity, growth and development in human societies over the centuries by many economists and historians (e.g. von Mises 1957). Economic theory indicates that economic freedom affects incentives, productive effort, and the effectiveness of resource use. Islam (1996) used the cross-sectional data from 98 low-, middle-, and high-income countries, and found that economic freedom has a direct relation with per capita income and economic growth rate. Bengoa and Sanchez-Robles (2003) argued that economic freedom in the emerging and developing economies is a positive determinant of FDI inflow and increasing economic freedom is a key priority of policy makers. Cole (2003) and Gwartney et al. (2009) have shown the countries with greater economic freedom – the protection of private operating markets and of private property with minimal government interference – receive increased FDI inflows and market entries by the MNCs, in comparison to the countries with lower levels of economic freedom.

It has been referred earlier in this paper that economic freedom in a country has also been shown to have strong linkages with increased FDI inflows and activities by MNCs in those economies (e.g. Friedman 1982; Krugman 1991, 2008). Therefore, we expect different aspects of economic freedom to affect FDI ownership mode strategy of the MNCs too. Following the classification of economic freedom in countries into five dimensions (pillars) by Gwartney et al. (2008, 2009), we discuss the impacts of each of these on FDI ownership mode strategy of MNCs as follows.

Size of Government: Government matters in business (Ring et al. 2005) and even more so in developing and emerging economies (Watkins 2003). The role of host country governments require much attention of MNCs since they are in control of resources and opportunities that shape their business and institutional environment (e.g. Baron 1995; Child & Tse 2001; Makino et al. 2004). The government size can be referred as a sum of government components, but intrinsically it refers to the power of the government. Most studies tend to focus on the inherent sense of government size. For example, Tridimasa & Winer (2005) pointed out that government size mainly depends on the use of fiscal policy for compulsory redistributions, the delivery of public services, the size of tax revenue and political influence.

The government plays an important role in economy, and its intervention is a double-edged sword (Yuan et al. 2010). On one hand, it can effectively allocate social resources and avoid market failure, which is good for reducing investment risks; policies can influence economy promptly. On the other hand, the regulation is not good for optimizing the use of resources; at the same time, with the increase of government size, burdens of the society increase due to its strong linkages with taxation (Wang et al. 2007). More money is needed to maintain the function, and many problems such as inefficiency and corruption occur (e.g. Kolstad & Villanger 2008). The governments have ownership of state-owned enterprises, have control on key infrastructure such as electricity, water system, railways and maintain large public sector employment. To finance all these expenses, taxation is the biggest tool available to the governments. Hence, it is one of the most important factors considered by foreign investors, and high tax rate means high costs and low profits. Thus, a country with low corporate tax rate is much more attractive to the investing MNCs. There has also been a considerable amount of quantitative research on the effect of corporate taxation on the level and direction of FDI (e.g. Hartman 1984; Slemrod 1990). Some studies have also used aggregate FDI data (Devereux & Freeman 1995; Doucouliagos & Ulubasoglu 2006) and firm-level data (Cummins & Hubbard 1995; Grubert & Mutti 2000) to evaluate the effect of corporate taxation on the FDI strategies and decisions of MNCs. These studies further reinforce the conclusion that movements in a country's level of corporate taxation are negatively associated with the country's level of inward FDI.

We argue that if the host country will have large size of government and public sector, it is highly probable that it can lead to high tax rates especially on foreign MNCs. These conditions can discourage the MNCs to commit more resources to that particular host economy by formation of WOSs, and hence they may prefer JVs in order to share risks and costs. Therefore, we hypothesize

Hypothesis 1: *Size of government in the host country is negatively associated with the preference of WOSs by the investing MNCs.*

Legal Structure and Security of Property Rights: According to Globerman and Shapiro (2003: 19), "the rule of law in a country is codified by its governance infrastructure, which represents attributes of legislation, regulation, and legal systems that condition freedom of transacting, security of property rights, and transparency of government and legal process". All these aspects of legal structure and security of property rights are important for both local and foreign business firms operating in a certain country. Therefore, the existing laws and rules in a particular host country tend to promote certain types of business behaviors and strategies, while restricting others. In many cases, the laws and legal regulations of countries are rather clearly stated (Scott 2008) and MNCs tend to understand them rather quickly (Xu & Shenkar 2002; Eden & Miller 2004).

The differences in the legal structures and security of property rights across countries create unique challenges for MNCs' business strategy, because institutions alter the costs of engaging in business activities in one host country compared with another host country (Henisz 2004). It has been empirically established in the literature that legal and institutional infrastructure of a host country attracts more foreign direct investment by the MNCs (North 1990; Hitt et al. 2006) because they guarantee the reduction of transaction hazards and opportunism (Henisz 2000; Meyer 2001). Many MNCs are entering emerging markets of Asia, where strength of legal structure and security of property is relatively low as compared to developed (mostly western) economies (Khanna & Palepu 1997, 2010; Hitt et al. 2004, 2005). However, the implementation of the market supporting economic reforms makes legal structure and security of property rights strong in a country (e.g. Peng 2003; Peng et al. 2008). Moreover, it also results in facilitating interactions that reduce uncertainties and lower transaction and operating costs (North 1990; Khanna et al. 2005) as well as provide a more favorable institutional context for MNCs, which in turn can motivate MNC to show more resource commitment to the host country (Bergara et al. 1998; Child & Tsai 2005). In this case, the investing firms may prefer WOSs formation because the strong legal structure and high security of property rights reduces their uncertainty (Chung & Beamish 2005) and motivate it to commit more resources to that particular host economy (Arslan 2012). Therefore, we hypothesize that

Hypothesis 2: *Strong legal structure and security of property rights in host country is positively associated with the preference of WOSs by the investing MNCs.*

Access to Sound Money: An efficient financial system is essential to an economy because better financial services enable an economy to function well and attract foreign investment (Klodko 2000). Levine (1997) refers to the functions of a financial system into five basic tasks: 1) facilitate the trading, hedging, diversify-ing, and pooling of risk, 2) allocate resources, 3) monitor managers and exert corporate control, 4) mobilize savings, and 5) facilitate the exchange of goods and services." Financial factors are also important in FDI decision of investing MNCs because they affect the cost structure of investment projects.

It has been referred in past studies that FDI flows are positively correlated with the development of the financial markets in host economies. For example, Claessens et al. (2001) examine the determinants of the growing migration of stock market activity to international financial centers for 77 countries between 1975 and 2000. They argue that FDI goes to countries with good institutions and fundamentals helps in development of the domestic financial system. Their results show that FDI is positively correlated with market capitalization and domestic value traded, suggesting that FDI is a complement and not a substitute of domestic stock market development. Agarwal & Mohtadi (2004) study the role of financial market development in the financing strategy of firms in 21 developing countries in different geographical regions over the period 1980–1997. They show that FDI as a proportion of GDP, and investment as a proportion of GDP are positively correlated with both the stock market variables and the banking variables. Finally, Jeffus (2004) examines the relationship between FDI and stock market development in four emerging economies for the period 1988–2002. The results show that stock market development and FDI are highly and positively correlated.

As, referred earlier that FDI inflows are positively correlated to development of financial sector especially stock market, it is also expected to have influences on FDI ownership mode strategy of MNCs. We expect access to sound money in host countries to be positively associated with preference of WOSs by the MNCs, as a healthy financial system can motivate them to commit more resources to that particular emerging economy. Therefore, we hypothesize that

Hypothesis 3: *Access to sound money in the host country is positively associated with the preference of WOSs by the investing MNCs.*

Freedom to Trade Internationally: One of the often cited problems by MNCs in the context of expanding to overseas markets relates to the restrictions on their market entry, expansion and trade strategies by the host country governments (e.g. Reardon et al. 1996; Trevino & Mixon 2004). International economics literature clearly mentions that the emerging and developing economies should attempt to remove restrictions on investments especially by the foreign MNCs, as it will result in higher entry rates (Krugman 1991; IMF 2003; Klapper et al. 2004; World Bank 2005; Flores & Aguilera 2007), which in turn also contributes to the economic development in those host countries (Aghion et al. 2008). Krugman (1991: 72) also sees economic regimes across countries "…defined by their restrictions". These restrictions are obstacles preventing entrant firms from being established in a particular market totally or putting restrictions on their establishment and entry options (Chen et al. 2009).

These restrictions by the governments can also result in pressures on MNCs to only use low equity collaborative ventures (like JVs) at time of market entry (Taylor et al. 2000; Deng 2003). The reduction of market entry and trade barriers i.e. abandonment of protectionist policies which often concern a country's strategic assets such as enterprises in key industries and industry deregulation attracts more FDI inflows and can encourage WOSs formation (e.g. Klapper et al. 2004; Arslan 2011). Moreover, previous research also shows that the institutional changes by the governments manifested by reduction of entry barriers and allowing more freedom to trade internationally can stimulate the adjustments of market entry strategies of the incumbent firms, as well as increase in number of entrants (e.g. Luo 2001; Pehrsson 2004; Trevino et al. 2008).

We argue that the host countries representing more freedom to trade internationally and lesser restrictions on the domestic as well as international operations of the foreign firms provide a favorable institutional context for MNCs. This in turn can motivate them to show more resource commitment to the host economy and prefer formation of WOSs rather than JVs. Therefore, we hypothesize that

Hypothesis 4: *Freedom to trade internationally in the host country is positively associated the preference of WOSs by the investing MNCs.*

Regulation of Credit, Business and Labor: It is generally assumed that the credit regulations, ease of doing business and stringent labor policies in the recipient country have significant effects on the entry strategy of the MNC's (e.g. Kelin et al. 2000). The stiffer capital requirements imposed by the central banks and higher cost of capital in the host country will restrict the availability of domestic credit for the wholly owned subsidiaries of foreign companies. Klein et al. (2000) show that differences across firms in the degree of their access to credit can be an important determinant of foreign direct investment. They argue that while firms may be constrained by their balance sheet positions, they may also be constrained by a reduction in the willingness of lenders to provide credit. This relative access to credit hypothesis (RAC) implies that firms' ability to engage in foreign direct investment will be influenced by their ability to raise external funds (Klein et al. 2000).

Some past researches (e.g. Haaland et al. 2005; Javorick & Spatareanu 2005) have analyzed the impacts of labor market regulation on MNC’s entry strategy and FDI flows. Their findings reveal that issues lack of flexibility in hiring and laying off workers, and termination period are one of the main concerns faced by MNCs considering entering emerging economies. Estrin et al. (2009) explore the complementary roles of institutional and human resource distances on foreign investors' entry strategies. They show empirically that the likelihood of first-time investors choosing wholly owned greenfield investment rather than a cooperative mode is expected to be curvilinear (inverse U-shaped) relative to the distance in labor markets regulations between the home and the host countries. However, it is important to note that the impact of distance differs between first and subsequent entries (Estrin et al., 2009). Busse and Groizard (2008) argue that countries may benefit from foreign investment inflows only if they have appropriate local government regulations and institutions in place. Excessive business regulations are likely to restrict FDI. For example, if starting and closing down a business are hindered by extensive and costly government regulations, which involve many bureaucratic procedures, the MNC’s will be discouraged to commit themselves for long term, thereby choosing low equity modes like JVs (Busse & Groizard 2008).

We argue that stringent credit regulations, tougher labor laws and high interference in business activities will discourage foreign MNC’s to adopt high equity entry mode. Therefore, we hypothesize

Hypothesis 5: *Tougher regulation of credit, business and labor in the host country is negatively associated with the preference of WOSs made by the investing MNCs.*

3 Research Methods and Empirical Analysis

Data Collection and Sample Description: The empirical data for the study is based on an internal databank of manufacturing foreign investments made by the Finnish firms in their international markets in both developed and emerging economies. The data is drawn from annual reports and press releases of the investing Finnish firms, and is also supplemented with the data gathered through direct contacts with these firms. The sample for this study consists of 118 FDIs made by 65 Finnish firms in selected emerging economies of Asia during the time period 1990–2006. The main aspects of the study sample are summarized in the Table 1.

Table 1. Characteristics of the Study Sample

Sample Characteristic	Description
Ownership Mode	WOSs (44), JVs (74)
Establishment Mode	Greenfield Investment (88), Acquisitions (30)
International Experience of investing MNCs	Minimum (No experience), Maximum (74 foreign investments)
Host Country Experience of Investing MNCs	Minimum (No experience), Maximum (17 years in the host country)
Number of FDIs in a particular year	1990 (4), 1991 (5), 1992 (4), 1993 (5), 1994 (8), 1995 (6), 1996 (8), 1997 (12), 1998 (7), 1999 (6), 2000 (8), 2001 (13), 2002 (7), 2003 (5), 2004 (3), 2005 (9), 2006 (8).
Largest recipient countries of FDI	China (67), India (13)

Operationalization of Study Variables: Economic freedom of the world annual reports measure country's openness to international business and trade by measuring and ranking them along five major pillars i.e. *Size of government, Legal structure and security of property rights, Access to sound money, Freedom to trade internationally and Regulations of credit, labor and business*. These pillars are further divided into different categories and finally country's summary ratings (1-10) are developed. Economic freedom of the world reports use the data from World Economic Forum, World Bank, International Monetary Fund, United Nations and World Trade Organization to measure these variables (Gwartney et al. 2008, 2009). The higher country score represents openness of the economy to international business, presence of strong market institutions, ease of business for foreign firms and sound financial and fiscal policies.

The data from economic freedom of the world reports has been used broadly in studies in the fields of international and institutional economics (see e.g. Ali 2003; Bengoa & Sanchez-Robles 2003; Doucouliagos & Ulubasoglu 2006; Dreher & Rupprecht 2007; Feldman 2007) and international political economy studies (e.g. Cole 2003; Boockmann & Dreher 2003; Xavier et al. 2005; Doucouliagos & Ulubasoglu 2006). Moreover, DiRienzo et al. (2007), Arslan (2011), Demirbag et al. (2011) and Arslan & Larimo (2012) also used this source to study different aspects of IB strategy of firms. Therefore, it can be argued that the use of data from economic freedom of the world annual reports is justified in this study due to its reliability and its close link with the current study's theoretical discussion and resultant hypotheses. The details of operationalization of study variables along with relevant references are provided in Table 2.

Table 2. Variables Operationalization

Variables	**Operationalization**
Ownership Mode	0=JV, 1=WOS (Source: Internal Databank)
Size of Government	Host country score (1–10) in this item in the year of investment or nearest available year (Source: Economic Freedom of the World Annual Reports).
Legal Structure and Security of Property Rights	Host country score (1–10) in this item in the year of investment or nearest available year (Source: Economic Freedom of the World Annual Reports).
Access to Sound Money	Host country score (1–10) in this item in the year of investment or nearest available year (Source: Economic Freedom of the World Annual Reports).
Freedom to Trade Internationally	Host country score (1–10) in this item in the year of investment or nearest available year (Source: Economic Freedom of the World Annual Reports).
Regulations of Credit, Labor and Business	Host country Score in this item in the year of investment or nearest available year (Source: Economic Freedom of the World Annual Reports).
Establishment mode	0=greenfield investment, 1=Acquisition (Source: Internal Databank)
International Experience of MNC	The number of earlier manufacturing FDIs made by the MNC (e.g.Larimo 2003; Arslan & Larimo 2010, 2011). (Source: Internal databank).
Host Country Experience of MNC	MNC’s length of earlier manufacturing experience in the host country calculated in number of years of operations (e.g. Hennart & Park 1993; Cho & Padmanabhan 2005; Arslan & Larimo 2011) (Source: Internal databank).
Host Country Risk	The Host country risk in the year before the year of investment based on Euro money country risk ratings (e.g. Cosset & Roy 1991; Arslan & Larimo 2010; Arslan 2011) (Source: Euromoney country risk ratings).
Economic Growth	The annual growth of GDP in the host country in the year of investment (e.g. Brouthers & Brouthers 2000; Arslan 2011) (Source: UNCTAD).
Parent MNC Size	Natural Log of Global sales of the parent MNC in the year preceding to investment changed to Euros (e.g. Hennart & Park 1993; Larimo 1997, 2003; Arslan & Larimo 2011) (Source: Internal databank)
China Dummy	1 for FDIs in China, 0 for other countries.
Timing	1 for FDIs in 1990s. 0 for FDIs in 2000s.

Statistical Method: We use binomial logistic regression to test our hypotheses because the dependent variable is dichotomous i.e. ownership mode preference can be either a JV or a WOS. The binomial logistic regression model is formally expressed as

$$P\ (y_i=1) = 1 \,/\, 1 + \exp\ (-a-X_iB)$$

Where y_i is the dependent variable, X_i is the vector of independent variables for the ith observation, a is the intercept parameter and B is the vector of regression coefficients. The statistical software PASW Statistics 20 is used for the binomial regression analysis in this study. The dependent variable has value 1 if the MNCs choose to form a WOS at time of market entry. Hence a positive regression coefficient indicates that a specific control or independent variable increases the probability of choosing WOS by the investing MNCs. Binomial logistic regression analysis has been used frequently in studies addressing FDI ownership mode strategy analysis (e.g. Gatignon & Anderson, 1988; Padmanabhan & Cho 1996; Pan 1996; Hennart & Larimo 1998; Dikova & van Witteloostuijn 2007; Kaynak et al. 2007; Arslan & Larimo 2010). Therefore, it can be referred as a useful and reliable statistical analysis technique to examine the ownership mode strategy of MNCs in different IB studies and our study also employs this statistical technique.

Appendix 1 shows the descriptive statistics and correlations of the variables used in the study. The table shows that correlations among certain variables are significant as is the case with most data sets. According to Pallant (2007), correlations above 0.7 indicate a potential for multicollinearity among variables that can influence the regression results. All the correlations in our dataset are lower than that cutoff point. Additional multicollinearity diagnostic (tolerance and variance inflation factor (VIF) were also conducted to increase the validity of our study findings. According to Belsley, Kuh & Welsch (1980) and Wetherill (1986)**,** the VIF value should not exceed 10. In this study the VIF values were even lower than 5 and thus, the potential collinearity among variables is not expected to influence the logistic regression results of this study.

4 Study Results and Discussion

Table 3 displays the results of the binomial regression analysis of our study. The explanatory power of all statistical models of the study is good, as their chi-square (x^2) values are significant at $p<0.001$ level. Moreover, the predictive ability of the statistical models can be assessed by the correct classification rate. Both statistical models of the study have a higher correct classification rate than the chance rate of 53 %, which is calculated using the *proportional chance criterion* which is a^2 +

(1-*a*) ^{2}, where *a* is a proportion of WOSs (37.3%) in our sample. The regression models show correct classification rates of 74.6% and 78%; therefore showing significant improvement of 21.6% and 25% in classification rates of statistical models compared to chance rate. Hence, our models fulfill the recommendation by Hair et al. (1998) concerning 25% improvement in classification rate compared to chance rate. Finally, good Nagelkerke R^2 values of both models (0.416 and 0.510) also show significant predictive capabilities of both models.

Table 3. Logistic regression estimates FDI Ownership Mode Strategy (WOS=1)

Variables	**Model 1: Control Variables**	**Model 2: Independent Variables**
Establishment mode	-0.900	-0.898
International Experience of MNC	-0.22**	-0.030*
Host Country Experience of MNC	0.148**	0.159*
Host Country Risk	-0.044**	-0.09*
Economic Growth	0.354*	0.0.454*
Parent MNC Size	0.193	0.145
China Dummy	-0.278*	-6.675***
Timing	-2.468***	-2.972**
Size of Government		-1.150**
Legal Structure and Security of Property Rights		0.185
Access to Sound Money		1.307*
Freedom to Trade Internationally		1.497**
Regulations of Credit, Labor and Business		-2.235**
N (WOS)	118 (44)	118 (44)
Model x^2	42.980***	55.274***
–2 Log likelihood	112.891	100.597
Nagelkerke R^2	0.416	0.510
Correctly classified (%)	74.6%	78.0%

Significance: $*p< 0.1$, $** p < 0.05$, $*** p <0.01$.

In Table 3, Model 1 presents the results of the binary logistic regression depicting the impacts of *control variables* on the FDI ownership mode strategy of Finnish

MNCs in Asia. The results show that host country risk, economic growth, international experience, host country experience, Timing and China dummy are statistically significant for ownership mode strategy of Finnish MNCs. The regression coefficients depict that high economic growth leads to the preference of WOSs by the investing Finnish MNCs. This finding is understandable and logical as high economic growth in emerging economies of Asia motivated Finnish firms to commit more resources to those economies to take advantage of the market growth (e.g. Makino & Neupert 2000) and formed WOSs, whenever it was possible for them. This finding also confirms findings of past studies addressing FDI ownership mode strategy of MNCs in general (e.g. Wilson 1980) as well as specifically Finnish MNCs (e.g. Arslan & Larimo 2010).

The regression results further show that high host country risk leads to the preference of JVs formation by the Finnish firms. This finding can be explained by referring to the fact that in high risk host countries, MNCs tend to limit their equity involvement by avoiding WOSs formation, which offers necessary flexibility and low switching costs (Erramilli & Rao 1993). Therefore, JVs emerge in previous empirical research as a favored choice for the MNCs to share costs and risks with a local partner in a risky business environment (e.g. Gatignon & Anderson 1988; Lou 2001; Brouthers & Brouthers 2003; Pak & Park 2004), similar to the finding of this study. An interesting finding concerns different influences of international and target country experience. The results show that general international experience leads to preference of JVs, which is in line with some past studies (e.g. Dikova & van Witteloostuijn 2007). However, increased target country experience leads to preference of WOSs, as MNCs become more familiar with local environment as well as are embedded more in local networks to survive as a fully owned establishment.

The results further show that FDIs made in China tended to be more JVs than WOSs. This finding can be explained by referring to the results of past studies where JVs have been referred as a preferred entry strategy for the foreign firms in China (e.g. Luo 2000; Li & Meyer 2010). Finally, the control variable timing of investment is highly significant and regression coefficient of timing of investment variable shows that FDIs made during 1990s tended to more JVs. This finding confirms the results of some earlier studies that MNCs prefer to form JVs in emerging economies that are in early stages of transition (e.g. Peng 2003; Meyer 2004). The host countries in our sample from emerging economies went through economic reforms and transition in 1990s (Luo 2000; Huang 2008; Trevino et al. 2008). Therefore, formation of JVs by the Finnish firms in 1990s is understandable and also proves importance of using this dummy in our analysis. The control variables like parent MNC size and establishment mode are non-significant in our study.

Model 2 in Table 3 reports the influences of *independent variables* on the FDI ownership mode strategy of Finnish firms in Asia. The regression results indicate that the variables freedom to trade internationally, size of government and regulation of credit, labor and business are highly significant at p<0.05 level, while the variable access to sound money are mildly significant at p<0.1 level. The regression coefficients indicate that in host economies with large size of government, Finnish firms preferred formation of JVs over WOSs. This finding can be explained by referring to our earlier discussion, where it has been argued that large government size corresponds to high taxation in the country which may discourage foreign firms to invest more resources and share costs with a local JV partner (Cummins & Hubbard 1995; Grubert & Mutti 2000; Makino et al. 2004).

The regression results also show that access to sound money in a host country is positively associated with the preference of WOSs by investing Finnish MNCs. This finding supports hypothesis 3 of the study, and it is therefore accepted. Some past studies have found that FDI inflows are positively correlated with the development of the financial markets in the host economies (Claessens et al. 2001). Moreover, FDI has also been found to be positively correlated with stock market capitalization and domestic value traded, suggesting that FDI is a complement and not a substitute of domestic stock market development. The relationship between FDI and stock market and financial system development has also been found to be highly and positively correlated (e.g. Jaffus 2002). Accordingly, we explain this finding by referring to the increased and relatively easy access to sound money in certain host countries motivated Finnish MNCs to form WOSs and commit to that emerging economy for the long term.

The regression coefficients also depict that freedom to trade internationally is positively associated with the preference of WOSs by the Finnish MNCs in Asian economies. The hypotheses 4 developed in section two is supported by this finding and is therefore accepted. We would like to refer to past studies where it has been found that restrictions on international trade can force MNCs to prefer low equity collaborative ventures (like JVs) at the time of market entry (Taylor et al. 2000; Deng 2003). Other studies have also shown that the reduction of market entry and trade barriers i.e. abandonment of protectionist policies and industry deregulation attract**s** more FDI to the host economy and also encourage**s** commitment of more resources by investing MNCs in form of WOSs (e.g. Luo 2001; Klapper et al. 2004). Moreover, previous research also shows that the institutional changes by the governments manifested by reduction of entry barriers inspire changes in market entry strategies of investing firms, as well as allowing them to enter markets, business sectors and industries that were previously inaccessible for them (e.g. Luo 2001; Pehrsson 2004; Trevino et al. 2008). Therefore, in those Asian countries that represent higher freedom to trade internationally and lesser restrictions on the domestic as well as international operations of the MNCs, Finnish firms formed WOSs to show their long term commitment to those economies, as well as follow their international strategy.

Finally, the results show that tough regulations of credit, labor and business lead to the preference of JVs rather than WOSs by investing Finnish MNCs and thereby, hypothesis 5 is accepted. This finding is line with past studies where it has been referred that if starting and closing down a business are hindered by extensive and costly government regulations, as well as there are stiff restrictions on capital movement and access, the MNC's will be discouraged to commit themselves for long term, thereby choosing low equity modes like JVs (Busse & Groizard 2008; Klein et al. 2010).

Based on the regression analysis results, we do not receive support for hypotheses 2, and therefore it is rejected. One reason for the non-significance of hypotheses 2 may be due to certain characteristics of our study sample. As the host countries in our sample consist of emerging economies from Asia that vary significantly among each other in characteristics relating to the development of market economy institutions especially in relation to legal structure and security of property rights. Some countries like Malaysia show good scores in economic freedom dimension of legal structure and security of property rights, some show significant improvement over study time period (e.g. India and Turkey), while others (e.g. China and Indonesia) still score very low depicting slow pace of reforms and transition. Hence, we expect this sample heterogeneity to offer partial explanation of non-significance of this independent variable.

5 Conclusions, Limitations, and Future Research Directions

The purpose of our study was to examine the impacts of different aspects of economic freedom on the FDI ownership mode strategy of the MNCs. Our dataset of 118 FDIs made by 54 Finnish MNCs Asian emerging economies allowed us to perform a vigorous analysis of our hypotheses. Our study contributes to IB and FDI literature by being one of the first ones to hypothesize ownership mode strategy of MNCs in relation to all five dimensions of economic freedom in the host countries, as categorized by "economic freedom of the world annual reports". It has been mentioned earlier that IB studies lack a detailed analysis of impacts of different dimensions of economic freedom on FDI ownership strategy of MNCs. Therefore, our paper fills this research gap as well as enhances the understanding of important ownership strategy of MNCs from these unique dimensions.

In the empirical part, stepwise binomial regression is used to test our hypotheses as the dependent variable is dichotomous in nature. Our results indicate that high economic growth in the host country lead to the preference of WOSs, while host country risk leads to the preference of JVs by the investing Finnish MNCs in Asia. We further found that while general international experience leads to a preference for JVs, increased host experience leads to a preference for WOSs by investing firms. Concerning the pillars of economic freedom, we found four out of five pillars to be statistically significant for FDI ownership mode strategy of Finnish MNCs based on regression analysis. Our results showed that while size of government and regulation of credit, labor and business lead to a preference for JVs; easy access to sound money and high freedom to trade internationally resulted in a preference for WOSs by investing Finnish MNCs.

The findings of the study have some useful implications for managers of firms from the Nordic region internationalizing their operations to Asian emerging economies. Based on our analysis and findings, different dimensions of economic freedom have different impacts on FDI ownership mode strategy of MNCs. Therefore, MNC managers need to consider the impacts of different dimensions of economic freedom on their proposed entry strategy in depth along with alternative options. Moreover, we found economic growth and country risk to be significant too; implying that managers should take into consideration these important factors before finalizing FDI ownership mode strategy at the time of market entry. Hence, the final FDI ownership mode strategy of Finnish MNCs entering these emerging economies in the future can be a balanced approach incorporating elements from level of economic freedom, growth prospects, and political risk in host country as well as the firm’s international strategy.

Our study also has certain limitations. Firstly, we only address FDI ownership mode strategy in relation to the decision between WOSs and JVs. We do not address majority, minority and 50/50 JVs separately. Moreover, our study concentrates on FDIs made by the Finnish firms in the emerging economies only in Asia, which can also be considered a limitation. On the other hand, the focus on FDIs made by Finnish MNEs provides an interesting opportunity to analyze the impacts of different dimensions of economic freedom on the FDI ownership strategy choice from the perspective of internationalizing firms from a highly internationalized Nordic country in fast growing and economically important Asia.

For the future research, it is suggested to expand the sample size by also including the FDIs made by the MNCs from other Nordic countries i.e. Sweden, Norway and Denmark, and study the impacts of dimensions of economic freedom on FDI ownership mode strategy. This kind of analysis is expected to offer a rather comprehensive understanding of the impacts of economic freedom on the FDI ownership mode strategy and would increase validity as well as generalizability of the study findings from a Nordic perspective.

References

Agarwal, S. & Mohtadi, H. (2004). Financial markets and the financing choice of firms: evidence from developing countries. *Global Finance Journal*, 15:1, 57–70.

Aghion, P., Burgess, R. Redding, S. & Zilibotti, F. (2008). The unequal effects of liberalization: Evidence from dismantling the license Raj in India. *American Economic Review* 98:4, 1397–1412.

Anderson, E. & Gatignon, H. (1986). Modes of entry: a transaction cost analysis and propositions, *Journal of International Business Studies* 17:1, 1–26.

Arslan, A. & Larimo, J. (2012). Partial or full acquisition: Influences of institutional pressures on acquisition entry strategy of multinational enterprises. In M. Demirbag & G. Wood (Eds), *Handbook of Institutional Approaches to International Business*. Edward Elgar, United Kingdom. 320–343.

Arslan, A. (2011). *Institutional Distance – Market Conforming Values in the Host Country and Foreign Direct Investment Choices of Multinational Enterprises*. Acta Wasaensia 245. University of Vaasa, Finland.

Arslan, A. & Larimo, J. (2011). Greenfield investments or acquisitions: Impacts of institutional distance on establishment mode choice of multinational enterprises in emerging economies. *Journal of Global Marketing* 24:4, 345–356.

Arslan, A. & Larimo, J. (2010). Ownership strategy of multinational enterprises and the impacts of regulative and normative institutional distance: Evidence from Finnish foreign direct investments in Central and Eastern Europe. *Journal of East – West Business* 16:3, 179–200.

Baron, D.P. (1995). Integrated strategy: market and nonmarket components. *California Management Review* 37:2, 47–65.

Belsley, D.A. Kuh, E. & Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Bengoa, M. & Sanchez- Robles, B. (2003). Foreign direct investment, economic freedom and growth: new evidence from Latin America. *European Journal of Political Economy* 19:3, 529–545.

Bergara, M.E., Henisz, W.J. & Spiller, P.T. (1998). Political institutions and electric utility investment: a cross-nation analysis. *California Management Review* 40:2, 18–35.

Boockmann, B. & Dreher, A. (2003). The contribution of the IMF and the World Bank to economic freedom. *European Journal of Political Economy* 19:3, 633–649.

Brouthers, K.D. (2002). Institutional, cultural and transaction cost influences on entry mode choice and performance. *Journal of International Business Studies* 33:2, 203–221.

Brouthers, K.D. & Brouthers, L.E. (2003). Why service and manufacturing entry mode choices differ: The influence of transaction cost factors, risk and trust. *Journal of Management Studies* 40:5, 1179–1204.

Busse, M. & Groizard, J.L. (2008). Foreign direct investment, regulations and growth. *The World Economy* 31:7, 861–886.

Chang, S.J. & Rosenzweig, P.M. (2001). The choice of entry mode in sequential foreign direct investment. *Strategic Management Journal* 22:8, 747–776.

Chen, Y-R. Yang, C. Hsu, S-M. & Wang, Y-D. (2009). Entry mode choice in China's regional distribution markets: Institution vs. transaction costs perspectives. *Industrial Marketing Management* 38:7, 702–713.

Child, J. & Tse, D.K. (2001). China’s transition and its implications for international business. *Journal of International Business Studies* 32:1, 5–17.

Cole, J. (2003). The contribution of economic freedom to world economic growth, 1980–1999. *Cato Journal* 23:2, 189–199.

Cummins, J.G., Trevor, H.S. & Hassett, K.A. (1995). Accounting standards, information flow, and firm investment behavior. In Martin Feldstein, James R. Hines, Jr. & R. Glenn Hubbard (Eds). *The Effects of Taxation on Multinational Corporations*. Chicago: University of Chicago Press.

Dikova, D. & Van Witteloostuijn, A. (2007). Foreign direct investment mode choice: entry and establishment modes in transition economies. *Journal of International Business Studies* 38:6, 1013–1033.

Dikova, D., Rao, S.P. & Van Witteloostuijn, A. (2010). Cross- border acquisition abandonment and completion: The effect of institutional differences and organizational learning in the international business service industry, 1981–2001. *Journal of International Business Studies* 41:2, 223–245.

DiRienzo, C.E., Jayoti, D., Cort, K.T. & Burbridge, J. (2007). Corruption and the role of information. *Journal of International Business Studies* 38:3, 320–332.

Doucouliagos, C. & Ulubasoglu, M. (2006). Economic freedom and economic growth: Does specification make a difference? *European Journal of Political Economy* 22:1, 60–81.

Delios, A. & Beamish, P.W. (1999). Ownership strategy of Japanese firms: Transactional, institutional and experience influences. *Strategic Management Journal* 20:8, 711–727.

Deng, P. (2003). Determinants of full-control mode in China: An integrative approach. *American Business Review* 21:1, 113–123.

Demirbag, M., Apaydin, M. & Tatoglu, E. (2011). Survival of Japanese subsidiaries in the Middle East and North Africa. *Journal of World Business* 46:4, 411–425.

Devereux, P.M. & Freeman, H. (1995). The impact of tax on foreign direct investment: Empirical evidence and the implication for tax integration schemes. *International Tax and Public Finance* 2:1, 85–106.

Dunning, J.H. (1993). *Multinational Enterprises and the Global Economy*. Wokingham: Addision-Wesley.

Dunning, J. H. (2000). The eclectic paradigm as an envelope for economic and business theories of MNE activity. *International Business Review* 9:2, 163–190.

Dunning, J.H. (2001). The eclectic (OLI) paradigm of international production: Past, present and future. *International Journal of the Economics of Business* 8:2, 173–190.

Dunning, J.H. (2004). An evolving paradigm of the economic determinants of international business activity. In J. L. C. Cheng, & A. Hitt (Eds). *Managing Multinationals in a Knowledge Economy: Economics, Culture and Human Resources*. Amsterdam: Elsevier. 3–28.

Dunning, J.H. & Lundan, S. (2008). *Multinational Enterprises and the Global Economy*. 2nd ed. Cheltenham: Edward Elgar.

Erramilli, K.M. & Rao, C.P. (1993). Service firms' international entry-mode choice: a modified transaction-cost analysis approach. *Journal of Marketing* 57:3, 19–38.

Estrin, S., Baghdasaryan, D. & Meyer, K.E. (2009). The Impact of Institutional and Human Resource Distance on International Entry Strategies. *Journal of Management Studies* 46:7, 1171–1196.

Flores, R.G. & Aguilera, R.V. (2007). Globalization and location choice: an analysis of US multinational firms in 1998 and 2000. *Journal of International Business Studies* 38:7, 1187–1210.

Friedman, M. (1982). *Capitalism and Freedom*. Chicago: University of Chicago Press.

Gatignon, H. & Anderson, E. (1988). The multinational corporation's degree of control over foreign subsidiaries: an empirical test of a transaction cost explanation. *Journal of Law, Economics and Organization* 4:2, 305–336.

Gaur, A.S. & Lu, W. (2007). Ownership strategies and survival of foreign subsidiaries: Impacts of institutional distance and experience. *Journal of Management* 33:1, 84–110.

Grubert, H., & Mutti, J. (2000). Do taxes influence where U.S. corporations invest? *National Tax Journal* 50:4, 825–839.

Gwartney, J.D., Lawson, R., Norton, S. & Hall, J. (2008). *Economic Freedom of the World: 2008 Annual Report*. Vancouver, BC: The Fraser Institute. Data retrieved from www.freetheworld.com.

Gwartney, J.D., Lawson, R., Grubel, H., de Haan, J. & Zandberg, E. (2009). *Economic Freedom of the World: 2009 Annual Report.* Vancouver, BC: The Fraser Institute. Data retrieved from www.freetheworld.com.

Haaland, J.I., Wooton, I. & Faggio, G. (2002). Multinational firms: Easy come, easy go? Finanz Archiv. *Public Finance Analysis* 59:1, 3–22.

Hair, J.F. Jr., Anderson, R.E., Tatham, R.L. & Black, W.C. (1998). *Multivariate Data Analysis* (5th Edition). Upper Saddle River, NJ: Prentice Hall.

Hartman, D.G. (1984). Tax policy and foreign direct investment in the United States. *National Tax Journal* 37:4, 475–488.

Henisz, W.J. (2000). The institutional environment for multinational investment. *Journal of Law, Economics and Organization* 16:2, 334–364.

Henisz, W.J. (2004). The institutional environment for international business. In Buckley, P.J. (Ed.). *What Is International Business?* Palgrave Macmillan: New York. 85–109.

Hitt, M.A.,Franklin, V. & Zhu, H. (2006). Culture, institutions and international strategy. *Journal of International Management* 12:2, 222–234.

Huang, Y. (2008). *Capitalism with Chinese Characteristics: Entrepreneurship and State During the Reform Era*. New York: Cambridge University Press.

IMF (2003). *World Economic Outlook*, Washington DC.

Islam, S. (1996). Economic freedom per capita income and economic growth. *Applied Economic Letters* 3:9, 595–597.

Javorcik, B.S. & Spatareanu, M. (2005). *Do Foreign Investors Care About Labour Market Regulations?* CEPR Discussion Papers 4839.

Jung, J.C., Beamish, P.W. & Goerzen, A. (2008). FDI ownership strategy: A Japanese-US MNE comparison, *Management International Review* 48:5, 491–524.

Khanna, T. & Palepu, K. (1997). Why focused strategies may be wrong for emerging markets. *Harvard Business Review* 75:4, 41–51.

Khanna, T., Palepu, K.G. & Sinha, J. (2005). Strategies that fit emerging markets. *Harvard Business Review* 83:6, 63–76.

Khanna, T. & Palepu, K. (2010). *Winning in Emerging Markets: A Roadmap for Strategy and Execution*. Cambridge, MA: Harvard University Press.

Klapper, L., Laeven, L. & Rajan, R. (2004). *Business Environment and Firm Entry: Evidence from International Data.* NBER Working Paper No. 10380.

Klein, M., Peek, J. & Rosengren, E. (2000). T*roubled Banks, Impaired Foreign Direct Investment: The Role of Relative Access to Credit.* NBER Working Paper No. 7845.

Kolstad, I. & Villanger, E. (2008). Determinants of foreign direct investment in services. *European Journal of Political Economy* 24:2, 518–533.

Krugman, P. (1991). *Geography and Trade.* Cambridge, MA: The MIT Press.

Krugman, P. (2008). *The Return of Depression Economics and the Crisis of 2008*. Penguin: London.

Lu, J.W. & Xu, D. (2006). Growth and survival of international joint ventures: An external-internal legitimacy perspective. *Journal of Management* 32:3, 426–448.

Luo, Y. (2000). *How to Enter China: Choices and Lessons.* Michigan: University of Michigan Press.

Luo, Y. (2001). Determinants of entry in an emerging economy: A multilevel approach. *Journal of Management Studies* 38:3, 443–472.

Mises, L.V. (1957). *Theory and History: An Interpretation of Social and Economic Evolution*. New Haven: Yale University Press.

Makino, S., Isobe, T. & Chan, C.M. (2004). Does country matter? *Strategic Management Journal* 25:10, 1027–1043.

Meyer, K.E. (2004). Perspectives on multinational enterprises in emerging economies. *Journal of International Business Studies* 35:4, 259–276.

North, D.C. (1981). *Structure and Change in Economic History*. New York: Norton.

North, D.C. (1990). *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press.

North, D.C. (2005). *Understanding the Process of Economic Change*. Princeton: Princeton University Press.

North, D.C. & Davis, L. (1970). Institutional change and American economic growth. *Journal of Economic History* 30:1, 131–149.

North, D.C., Wallis, J.J. & Weingast, B.R. (2009). *Violence and Social Order: A Conceptual Framework for Interpreting Recorded Human History*. New York: Cambridge University Press.

Pak, Y. & Park, Y. (2004). A framework of knowledge transfer in cross-border joint ventures: An empirical test of the Korean context. *Management International Review* 44:4, 435–455.

Pallant, J. (2007). *SPSS Survival Manual: a Step by Step Guide to Data Analysis Using SPSS for Windows*. Buckingham: Open University Press.

Peng, M. (2002). Towards an institution-based view of business strategy. *Asia Pacific Journal of Management* 19:2/3, 251–266.

Peng, M.W. (2003). Institutional transitions and strategic choices. *Academy of Management Review* 28:2, 275–296.

Peng, M., Wang, D. & Jiang, Y. (2008). An institution-based view of international business strategy: A focus on emerging economies. *Journal of International Business Studies* 39:5, 920–936.

Pehrsson, A. (2004). Strategy competence: a successful approach to international market entry. *Management Decision* 42:6, 758–768

Reardon, J.R., Erramilli, M.K. & Dsouza, D. (1996). International expansion of service firms: Problems and strategies. *Journal of Professional Services Marketing* 15:1, 31–46.

Ring, P.S., Bigley, G.A., D'Aunno, T. & Khana, T. (2005). Perspectives on how governments matter. *Academy of Management Review* 30:2, 308–320.

Slangen, A.H. & Van Tulder, R.J. (2009). Cultural distance, political risk, or governance quality towards a more accurate conceptualization and measurement of external uncertainty in foreign entry mode research. *International Business Review* 18:2, 276–291.

Slemrod, J. (1990). Optimal taxation and optimal tax systems. *Journal of Economic Perspectives* 4:1, 157–178.

Taylor, C.R., Zou, S. & Osland, G.E. (2000). Foreign market entry strategies of Japanese MNCs. *International Marketing Review* 17:2, 146–163.

Trevino, L.J. & Mixon, F.G. (2004). Strategic factors affecting foreign direct investment decisions by multinational enterprises in Latin America. *Journal of World Business* 39:2, 233–243.

Trevino, L.J., Thomas, D.E. & Cullen, J. (2008). The three pillars of institutional theory and FDI in Latin America: An institutionalization process. *International Business Review* 17:1, 118–133.

Tridimasa, G. & Winer, S.L. (2005). The political economy of government size. *European Journal of Political Economy* 21:3, 643–666.

Wang, D., Zhang, Z. & Bai, C. (2007). Size of government, rule of law, and the development of services sector. *Economic Research Journal* 8:1, 51–64.

Watkins, M.D. (2003). Government games. *MIT Sloan Management Review* 44:2, 91–95.

Wetherill, G.B. (1986). *Regression Analysis with Applications*. London: Chapman and Hall.

World Bank (2005). *A Better Investment Climate for Everyone*. Oxford, UK: Oxford University Press.

Xu, D. & Shenkar, O. (2002). Institutional distance and multinational enterprise. *Academy of Management Review* 27:4, 608–618.

Xu, D., Pan, Y. & Beamish, P.W. (2004). The effect of regulative and normative distances on MNE ownership and expatriate strategies. *Management International Review* 44:3, 285–307.

Appendix

Appendix 1. Descriptive Statistics and Pearson Correlation Matrix

	Mean	Std.dev.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1. Ownership mode	**0.37**	**0.486**	1													
2. Establishment Mode	**0.25**	**0.437**	-.128	1												
3. International Experience	**26.87**	**24.88**	-.180	.147	1											
4. Host Country Experience	**1.38**	**2.80**	.184*	.074	.335*	1										
5. Country Risk	**63.13**	**10.99**	-0.148	-.319*	.109	.027	1									
6. Economic Growth	**6.00**	**2.41**	.072	-.311*	-.068	.045	.551*	1								
7. Parent Firm Size	**6.36**	**1.88**	-.123	-.205	-.650*	.273*	.117	-.185*	1							
8. Timing	**0.55**	**0.50**	-.502*	-.021	.214	-.195	.310*	.183	.127	1						
9. China Dummy	**0.57**	**0.497**	.284*	-.080	-.118	.101	-.137	.320	-.135	-.444	1					
10. Size of Govt.	**5.69**	**1.02**	-.218	.042	.118	0.005	.160	-.086	.097	.415*	-.683*	1				
11. Legal Structure and security of property rights	**6.52**	**2.93**	-0.243	-0.096	0.044	-.097	.461*	.217	-.036	.371*	-.408*	.317*	1			
12. Access to Sound Money	**7.59**	**1.27**	-.165	-.154	.031	-.033	.375*	.005	.018	-.187*	-.130	.054	.098	1		
13. Freedom to Trade Internationally	**6.93**	**1.01**	-.007	-.163	-.097	.132	.188	.064	.037	-.266*	.002	.177	.146	.483*	1	
14. Regulation of Credit Labor and Business	**5.32**	**0.83**	-.232	.031	.032	-.044	.289*	-.216	.127	.274*	-.616*	.633*	.460*	.453*	.454*	1

* Correlation is significant at 0.05 level (2-tailed)

REFLECTIONS ON THE NEED FOR STRUCTURAL REFORM OF BANKS

Esa Jokivuolle[1]
Bank of Finland

1 Introduction

In this article I review the proposals of the High-Level Expert Group (henceforth the Group) on reforming the structure of the EU banking sector (see Liikanen, 2012a), and their rationale. I briefly compare the Group's proposals with other structural proposals, and comment on the public discussion that has followed. I also touch upon other key areas of regulatory reforms of banking, notably recovery and resolution regimes and capital adequacy requirements, which were endorsed by the G20 in response to the global financial crisis.[2] I will comment on their interplay with the structural reforms.

2 The Group's proposals

The Group made five proposals to complement the already on-going regulatory reforms.

First, there should be a mandatory separation of deposit banking and a trading entity into different subsidiaries within a banking group, if the trading activities to be separated are sufficiently large. In practice, separation would apply only in the larger institutions. Only the deposit bank can raise deposits and provide retail payment services. Respectively, only the trading entity can do proprietary trading and market making, although certain low-risk, client-oriented market making for risk management purpose is allowed also to the deposit bank.

---

[1] I thank Hanna Westman for valuable comments. All views expressed in this article are my own and do not necessarily reflect the views of the Bank of Finland. All remaining errors are my responsibility.

[2] Accounts of the causes of the global financial crisis are plenty; see e.g. Brunnermeier (2009), Liikanen (2012b), Lo (2012), and Rajan (2010).

Exposures to hedge funds and alike are assigned to the trading entity as a rule. Both the deposit bank and the trading entity have to be capitalized and funded on a stand-alone basis, and all financial links between them are restricted.

Second, there should be a possibility for supervisors to require a more extensive separation of activities, if a bank's recovery and resolution plans are otherwise not credible. The idea is that a bank's resolvability could be further improved by means of ex ante structural changes. This proposal underscores the interplay of structural measures and bank recovery and resolution measures as complements (I will return to this issue below). Third, designated bail-in instruments with clearly pre-defined terms, subject to holding restrictions by other banks, should be introduced. Fourth, the Group acknowledged the on-going review of capital requirements on trading assets done by the Basel Committee, but recommended the Commission to ensure that the outcome would suffice for the EU. The Commission should also carry out a review of capital requirements on real estate related lending. Fifth, the Group made suggestions to strengthen the corporate governance of banks. For example, the designated bail-in instrument was seen as an appropriate remuneration mechanism to management, serving the purpose of extending the effective decision-making horizon within banks.

3 Rationale for the proposals

What is the rationale for the Group's key proposal, mandatory separation of deposit banking and trading? I will consider four arguments.

First, separation is a way to limit opportunities for risk-taking in deposit banks. Incentives for excessive risk-taking (i.e., moral hazard) arise in the presence of deposit insurance and expectations of other government guarantees, especially under tough banking competition. Separation complements risk-based capital requirements which may be ineffective in limiting this moral hazard if risks are hard to measure, and if risky positions can change rapidly, as in trading (cf. e.g. Matutes & Vives 2000, and Boot & Ratnovski 2012). In short, activity restrictions imposed by structural separation can be a robust complement to capital requirements against banks' excessive risk taking incentives.[3]

[3] Activity restrictions are about "narrowing down" banking. A recent case presented for narrow(er) banking is the review by Pennacchi (2012).

Separation can curb risk-taking also in the trading entity. This is because the trading entity is not allowed to finance itself with insured deposits, nor rely on the internal capital market.[4] Hence the cost of its funding will reflect its risks so there will be no immediate cause for moral hazard.

Second, separation of activities is a direct way to tackle banks' complexity and interconnectedness. If banks become simpler in structure, a troubled banking group's recovery and ultimately resolution in a crisis will be easier. Hence structural proposals can support the EU's Bank Recovery and Resolution Directive by enhancing the ex ante credibility of resolution on a national level. Further, the aim of the Single Resolution Mechanism (SRM) is to make timely and efficient resolution also of large cross-border banks possible in the European banking union. Structural changes in the largest European banks would facilitate the task of the SRM. In sum, sufficient ex ante credibility of a large bank's resolution is necessary to make resolution a true threat point to bank financiers, especially the debt holders. This is crucial to ending the too-big-to-fail expectations and hence bringing market discipline back to work on large banks.[5]

To really achieve the above aim of simplicity in structure, proper implementation is essential. The separating line within the banking group should be kept clean and simple, with no significant exceptions to the rules allowed.

Further, as intra-group links between the deposit bank and the trading entity will be strictly limited, there will be fewer channels of contagion. Limits on trading activities will also reduce the counterparty risks of deposit banks, including links to shadow banks.

Third, simpler structures make it easier to manage, monitor and supervise banks. This further enhances market discipline. If the bank recovery and resolution reform is successful in restoring market discipline on large banks, market forces should automatically be driving banks towards less complicated conglomerate structures if these are seen to inhibit efficient management and reduce the necessary transparency for investor monitoring. But the banking industry may be stuck

---

[4] Blinder (2013) emphasizes that cutting the trading entity entirely from the internal capital market of the banking group is crucial. In particular, he would make sure that "downstreaming capital" from the parent to the trading subsidiary to cover the latter's losses is not allowed under any circumstances.

[5] Gandhi & Lustig (2013) show that the repeal of the Glass-Steagall Act in the US in 1999 greatly increased large commercial banks' and investment banks' expected government subsidy, implicit in bank stock returns. This indicates that (de)regulating bank structures may have an important bearing on the too-big-to-fail problem.

in a “bad equilibrium”: complicated conglomerate structures prevail without market discipline and market discipline does not work if large banks are considered too complicated to fail, even in the presence of regulatory recovery and resolution tools. A push from structural reforms imposed by the regulators may be needed to bring about the shift to a good equilibrium with sufficiently simple structures and effective market discipline (see Liikanen 2013).

Fourth, separating deposit banking and trading entities may reduce the mixing of the different management cultures of traditional banking and investment banking. Some observers say this played a role in the run up to the crisis, especially by spreading the bonus culture to lending business where it may have created hazardous incentive effects.[6]

Yet separation would maintain all financial services within one group and hence preserve potential economies of scope from “one-stop-shopping”. Hence the universal bank concept would remain, but in a more structured form. One could also argue that the universal bank model would be taken back to its roots – the focus would once more be on customer relationships, and the excessive expansion we saw prior to the crisis in e.g. intra-financial transactions would be reversed.

Why did the Group make a proposal concerning designated bail-in instrument? The bail-in proposal could be seen as a complement to the European Commission’s recovery and resolution regime involving the bail-in of debt instruments in order to recapitalize the vital parts of a troubled bank and cover losses as part of a resolution. In the Group’s view, a layer of designated bail-in debt instruments with clear contractual terms on what would trigger the bail-in would facilitate the creation of a market for such instruments. Such instruments should be held by investors outside the banking sector in order to reduce contagion and hence systemic risks.

Why did the Group propose a review of capital requirements in the EU? Uncertainty concerning risk measurement is particularly severe in trading where liquidity risk and systemic risk are intertwined. Separation of the riskiest trading activities from deposit banking is an important step to limit the impact of this uncertainty. Robust, additional capital requirements which do not rely on risk models may be another.

---

[6] See Financial Times, September 8, 2013: “Culture clash means banks must split, says former Citi chief”.

The Group acknowledged the on-going work of the Basel Committee on Banking Supervision in reviewing the trading book capital requirements, and recommended that the European Commission should carry out an evaluation of whether the resultant amendments would be sufficient at the EU level. Similar review was recommended for capital requirements on real estate-related lending, obviously for the reason that such lending was, as so often in the history, also at the centre of the global financial crisis. The Group did not make further suggestions related to systemic risks stemming from traditional banking and real estate-related lending in particular. Rather, the Group acknowledged the work currently being done to create macroprudential tools which would aim at dampening the credit cycle and its effects.

4 Comparison with other structural proposals

How does the mandatory separation compare with other structural proposals, especially its two influential predecessors, the Volcker Rule (as part of the Dodd-Frank Act) in the US, and the Vickers report for the UK? Their close relationships are obvious as e.g. Blinder (2013) calls them “cousins”.

The Group, having the “last mover advantage” with respect to the US and UK proposals, wanted to avoid the decoupling of proprietary trading and market making because it is difficult and hence easy to test by the industry. Hence, both proprietary trading and market making are to be separated. Moreover, unlike in the US Volcker Rule, proprietary trading is not banned; it can reside within the trading entity. This may reduce incentives to take prop trading to the “shadows” outside regulated banking; something for which the Volcker Rule has been criticized (see Duffie 2012).

In comparison to the UK reform, the Group would allow securities underwriting within the deposit bank as it is seen as highly complementary to corporate finance. The Vickers model is narrower than this, and it also imposes higher capital requirements on deposit banks operating on the domestic market. The Group instead contemplated strengthening the capital buffers of the trading entity.

Interestingly, France and Germany have already made national moves to structural banking reforms after the publication of the Group’s proposals for the whole EU, not having waited for the European Commission’s view. These national processes draw heavily from the Group’s proposal but are generally viewed as milder versions of it. Similar legislative proposals are discussed in Belgium and the Netherlands.

5 Evaluation of some critical comments

I will next address some of the arguments in the public discussion concerning the Group's proposals and structural reforms more generally.

It has been suggested that separation does not prevent systemic risks in the separated trading entities (see e.g. Admati & Hellwig 2013). Lehman Brothers is mentioned as a warning example. However, this argument is not entirely immune to Lucas critique: Lehman might have been less systemic if the financial network had been less interconnected, and the general level of risk-taking should have been smaller, had structural restrictions been in place.[7]

Secondly, there is the question whether market making should be allowed to deposit banks more widely than proposed by the Group. The counter question should be, however, whether there is a market failure in the supply of liquidity through market making, which justifies use of insured deposits to fund the market making inventory. This is not obvious. On the contrary, one could imagine that superfluous liquidity of securities supported by an indirect government subsidy could encourage excessive risk-taking.

Third, there has also been the concern that structural reforms, especially if not globally coordinated, may increase rather than reduce complexity. It is true that separation may add legal units within banking groups and hence increase ostensible complexity. However, because separation also reduces financial linkages between the units, it facilitates their resolvability, and hence reduces complexity in this regard.

Fourth, it has been argued that separation reduces benefits from risk diversification between trading and deposit banking. Banks would hence need more equity capital, which they consider relatively costly, to maintain their desired level of solvency.

To evaluate this criticism we can start by asking why banks, or any private corporations, need to diversify, if the (bank) stockholders can do the diversification in their own portfolios. More generally, why do banks or corporations need risk management in the first place?

[7] Low level of bank equity capital in the global financial network, especially in many investment banks, was certainly a key problem causing the amplification of the global financial crisis after Lehman's failure, so one cannot overemphasize the importance of capital adequacy reforms either. The Basel III reform has done a lot but the issue still remains whether the capital adequacy reform so far is sufficient (see e.g. Admati & Hellwig 2013).

The standard answer given by corporate finance theory is that risk management has value because it can save expected bankruptcy costs. And indeed, in the case of large banks bankruptcy costs are high because they have particularly large spillover effects on the economy.

By developing the bank recovery and resolution framework, regulators in effect reduce banks' bankruptcy costs. It follows that mandatory separation will have two opposing effects on the need of additional equity capital: the loss of diversification increases the need for equity, but by facilitating recovery and resolution and hence helping to lower private and public bankruptcy costs mandatory separation can reduce the need for additional equity.

One should also consider that when we talk about benefits of diversification from putting together two very different profit and loss probability distributions, such as those of trading activities and deposit banking, care must be taken. If these distributions involve considerable tail risks, the benefits of diversification may be in doubt.

Finally, John Vickers, the chairman of the UK's Independent Commission of Banking, has commented that simply setting capital requirements sufficiently high could have been the alternative (better?) way to increase bank stability than regulating bank structures. According to him, structural measures can come into play, however, if for any reason it is not possible to impose sufficiently high capital requirements (see Vickers 2012).

However, a combination of high, but not too high, capital requirements and structural reforms might also have genuine benefits. One perspective is offered by Richardson (2012). The starting point is that financial crises appear to coincide with catastrophic losses which materialize with a low-probability. Most of the time losses in banking are relatively moderate.

In order to stay solvent in such rare "tail-risk" scenarios, banks need to hold a sizeable amount of equity capital which for most of the time is not really needed on their balance sheet. It may even restrict financial intermediation, and hence not fully support economic growth opportunities (although this effect should not be exaggerated either, see e.g. Admati & Hellwig 2013, and Thakor 2013). We might then obtain socially more efficient regulation in terms of balancing between financial stability and the amount of financial intermediation by combining high, but not too high, capital requirements with structural measures.

6 Conclusions

To conclude, the structural proposals for the EU maintain the idea of supplying all financial services within a single banking group.

The proposals aim to reduce scope for conflicts of interest and distorted incentives which could endanger socially efficient business decisions within banking groups, and financial stability.

Although capital adequacy is the key to support the stability of banking and hence sustainable long-term economic growth, structural regulations which in effect "narrow down" deposit banking can further stabilize banking in a robust manner and facilitate orderly failure of large banks in order to restore market discipline on them. Clean and simple implementation of structural changes is necessary to reap the intended benefits and to avoid a backlash from unintended increase in complexity.

References

Admati, A. & Hellwig, M. (2013). *The Bankers' New Clothes: What's Wrong with Banking and What to Do about It?* Princeton University Press.

Blinder, A. (2013). Guarding Against Systemic Risk: The Remaining Agenda. forthcoming in *SUERF Studies*.

Boot, A.W. & Ratnovski, L. (2012). Banking and trading. IMF Working Paper 12/238.

Brunnermeier, M. (2009). Deciphering the liquidity and credit crunch 2007–08. *Journal of Economic Perspectives* 23:1, 77–100.

Duffie, D. (2012), Market making under the proposed Volcker Rule. Rock Center for Corporate Governance Working Paper Series No. 106.

Gandhi, P. & Lustig, H. (2014). Size anomalies in U.S. bank stock returns *Journal of Finance* (forthcoming).

Liikanen, E. (2012a). *High-level Expert Group on Reforming the Structure of the EU Banking Sector.* Brussels, 2 October 2012.

Liikanen, E. (2012b). On the structural reforms of banking after the crisis. Speech at CEPS, Brussels, 23 October 2012.

Liikanen, E. (2013). How to improve financial stability and resilience of systemically important financial institutions after the crisis? Bank of Finland speeches, 14 November 2013.

Lo, A.W. (2012). Reading about the financial crisis: A twenty-one-book review. *Journal of Economic Literature* 50:1, 151–178.

Matutes, C., & X. Vives (2000). Imperfect competition, risk taking, and regulation in banking. *European Economic Review* 44:1, 1–34.

Pennacchi, G. (2012). Narrow banking. *Annual Review of Financial Economics* 4.

Rajan, R.G. (2010). *Fault Lines: How Hidden Fractures Still Threaten the World Economy.* Princeton University Press.

Richardson, M. (2012). *Why the Volcker Rule Is a Useful Tool for Managing Systemic Risk.* NYU Stern School of Business. White Paper.

Thakor, A.V. (2013). Bank capital and financial stability: An economic tradeoff of a Faustian bargain? Olin School of Business, Washington University in St. Louis.

Vickers, J. (2012). Some economics of banking reform. Oxford University, Department of Economics Discussion Paper, November 2012.

ARBITRAGE AT THE RACETRACK

Gunnar Rosenqvist
Hanken School of Economics

1 Introduction

Parimutuel racetrack betting resembles in many ways investment in securities like common stocks. In both cases the bettors/investors bid against each other. At the racetrack, the empirical payoff odds, and hence also the return on each bet, are formed mutually by the bettors, like the prices on a market for common stocks. Racetrack betting thus provides interesting opportunities for research in the functioning of financial markets and in human behavior in decision making under risk and uncertainty. Consequently a vast literature has emerged on the efficiency of racetrack betting markets and the rationality and behavior of race track bettors. This literature is collected and surveyed by Hausch, Lo & Ziemba (1994, 2008) up to 1994. More recent contributions are comprehensively reviewed in the volume by Hausch & Ziemba (2008). Economists, financial economists and statisticians have paid much interest to this area of study; mathematicians, psychologists, decision analysts and others have also contributed.

A particular advantage with studying racetrack betting markets compared e.g. with markets for common stocks is that there is a termination point in time for each race at which the value of the tote ticket becomes certain. This is in contrast to investments in common stocks and the similar where the value of a security at each point in time is the present value of infinite future prospects and their expectations, with a final value of the investment emerging not before infinity (except in case of bankruptcy whereby the whole investment is lost), hence involving complicated dynamics already by definition in the outset.

A particularly outstanding and well known result from the literature on efficiency of racetrack betting markets is the favorite-longshot bias. A number of studies have found that favorites with low payoff odds are systematically underbet and longshots with high payoff odds overbet by the general public, e.g. Weitzman (1965), Ali (1977), Snyder (1978), Ash, Malkiel & Quandt (1982). Jullien & Salanié (2008) call this ”the most salient stylized fact of racetrack betting data”.

The utility function of racetrack bettors has also been estimated, e.g. by Weitzman (1965), Ali (1977); and with Finnish data by Kanto, Rosenqvist & Suvas (1992). The literature on empirical estimation of the bettor´s utility function is surveyed by Jullien & Salanié (2008). The favorite-longshot bias turns out in a risk loving utility function.

Ottaviani & Sorensen (2008), Jullien & Salanié (2008) and Snowberg & Wolfers (2008) review the main theoretical explanations for the favorite-longshot bias that have appeared in the literature. Sobel & Ryan (2008) provide an explanation of the favorite-longshot bias in the form of changing combinations of serious and casual bettors. Hausch, Lo & Ziemba (2008), in the preface to the 2008 Edition of their book Efficiency of Racetrack Betting Markets, note that the extent of the favorite-longshot bias has become clearly weaker in recent years. Smith & Vaughan Williams (2008) report favorite-longshot bias in exchange markets to a significantly smaller degree than in corresponding bookmaker odds.

For each race at the track different types of bets are typically offered, with separate pools. For example, at one and the same race one can participate in win (pick the winner), place (your selection is first, second or third) and double (your selections are first and second irrespective of the order) betting. (The place and double bets are called show and quinella in North America).

For example, if W_i is the amount bet on horse i to win and X_{ij} the amount bet on the pair i, j in the double pool, (i ≠ j), the observed payoff odds in the win and double pools are, respectively,

$$o_i = \frac{q \sum_{i=1}^{n} W_i}{W_i} \text{ and } o_{ij} = \frac{q \sum_{i<j} X_{ij}}{X_{ij}},$$

where n is the number of horses, and q is the track payback.

The development of betting strategies and testing of market efficiency go hand in hand as betting strategies with positive profit imply rejection of the efficient market hypothesis.

There is a distinction between fundamental and technical betting strategies. While technical betting strategies are based only on current publicly available betting data, i.e. the empirical odds described above, fundamental analysis utilizes also other knowledge about the competing horses, e.g. past performance data available from racing forms and special sources, to ‘handicap’ races. There is a vast literature on this, with Figlewski (1979), Bolton & Chapman (1986), Chapman (1994), Benter (1994) and Ludlow (1994) as a few examples. Bain, Hausch & Ziemba (2006) and Gramm & Ziemba (2008), combine fundamental information with win odds. - The current paper focuses on technical strategies, in particular strategies which compare different bets on the same horses in the same race. Thus, we are discussing (risk) arbitrage within and between betting pools on the same horses.

Hausch, Ziemba & Rubinstein (1981) and others have developed and applied betting strategies where the probabilities of place, show (your selection at least second or third; North American terminology) and exotic (predicting the outcome of two or more horses) bets are estimated on basis of winning probabilities obtained from the win bets for the same horses in the same race. This requires means of estimating probabilities of place, show and exotic bets on the basis of win bets. Originally formulas presented by Harville (1973) were used for this purpose. Hausch et al. (1981) and others argued that the market for win bets can basically be taken as efficient in weak form, despite the favorite-longshot bias, alternatively that win probabilities obtained from empirical win odds can be corrected for this bias, and hence used to predict the outcomes of place, show and exotic bets.

The Hausch et al. (1981) betting system was developed and operationalized further by the same authors, e.g. Hausch & Ziemba (1985), Ziemba & Hausch (1987). Also see the Introduction to Part 6 on Efficiency of Exotic Wagering Markets, pp. 445-446, in Hausch, Lo & Ziemba (1994, 2008), Hausch, Lo & Ziemba (1994) and Sung & Johnson (2008). As noted, the Hausch et al. (1981) betting scheme originally utilized the formulas presented by Harville (1973), but Hausch, Lo & Ziemba (1994), as well as Lo, Bacon-Shore & Busche (1995), also considered alternatives to Harville's formulas.

As noted e.g. by Dolbear (1993), a necessary though not sufficient condition for efficiency in racetrack betting markets is that bettors as a group identify and then eliminate any implicit probability inconsistencies between related betting pools. The same argument also holds for related bets within the same betting pool. Kanto & Rosenqvist (1994) noted that in double betting the number of bets $\binom{n}{2}$ exceeds, often by large, the number of horses n. For example with 10 horses there are 45 different double bets. Then, on an efficient market there should be consistency between payoff odds for the different double bets reflecting that the same horses appear in the various double bets.

Harville's formulas were also employed by Tuckwell (1981), Ash & Quandt (1987), Dolbear (1993), Kanto & Rosenqvist (1994) and Lo & Busche (1994), among others. Tuckwell (1981) with bookmaker odds from Australia, Sydney and Melbourne, as it seems independently of Hausch et al. (1981), used Harville's formulas for testing efficiency of place (show) betting markets utilizing win probabilities estimated from win bets.

Testing for efficiency of financial markets is often classified as being in weak, semi-strong or strong form, depending on whether profits can be earned using historical prices (current payoff odds), other publicly available information, and all (including inside) information. Within this categorization we are thus dealing with weak form efficiency, although Sung & Johnson (2008), deviating from other sources like Ziemba (2008), classify arbitrage between betting pools as an issue of semi-strong form efficiency.

This paper is concerned with the possibility to utilize different bet types from parallel pools, or alternative bets within the same pool, on the same horses in the same race. The aim of the paper is to discuss the possibilities to relate and compare to each other the odds for these respective bets in these types of situations.

A successful betting system also needs a principle for determining the amount to bet when a bet appears with positive expected value. The Kelly criterion is typically employed for this purpose. For a brief review, see e.g. Ziemba (2008) section 3.5, pp. 201-203. This criterion maximizes the expected growth of capital (e.g. minimizes the expected time to reach a specified wealth growth). In practice it amounts to act at each stage *as if* your utility function were logarithmic. A number of betting systems, including those of Hausch et al. (1981) and Kanto & Rosenqvist (1994), utilize this principle.

This paper thus deals with risk arbitrage. With betting on horses, possibilities for risk free arbitrage are rare, and challenging to utilize in practice. Risk free arbitrage at racetrack betting was presented by Hausch & Ziemba (1990a and 1990b), and further discussed by Rosenbloom (1992), Edelman & O'Brian (2004) and Ashiya (2013). Hausch, Lo & Ziemba in the preface to the 2008 Edition of their book Efficiency of Racetrack Betting Markets note that recent developments, like cross-track betting and betting exchanges, give new opportunities to seek bet arrangements with guaranteed profit. Smith & Vaughan Williams (2008) provide a survey of Betfair and other betting exchanges. As noted, here we restrict ourselves to risk arbitrage, comparing the empirical odds for different betting pools, or different bets within the same pool, for the same race.

2 Harville's formulas

Harville's (1973) formulas can be described as follows. Assume p_i is the probability that horse i wins. Then the probability that i is first and j second is

$$\frac{p_i p_j}{1 - p_i},$$

and the probability that i is first, j second and k third is

$$\frac{p_i p_j p_k}{(1-p_i)(1-p_i-p_j)}.$$

Henery (1984), among others, noted that assuming independent exponentially distributed running times implies Harville's formulas. Henery (1984) also showed that assuming independent extreme-value distributed running times leads to the same result.

Harville (1973) himself, applying his formulas on subjective probabilities from win odds, found a tendency to overestimate the chances of second or third place finish for horses with high probabilities of such finishes and to underestimate the chances of those with low theoretical probabilities.

3 A general framework for ordering probabilities of multi-entry competitions

As pointed out by e.g. Henery (1981), one feature of Harville's formulas is that they do not depend on the number of horses in the race. Alternative models were suggested by Henery (1981) and Stern (1990).

Let T_1, T_2, ..., T_n be n independent random variables, e.g. running times of horses in a race, with probability density functions $f(t;\alpha_i)$ and distribution functions $F(t;\alpha_i)$. Then, as pointed out e.g. by Ali (1998), the probability of horse i winning the race is

$$p(i) = \int f(x;\alpha_i) \prod_{\substack{j=1 \\ j\neq i}}^{n} \left[1 - F\left(x;\alpha_j\right)\right] dx.$$

Similarly the probability, say, of i ending first and j second is

$$P(i,j) = \int f\left(x;\alpha_j\right) F(x;\alpha_i) \prod_{\substack{l=1 \\ l\neq i,j}}^{n} [1 - F(x;\alpha_l)] dx,$$

Henery (1981) suggested independent normally distributed running times $T_i \sim N(\alpha_i, 1)$, while Stern (1990) proposed independent gamma distributions with scale parameters α_i and common fixed shape parameter r, $T_i \sim gamma(\alpha_i, r)$. In other words, with Stern's model,

$$f(x;\alpha_i, r) = \frac{\alpha_i^r}{\Gamma(r)} x^{r-1} \exp(-\alpha_i x)\, , x \geq 0.$$

In the formula for p(i), change of variable in the integral, e.g. for the normal distribution, for T_i from x to $x - \alpha_i$, may imply that $F(x; \alpha_j)$ depends also on α_i.

The shape parameter r is here taken as fixed. Taking $r = 1$ gives the exponential distribution and Harville's formulas, while Stern's model converges to Henery's for r $\rightarrow \infty$. Stern (1990) also applied $r = 2$ and Ali (1998) a range of values $(r = 0.5,\ 0.75, 1, 2, 5, 10, 20)$.

Several authors, among them Henery (1981) and Stern (1990) themselves, noted that the normal and gamma models (except when r = 1) are too complicated to use in practice. Henery (1981) also suggested an approximation for the normal model, put it is claimed to be inaccurate (Ziemba, 2008, p. 200). Dansie (1986) extended Henery's (1981) approach to a multivariate normal distribution. A particular approximation of ordering probabilities for the Henery and Stern models were suggested by Hausch, Lo & Ziemba (1994) and Lo & Bacon-Shone (2008).

Henery (1984) fitted extreme value distributions to running times, concluding that the tail with the fastest running times is consistent with the model. Stern (1990) reported r > 1 to be superior to Harville's r=1. Lo & Bacon-Shone (2008) reported on the Henery and Stern models fitting better than the Harville model for particular horse racing datasets, but also mentioned the Stern model exceptionally found to perform better than the Henery and Harville models for a particular Japanese data set (see also Lo, 1994). Ali (1998) concluded the Henery model best fitted his data, consistently with some earlier findings by Lo & Bacon-Shore. Ziemba (2008) gives an insightful review of ordering probabilities (section 3.4, pp 196-201). He concludes that "while one running-time distribution model does not appear to hold universally, there is limited empirical support for Harville's model", and further that ”While Henery's (1981) and Stern's (1987) ordering probabilities are superior to Harville's, the complex numerical calculations that they both require essentially precludes them from being used on-track", and finally that ”Harville's model still is useful; in particular for place and show probabilities at tracks where the favorite-longshot bias is exhibited in the win market".

Gibson & Rosenbloom (2005) empirically compared exacta pool probabilities generated by Harville's and Henery's (1981) formulas with those produced by the exacta pool. They concluded that exacta probabilities calculated from the exacta pool are more accurate than those calculated from the win pool and that the Harville model did better than the Henery model.

McCulloch & van Zijl (1986) performed a direct test of Harville's formulas by utilizing that show betting in New Zealand allows direct estimation of show probabilities which on the other hand can be estimated by Harville's formulas from win odds in win betting.

4 Kanto, Rosenqvist and Suvas (1991)

Kanto, Rosenqvist & Suvas (1991), hereafter KRS, studied trifector betting, where the three first coming horses have to be selected in the correct order. They analyzed data from 80 races with a total of 1115 horses at Vermo, the Helsinki racetrack. The total amount of money bet was FIM 35,973,356 (≈ 10.2 milj euro equivalent 2010).

Let x_{ijk} be the amount of money bet on the combination {ijk} with horse i finishing first, j second and k third. Estimates of the probabilities for i winning, j coming second and k third are then obtained as

$$\hat{p}(i\ first) = \frac{1}{x}\sum_{j=1}^{n}\sum_{k=1}^{n} x_{ijk}\,, i = 1, \dots, n,$$

$$\hat{p}(j\ second) = \frac{1}{x}\sum_{i=1}^{n}\sum_{k=1}^{n} x_{ijk}\,, j = 1, \dots, n,$$

$$\hat{p}(k\ third) = \frac{1}{x}\sum_{i=1}^{n}\sum_{j=1}^{n} x_{ijk}\,, k = 1, \dots, n,$$

where n is the number of horses in the race, and

$$x = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} x_{ijk}$$

is the total amount of money bet in the trifector pool at the race. KRS noted that these estimates can be considered as the bettors' aggregate subjective probabilities for each horse ending first, second and third. On the other hand the probabilities of j and k finishing second and third, respectively, can be estimated by Harville's formulas, assuming they hold, on the basis of the estimates $\hat{p}(i\ first)$:

$$\tilde{p}(j\ second) = \sum_{i \neq j}^{n} \frac{\hat{p}(i\ first)\hat{p}(j\ first)}{1 - \hat{p}(i\ first)}$$

and

$$\tilde{p}(k\ third) = \sum_{i \neq k}^{n} \sum_{\substack{j \neq i \\ j \neq k}}^{n} \frac{\hat{p}(i\ first)\hat{p}(j\ second)\hat{p}(k\ third)}{\big(1 - \hat{p}(i\ first)\big)\big(1 - \hat{p}(i\ first) - \hat{p}(j\ second)\big)}.$$

Hence there are two sets of estimates, $\hat{p}(j\ second)$ and $\tilde{p}(j\ second)$, j=1,2,…,n, for the horses to end second. Of these $\tilde{p}$ are based on Harville's formulas, but $\hat{p}$ are not. KRS therefore set out to test Harville's formulas by comparing these two sets of estimates.

For each race, KRS calculated Shannon's entropy measure

$$E = -\sum_{j=1}^{n} p_j \log_2 p_{j,}$$

where p_j denotes either $\hat{p}(j\ second)$ or $\tilde{p}(j\ second)$. They found that in 79 out of 80 cases, $\tilde{p}$ had smaller entropy than $\hat{p}$. In a two-sided sign test this gives an effective significance level 1.34 10^{-22}! This is very clear significance although the number of races (80) is not overly high. KRS accordingly concluded it is quite clear that the distributions are not similar. The dispersion of $\tilde{p}$ is smaller than that of $\hat{p}$. They concluded that should the bettors act as Harville's formulas suggest, the amounts of money laid on horses expected to be second or third would be less dispersed than they are in practice and that bettors thus are more unsure about which horse will end second than might be inferred from the winning probabilities derived from the trifector bets.

KRS estimated also $p(k\ third)$ in two different ways and compared them with similar results. The entropy calculated using Harville's formula turned out smaller in all 80 cases. In conclusion KRS showed that the trifecta bets studied by them are not internally consistent with Harville's formulas.

Acknowledgements: Section 4 in this paper is based on Kanto, Rosenqvist and Suvas (1991).

References

Ali, M.M. (1977). Probability and utility estimates for racetrack bettors. *Journal of Political Economy* 85, 803–815.

Ali, M.M. (1998). Probability models on horse-race outcomes. *Journal of Applied Statistics* 25, 221–229.

Ash, P., Malkiel, B.G. & Quandt, R.E. (1982). Racetrack betting and informed behavior. *Journal of Financial Economics* 10, 187–194.

Ash, P. & Quandt, R.E. (1987). Efficiency and profitability in exotic bets. *Economica* 54, 289–298.

Ashiya, M. (2013). *Lock! Risk-Free Arbitrage in the Japanese Racetrack Betting Market*. Available: http://ideas.repec.org/p/koe/wpaper/1301.html accessed 21.1. 2014.

Bain, R.S., Hausch, D.B. & Ziemba, W.T. (2006). An application of expert information to win betting on the Kentucky derby, 1981–1985. *The European Journal of Finance* 12, 283–301.

Benter, W. (1994). Computer Based horse race handicapping and wagering systems: A report. In D.B. Hausch, V.S.Y. Lo & W.T. Ziemba (1994, 2008). *Efficiency of Racetrack Betting Markets*. Academic Press; Reprinted as *Efficiency of Racetrack Betting Markets. 2008 Edition*. Singapore: World Scientific. 183–198.

Bolton, R.N. & Chapman, R.G. (1986). Searching for positive returns at the track: A multinomial logit model for handicapping horse races. *Management Science* 32, 1040–1059.

Chapman, R.G. (1994). Still searching for positive returns at the track: Empirical results from 2000 Hong Kong races. In D.B. Hausch, V.S.Y. Lo & W.T. Ziemba (Editors). *Efficiency of Racetrack Betting Markets*, 173–181. Academic Press.

Dansie, B.R. (1986). Normal order statistics as permutation probability models. *Applied Statistics* 35, 269–275.

Dolbear, T. Jr. (1993). Is racetrack betting on exactas efficient? *Economica* 60, 105–111.

Edelman, D.C. & O'Brian, N.R. (2004). Tote arbitrage and lock opportunities in racetrack betting. *The European Journal of Finance* 10, 370–378.

Figlewski, S. (1979). Subjective information and market efficiency in a betting market. *Journal of Political Economy* 87, 7–88.

Gibson, L.M. & Rosenbloom, E.S. (2005). The Best Probability Model for the Exacta. ASAC 2005 Conference. Available: http://www.google.fi/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0CD8QFjAB&url=http%3A%2F%2Fluxor.acadiau.ca%2Flibrary%2FASAC%2Fv26%2F02%2F26_02_p032.pdf&ei=n7_BUs-ZD-im4ASM54DgAg&usg=AFQjCNGAI2mYv1bIYOUYdjV2elLsa_oDiQ&bvm=bv.58187178,d.bGE&cad=rja Accessed 10.12.2013.

Gramm, M. & Ziemba, W.T. (2008). The dosage breeding theory for horse racing predictions. In D.B. Hausch & W.T. Ziemba (Editors), *Handbook of Sports and Lottery Markets*, Chap. 15, 307–340. North-Holland.

Harville, D.A. (1973). Assigning probabilities to the outcomes of multi-entry competions. *Journal of the American Statistical Association* 68, 312–316.

Hausch, D.B., Lo V.S.Y. & Ziemba, W.T. (1994, 2008). *Efficiency of Racetrack Betting Markets*, Academic Press; Reprinted as *Efficiency of Racetrack Betting Markets. 2008 Edition*. Singapore: World Scientific.

Hausch, D.B., Lo V.S.Y. & Ziemba, W.T. (1994). Pricing exotic racetrack wagers. In D.B. Hausch, V.S.Y. & W.T. Ziemba (1994, 2008). *Efficiency of Racetrack Betting Markets*, Academic Press; Reprinted as *Efficiency of Racetrack Betting Markets. 2008 Edition*. Singapore: World Scientific. 469–483.

Hausch, D.B. & Ziemba, W.T. (1985). Transaction costs, extent of inefficiencies, entries and multiple wagers in a racetrack betting model. *Management Science* 31, 381–394.

Hausch, D.B. & Ziemba, W.T. (1990a). Arbitrage strategies for cross-track betting at major horse races. *Journal of Business* 63, 61–78.

Hausch, D.B. & Ziemba, W.T. (1990b). Locks at the racetrack. *Interfaces* 20, 41–48.

Hausch, D.B. & Ziemba, W.T., Editors (2008). *Handbook of Sports and Lottery Markets*. North-Holland.

Hausch, D.B., Ziemba, W.T. & Rubinstein, M. (1981). Efficiency of the market for racetrack betting. *Management Science* 27, 1435–1452.

Henery, R.J. (1981). Permutation probabilities as models for *horse* races. *Journal of the Royal Statistical Society B* 43, 86–91.

Henery, R.J. (1984). An extreme-value model for predicting the results of horse races. *Applied Statistics* 33, 125–133.

Jullien, B. & Salanié, B. (2008). Empirical evidence on the preferences of racetrack bettors. In D.B. Hausch & W.T. Ziemba (Editors). *Handbook of Sports and Lottery Markets*, Chap. 3, 27–49. North-Holland.

Kanto & Rosenqvist (1994). On the efficiency of the market for double (quinella) bets at a Finnish racetrack. In D.B. Hausch, V.S.Y. Lo & W.T. Ziemba (1994, 2008). *Efficiency of Racetrack Betting Markets*, Academic Press; Reprinted as *Efficiency of Racetrack Betting Markets. 2008 Edition*. Singapore: World Scientific. 485–498.

Kanto, A.J., Rosenqvist, G. & Suvas, A. (1991). *Testing Harville's Formulas for Racetrack Betting*. Meddelanden från Ekonomisk-statsvetenskapliga fakulteten vid Åbo Akademi, Statistiska institutionen. Ser. A:334.

Kanto, A.J., Rosenqvist, G. & Suvas, A. (1992). On utility function estimation of racetrack bettors. *Journal of Economic Psychology* 13, 491–498.

Lo, V.S.Y. (1994). Application of running time distribution models in Japan. Pp. In D.B. Hausch, V.S.Y. Lo & W.T. Ziemba (1994, 2008). *Efficiency of Racetrack Betting Markets*, Academic Press; Reprinted as *Efficiency of Racetrack Betting Markets. 2008 Edition*. Singapore: World Scientific. 237–247.

Lo, V.S.Y., Bacon-Shone, J. & Busche, K. (1995). The application of ranking probability models to racetrack betting. *Management Science,* 41, 1048–1059.

Lo, V.S.Y. & Bacon-Shone, J. (2008). Approximaing the Ordering Probability of Multi-Entry Competitions by a Simple Method. In D.B. Hausch & W.T. Ziemba (2008) (Editors). *Handbook of Sports and Lottery Markets,* Chap 4, 51–65. North-Holland.

Lo, V.S.Y. & Busche, K. (1994). How accurately do bettors bet in doubles? In D.B. Hausch, V.S.Y. Lo & W.T. Ziemba (1994, 2008). *Efficiency of Racetrack Betting Markets*, Academic Press; Reprinted as *Efficiency of Racetrack Betting Markets. 2008 Edition*. Singapore: World Scientific. 465–468.

Ludlow, L.H. (1994). An empirical cross-validation of alternative classification strategies applied to harness racing data for win bets. In D.B. Hausch, V.S.Y. Lo & W.T. Ziemba (1994, 2008). *Efficiency of Racetrack Betting Markets*, Academic Press; Reprinted as *Efficiency of Racetrack Betting Markets. 2008 Edition*. Singapore: World Scientific. 199–212.

McCulloch, B. & van Zijl, T. (1986). Direct test of Harville's multy-entry competitions model on race-track betting data. *Journal of Applied Statistics* 13, 213–220.

Ottaviani, M. & Sorensen, P.N. (2008). The Favorite-Longshot Bias: An Overview of the Main Explanations. In D.B. Hausch & W.T. Ziemba (Editors). *Handbook of Sports and Lottery Markets,* Chap 6, 83–101. North-Holland.

Rosenbloom, E.S. (1992). Picking the lock: A note on "Locks at the Racetrack". *Interfaces* 22, 15–17.

Smith, M.A. & Vaughan Williams, L. (2008). Betting Exchanges: A Technological Revolution in Sports Betting. In D.B. Hausch & W.T. Ziemba (Editors). *Handbook of Sports and Lottery Markets,* Chap 19, 403–418. North-Holland.

Snowberg, E. & Wolfers, J. (2008). Examining Explanations of a Market Anomaly: Preferences or Perceptions? In D.B. Hausch & W.T. Ziemba (Editors). *Handbook of Sports and Lottery Markets,* Chap 7, 103–136. North-Holland.

Snyder, W.W. (1978). Horse racing: testing the efficient markets model. *The Journal of Finance* 33, 110–1118.

Sobel, R.S. & Ryan, M.E. (2008). Unifying the favorite-longshot bias with other market anomalies. In D.B. Hausch & W.T. Ziemba (Editors). *Handbook of Sports and Lottery Markets,* Chap 8, 137–160. North-Holland.

Stern, H. (1990). Models for distributions on permutations. *Journal of the American Statistical Association* 85, 558–564.

Sung, M. & Johnson, J.E.V. (2008). Semi-strong form information efficiency in horse race betting markets. In D.B. Hausch & W.T. Ziemba (Editors). *Handbook of Sports and Lottery Markets,* Chap 14, 275–306. North-Holland.

Tuckwell, R.H. (1981). Anomalies in the gambling market. *Australian Journal of Statistics* 23, 287–295.

Weitzman, M. (1965). Utility analysis and group behavior: An empirical study. *Journal of Political Economy* 73, 18–26.

Ziemba, W.T. (2008). Efficiency of racing, sports andf lottery betting markets. In D.B. Hausch & W.T. Ziemba (Editors). *Handbook of Sports and Lottery Markets,* Chap 10, 183–222. North-Holland.

Ziemba, W.T. & Hausch, D.B. (1987). *Beat the Racetrack*. Harcourt Brace Williams Morrow, New York.

FRIDAY THE THIRTEENTH AND STOCK INDEX RETURNS

Jarkko Peltomäki
Stockholm University

Emilia Vähämaa
Hanken School of Economics

1 Introduction

Many calendar anomalies are considered to be persistent (see Lakonishok & Smidt, 1988). Friday the thirteenth, however, even though being a very distinguishable calendar day, is not among these persistent anomalies. This potential calendar effect may be expected to exist as lower returns on Friday the thirteenth than normally, given that people are superstitious, and that Friday the thirteenth is attributed to bad luck. Investigating the existence of Friday the thirteenth effect on the stock market is important, because an abnormal market behavior would imply that the stock market is affected by human superstition, as indicated by Chamberlain, Cheung & Kwan (1991).

The stock market participants, even though having said to be subject to fear and greed, do not seem to be affected by Friday the thirteenth. The early evidence on the existence of this possible anomaly by Kolb & Rodriguez (1987) is rejected by many studies, including Dyl & Maberly (1988), Chamberlain et al. (1991), Coutts (1999), Lucey (2001), and Patel (2009). However, Zweig (2009) documents that Friday the thirteenth have generated an average return of 0.28%, whereas, on average, the market goes up by 0.2% per day in a long run.

In this study, we re-examine whether Friday the thirteenth affects the stock market returns by taking a new route in investigating the problem, in relation to the previous literature presented above. We argue that the superstition at Friday the thirteenth does not necessarily affect the Friday returns, but can instead be seen on the day close to Friday the thirteenth. Therefore, we also focus on the days before and after Friday the thirteenth. In our analyses, we use the S&P 500 index returns over the period 1/1950–7/2009 and the Dow Jones Industrial average index returns over the period 10/1928–7/2009.

We expect the stock returns to be lower on the trading day before Friday and higher on the trading day after Friday the thirteenth than normally. This is explained by the fear of Friday the thirteenth, which should cause selling during the day before Friday the thirteenth and, once the new trading day after Friday the thirteenth starts, the investors in relief should return to buy stocks. This analogy is simple; if one is afraid of a monster under the bed, one is probably not willing to look there to realize how afraid of the monster one can be. So, if one fears Friday the thirteenth and owns stocks, one dislikes holding them until Friday and, since no disaster existed, buys them back after Friday the thirteenth.

We also expect the relation between the returns on Friday the thirteenth and the following Monday to be negative, given that poor returns are considered as a buying opportunity on Monday due to Gambler's fallacy. This fallacy states that investors anchor on the information on prior Friday the thirteenth closing prices, and may thereby they see an opportunity after worse than expected returns on Friday the thirteenth. And this Friday is just considered as an unlucky day. This should also cause the returns on Monday after Friday the thirteenth to be higher than normally, as a result of relief. Since the relation between the returns of Friday the thirteenth and the following Monday is considered, our paper also contributes to research on the weekend effect (see e.g. French 1980: Gibbons & Hess 1981; Keim & Stambaugh 1984; Lakonishok & Smidt 1988) in addition to the research on Friday the thirteenth and stock returns (see e.g. Lucey 2000; Patel 2009).

Our results suggest that Friday the thirteenth affects stock returns, thus further implying that the stock market is affected by superstition. We find that, for most periods, the returns prior Friday the thirteenth are lower than on average, and the returns after the day are higher than on average. The relation between Monday and Friday returns is also affected by Friday the thirteenth, since the relation between the returns on Friday the thirteenth and the following Monday are found to be negative, while normally the relation is positive.

The remainder of this paper is organized as follows. Section 2 revises the earlier literature on related financial anomalies and introduces the research hypotheses and their development. Section 3 proposes the data and methodologies employed in the statistical analyses, while Section 4 introduces the estimation results. Finally, the last section provides concluding remarks.

2 Literature review and hypotheses

2.1 Financial anomalies

Previous literature has reported several irrational investment behavior patterns, so called anomalies. In financial context, an anomaly is commonly determined as a documented pattern of price behavior, which is in contrast with the efficient market hypothesis (see e.g. Brav & Heaton 2002). Among the most studied financial anomalies are the ones of size, season, and price ratio (see e.g. Zivney & Thompson 1987; Chen & Chan 1997; Dobbs 1999; Gaunt 2004). The research on calendar anomalies is the avenue of research on financial anomalies which is the most closely related to this study.

Perhaps the best known anomaly is the January effect, which predicts that the January returns, particularly those of the small stocks, are higher than in the other months. This anomaly is evinced by many studies, such as Keim (1983), Reinagum (1983), and Lakonishok et al. (1988).

For monthly returns, Ariel (1987) finds that stock returns are on average positive only for the days immediately before and during the first half of calendar months. Cadsby & Ratner (1992) find further and statistically significant evidence for high returns around the turn of the month in Canada, the U.K., Australia, Switzerland, and West-Germany.

For weekdays, a significant stream of literature has reported the existence of a Monday effect, i.e. the average stock returns of Mondays being negative (see e.g., French 1980; Gibbons & Hess 1981; Keim & Stambaugh 1984; Lakonishok & Smidt 1988). In addition to stock markets, a similar Monday effect has been reported to exist e.g. in the gold price, exchange rates, and real estate investment trusts (see e.g. Ball et al., 1982; McFarland et al. 1982; Ma 1986; Redman et al. 1997; Thatcher & Blenman 2001). There are various possible explanations for the Monday effect, but one the researchers could agree upon is still to be found. The so far suggested explanations include, for example, a misapplication of statistical methods, a response to market arrangements or to micro or macro information, and the trading patterns of market participants (see Pettengill 2003 for a review).

The Monday anomaly, however, has recently been found found to be nearly nonexistent. Connolly (1989) uses robust econometric methods to study the weekday returns. His results suggest that the Monday effect has disappeared by 1975. Brusa et al. (2000) use a sample period over the period January 1, 1990 to December 31, 1994 and find evidence for a reverse Monday effect. Mehdan & Perry

(2001) consider more recent stock returns and find that the Monday effect is unstable and time-varying over the period 1964–1998. The evidence for the Monday effect is followed by Brusa et al. (2005) and Cho et al. (2007). Yet, Brusa et al. (2003) still find evidence for Monday effect in foreign markets from U.S. investors' point of view. It is also objectionable whether the Monday day in itself is the cause for the Monday effect. Draper and Paudyal (2002) study Monday effect by controlling for various effects, such as trading activity, which may be associated with the Monday returns. Their results suggest that, after controlling for these effects, the Monday returns become positive.

In addition to the day of the week, several other types of factors have been reported to have an effect on the stock returns. For example, Kim and Park (1994) report abnormally high returns on the trading days preceding holidays. Moreover, also rain and time changes around daylight savings have been documented to effect the investor mood and, thus, also the stock returns (see e.g. Dowling & Lucey 2005).

The evidence for calendar anomalies also includes the Halloween effect. Bouman and Jacobsen (2002) report evidence suggesting the Halloween indicator 'sell in May and go Away' to be profitable, as the November–April returns are found to be higher than the May–October returns. Yet, Maberly & Pierce (2004) suggest that this anomaly is rather subject to the Crash of October 1987 and the Collapse of Long Term Capital Management in August 1998, than a phenomenon caused by the Halloween.

Friday the thirteenth is also considered as a potential calendar anomaly. In an early paper, Kolb & Rodriguez (1987) examine whether the returns for Friday the thirteenth are significantly lower than the returns for other Fridays using a sample of CRSP value- and equally weighted indexes over the period July 1, 1962 – December 31, 1985. The authors find results suggesting that the market returns for Friday the thirteenth are significantly lower than the returns for other Fridays in general.

Dyl & Maberly (1988) continue examining whether the returns for Friday the thirteenth are lower than for other Fridays using a Standard & Poor's 500 index over the period 1940–1987, and consider five sub-periods during this period. The authors state that there is no so-called Friday the thirteenth effect. In fact, their evidence indicates that the mean return of Friday the thirteenth is higher than the average Friday returns.

Chamberlain et al. (1991) also examine the S&P composite index and find evidence suggesting that the so-called Friday the thirteenth effect does not have an effect on investment behavior. They conclude that, after considering the turn of the month anomaly, there is no evidence implying that Friday the thirteenth would influence market returns.

Coutts (1999) examines the issue for the period 1935–1994 and reports that the returns for Friday the thirteenth are higher than on other Fridays. In a similar vein, by examining the FTSE world indices, Lucey (2000) proposes that the Friday the thirteenth returns are generally greater than the returns on other Fridays. In a recent study, Patel (2009) suggests that the U.S. stock returns do not show evidence supporting Friday the thirteenth phenomenon. By examining the S&P 500 firms over the period 1950–2007 he provides evidence indicating that Friday the thirteenth returns are not significantly different from the other Friday returns.

2.2 Hypothesis development

People are, in general, prone to superstitious behavior. For example, Risen (2008) proposes that people tend to have an intuition that tempting the fate increases the possibility of a negative outcome. Friday the thirteenth is commonly believed to be an unlucky day and, thus, many investors may expect these Fridays to be unprofitable also in the stock market. Social effect of Friday the thirteenth is evinced by many studies, and the fear of Friday the thirteenth is known as the paraskevidekatriaphobia. However, the Dutch Centre for Insurance Statistics reports that less accidents tend to happen on Friday the thirteenth, although the difference is statistically insignificant (Dutch Centre for Insurance Statistics, 2008 p. 14) . The evidence for the stock market is similar to the accidents reported by the Dutch Centre for Insurance Statistics. Dyl et al. (1988), Chamberlain et al. (1991), Coutts (1999), Lucey (2001), and Patel (2009) suggest that Friday the thirteenth returns are not lower, and may even be higher, than the other Friday returns. However, if investors fear Friday the thirteenth, they may anticipate bad luck on Friday and sell stocks on the trading day before Friday, thus affecting stocks returns negatively. As a result, the following hypothesis is presented:

H1: Returns on the day before Friday the thirteenth are lower than on average.

Cross (1973), and Kleim & Stambaugh (1984) present evidence for a positive correlation between Friday and Monday returns. If Friday the thirteenth would be information used by individuals, it should be seen in the returns following the day, thereby implying that it would also affect the relation between Friday and Monday returns. As the returns for Friday the thirteenth may be considered as bad luck, the Gambler's fallacy (see Kahneman & Trevsky, 1974) may act as a catalyst among investors to buy stocks after Friday the thirteenth if the downward deviations from the expected returns for Friday the thirteenth are considered to be bad luck. After Friday the thirteenth, people may believe that the bad luck was supposed to be present on Friday, but not anymore on Monday. The returns for Friday the thirteenth are also pieces of past information on which people may anchor (see Kahneman & Trevsky, 1974). Thursday price level may act as a motivator for superstitious investors to return to the market, thereby affecting Monday returns positively if the price level is lower, and vice versa. Thus, Friday the thirteenth may affect both the serial correlation between Friday and Monday, and the Monday returns. As a result, we present the following hypotheses:

H2: The returns on Monday after Friday the thirteenth are higher than on other Mondays.

H3: The returns on Friday the thirteenth have a negative relation to Monday returns.

3 Data and methodology

For the empirical analyses of this study, we use the returns for the S&P 500 index and the Dow Jones Industrial Average index. To test Hypotheses 1 and 2, we carry out our analysis using time-series ordinary least squares (OLS). To test Hypothesis 3, we carry out our analysis using a pooled OLS analysis of the returns on Fridays and Mondays. Our model for the S&P 500 and the Dow Jones Industrial Average indexes (all returns used are expressed as percentages) to test Hypotheses 1 and 2 is the following:

$$R_{i,t} = \alpha_0 + \alpha_1(\mathrm{MONDAY})_t + \alpha_2(\mathrm{THURSDAY})_t + \alpha_3(\mathrm{FRIDAY})_t$$

$$+ \alpha_4(\mathrm{BEFORE})_t + \alpha_5(\mathrm{THE13TH})_t + \alpha_6(\mathrm{AFTER})_t + \alpha_7 R_{i,t-1} + e_{i,t}, \quad (1)$$

where $R_{i,t}$ defines the return for an index; $(MONDAY)_t$ indicates a dummy variable for a Monday return; $(THURSDAY)_t$ denotes a dummy variable for a Thursday return; $(FRIDAY)_t$ defines a dummy for a Friday return; $(BEFORE)_t$ is a dummy variable indicating the trading days before Friday the thirteenth; $(THE13TH)_t$ is a dummy variable indicating the trading days on Friday the thirteenth, and $(AFTER)_t$ is a dummy variable indicating the trading days after Friday the thirteenth. Hypotheses 1 and 2 predict that α_4 is negative and α_6 positive.

To test Hypothesis 3, data for Friday the thirteenth and the following trading day are pooled and investigated using the following model:

$$R_{Monday,t}=\alpha_0+\alpha_1(AFTER)_t+\alpha_2 R_{Friday,t}+\alpha_3(AFTER)_t{*}R_{Friday,t}+e_{Monday,t} \qquad (2)$$

where $R_{Monday,t}$ defines the return for Monday, and $R_{Friday,t}$ defines the return for Friday (or the last trading day prior to Friday). As the empirical evidence suggests that there is a correlation between the returns for Fridays and Mondays (see e.g., Kleim & Stambaugh, 1984), it is reasonable to control for this correlation. Therefore, a relative correlation between Friday the thirteenth and Monday returns is investigated using Equation 2, for which Hypothesis 3 predicts that α_3 has a negative value. Equation 2 can also be used to test Hypothesis 2 which predicts that α_1 is positive.

4 Results

Table 1 presents the ordinary least squares (OLS) analysis for Friday the thirteenth and stock market returns using Equation 1. The results do not provide any statistically significant evidence for the Friday the thirteenth return being lower than on other Fridays. This evidence is consistent with previous evidence by Dyl et al. (1988), Chamberlain et al. (1991), Coutts (1999), Lucey (2001), and Patel (2009).

The results reported in Table 1 provide partial support for Hypothesis 1, which predicts that the returns on the day before Friday the thirteenth are lower than otherwise. However, the evidence only concerns the subperiod 1/1950–1/1980 for the S&P 500 sample, and the subperiods 10/1928–1/1950 and 1/1950–12/1980 for the Dow Jones sample. As such, Hypothesis 1 is not supported for the most recent sample period.

Table 1, Panel A. Friday the thirteenth and stock market returns. Regression results for the S&P500 sample. The table presents Ordinary Least Squares (OLS) analysis statistics of the analyses of Monday and Friday returns for the following regression model.

$$R_{i,t}=\alpha_0+\alpha_1(\mathrm{MONDAY})_t+\alpha_2(\mathrm{THURSDAY})_t+\alpha_3(\mathrm{FRIDAY})_t$$
$$+\alpha_4(\mathrm{BEFORE})_t+\alpha_5(\mathrm{THE13TH})_t+\alpha_6(\mathrm{AFTER})_t+\alpha_7 R_{i,t-1}+e_{i,t},$$

where $R_{i,t}$ defines the return for an index; *(MONDAY)*$_t$ indicates a dummy varia-ble for a Monday return; *(THURSDAY)*$_t$ denotes a dummy variable for a Thursday return; *(FRIDAY)*$_t$ defines a dummy for a Friday return; *(BEFORE)*$_t$ is a dummy variable indicating the trading days before Friday the thirteenth; *(THE13TH)*$_t$ is a dummy variable indicating the trading days on Friday the thirteenth, and *(AFTER)*$_t$ is a dummy variable indicating the trading days after Friday the thirteenth. The t-statistics in brackets use Newey-West heteroskedasticity and autocorrelation robust standard errors (lag=12).

Variable	Exp. sign	whole sample		1/1950–1/1980		1/1981–7/2009	
Constant	+	0.001	***	0.001	***	0.001	***
		(4.95)		(5.19)		(2.95)	
Variables:							
Lag return	+	0.038	***	0.180	***	-0.028	*
		(2.79)		(11.63)		-(1.77)	
THE13TH	?	0.000		-0.001		0.002	
		(0.13)		-(1.02)		(0.83)	
BEFORE	-	0.000		-0.002	**	0.003	
		(0.12)		-(2.44)		(1.48)	
AFTER	+	0.002		0.000		0.004	***
		(1.62)		(0.02)		(2.86)	
MONDAY	-	-0.001	***	-0.002	***	-0.001	
		-(5.30)		-(8.14)		-(1.58)	
THURSDAY	?	0.000		0.000		0.000	
		-(1.15)		-(1.05)		-(1.25)	
FRIDAY	?	0.000		0.000	*	0.000	
		(0.67)		(1.80)		-(1.03)	
Adjusted R^2		0.004		0.045		0.001	
F-stat.		9.850	***	52.862	***	2.314	**
n		14 980		7 777		7 203	

Table 1, Panel B. Friday the thirteenth and Stock Market Returns. Regression results for the Dow Jones sample.

Variable	Exp. sign	whole sample		10/1928–12/1949		1/1950–12/1980		1/1981–7/2009	
Constant	+	0.058	***	0.060	*	0.061	***	0.062	***
		(4.72)		(1.88)		(4.67)		(3.16)	
Variables:									
Lag return	+	0.016		-0.008		0.169	***	-0.026	*
		(1.15)		-(0.34)		(11.38)		-(1.74)	
THE13TH	?	-0.064		-0.260		-0.023		0.071	
		-(0.64)		-(1.11)		-(0.25)		(0.37)	
BEFORE	-	-0.172	*	-0.630	***	-0.232	**	0.219	
		-(1.76)		-(2.92)		-(2.37)		(1.18)	
AFTER	+	0.090		-1.810		0.007		0.391	**
		(0.75)		-(0.54)		(0.05)		(2.40)	
MON	-	-0.133	***	-0.180	**	-0.220	***	-0.027	
		-(4.86)		-(2.54)		-(8.16)		-(0.58)	
THU	?	-0.024		0.003		-0.029		-0.060	*
		-(1.09)		(0.05)		-(1.24)		-(1.68)	
FRI	?	-0.009		-0.053		0.043	*	-0.052	
		-(0.46)		-(1.00)		(1.89)		-(1.55)	
Adjusted R^2		0.002		0.002		0.040		0.001	
F-stat.		6.318	***	2.461	**	47.508	***	2.244	**
n		20 288		5 305		7 779		7 204	

Table 1 also presents relevant results for testing Hypothesis 2, which predicts that the returns on Mondays after Friday the thirteenth are higher than during other Mondays. This hypothesis is only supported for the most recent sample period 1/1981–7/2009. In relation to the Monday anomaly (abnormally low returns on Mondays), it is an interesting observation that the Monday returns are not significantly lower in relation to the other days for the most recent sample period. For the earlier Dow Jones samples, the dummy variable for pre-Friday the thirteenth returns has relatively large values in comparison to the dummy variable for the Monday returns (-0.630 vs. -0.180 and -0.232 vs. -0.220), thus implying that Friday the thirteenth anomaly was economically significant when compared to the Monday effect. However, the value for the dummy variable for pre-Friday the thirteenth returns turns out to be positive for the latest sample period, while the value for the dummy variable for Monday returns is still negative, although statistically insignificant. Consistently, the value of the dummy variable for the Friday returns also appears to change from positive to negative, when moving to evaluate the latest period. This may be an effect of investors avoiding to adverse the previously documented low Monday returns, and the Monday effect has therefore shifted to Friday returns.

One more interesting finding in Table 1 related to the shift in the above mentioned characteristics over time is the serial correlation; for the period 1/1950–12/1980 it is positive and statistically significant, while for the period 1/1981–7/2009 it is negative and statistically significant, using both indexes. This shift and the shift in the Monday anomaly are contemporaneous with the shift from the support for Hypothesis 1 to support for Hypothesis 2. To sum up the evidence presented in Table 1, there appears to be a change in the anomalous behavior of the stock market after the year 1980.

Table 2 presents the results of Monday and Friday returns to test Hypotheses 2 and 3. Hypothesis 2 is still supported as the coefficients for after Friday the thirteenth returns are statistically significant and positive. The results presented suggest that there is a statistically significant and positive serial correlation between the returns for Friday and Monday. The result is consistent with the study by Kleim & Stambaugh (1984).

In addition to the analyses above, we tested whether Friday the thirteenth affects changes in trading volumes of the S&P 500 and Dow Jones Industrial Average indexes. Our results do not indicate statistically significant evidence on the effect of Friday the thirteenth on the index volumes.

Table 2. Analysis of the Friday and Monday Returns. The table presents Ordinary Least Squares (OLS) analysis statistics of the analyses of Monday and Friday returns. The models used are the following:

$$R_{Monday,t}=\alpha_0+\alpha_1(AFTER)+\alpha_2 R_{Friday,t}+\alpha_3(AFTER)_t*R_{Friday,t}+e_{Monday,t}$$

where $(AFTER)_t$ is a dummy variable indicating the trading days after Friday the thirteenth; $R_{Monday,t}$ defines the return for Monday and $R_{Friday,t}$ defines the return for Friday. The t-statistics in brackets use Newey-West heteroskedasticity and autocorrelation robust standard errors.

Variable	Exp. sign	Mon S&P500		Mon Dow	
Constant (α_0)		-0,097	***	-0,073	***
		-(4,22)		-(3,20)	
Fri. Return (α_2)	+	0,259	***	0,227	***
		(4,15)		(3,84)	
AFTER (α_1)	-	0,200	**	0,253	**
		(2,10)		(2,48)	
Interaction (α_3)	-	-0,480	***	-0,536	***
		-(3,82)		-(4,57)	
Adjusted R^2		0,039		0,033	
F-stat.		39,551	***	33,271	***

5 Conclusions

The objective of our paper is to re-examine Friday the thirteenth effect by focusing on the trading days which precede and follow it. As such, this paper also focuses on the Monday effect, as Monday is normally the following trading day after Friday the thirteenth. We show that the earlier studies on the topic by Dyl et al. (1988), Chamberlain et al. (1991), Coutts (1999), Lucey (2001), and Patel (2009) do not find Friday the thirteenth to affect stocks markets because they do not investigate the returns for the preceding and following days. The preceding returns are statistically significant and negative before 1981, and the following returns are statistically significant and positive after 1980. The change in the Friday the thirteenth effect appears to be somewhat contemporaneous to the Monday effect, for which statistically significant and negative returns disappear, while statistically significant and positive Friday returns also disappear. This evidence is in line with the studies by Connolly (1989), Brusa et al. (2000), Mehdan et al. (2001), Brusa et al. (2005) and Cho et al. (2007) indicating that the Monday effect is not a robust anomaly. Change from statistically significant and positive serial correlation to statistically significant and negative correlation in stock returns is also in line with the structural change.

We argue that our study provides evidence indicating that anomalies are not robust, as they disappear, are unpredictable, and therefore cannot offer certain returns. However, as we find that the Friday the thirteenth appears to affect the stock market, we cannot conclude that the stock market would run rationally either. Moreover, the anomalies considered in this paper appear to be persistent over short periods of time and change their location from one to another. Turn of the 1980s decade also appears to be associated with a significant change in stock return behavior, and this scope is left open for future research.

References

Ariel, R.A. (1987). A monthly effect in stock returns. *Journal of Financial Economics* 18, 161–174.

Ball, C., Torous, W. & Tschoegl, A. (1982). Gold and the weekend effect. *Journal of Futures Markets* 2, 175–182.

Brav, A. & Heaton, J.B. (2002). Competing theories of financial anomalies. *The Review of Financial Studies* 15, 575–606.

Brusa, J., Liu, P. & Schulman, G. (2000). The “reverse” weekend effect: the U.S: market versus international markets. *International Review of Financial Analysis* 12, 267–286.

Brusa, J., Liu, P. & Schulman, G. (2003). The weekend effect, ‘reverse’ weekend effect, and firm size. *Journal of Business Finance & Accounting* 27, 555–574.

Brusa, J., Liu, P. & Schulman, G. (2005). The weekend effect, ‘reverse’ weekend effect, and investor trading activities. *Journal of Business Finance & Accounting* 32, 1495–1517.

Bouman, S. & Jacobsen B. (2002). The Halloween indicator, “Sell in May and go away”: Another puzzle. *American Economic Review* 92, 1618–1635.

Cadsby, C.B. & Ratner, M. (1992). Turn-of-the month and pre-holiday effects on stock returns: Some international evidence. *Journal of Banking & Finance* 16, 497–509.

Connolly, R.A. (1989). An examination of the robustness of the weekend effect. *Journal of Financial and Quantitative Analysis* 24, 133–169.

Chamberlain, T.W., Cheung, C.S. & Kwan, C.C.Y. (1991). The Friday the thirteenth effect: Myth or reality? *Quarterly Journal of Business and Economics* 30, 111–117.

Chen, C.R. & Chan, A. (1997). From T-Bills to stocks: Seasonal anomalies revisited. *Journal of Business Finance and Accounting* 24, 573–592.

Cho, Y.-H., Linton, O. & Whang, Y.-J. (2007). Are there Monday effects in stock returns: A stochastic dominance approach. *Journal of Empirical Finance* 14, 736–755.

Coutts, J.A. (1999). Friday the thirteenth and the Financial Times Industrial Ordinary Shares Index 1935–1994. *Applied Economics Letters* 6, 35–37.

Cross, F. (1973). The behavior of stock prices on Fridays and Mondays. *Financial Analysts Journal* 29, 67–69.

Draper, P. & Paudyal, K. (2002). Explaining Monday returns. *Journal of Financial Research* 25, 507–520.

Dobbs, I.M. (1999). The anomaly of size: Does it really matter? *International Journal of Finance and Economics* 4, 179–192.

Dowling, M. & Lucey, B.M. (2005). Weather, biorhythms, beliefs and stock returns – Some preliminary Irish Evidence. *International Review of Financial Analysis* 14, 337–355.

Dutch Centre for Insurance Statistics (2008). *Verzekerd!* 12, 6–12.

Dyl, E.A. & Maberly, E.D. (1988). The anomaly that isn't there: A comment on Friday the thirteenth. *Journal of Finance* 43, 1285–1286.

French, K.R. (1980). Stock returns and the weekend effect. *Journal of Financial Economics* 8, 55–70.

Gaunt, C. (2004). Size and book to market effects and the Fama French three factor asset pricing model: Evidence from the Australian stock market. *Accounting and Finance* 44, 27–44.

Gibbons, M. & Hess, P. (1981). Day of the week effect and asset returns. *Journal of Business* 54, 579–596.

Kahneman, D. & Trevsky, A. (1974). Judgment under uncertainty: Heuristics and biases. *Science* 185, 1124–1131.

Keim, D.B. (1983). Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of Financial Economics* 12, 13–32.

Keim, D.B. & Stambaugh, R.F. (1984). A further investigation of the weekend effect in stock returns. *Journal of Finance* 39, 819–840.

Kim, C.-W. & Park, J. (1994). Holiday effects and stock returns: Further evidence. *Journal of Financial and Quantitative Analysis* 29, 145–157.

Kolb, R.W. & Rodriguez, R.J. (1987). Friday the Thirteenth: "Part VII" – A note. *Journal of Finance* 42, 1385–1387.

Lakonishok, J. & Smidt, S. (1988). Are seasonal anomalies real? A ninety year perspective. *Review of Financial Studies* 403–425.

Lucey, B.M. (2000). Friday the 13th & the philosophical basis of financial economics. *Journal of Economics & Finance* 24, 294–301.

Ma, C. (1986). A further investigation of the day-of-the-week effect in the gold market. *Journal of Futures Markets* 409–420.

Maberly, E.D. & Pierce, R.M. (2004). Stock market efficiency withstands another challenge: Solving the "Sell in May/Buy after Halloween." *Econ Journal Watch* 1, 29–46.

McFarland, J., Pettit, R. & Sung, S. (1982). The distribution of foreign exchange prices changes: Trading day effects and risk measurement. *Journal of Finance* 37, 693–715.

Mehdan, S. & Perry, M.J. (2001). The reversal of the Monday effect: New evidence from US equity markets. *Journal of Business Finance & Accounting* 28, 1043–1065.

Patel, J.B. (2009). Recent evidence on Friday the thirteenth effect in U.S. stock returns. *Journal of Business & Economics Research* 7, 55–58.

Pettengill, G.N. (2003). A survey of the Monday effect literature. *Quarterly Journal of Business and Economics* 42, 3–27.

Redman, A., Manakyan, H. & Liano, K. (1997). Real estate investment trusts and calendar anomalies. *Journal of Real Estate Research* 14, 19–28.

Reinagum, M.R. (1983). The anomalous stock market behavior of small firms in January: Empirical tests for tax-loss selling effects. *Journal of Financial Economics* 12, 89–104.

Risen, J.L. (2008). Why people are reluctant to tempt fate. *Journal of Personality and Social Psychology* 95, 293–307.

Thatcher, J. & Blenman, L. (2001). Synthetic trades and calendar day patterns: The case of the Dollar/Sterling markets. *Financial Review* 3, 177–200.

Zivney, T.L. & Thompson, D.J. (1987). Size, seasonal, and price ratio anomalies in common stock returns. *Financial Review* 22, 123–123.

Zweig, J. (2009). Friday the 13th phobia? For investors, it's usually a good day. *The Wall Street Journal* 2009; Feb. 13.

TWO SPECULATIVE RISK MEASURES IN STOCK MARKETS

Martti Luoma
University of Vaasa

1 Introduction

Risk is a very crucial concept when investing in stocks or trading them. Therefore it is important for the investor and speculator to have a risk measure which measures the total risk of a stock, preferably for speculative use i.e. one which measures the risk in the future. In this article it will be shown that there is a lot of work to do for finding a risk measure which measures future risk well enough. That is what every investor and speculator would want to have. The aim is especially to demonstrate that the well known and much used volatility is a rather incomplete measure of risk

2 Measures of Risk

Volatility is one of many risk measures, perhaps the most used one. But does it work well enough? Nowadays there is some criticism against volatility as a measure of risk for stocks. The criticism has been on both theoretical and practical grounds. Volatility is by definition a measure for the price variation of a financial instrument over time. As a matter of fact volatility is defined as the standard deviation of (logarithmic) returns. Historic volatility is derived from time series of past market prices. In finance it is well known that risks of stock markets and stocks are changing all the time. It is not very logical to measure risk from time series with changing variation. It is rather evident that all remarkable single risks will not always be reflected in volatility. One alternative risk measure is a payback period. It is defined by the time it takes for a stock to earn back its share price. Formally we define the Earn Back period by the equation

$$\int_0^{EBP} f(t)dt = P,$$

where EBP denotes the Earn Back period, f(t) denotes future earnings inflows as determined by financial analysts' consensus forecasts, and P denotes the current stock price.

A stock with high risk should have a shorter payback period than a low risk stock. Luoma & Sahlström (2009) present the measure more thoroughly, see also Luoma & Ruuhela (2001), and Luoma, Sahlströn & Ruuhela (2006).

In order to establish superiority of either volatility or EBP it is necessary to compare their abilities to measure risk. But it is not easy to compare two totally different risk measures. There exists no true risk measure which would be a benchmark measure, neither it is possible to know the true amount of risk the stock in question has. The obvious question is why investors and traders use risk measures. The answer is also obvious: to make money. According to financial theories it is well-known that more risk stands usually for higher returns. Therefore it is logical to choose that measure of these two, which is better in making money. Our criterion is both empirical and practical.

3 Empirical Analysis

To find the answer we examine the twenty largest companies listed on the Helsinki Stock Exchange using historical data primarily from 2011-2012. The hypothetical buy date is February 13, 2012. We use several sell dates to get time periods of different lengths. The last hypothetical sell date is January 16, 2013. The time periods used are 1, 2, 3, 6, 9 and 12 months. For all six holding periods we compare the average returns of the ten most risky stocks as measured by EBP with those of the ten most risky stocks as measured by the volatility over the respective holding period. Table 1 presents all volatilities and EBP's as calculated on January 13, 2012. The volatilities are converted to yearly volatilities, which allows for comparisons between periods of different duration. The returns are corrected for dividends.

The characteristics EBP as used here is an ex ante measure, it is looking into the future. Therefore it is calculated using data available in the buy date including ana-lysts' consensus forecasts. Small values of EBP indicate high risk and large values indicate low risk. Now we take for every time period the average return of the most risky stocks using both risk measures. That risk measures - volatility or EBP – which results in higher average returns, is better.

Table 1. Risks and returns for the twenty largest companies listed on Helsinki Stock Exchange.

	Risk							Return					
Stock	Vola1m	Vola2m	Vola3m	Vola6m	Vola9m	Vola12m	EBP	Ret 1m	Ret 2m	Ret 3m	Ret 6m	Ret 9m	Ret 12m
Cargotec	41.9 %	59.0 %	70.4 %	67.3 %	59.1 %	53.8 %	9.5	16.2 %	20.1 %	9.1 %	-28.1 %	-25.9 %	-15.9 %
Elisa	10.1 %	17.0 %	18.3 %	21.6 %	20.4 %	19.8 %	12.6	3.8 %	7.2 %	7.0 %	11.9 %	16.9 %	12.3 %
Fortum	25.8 %	31.0 %	29.9 %	34.7 %	32.2 %	29.6 %	8.6	11.5 %	17.4 %	9.7 %	-2.7 %	-6.6 %	-5.9 %
Kesko B	32.4 %	37.2 %	38.0 %	41.2 %	36.6 %	34.1 %	11.1	-5.7 %	-4.4 %	-9.7 %	-19.2 %	-11.2 %	0.3 %
Kone B	21.8 %	27.2 %	30.9 %	35.3 %	30.8 %	28.9 %	10.2	4.0 %	7.3 %	2.4 %	22.6 %	39.3 %	42.3 %
Metso	34.7 %	43.4 %	52.7 %	61.1 %	52.9 %	47.7 %	8.7	10.8 %	18.8 %	9.3 %	-6.8 %	-5.1 %	19.0 %
Neste	32.2 %	43.1 %	53.5 %	58.6 %	49.8 %	45.5 %	7.1	14.5 %	10.7 %	7.8 %	9.2 %	34.2 %	31.0 %
Nokia	44.0 %	47.5 %	48.1 %	51.7 %	50.8 %	48.4 %	10.5	-10.1 %	-11.8 %	-29.9 %	-62.4 %	-49.4 %	-15.1 %
Nokia Tyres	39.4 %	52.1 %	53.1 %	61.3 %	52.7 %	48.2 %	7.9	24.0 %	35.4 %	34.2 %	16.0 %	29.4 %	20.4 %
Nordea	28.2 %	41.2 %	45.9 %	52.3 %	45.4 %	41.5 %	6.3	12.1 %	19.6 %	6.3 %	18.1 %	21.4 %	27.1 %
Orion B	20.1 %	21.2 %	22.7 %	26.1 %	23.8 %	22.1 %	11.6	9.1 %	11.2 %	7.7 %	18.0 %	28.5 %	51.6 %
Outotec	39.8 %	46.8 %	56.2 %	61.4 %	53.8 %	48.7 %	10.1	15.9 %	19.1 %	2.6 %	2.5 %	5.9 %	28.5 %
Pohjola Bank	36.1 %	43.8 %	47.5 %	45.2 %	40.5 %	37.2 %	7.0	7.7 %	10.0 %	1.6 %	23.7 %	38.7 %	60.7 %
Sampo	26.0 %	29.1 %	34.4 %	38.9 %	33.7 %	30.9 %	7.4	9.2 %	15.0 %	9.7 %	18.2 %	35.5 %	41.6 %
Sanoma	33.9 %	38.1 %	38.6 %	44.0 %	38.7 %	34.9 %	9.4	5.9 %	8.6 %	-9.1 %	-24.2 %	-22.7 %	-15.4 %
Stora Enso R	42.5 %	49.6 %	52.9 %	53.6 %	47.1 %	43.2 %	7.3	2.7 %	6.0 %	-1.9 %	-0.3 %	-5.0 %	7.9 %
TeliaSonera	21.0 %	28.0 %	29.9 %	34.4 %	31.3 %	28.9 %	8.3	3.1 %	3.3 %	-0.7 %	9.5 %	9.8 %	7.5 %
UPM-Kymmene	30.5 %	39.7 %	45.7 %	50.3 %	45.0 %	40.9 %	8.1	5.9 %	12.5 %	3.0 %	6.2 %	0.1 %	6.5 %
Wärtsilä	30.1 %	39.3 %	45.3 %	53.0 %	48.2 %	45.0 %	10.9	4.3 %	12.4 %	14.6 %	12.2 %	7.7 %	47.8 %
Yit Group	33.7 %	43.2 %	53.3 %	53.8 %	47.6 %	45.0 %	7.4	21.1 %	33.2 %	21.9 %	11.7 %	22.2 %	27.9 %

Table 2. Returns of risky stocks by different risk measures and time periods.

Risk measure	Ret 1m	Ret 2m	Ret 3m	Ret 6m	Ret 9m	Ret 12m
Volatility	8.9 %	16.1 %	6.1 %	-2.8 %	3.5 %	17.9 %
EBP	11.2 %	16.3 %	9.2 %	11.0 %	18.0 %	22.5 %

Table 2 shows that the most risky stocks selected by the EBP method render for all six holding periods larger returns than the method using volatility. The return difference for a holding period of two months is rather small, anyway. It is only 0.2 percent units. The differences for the other cases are rather undisputed. The largest differences appear at six and nine months, the maximum difference is 14.5 percent unit for a holding period of nine months.

3 Conclusion

According to this simple example it is anyway very clear: It can’t be taken for granted that volatility is the best or not even a good enough speculative risk measure. Risk and return are core concepts in investing and trading. Therefore all additional research in this area is very important and advisable.

References

Luoma, M. & Ruuhela, R. (2001). Screening for misvalued stocks. *Technical Analysis of Stocks & Commodoties*, May 2001, 34–38

Luoma, M & Sahlström, P. (2009). *A New Method for Estimation of Ex ante Equity Risk Premiums.* Acta Wasaensia 206. Universitas of Wasaensis.

Luoma M. & Sahlström P. & Ruuhela R. (2006). An alternative estimation method of the equity risk premium using financial statements and market data. *Advances in Accounting* 22, 229–238.