Reprinted from

OMEGA

Vol. 15, *No.* 2, *pp.* 145–155

FORECASTING STOCK MARKET PRICES IN A THIN SECURITY MARKET

ILKKA VIRTANEN and PAAVO YLI-OLLI

School of Business Studies, University of Vaasa, P.O. Box 297, SF-65101 Vaasa, Finland

PERGAMON PRESS

OXFORD · NEW YORK · BEIJING · FRANKFURT

SÃO PAULO · SYDNEY · TOKYO · TORONTO

1987

OMEGA Int. J. of Mgmt Sci., Vol. 15, No. 2, pp. 145–155, 1987
Printed in Great Britain.

0305-0483/87 $3.00 + 0.00

Forecasting Stock Market Prices in a Thin Security Market

ILKKA VIRTANEN
PAAVO YLI-OLLI

University of Vaasa, Finland

(*Received July 1986; in revised form September 1986*)

Stock market efficiency is a crucial concept when forecasting of future stock price behaviour is discussed. In the literature, a distinction is made between three potential levels of efficiency. Under a weak form of efficiency, information on historical price movements is of no value for predicting the future price development. Similarly, a semi-strong form of efficiency holds that no publicly available information can be successfully used in the prediction of prices. And finally, a strong form of efficiency means that the share prices fully reflect all relevant information including data not yet publicly available. Stock market efficiency has been extensively studied in different countries. On a thin security market, like in the Helsinki Stock Exchange, many anomalies and deviations from market efficiency have been obtained. This paper aims to contribute to that discussion. It is shown in the paper that both the monthly and quarterly stock market prices (the general stock market index) can be adequately forecasted using either univariate time-series analysis or multivariate econometric modelling. The univariate ARIMA-models seem to be slightly outperformed by the econometric models. It is further shown that the forecasting accuracy of the models can be improved when time-series and econometric forecasts are combined into a composite forecast. The empirical results obtained indicate an absence of efficiency on the Finnish security market.

1. INTRODUCTION

IN AN EFFICIENT market, a security's price is a good estimate of its investment value defined as discounted future cash flow. Any substantial disparity between price and value would reflect market inefficiency. In the literature, a distinction is made between three potential levels of efficiency. The market is efficient in a weak sense if share prices fully reflect the information implied by all prior price movements. The market is efficient in a semi-strong sense if share prices respond instantaneously and without bias to newly published information. And finally, the market is efficient in a strong sense if share prices fully reflect all relevant information including data not yet publicly available.

Market efficiency is a crucial concept when predictions of stock price behaviour are discussed. Under a weak form of efficiency, information on historical price trends is of no value for the prediction of either the magnitude or direction of price changes. As such the weak form is directly opposed to the basic premises of technical analysis or univariate time-series analysis (e.g. that presented by Box and Jenkins [4]) where the behaviour of the series is explained by its own past variability. Similarly, the semi-strong efficiency holds that all publicly available information is of no value in the prediction of future prices [16, p. 11–36]. Thus, the semi-strong form of efficiency is analogously directly opposed to the concept of fundamental analysis, e.g. multivariate econometric models for estimating future levels of stock prices.

Stock market efficiency has been tested extensively in recent years in the U.S. and Europe, [1, 8, 9, 18, 20]. The results of these studies are, as a rule, in support of the weak form of efficiency. However, the price changes of some German and Scandinavian stocks exhibit statistically significant dependence over time, [1, p. 181, 18, p. 8–25].

According to the empirical results of many studies made in the U.S. the stock market is efficient also in the semi-strong form, [18, p. 12–84]. However, we can also find opposite results [28, p. 427–441]. A survey of quite a small number of empirical studies performed on European data indicates that some European stock markets are also efficient in the semi-strong form [18, p. 46–49]. On a thin security market there may exist many anomalies and deviations from market efficiency even in the weak form, as found by Berglund [1]. These anomalies and deviations from market efficiency are the starting point of this research.

The purposes of this study are:

(1) To analyze whether, to what extent, and in which form, a general monthly and quarterly stock market price index is predictable on a thin security market like the Helsinki Stock Exchange.

(2) To compare the forecasting results based on univariate time-series analysis and multivariate econometric models with each other.

(3) To develop composite forecasting models for stock prices and examine the forecasting improvement of these models relative to the time-series models and econometric models.

2. DATA AND MODEL BUILDING

2.1 Data and empirical variables

In a statistical investigation concerning the predictability of stock prices we can forecast, first, the price level (see e.g. [17]), second, price changes [13, p. 58–59] and third, total returns where current dividend yield is added to the price changes [28].

From a theoretical point of view, total returns are the best criterion variable when the market efficiency is considered. In Finland, on the Helsinki Stock Exchange we have in public use the so-called Unitas and KOP stock market indices published by two Finnish commercial banks (the Unitas index by SYP and the KOP index by KOP, respectively). KOP and Unitas indices do not include the dividend component of market returns. In practice the dividend component of market returns in Finland is very stable and variation in total returns from month to month is almost entirely due to price fluctuations. The correlation coefficient between the theoretically correct total return index, developed by Berglund, Wahlroos and Grandell [2] and the Unitas stock market index is higher than the correlation coefficient between the total return index and the KOP index (the correlations between the changes of indices are 0.991 and 0.978 respectively [2, p. 39]. Therefore, the publicly available Unitas index is used in this research.

The predicted general index has been measured in the level form and as natural logarithms. There are two main reasons for the use of the logarithmic transformation. The first one is the empirical fact that there have been considerable changes in the value of the index, as there are in most economic time series, which tend to invalidate the assumption of a constant relationship between the original values of variables. The second reason is that, when using logarithms, the efficiency of the estimates is increased because heteroscedasticity in regression analysis is reduced [6, p. 11–12]. In the case of time series analysis, stationarity in variance can be achieved, respectively [22, p. 439].

The models will be estimated using both monthly and quarterly data. The monthly data contain more detailed information and they are, therefore, expected to produce more accurate models. However, the quarterly models will also be estimated in order to control and confirm the results based on monthly data, because some independent variables in our econometric analysis are observed only quarterly. The monthly values of those variables are interpolated. Further, the quarterly models generate one-step forecasts for three months, instead of forecasts for one month in monthly models.

2.2 Univariate time-series model building

Auto Regressive Integrated Moving Average (ARIMA) models have been studied extensively by Box and Jenkins [4], and their names have frequently been used synonymously with general ARIMA processes applied to time-series analysis and forecasting. For several years now, accounting and finance researchers have also applied this methodology to investigate, for instance, the behaviour of stock market prices.

Univariate modelling involves a single series observed at equally spaced intervals (e.g. monthly, quarterly or annually), and assumed to be generated by an autoregressive integrated moving average process. This technique examines the interstructure of the series, that is, the behaviour of the series explained by its own past variability. ARIMA models express the current value of a series (y_t) as a function of its past values as well as current and past values of a noise or error series (e_t):

$$\phi^p(B)(1-B)^d y_t = \theta^q(B)e_t. \quad (3.1)$$

In (3.1) B is the backward shift or lag operator such that

$$B^k y_t = y_{t-k}, \quad k = 1, 2, \ldots, \quad (3.2)$$

$(1-B)^d$ is the differencing operator to produce differences of order d for the series, $\phi^p(B)$ is the autoregressive polynomial in B of order p, $\theta^q(B)$ is the moving average polynomial in B of order q, and e_t is the noise series, that is assumed to be independent and normally and identically distributed over time. Model (3.1) is also called ARIMA (p, d, q)-model.

Model (3.1) is for non-seasonal processes. If we have to add seasonal components to the model, we get an ARIMA $(p, d, q)(P, D, Q)^S$-model, which in terms of the polynomials becomes

$$\phi^p(B)\Phi^P(B^S)(1-B)^d(1-B^S)^D y_t = \theta^q(B)\theta^Q(B^S)e_t. \quad (3.3)$$

In (3.3) S is the number of periods per season, Φ^P and θ^Q, the seasonal autoregressive and seasonal moving average polynomials, respectively, as well as the seasonal differencing operator $(1-B^S)^D$ are now all expressed in powers of B^S.

In the following the univariate Box-Jenkins method will be applied to model both monthly and quarterly stock prices on the Helsinki Stock Exchange. The data to be used for identifying, estimating and testing the model consist of the values of the Unitas Index in years 1975–84 (120 monthly values or 40 quarterly values). The rest of the data (from January 1985 to March 1986 in monthly data or from I/1985 to I/1986 in quarterly data) will be used for measuring the forecasting accuracy of the models.

2.2.1 Identification of tentative models. We start our analysis with the monthly data. The data indicate that the changes in the values of the Unitas index increase as one moves from the beginning to the end of the period. Until December 1981, the value of the index was low and so were the changes. From January 1982 until April 1984, prices increased and so did their variations from one month to the next (the value of the Unitas index tripled). The rest of the year 1984 shows a clear decrease in the prices.

The variation in the magnitude of the price changes with time is referred to as non-stationarity in the variance of the data. A stationary variance must be achieved, however, before fitting an ARIMA model to the series. As stated earlier, we apply the natural logarithm transformation to the Unitas index. This transformation stabilizes the variance quite well. It also appears reasonable from a theoretical point of view since the first differences of the natural logarithms will closely approximate the percentage change for the month (the data to be analyzed consist of index numbers, where only percentage, not absolute, changes have any empirical meaning).

Due to a strong trend even in the logarithmic transformed series, the data still have non-stationarity in the mean. This can be seen for example from the autocorrelation function which does not dampen out quickly (in the partial autocorrelation function the first partial is very dominant, too). To achieve stationarity, we must difference the series. The trend in the transformed series is stronger than linear. Therefore, taking of the first differences of the data is not enough to fully stationarize it: the graph of the series still has a slight (but statistically significant) trend, the autocorrelations dampen out quite slowly, and the first partial is dominant. That is to say, we must take second differences of the data, and reanalyze.

As a result, we obtain that the logarithmic data are now, after double differencing, stationary in both the mean and variance. Further, there is a suggestion of a MA(1) process for this twice differenced series, because only the first autocorrelation is large and significant and because the first few partials decay exponentially. With this interpretation, a tentative model for the logarithms of the Unitas index would be ARIMA (0, 2, 1), or

$$(1-B)^2 y_t = (1-\theta B)e_t. \quad (3.4)$$

A closer look at the index series shows, however, that there also exists a clear, although not very strong, seasonality pattern in the data. The value of the stock market index is, for example, high at the beginning of each year and low in early summer and late autumn, indicating a 12-month seasonal pattern. The autocorrelation function of the stationarized series supports this impression fairly well.

To take also the seasonal effect into account, we take, in addition to the logarithmic transformation and double differencing of the raw data, a 12-month seasonal difference. The resulting series has two spikes in its autocorrelation function (with lags of one and twelve months), and a partial autocorrelation function with exponentially decaying values. This all suggests a seasonal MA(1)-component to be added to the model. In comparison with our non-seasonal model (3.4) we thus have a tentative seasonal model ARIMA $(0,2,1)(0,1,1)^{12}$, or

$$(1-B)^2(1-B^{12})y_t=(1-\theta B)(1-\Theta B^{12})e_t. \qquad (3.5)$$

The behaviour of the quarterly Unitas index is very similar to that of the monthly index (the quarterly index is computed as arithmetic means of the relevant monthly values). The main difference is that the seasonal variation becomes much more clear-cut, due to reduced random variation in the more aggregated and thus smoothened raw data.

Derivation of a tentative model for the quarterly index proceeds as in the case of the monthly data. The transformed data, after taking the natural logarithm, the second non-seasonal and the first (four-quarter) seasonal difference, are stationary in both the mean and variance. The autocorrelations and partials suggest a seasonal MA(1)-process for the transformed series (there is now no sign of any non-seasonal MA(1)-process as was the case for the monthly series, the fourth autocorrelation is the only large and significant one). We thus have an ARIMA $(0,2,0)(0,1,1)^4$-model, or

$$(1-B)^2(1-B^4)Y_t=(1-\Theta B^4)E_t, \qquad (3.6)$$

where we have used capital letters Y and E referring to quarterly values of the logarithm of the Unitas index and error term, respectively.

2.2.2 Estimating the parameters. The parameters have been estimated by a nonlinear Gauss–Marquardt algorithm using the backcasting method [5]. The results from the estimation are presented in Table 1.

We can see that the results are very stable and consistent. The first-order moving average parameter θ gets in both of the monthly models (3.4) and (3.5) a value near to 0.75. The seasonal MA(1)-parameter, both in the monthly model (3.5) and in the quarterly model (3.6), is of the size 0.83. We can further see that in all the models the true values of the parameters fall with probability (higher than) 0.95 into the interval $0 < \text{parameter} < 1$, as desired.

Table 1. Summary of estimation and test statistics associated with the tentative ARIMA-model

Model	Parameter	St. error	t-value	D.o.f.
(3.4)	$\hat{\theta}=0.7591$	0.0614	12.37	117
(3.5)	$\hat{\theta}=0.7529$	0.0627	12.01	104
	$\hat{\Theta}=0.8396$	0.0359	23.40	104
(3.6)	$\hat{\Theta}=0.8332$	0.0556	14.98	33

2.2.3 Diagnostic checking and testing of the models. After having estimated the parameters of tentatively identified ARIMA-models, it is necessary to verify that the models are adequate. There are basically two ways of doing this, [22, p. 446]:

(1) Study the residuals—to see if any pattern remains unaccounted for, and
(2) study the sampling statistics of the current solution—to see if the model could be simplified.

We have three models to consider, two based on monthly data (the non-seasonal model (3.4) and seasonal model (3.5)) and one on quarterly data (model (3.6)). The optimum values for the parameters of the above models were given in Table 1.

The autocorrelation coefficients for the residuals of all the three models were examined for several time lags to see if any of them were significantly different from zero. The analysis showed that all the models (3.4)–(3.6) have produced residuals for which the autocorrelations are essentially random: none of the individual autocorrelations was significantly different from zero and the Ljung-Box statistics (see [5, p. 690]) revealed no evidence of inadequacy of fit for any of the models.

In Section 2.2.2 we already saw that the estimates of the parameters are very stable. The standard errors of the estimates are small producing quite narrow confidence intervals for the parameters. In testing significance this means that the parameter values are statistically significant. The relevant test statistic is also included in Table 1. All the t-values obtained are highly significant, the probabilities associated with them are all less than 0.001.

Another way to evaluate the appropriateness

Table 2. The residual statistics for the estimated ARIMA-models

Model	Residual sum of squares	D.o.f.	Residual mean square (RMS)	Square root of RMS
(3.4)	0.094177	117	0.000805	0.0284
(3.5)	0.076103	104	0.000732	0.0271
(3.6)	0.061333	33	0.001859	0.0431

of a model is to figure out how well the model fits with the data. Residual mean square (mean of squared errors of the fitted values) and its square root are common measures for this purpose. These residual statistics for different models are presented in Table 2.

To judge the size of the different RMS's it is worth noting that the statistics associated with the explained variable y_t or Y_t (the natural logarithm of the Unitas stock market index) are the following: mean = 4.688, standard deviation = 0.400, standard error of the mean = 0.0365 (for y_t) and 0.0632 (for Y_t). The adequacy of the fit is evident.

2.2.4 Discussion. The results of the preceding analysis can now also be scrutinized in the light of the question about market efficiency. Under the weak form of efficiency, information on historical price trends is of no value when predicting future stock prices. In terms of Box–Jenkins methodology it means that the prices follow a random walk model. A random walk in the general price level (in the logarithm of the general Unitas index, to be exact) would take the following form, see e.g. [29, p. 258]:

$$y_t = y_{t-1} + e_t + \delta \tag{3.7}$$

or

$$(1 - B)y_t = e_t + \delta, \tag{3.7}$$

where δ is a possible trend or drift in the series.

The preceding analysis showed that the Finnish stock market prices do not follow, either in the level form (y_t and Y_t) or as percentage price changes ($y_t - y_{t-1} = (1 - B)y_t$ and $Y_t - Y_{t-1} = (1 - B)Y_t$), the general random walk model. At least one moving average component must be included in the ARIMA-model. This can be interpreted as a violation against the existence of equity market efficiency on the Finnish market, even in the weak sense (transaction costs have been omitted in the analysis). The interpretation must be seen, however, in the light that it is based on the use of average values (monthly or quarterly) in the analysis [30, p. 916].

2.3 Econometric models

2.3.1 The selection of explanatory variables. The results of the preceding section show that a random walk model is not a tenable representation of general price movements on the thin Finnish security market. The results seem to indicate deviations from the weak form of the efficient market hypothesis.

If the market is not efficient in the weak sense it can not be efficient in the semi-strong sense either. A logical step is now to analyze the usefulness of information other than the past historical data of prices to predict the monthly and quarterly stock market prices.

In this section we develop a forecasting model based on multiple regression analysis, in order to predict the future development of stock market prices. Our econometric model includes six groups of explanatory variables or 'the kinds of information' which *a priori* can be supposed to affect the development of stock market prices. The classification of the variables is the following:

(1) Lagged endogenous variable. The first explanatory variable is, according to the results of univariate time-series models, the endogenous variable itself lagged one period (regression coefficient is *a priori* positive). This variable does not totally exhaust the effect of past historical development on the variable itself. However, we do not make experiments with very many kinds of distributed-lag models because of other explanatory variables. Further, composite stock price forecasting models to be presented in the following section include this 'full' past history.

(2) The aggregated future cash-flow of the firms. As a crude surrogate of this exogenous variable we use the anticipated order stock next period as compared to now in Finnish industry ('decreases' answers; the regression coefficient is *a priori* negative). The data originate from the business surveys by the Confederation of the Finnish Industries (CFI, see in detail Teräsvirta [27, pp. 3–4 and p. 21]). In an efficient market stock prices change significantly only in response to unanticipated changes in prospects for future cash-flow.

(3) Interest rates of bank deposits or the return of the state bonds. Our hypothesis is that

if the return of bank deposits or bonds decreases, *ceteris paribus*, stock prices will rise and the lag is some months, or one year and some months (time deposits).

(4) The supply of money. According to Sprinkel's forecasting framework, changes in the stock of money and changes of stock market index are leading indicators for an economy. The lead time for monetary supply is longer than the lead time for stock market prices. Therefore, the money stock can be used to predict changes in the level of the market index [3, p. 229–230]. According to our hypothesis the rise in money supply will raise the stock market prices.

(5) Inflation. The classical Fisherian theory implies that common stocks of unlevered firms serve as an effective inflation hedge during anticipated inflation [21, p. 270]. However, the results in [10, 11, 19, 21, 24 and 25], show that inflation (both anticipated and unanticipated) has a variety of effects on the real earnings of firms depending on the net monetary position of the firms (see [26, p. 251–252]) and on the prevailing tax system. Most empirical results do not support the hypothesis that common stock serves as an effective inflation hedge. However, the results of Kanniainen and Kurikka [19] suggest that inflation seems to be good rather than bad news for the stock market in Finland. This is also our *a priori* hypothesis.

(6) Psychological aspects. According to empirical experience we know that stock price fluctuations are parallel in different countries. We are especially interested in the price fluctuations of the Stockholm Stock Exchange for example whether the prices in the Finnish stock market follow the prices of the Swedish stock market. During the period to be examined the exchanges of Stockholm and Helsinki were isolated from each other. Only at the end of the period were there some few firms whose stocks were quoted on both exchanges. However, the economy of both these countries is very open, export quite similar and trade between them on a high level. Thus, it is possible that economic time series including stock price fluctuations are to some extent similar in these countries. According to our hypothesis Finnish stock prices follow Swedish stock prices.

2.3.2 Empirical results. The least-squares results for estimated econometric models using monthly and quarterly data, respectively, are presented in Table 3. The results show that the signs of all coefficients—both in the monthly and quarterly equations—are parallel to the hypotheses. The sign of inflation coefficient is opposed to most empirical results but parallel to recent Finnish results [19]. The coefficients of all explanatory variables are also significant at least at 5% level.

The Durbin–Watson statistic shows that positive autocorrelation is actually not a problem in the models (in the econometric model 3 the test is inconclusive).

The Durbin–Watson statistic is biased if lagged endogenous variable (e.g. y_{t-1}) appears as an explanatory variable. The bias tends to decrease if, apart from y_{t-1}, there are also exogenous explanatory variables in the model [23, p. 460–465]. However, the absence of any positive autocorrelation was verified also by Durbin's method, which allows the lagged endogenous variable as an explanatory variable [7, p. 410–421].

The models in Table 3 include explanatory variables from all the groups—excluding the supply of money—presented in the preceding section. The supply of money and inflation actually proved to be the alternative explanatory variables. However, the statistical features of the models were better when inflation was included.

The history of the predicted variable was clearly the most important explanatory variable especially in monthly but also in quarterly models.

The price fluctuations of the Stockholm Stock Exchange seem to be a leading indicator to the prices of the Helsinki Stock Exchange. Alternatively, we also experimented with Dow–Jones and Standard & Poor indices but they had no statistically significant effect on the development of the Unitas stock market index.

The surrogate of the aggregated future cash-flow (the anticipated order stock in Finnish industry) was also a necessary variable in the model. So, in the Finnish stock market anticipated changes in future cash-flow also affect the development of stock market prices.

The lag-structure between the endogenous variable and the inflation variable is quite long; 13 months when monthly data was used and five

Table 3. OLS regression summary for monthly and quarterly econometric models

	Parameter estimates and their t-values (in parentheses)								
Monthly data	Constant	y_{t-1}	$x_{1,t-1}$	$x_{2,t-1}$	$x_{3,t-13}$	$x_{4,t-13}$	$x_{5,t-4}$	$100R^2$	D-W	RMS
Econ. model 1	−0.2371 (−2.372)	0.8425 (37.741)	0.1001 (4.936)	−0.0012 (−3.115)	0.0605 (2.742)	−0.0102 (−2.240)	—	99,70	1.830	0.000555
Econ. model 2	−0.3348 (−4.026)	0.8672 (42.160)	0.0798 (4.799)	−0.0011 (−2.514)	0.0839 (4.168)	—	−0.0128 (−2.177)	99.70	1.817	0.000556
Quarterly data	Constant	Y_{t-1}	$X_{1,t-1}$	$X_{2,t-1}$	$X_{3,t-5}$	$X_{4,t-5}$	$X_{5,t-3}$	$100R^2$	D-W	RMS
Econ. model 3	−0.7563 (−2.597)	0.5529 (7.783)	0.2858 (4.554)	−0.0023 (−2.226)	0.1776 (2.812)	−0.0286 (−2.242)	—	99.31	1.526	0.001495
Econ. model 4	−1.0652 (−4.433)	0.5929 (9.101)	0.2464 (4.509)	−0.0020 (−1.859)	0.2560 (4.355)	—	−0.0393 (−2.087)	99.30	1.664	0.001525

Legend:
y_t; Y_t: natural logarithm of the Unitas stock market index (Helsinki Stock Exchange).
$x_{1,t}$; $X_{1,t}$: natural logarithm of the Veckans Affär general index (Stockholm Stock Exchange).
$x_{2,t}$; $X_{2,t}$: anticipated order stock in Finnish Industry (next period as compared to now; the proportion of "decreases" answers).
$x_{3,t}$; $X_{3,t}$: natural logarithm of the wholesale price index.
$x_{4,t}$; $X_{4,t}$: average interest rate of bank deposits.
$x_{5,t}$; $X_{5,t}$: average return of state bonds.
$100R^2$: coefficient of determination (in per cent).
D-W: Durbin–Watson statistics.
RMS: residual mean square.

quarters when quarterly data was used. However, it was very interesting to find out that when we excluded inflation or some other explanatory variables or supposed other lag-schemes we usually had a serious positive autocorrelation in the models.

Interest rates of bank deposits and the return of the state bonds were alternative explanatory variables. We experimented with the average interest rate of all banks and commercial banks without concrete difference. The lag-structure between the endogenous variable and interest rate variable was very clear-cut and the results show that time deposits are a noticeable alternative to the common stocks. The lag between the endogenous variable and the return of the state bonds was four months. Eight months lag gives about as good results as four months lag that is parallel to the results using quarterly data.

Finally, it seems appropriate to conclude that the results of the econometric models support the results presented in the preceding section. We can predict both by using univariate time-series analysis and econometric models the monthly and quarterly stock market prices in Helsinki Stock Exchange. According to the empirical results of this study the stock market in Finland is neither efficient in the semi-strong nor in the weak form when transaction costs are excluded.

2.4 Procedure for obtaining improved composite forecasts

When two sets of one-step forecasts are available, it is well known that a linear combination of the two forecasts may outperform both of them (e.g. [14], for an application in financial analysis see [15]). In this section we develop a procedure for generating a composite forecast from an ARIMA-forecast and an econometric forecast. By doing so we expect to obtain a further improved forecasting procedure, where all the explanatory components included in ARIMA and econometric models have been incorporated. And as one outcome of this combining procedure, we had no need to try to include either exogenous variables in ARIMA models (via transfer function modelling) or any past history of the Unitas index in econometric models (e.g. via distributed lags).

Consider the case where for y_t (the natural logarithm of the Unitas index) we have two unbiased one-step forecasts f_t^{BJ} (from an

Table 4. Estimated composite models for monthly data

Model	Weights for components and their t-values: ARIMA (0, 2, 1)	ARIMA (0, 2, 1)(0, 1, 1)^{12}	Econometric model 1	Econometric model 2	Residual mean square
Component models					
ARIMA (0, 2, 1)	1.000	—	—	—	0.000805
ARIMA (0, 2, 1)(0, 1, 1)^{12}	—	1.000	—	—	0.000732
Econ. model 1	—	—	1.000	—	0.000555
Econ. model 2	—	—	—	1.000	0.000556
Composite models					
Model I	0.107 (0.878)	—	0.893 (7.340)	—	0.000520
Model II	0.124 (1.042)	—	—	0.876 (7.344)	0.000520
Model III	—	0.270 (2.509)	0.729 (6.744)	—	0.000496
Model IV	—	0.283 (2.679)	—	0.716 (6.763)	0.000495

ARIMA-model) and f_t^E (from an econometric model). The (linear) composite model is now of the form:

$$y_t = a + b^{BJ} f_t^{BJ} + b^E f_t^E + e_t, \tag{3.8}$$

where a is a constant, b^{BJ} and b^E are weights of the two forecast series and e_t is a random error term. Using ordinary least squares in estimation, there still remain three methods to fit the model, [14, p. 199–201]: (i) no constant term, weights restricted to sum to unity, (ii) no constant term, unconstrained weights, and (iii) unconstrained weights with constant term. In their paper Granger and Ramanathan [14, p. 201] conclude that method (iii) is theoretically the best and the common practice of obtaining composite forecasts as weighted averages (method (i)) should be abandoned in favour of an unrestricted linear combination including a constant term (method (iii)). This recommendation has been recently applied, e.g. in finance by Guerard and Beidleman [15].

We started the estimation of our composite model (the estimation period consisted of the years 1975–84 as in the case of individual models) with method (iii). In all cases the value of the constant term became, however, near zero and was far from being statistically significant. Therefore, in the final estimation method (ii) was applied (unconstrained weights without the constant term). The results of the estimation are presented in Table 4 (monthly data) and Table 5 (quarterly data).

From the results we can see that in all cases combining an ARIMA-model with an econometric model reduces the residual as compared with those of the component models. This reduction is largest in the two quarterly models, whereas combining the ARIMA (0, 2, 1)-model with any of the econometric models (composite monthly models I and II) reduces the RMS's of the econometric models only slightly. This is also revealed in that the t-values of the ARIMA-component weights are not significant. We see that the main contribution with which an ARIMA-component can improve the econometric models is in the seasonal pattern included.

Further, it is interesting to note that in all models the weights add nearly to one, although it was not explicitely required. This, together with the non-significant constant term obtained by method (iii), indicates that, for our models and data, the three different methods produce

Table 5. Estimated composite models for quarterly data

Model	Weights for components and their t-values: ARIMA (0, 2, 0)(0, 1, 1)^4	Econometric model 3	Econometric model 4	Residual mean square
Component models				
ARIMA (0, 2, 0)(0, 1, 1)^4	1.000	—	—	0.001859
Econ. model 3	—	1.000	—	0.001495
Econ. model 4	—	—	1.000	0.001525
Composite models				
Model V	0.365 (3.602)	0.635 (6.261)	—	0.000956
Model VI	0.376 (3.563)	—	0.623 (5.899)	0.000897

Table 6. Forecasting accuracy summary (monthly data)

Accuracy measure	ARIMA (0, 2, 1)	ARIMA $(0, 2, 1)(0, 1, 1)^{12}$	Econometric model 1	Econometric model 2	Composite model III	Composite model IV
ME	0.01152	0.00840	0.00655	−0.00795	0.00885	−0.00151
MAE	0.02366	0.02576	0.02996	0.02808	0.02674	0.02461
MSE	0.00120	0.00157	0.00141	0.00126	0.00132	0.00112
RMSE	0.03462	0.03964	0.03760	0.03547	0.03629	0.03352
MPE (%)	0.211	0.155	0.113	−0.151	0.157	−0.031
MAPE (%)	0.434	0.472	0.548	0.516	0.489	0.451

almost identical procedures for combining univariate time series forecasts and econometric forecasts.

3. FORECAST RESULTS

The object of this section is to compute, using each of the derived models, forecasts for the Unitas stock market index from January 1985 to March 1986 (monthly data) or from the first quarter of 1985 to the first quarter of 1986 (quarterly data), and to compare forecast accuracy of the models.

The development of the Unitas index was quite interesting during that period. The decrease in the share prices which began at the end of the estimation period, in spring/summer 1984, continued until summer 1985 after which the prices began to rise again. The increase in the value of the index was very rapid, especially at the end of the forecasting period.

The forecasts were computed as one-step forecasts for the natural logarithm of the Unitas index. The monthly forecasts were computed using the two univariate time series models, non-seasonal ARIMA (0, 2, 1)-model and seasonal ARIMA $(0, 2, 1)(0, 1, 1)^{12}$-model (with parameters given in Table 1), the two multivariate econometric models (models 1 and 2 in Table 3), which differ in the 'last' explanatory variable, and two composite models (models III and IV in Table 4; composite models I and II were not used in forecasting, because their ARIMA-component showed no statistical significance in the analysis). The quarterly forecasts were computed with all the derived quarterly models: ARIMA $(0, 2, 0)(0, 1, 1)^4$ time-series model, multivariate econometric models 3 and 4 given in Table 3, composite models V and VI given in Table 5.

For evaluating the forecasting ability of the presented models, a variety of more common accuracy measures will be used. That is because different measures give weight to different aspects in the error series. The following measures were computed (for definition of the accuracy measures see e.g. [22, p. 43–54] and [12]): mean error (ME), mean absolute error (MAE), mean of squared errors (MSE, analogical to residual mean square, RMS, used in diagnostic checking of the models), root mean of squared errors (RMSE), mean percentage error (MPE) and mean absolute percentage error (MAPE).

The results of forecasting procedure for monthly and quarterly data are summarized in Tables 6 and 7, respectively.

When we look at the statistics associated with the monthly forecasts, we see that all the models perform quite well. None of the mean error statistics are statistically significant, all the forecasts can be regarded as unbiased. The error variance measures have not increased alarmingly, either. For most of the models, the increase in the forecast error variance, as compared with the error variance during the fitting phase, is not statistically significant at 1% level. For econometric model 1 and composite

Table 7. Forecasting accuracy summary (quarterly data)

Accuracy measure	ARIMA $(0, 2, 0)(0, 1, 1)^4$	Econometric model 3	Econometric model 4	Composite model V	Composite model VI
ME	0.02230	0.01785	−0.01328	0.00299	−0.01713
MAE	0.08516	0.07041	0.06486	0.06519	0.06075
MSE	0.00908	0.00676	0.00494	0.00550	0.00477
RMSE	0.09528	0.08220	0.07028	0.07415	0.06904
MPE (%)	0.417	0.305	−0.261	0.038	−0.327
MAPE (%)	1.552	1.280	1.188	1.188	1.113

model III, which includes econometric model I as one component, the F-statistic slightly exceeds the 1% value. However, when we compare the numerical values of the accuracy measures of different models we see that composite model IV, which has the seasonal ARIMA $(0, 2, 1)(0, 1, 1)^{12}$-model and multivariate econometric model 2 (including the return of the state bonds as one of the explanatory variables) as its components, has the lowest error statistics: the forecast is actually unbiased ($ME = -0.00151$, $MPE = -0.031\%$), the error has low variation ($MSE = 0.00112$, $RMSE = 0.03352$) and is low in absolute values ($MAE = 0.02461$, $MAPE = 0.451\%$). We can further note that the non-seasonal ARIMA (0, 2, 1)-model, which was the least satisfactory model in the fitting phase, also has quite a low error statistic. It has, however, slightly more biased forecasts than the other models. The other four models are almost equal to each other in accuracy.

As expected, forecasting accuracy of the quarterly models is not as high as that of monthly models. In particular, the forecast error variances, except for econometric model 4, are greater (at 1% significance level) than the variance in the estimation period. This is natural since useful information is lost due to aggregation in the modelling. The one-step forecasts are also computed further ahead, for one quarter or three months ahead. The results can be regarded, however, as acceptable. Both econometric models seem to outperform the univariate time series model. However, combining this time-series forecast with an econometrically produced forecast (composite forecasts V and VI) still reduces the error statistics. Analogically to monthly modelling, composite model VI (seasonal ARIMA-model and the econometric model including the return of state bonds as its components) possesses the lowest error statistics in general.

4. SUMMARY

The purposes of this study were, first, to analyze to what extent, and in which form the monthly and quarterly stock market prices are predictable on a thin security market like the Helsinki Stock Exchange. Second, to compare the forecasting results based on univariate time-series analysis and multivariate econometric models with each other. Third, to develop composite stock price forecasting models and examine the forecasting improvement of these models relative to the time-series and econometric models.

At the beginning of the paper we presented market efficiency as a crucial concept when the predictability of stock market prices is analyzed. Under the weak form of efficiency, information of past prices is of no value for predicting future stock prices. As such the weak form of efficiency is directly opposed to the basic premises of univariate time-series analysis to forecast stock prices. Analogously, the semi-strong form of efficient market hypothesis is diametrically opposed to the basic premises of multivariate econometric analysis to forecast stock prices. Many anomalies and deviations from market efficiency even in the weak form serve a possibility to forecast Finnish stock market prices by using both univariate time-series analysis and multivariate econometric analysis.

The empirical reults showed that the Unitas general index did not follow a random walk model. This can be interpreted as a violation against the existence of equity market efficiency in the weak sense. If the market is not efficient in the weak sense it can not be efficient in the semi-strong sense either. The empirical results of multivariate econometric analysis were parallel to this deduction. We found both theoretically and statistically satisfactory econometric models to predict stock prices. After that we developed a procedure to generate composite forecasts from univariate time-series models and econometric models.

Finally, as expected, forecasting accuracy of monthly models was better than that of quarterly models. Econometric models provide slightly more accurate forecasts than the univariate time-series models. However, the composite models substantially reduced the mean square forecasting error compared to the results of univariate or econometric models. So, the empirical results strongly support the use of composite models to predict Finnish stock prices.

ACKNOWLEDGEMENT

The authors wish to thank an anonymous referee and Editor for constructive and helpful comments in preparing the final version of the manuscript. Financial support from the Research Foundation of OKO-banks, Marcus Wallenberg Foundation and Paulo Foundation is gratefully acknowledged.

REFERENCES

1. Berglund T (1986) Anomalies in stock returns on a thin security market. Publications of the Swedish School of Economics and Business Administration No. 37.
2. Berglund T, Wahlroos B and Grandell L (1983) The KOP and the UNITAS indexes for the Helsinki Stock Exchange in the light of a new value weighted index. *Finnish J. Bus. Econ.* **32**(1), 30–41 (in Swedish).
3. Bicksler JL (1972) A cross-spectral analysis of the lead-lag structure of money supply-stock prices. In *Studies in Business, Technology, and Economics* (Edited by Bicklser JL). Heath, Lexington, MA.
4. Box GEP and Jenkins GM (1970) *Time-Series Analysis, Forecasting and Control.* Holden–Day, San Francisco, CA.
5. Dixon WJ (1983) (Ed.) *BMDP Statistical Software*, 1983 printing with addition. University of California Press.
6. Driehuis W (1972) *Fluctuations and Growth in a Near Full Employment Economy.* Rotterdam University Press.
7. Durbin J (1970) Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica* **38**(2), 410–421.
8. Dyckman TR, Downes DH and Magee RP (1975) *Efficient Capital Markets and Accounting: A Critical Analysis.* Prentice–Hall, Englewood Cliffs, NJ.
9. Fama EF (1970) Efficient capital markets: a review of theory and empirical work. *J. Fin.* **25**(3), 383–417.
10. Fama EF and Schwert GW (1977) Asset returns and inflation. *J. Fin. Econ.* **5**(2), 115–146.
11. Feldstein M (1980) Inflation, tax rules and the stock market. *J. Monet. Econ.* **6**(3), 309–331.
12. Flores BE (1986) A pragmatic view of accuracy measurement in forecasting. *Omega* **14**(2), 93–98.
13. Granger CWJ and Morgenstern O (1970) *Predictability of Stock Market Prices.* Heath, Lexington, MA.
14. Granger CWJ and Ramanathan R (1984) Improved methods of combining forecasts. *J. Forecasting* **3**(2), 197–204.
15. Guerard JB Jr and Beidleman CR (1986) A new look at forecasting annual corporate earnings in the U.S.A. *European J. Opl Res.* **23**(3), 288–293.
16. Hagin RL (1979) *The Dow Jones–Irwin Guide to Modern Portfolio Theory.* Dow Jones–Irwin, New York.
17. Hansmann KW and Zetsche W (1985) Forecasting for portfolio selection with a microcomputer. The Institute of Industrial Research of the University of Armed Forces Hamburg, Discussion Papers No. 7.
18. Hawawani GA and Michel PA eds. (1984) *European Equity Markets; Risk, Return and Efficiency.* Garland Publishing.
19. Kanniainen V and Kurikka V (1984) On the effects of inflation in the stock market: empirical evidence with Finnish data 1968–1981. *J. Bus. Fin. Acctng* **11**(2), 139–150.
20. Korhonen A (1977) Stock prices, information and the efficiency of the Finnish stock market, empirical tests. *Acta Academiae Oeconomicae Helsigniensis*, Series A: 23.
21. Lintner J (1975) Inflation and security returns. *J. Fin.* **30**(2), 259–280.
22. Makridakis S, Wheelwright SC and McGee V (1983) *Forecasting: Methods and Applications*, Second edn. John Wiley, New York.
23. Malinvaud E (1966) *Statistical Methods in Econometrics.* North Holland, Amsterdam.
24. Modigliani F and Cohn RA (1979) Inflation, rational valuation and the market. *Fin. Anal. J.* (2), 24–44.
25. Pearce DK and Roley VV (1985) Stock prices and economic news. *J. Bus.* **58**(1), 49–67.
26. Sharpe WF (1985) *Investments.* Third edn. Prentice–Hall, Englewood Cliffs, NJ.
27. Teräsvirta T (1984) Forecasting the output of the Finnish metal industry using business survey data. The Research Institute of the Finnish Economy. Discussion Papers No. 157.
28. Umstead DA (1977) Forecasting stock market prices. *J. Fin.* **32**(2), 427–441.
29. Watts RL and Leftwich RW (1977) The time series of annual accounting earning. *J. Acctng Res.* **15**(3), 253–271.
30. Working H (1960) Note on the correlation of first differences of averages in a random chain. *Econometrica* **28**(4), 916–918.

ADDRESS FOR CORRESPONDENCE: *Professor Ilkka Virtanen or Associate Professor Paavo Yli-Olli, School of Business Studies, University of Vaasa, P.O. Box 297, SF-65101 Vaasa, Finland.*