Bootstrap Tests of Cointegration Rank with Financial Time Series

Niklas Ahlgren and Jan Antell
Swedish School of Economics, Department of Finance and Statistics
PO Box 479 (Arkadiagatan 22)
00101 Helsingfors, Finland
E-mail: niklas.ahlgren@shh.fi and jan.antell@shh.fi

November 15, 2005

Abstract

The likelihood ratio test of cointegration rank is the most widely used test for cointegration. Many studies have shown by simulation that the small sample distribution is not well approximated by the limiting distribution. We suggest using the bootstrap to generate small sample critical values instead of correcting the test statistics. The idea of bootstrapping the trace test of cointegration rank is of course not new, but it has mainly been studied in the simple context of first order vector autoregressive models with cointegrating rank equal to zero. For higher order VAR processes with cointegrating rank greater than zero little is known about the size and power properties of different methods. The main contribution of this article is to investigate the behaviour of different bootstrap approaches under the null of cointegration rank zero and fixed alternatives of cointegration rank greater than zero, and to compare their relative performance with small sample corrections based on correcting the test statistics. In particular, it is shown that in most cases the bootstrap is superior to small sample corrections. Two applications to financial time series are considered. The first application is to US interest rates and the second application is to international stock prices.

Key words: Bootstrap, Small sample correction, Cointegration Rank, Trace test.

1 INTRODUCTION

The likelihood ratio test of cointegration rank (Johansen 1996) is the most widely used test for cointegration. Its popularity stems from the fact that it is conceptually simple since it is an LR test, test statistics are easy to compute by reduced rank regression and limiting distributions are free from nuisance parameters. Many studies have shown by simulation that the small sample distribution is not well approximated by the limiting distribution (see, e.g., Johansen 2002, and the references cited therein). Ahn and Reinsel (1990) suggested to use a degrees of freedom small sample correction. Johansen (2002) proposed a correction factor which uses the idea of the Bartlett correction. An alternative approach is to use the bootstrap to simulate small sample critical values.

Bootstrapping likelihood ratio tests of cointegration rank has mainly been studied in the simple context of first order vector autoregressive models with cointegration rank equal to zero or very special models (see van Giersbergen 1996, Harris and Judge 1998, and Mantalos and Shukur 2001). For high dimensional and high order systems, little is known about the size and power properties of bootstrap likelihood ratio tests of cointegration rank. Results in Johansen (2002) show that in these cases the limiting distribution is a poor approximation to the finite sample distribution, and the correction factor manages to correct some of the size distortion. It is therefore important to enquire into whether the bootstrap gives an improved approximation.

Different bootstrap alternatives for testing the null hypothesis of cointegration rank have been proposed in the literature. One approach commonly followed is to resample residuals obtained by imposing the null hypothesis. This approach has been followed by van Giersbergen (1996), Harris and Judge (1998), and Mantalos and Shukur (2001). A different approach to bootstrap likelihood ratio tests of cointegration rank is to resample unrestricted residuals obtained without imposing the null hypothesis. The approach has been suggested and investigated by Swensen (2004).

The main contribution of this article is to study the behaviour of different bootstrap approaches under the null of cointegration rank zero and fixed alternatives of cointegration rank greater than zero, and to compare their relative performance with small sample corrections. We investigate the two main approaches for bootstrapping the likelihood ratio test of cointegration rank. The first bootstrap scheme is based on restricted residuals obtained by imposing the null hypothesis and the second bootstrap scheme is based on unrestricted residuals. Davidson and MacKinnon (2001) develop a fast double bootstrap (FDB) based on the double bootstrap proposed by Beran (1988). The FDB requires only about twice as much computation as the ordinary single bootstrap. The only existing results for the performance of the FDB in a time series context appear to be those of Omtzigt and Fachin (2002) who are concerned with testing restrictions on the cointegration vectors. In the article it is shown by simulation that in most cases bootstrap tests are superior to small sample corrections in terms of size of tests. The FDB produces a further improvement in cases where the performance of the asymptotic test depends on the parameter values and which are very unfavourable for the finite sample performance of the likelihood ratio test of cointegration rank. The simulation results show that bootstrap methods based on unrestricted residuals are more powerful than bootstrap methods based on restricted residuals (i.e., imposing the null hypothesis), although both methods are asymptotically consistent.

The likelihood ratio test for cointegration rank has been widely used with financial time series. A further aim of this paper is therefore to study the performance of bootstrap tests applied to financial time series. Ruiz and Pascual (2002) review the application of bootstrap procedures for financial time series, but they only consider univariate time series and unit root tests. To this end we consider two applications to US interest rates and international stock prices.

The article is organised as follows. Section 2 briefly discusses small sample corrections and bootstrap procedures for the likelihood ratio test of cointegration rank. Some simulation results on the relative performance of small sample corrections and bootstrap tests are presented in Section 3. Section 4 contains two empirical applications to financial time series. Section 5 summaries our findings and contains our conclusions.

2 SMALL SAMPLE CORRECTIONS AND BOOTSTRAP PROCEDURES

We consider the n-dimensional vector autoregressive (VAR) model in error correction form

$$\Delta \mathbf{X}_t = \mathbf{\Pi} \mathbf{X}_{t-1} + \sum_{i=1}^{k-1} \mathbf{\Gamma}_i \Delta \mathbf{X}_{t-i} + \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T, \tag{1}$$

where the errors $\boldsymbol{\varepsilon}_t$ are $\mathrm{IID}(\mathbf{0}, \mathbf{\Omega})$.

The null hypothesis of cointegration is

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}',$$

and the hypothesis about the deterministic terms is

$$\boldsymbol{\mu}_1 = \boldsymbol{\alpha}\boldsymbol{\rho}',$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times r$ matrices of rank r, and $\boldsymbol{\rho}$ is $1 \times r$. The model (1) corresponds to the model with the trend t being restricted to the cointegration space. This is the most general model recommended by Doornik, Hendry and Nielsen (1998) for determining the cointegration rank. We note that other models for the deterministic terms are also possible (see Johansen 1996, Section 5.7).

The likelihood ratio statistic for the hypotheses

$$H_0 : \text{rank}(\mathbf{\Pi}) = r \quad \text{against} \quad H_1 : \text{rank}(\mathbf{\Pi}) > r, \quad r \leq n-1$$

is

$$Q = -T \sum_{i=r+1}^{n} \ln(1 - \widehat{\lambda}_i),$$

where the eigenvalues $1 > \widehat{\lambda}_1 > \cdots > \widehat{\lambda}_n > 0$, $\widehat{\lambda}_{n+1} = 0$ solve

$$|\lambda \mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}| = 0,$$

and

$$\mathbf{S}_{ij} = T^{-1} \sum_{t=1}^{T} \mathbf{R}_{it}\mathbf{R}'_{jt}, \quad i, j = 0, 1$$

are the product moment matrices of the residuals $\mathbf{R}_{0t}$ and $\mathbf{R}_{1t}$ from regressing $\Delta\mathbf{X}_t$ and $(\mathbf{X}'_{t-1}, t)'$ on the lagged differences $\Delta\mathbf{X}_{t-1}, \ldots, \Delta\mathbf{X}_{t-k+1}$ and the constant (see Johansen 1996, for details).

In finite samples the distribution of the likelihood ratio statistic depends on T and $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ denotes the parameters under the null hypothesis. For $T \to \infty$ the dependence on $\boldsymbol{\theta}$ disappears, but not uniformly in $\boldsymbol{\theta}$ (see Nielsen 1997 for a discussion). The limiting distribution was derived by Johansen (1996), and can be expressed in terms of a Brownian motion, which depends on $n - r$ and the model for the deterministic terms.

2.1 Small Sample Corrections

Ahn and Reinsel (1990) and Reimers (1992) suggested to use the small sample correction $(T - kn)/T$. The idea is simply to adjust the test statistics for the number of estimated parameters.

The correction factor in Johansen (2002) uses the idea of the Bartlett correction. The idea of the Bartlett correction is to find the expectation of the likelihood ratio statistic and correct it to have the same mean as the limiting distribution. Let $\boldsymbol{\theta}$ denote the parameters of the model (1) under the assumption that $\mathbf{\Pi} = \boldsymbol{\alpha\beta}'$ and $\boldsymbol{\mu}_1 = \boldsymbol{\alpha\rho}'$. Johansen (2002) derived an approximation to $\mathrm{E}_{\boldsymbol{\theta}}(-2\log LR\{\mathbf{\Pi} = \boldsymbol{\alpha\beta}', \, \boldsymbol{\mu}_1 = \boldsymbol{\alpha\rho}'\})$. In Johansen (2002) it is suggested to use the correction factor

$$a(T, n-r, n_d)(1 + T^{-1}b(\widehat{\boldsymbol{\theta}})),$$

and the corrected statistic

$$\frac{-2\log LR}{a(T, n-r, n_d)(1 + T^{-1}b(\widehat{\boldsymbol{\theta}}))} = \frac{Q}{a(T, n-r, n_d)(1 + T^{-1}b(\widehat{\boldsymbol{\theta}}))}. \qquad (2)$$

The function $a(T, n-r, n_d)$ depends on T, $n-r$ and the model for the deterministic terms n_d, and the function $b(\widehat{\boldsymbol{\theta}})$ depends on the estimated parameters. The model (1) corresponds to the model with $n_d = 1$. They are both complicated functions of a random walk. Johansen (2002) tabulated $a(T, n-r, n_d)$ by simulation and derived an analytic expression for $b(\boldsymbol{\theta})$. It is worth pointing out that there is an error in the simulated functions in Johansen (2002). The implementation of the correction factor in this paper is based on the corrigendum in Johansen et al. (2005).

The main difference between the small sample correction and the correction factor is that the former takes account of the lag length or the number of estimated parameters, whereas the latter takes account of both the lag length and the dependence of the small sample distribution on the parameters. In the simulations in Section 3 we find that the small sample correction is dominated by the correction factor.

2.2 Bootstrap and Fast Double Bootstrap Tests

The first bootstrap scheme is based on the estimated parameters and residuals of the restricted model, i.e., on the model (1) with the null hypothesis $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ and $\boldsymbol{\mu}_1 = \boldsymbol{\alpha}\boldsymbol{\rho}'$ imposed. The bootstrap test can be described by the following three steps (see, e.g., Swensen 2004):

1. Estimate the model (1) by reduced rank regression to obtain the estimates $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\rho}}$. For the remaining parameters $\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{k-1}$ and $\boldsymbol{\mu}_0$ use the unrestricted estimates $\widehat{\boldsymbol{\Gamma}}_1, \ldots, \widehat{\boldsymbol{\Gamma}}_{k-1}$ and $\widehat{\boldsymbol{\mu}}_0$. Compute the restricted residuals

$$\widehat{\boldsymbol{\varepsilon}}_t = \Delta\mathbf{X}_t - \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'\mathbf{X}_{t-1} - \sum_{i=1}^{k-1} \widehat{\boldsymbol{\Gamma}}_i \Delta\mathbf{X}_{t-i} - \widehat{\boldsymbol{\mu}}_0 - \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\rho}}' t, \quad t = k+1, \ldots, T.$$

2. Generate $\mathbf{X}_1^*, \ldots, \mathbf{X}_T^*$, where $\mathbf{X}_t^* = \mathbf{X}_t$, $t = 1, \ldots, k$, as

$$\Delta\mathbf{X}_t^* = \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'\mathbf{X}_{t-1}^* + \sum_{i=1}^{k-1} \widehat{\boldsymbol{\Gamma}}_i \Delta\mathbf{X}_{t-i}^* + \widehat{\boldsymbol{\mu}}_0 + \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\rho}}' t + \boldsymbol{\varepsilon}_t^*.$$

   Here $\boldsymbol{\varepsilon}_t^*$, $t = k+1, \ldots, T$, is an IID sequence with $\boldsymbol{\varepsilon}_t^* \sim \widehat{F}_\varepsilon$, where $\widehat{F}_\varepsilon$ is the empirical distribution of the rescaled $\widehat{\boldsymbol{\varepsilon}}_t$'s,

$$\left(\frac{T}{T-kn}\right)^{1/2} \widehat{\boldsymbol{\varepsilon}}_t.$$

3. Let Q^* be the LR statistic computed using the data $\mathbf{X}_1^*, \ldots, \mathbf{X}_T^*$. Use the empirical distribution of Q^* to approximate the distribution of Q under the null hypothesis.

The fast double bootstrap (FDB) adds a fourth step:

4. For each $\mathbf{X}_1^*, \ldots, \mathbf{X}_T^*$ a second level bootstrap $\mathbf{X}_1^{**}, \ldots, \mathbf{X}_T^{**}$ is generated and a second level bootstrap LR statistic Q^{**} is computed using the data $\mathbf{X}_1^{**}, \ldots, \mathbf{X}_T^{**}$. Compute the quantile $1-\alpha^*$ of Q^{**}, which solves
$$Q^{**}(1-\alpha^*) \geq Q^*(1-\alpha). \tag{3}$$
For a finite number of bootstrap replications there will not be a value of $1-\alpha^*$ that satisfies (3) with equality, and we choose the smallest value of $1-\alpha^*$ which satisfies (3) with inequality. The quantile $Q^*(1-\alpha^*)$ is used to approximate the $1-\alpha$ quantile of Q under the null hypothesis.

The FDB (Davidson and MacKinnon 2001) can be thought of as a computationally inexpensive approximation to the double bootstrap of Beran (1988). The intuition behind the FDB procedure is the following. Suppose, for concreteness, that the Q^{**} tend to be smaller than the Q^*. This suggests that the Q^* will tend to be smaller than they would be if they were drawn from the true unknown DGP defined by the parameters $\boldsymbol{\theta}$ rather than from the bootstrap DGP defined by the estimated parameters $\widehat{\boldsymbol{\theta}}$. Therefore, the ordinary bootstrap quantile $Q^*(1-\alpha)$ will be too small, and the bootstrap test will overreject. In this situation, $Q^*(1-\alpha^*)$ will be larger than $Q^*(1-\alpha)$, and the rejection probability of the FDB test will be smaller than the rejection probability of the ordinary bootstrap test. The FDB can thus be though of as delivering a bias correction to the ordinary bootstrap test.

The second bootstrap scheme is based on the estimate $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\rho}}$ under the null hypothesis $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ and $\boldsymbol{\mu}_1 = \boldsymbol{\alpha}\boldsymbol{\rho}'$, and residuals of the unrestricted model, i.e., on the model (1) with $\mathbf{\Pi}$ and $\boldsymbol{\mu}_1$ unrestricted. The bootstrap test consists of the following three steps:

1. Estimate the model (1) by reduced rank regression to obtain the estimates $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\rho}}$. For the remaining parameters $\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_{k-1}$ and $\boldsymbol{\mu}_0$ use the unrestricted estimates $\widehat{\mathbf{\Gamma}}_1, \ldots, \widehat{\mathbf{\Gamma}}_{k-1}$ and $\widehat{\boldsymbol{\mu}}_0$. Compute the unrestricted residuals
$$\widetilde{\boldsymbol{\varepsilon}}_t = \Delta\mathbf{X}_t - \widehat{\mathbf{\Pi}}\mathbf{X}_{t-1} - \sum_{i=1}^{k-1} \widehat{\mathbf{\Gamma}}_i \Delta\mathbf{X}_{t-i} - \widehat{\boldsymbol{\mu}}_0 - \widehat{\boldsymbol{\mu}}_1 t, \quad t = k+1, \ldots, T.$$

2. Generate $\mathbf{X}_1^+, \ldots, \mathbf{X}_T^+$, where $\mathbf{X}_t^+ = \mathbf{X}_t$, $t = 1, \ldots, k$, as
$$\Delta\mathbf{X}_t^+ = \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'\mathbf{X}_{t-1}^+ + \sum_{i=1}^{k-1} \widehat{\mathbf{\Gamma}}_i \Delta\mathbf{X}_{t-i}^+ + \widehat{\boldsymbol{\mu}}_0 + \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\rho}}' t + \boldsymbol{\varepsilon}_t^+.$$
Here $\boldsymbol{\varepsilon}_t^+$, $t = k+1, \ldots, T$, is an IID sequence with $\boldsymbol{\varepsilon}_t^+ \sim \widetilde{F}_\varepsilon$, where $\widetilde{F}_\varepsilon$ is the empirical distribution of the rescaled $\widetilde{\boldsymbol{\varepsilon}}_t$'s,
$$\left(\frac{T}{T-kn}\right)^{1/2} \widetilde{\boldsymbol{\varepsilon}}_t.$$

3. Let Q^+ be the LR statistic computed using the data $\mathbf{X}_1^+, \ldots, \mathbf{X}_T^+$. Use the empirical distribution of Q^+ to approximate the distribution of Q under the null hypothesis.

The fast double bootstrap (FDB) adds a fourth step:

4. For each $\mathbf{X}_1^+, \ldots, \mathbf{X}_T^+$ a second level bootstrap $\mathbf{X}_1^{++}, \ldots, \mathbf{X}_T^{++}$ is generated and a second level bootstrap LR statistic Q^{++} is computed using the data $\mathbf{X}_1^{++}, \ldots, \mathbf{X}_T^{++}$. Compute the quantile $1-\alpha^+$ of Q^{++}, which solves

$$Q^{++}(1-\alpha^+) \geq Q^+(1-\alpha). \tag{4}$$

   For a finite number of bootstrap replications there will not be a value of $1-\alpha^+$ that satisfies (4) with equality, and we choose the smallest value of $1-\alpha^+$ which satisfies (4) with inequality. The quantile $Q^+(1-\alpha^+)$ is used to approximate the $1-\alpha$ quantile of Q under the null hypothesis.

Swensen (2004) suggested to check in step 2 that the roots of the characteristic polynomial $|\widehat{\mathbf{\Pi}}(z)| = 0$, where $\widehat{\mathbf{\Pi}}(z) = (1-z)\mathbf{I}_n - \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'z - \widehat{\mathbf{\Gamma}}_1(1-z)z - \cdots - \widehat{\mathbf{\Gamma}}_{k-1}(1-z)z^{k-1}$, are equal to one or outside the unit circle in order to ensure that the bootstrap observations are $I(1)$, although it happens most often that the estimates of the roots have the property (Johansen 1996, p. 71). We have checked the condition in the simulations in Section 3.

There are not many theoretical results for the asymptotic behaviour of bootstrap methods for the likelihood ratio test of cointegration rank. On exception is Swensen (2004), who showed that the asymptotic distribution of the bootstrap likelihood ratio test based on 1–3 above is the same as for the likelihood ratio test for both bootstrap schemes, and that the bootstrap procedure for determining the cointegration rank is asymptotically consistent. Swensen (2004) also shows that the bootstrap tests based on restricted and unrestricted residuals are asymptotically equivalent under the null hypothesis. No results exist for the relative power and the power behaviour of the bootstrap methods under the alternative hypothesis. The results in Paparoditis and Romano (2005) suggest that the bootstrap test based on unrestricted residuals is more powerful. We would therefore expect little or no difference in size properties of the tests and the test based on unrestricted residuals to have higher power.

3 SIMULATIONS

We present simulation results from three experiments. The experiments correspond to the special cases in Johansen (2002) and are chosen to illustrate the dependence on the parameters, the lag length and T. The simulations are considerably more general than those cited in the introduction. Five versions of the likelihood ratio test are computed: the asymptotic test (denoted Q), the test using the small sample correction $(T - kn)$ (Ahn and Reinsel 1990, and Reimers 1992, denoted Q_S), the corrected test (Johansen 2002, denoted Q_C), the bootstrap test based on restricted and unrestricted residuals (denoted Q^* and Q^+) and the FDB test (Davidson and MacKinnon 2001, denoted Q^{**} and Q^{++}). Since the small sample correction is dominated by the correction factor, the results for the test using the small sample correction are not discussed, but they are included for completeness and for comparisons with previous papers in which it has been extensively used. It is worth noting that the implementation of the correction factor is based on estimated parameters, so a direct comparison with the results in Johansen (2002) who used the known parameters of the DGP cannot be made. We refer to the single bootstrap test as the bootstrap test and the fast double bootstrap test as the FDB test. The series lengths are $T = 50, 100$, and in the third experiment also $T = 200$, $T = 500$.

Monte Carlo experiments involving the bootstrap become very expensive since the number of replications is MB, where M is the number of Monte Carlo replications and B is the number of bootstrap replications. In the experiments in Sections 3.1–3.3 we use $MB = 10,000 \times 1000 = 10,000,000$ and in the empirical applications in Section 4 we use $B = 1,000,000$.

3.1 The Test for No Cointegration in the Model with Two Lags

In the first experiment we investigate the size of the test of $\mathbf{\Pi} = \mathbf{0}$ and $\boldsymbol{\mu}_1 = \mathbf{0}$ in the model (1) with $k = 2$ and $n_d = 1$:

$$\Delta \mathbf{X}_t = \mathbf{\Pi} \mathbf{X}_{t-1} + \mathbf{\Gamma}_1 \Delta \mathbf{X}_{t-1} + \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \boldsymbol{\varepsilon}_t. \tag{5}$$

The distribution of the test statistic does not depend on the parameters $\boldsymbol{\mu}_0$ and $\mathbf{\Omega}$, so in the data generation process (DGP) we can take $\boldsymbol{\mu}_0 = \mathbf{0}$ and $\mathbf{\Omega} = \mathbf{I}_n$. In the DGP we assume that $\mathbf{\Gamma}_1$ is a scalar matrix, $\mathbf{\Gamma}_1 = \xi \mathbf{I}_n$, $-1 < \xi < 1$. In the simulations $n = 5$ and the values for ξ considered are $0, 0.3, 0.5, 0.6, 0, 7, 0.9$. Table 1 reports the rejection probabilities of a nominal 5% level test. The asymptotic test overrejects severely for all values of ξ, and the correction factor manages to correct some of the size distortion of the asymptotic test. The size of the asymptotic test is very close to 1, since the rejection probability for large values of ξ tends to 1. Note that when ξ tends to 1 the process becomes $I(2)$. When $\xi \leq 0.5$ the bootstrap test works very well and the FDB test only slightly better. When $\xi \geq 0.6$ the bootstrap test overrejects. For large values of ξ the FDB test also overrejects but much less so and performs as well as any test could be expected to perform, given that when ξ is large, the process is close to being $I(2)$. For example, when $\xi = 0.9$ the rejection probability of the FDB test is only about half that of the bootstrap test. Thus the bias correction in the FDB procedure can compensate for some of the overrejection of the bootstrap test.

Table 1: Test of $\mathbf{\Pi} = \mathbf{0}$, $\boldsymbol{\mu}_1 = \mathbf{0}$ in model (5). The nominal significance level is 5%. In the simulations $n = 5$, $\mathbf{\Pi} = \mathbf{0}$, $\boldsymbol{\mu}_1 = \mathbf{0}$, $\mathbf{\Gamma}_1 = \xi \mathbf{I}_n$, $\boldsymbol{\mu}_0 = \mathbf{0}$, $\mathbf{\Omega} = \mathbf{I}_n$. The number of simulations is $10,000$ and the number of bootstrap replications is 1000. Q denotes the likelihood ratio test based on asymptotic critical values, Q_S the test using the small sample correction, Q_C the corrected test, Q^* the bootstrap test based on restricted residuals, Q^{**} the FDB test based on restricted residuals, Q^+ the bootstrap test based on unrestricted residuals and Q^{++} the FDB test based on unrestricted residuals.

ξ	0	0.3	0.5	0.6	0.7	0.9
			$T = 50$			
Q	0.405	0.595	0.790	0.887	0.954	0.998
Q_S	0.034	0.085	0.190	0.300	0.466	0.855
Q_C	0.094	0.102	0.095	0.086	0.071	0.000
Q^*	0.055	0.061	0.075	0.089	0.108	0.193
Q^{**}	0.054	0.054	0.058	0.061	0.066	0.102
Q^+	0.055	0.061	0.075	0.085	0.108	
Q^{++}	0.055	0.055	0.057	0.058	0.067	
			$T = 100$			
Q	0.183	0.262	0.378	0.477	0.628	0.975
Q_S	0.047	0.080	0.143	0.207	0.327	0.861
Q_C	0.074	0.074	0.075	0.076	0.071	0.026
Q^*	0.052	0.054	0.059	0.066	0.076	0.154
Q^{**}	0.052	0.052	0.053	0.056	0.059	0.082
Q^+	0.054	0.056	0.060	0.065	0.074	0.152
Q^{++}	0.053	0.054	0.054	0.056	0.057	0.082

There were no simulations with a root of the characteristic polynomial outside the unit circle

3.2 The Test for Rank One in the Model with One Lag

In the second experiment we consider the tests of $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, $\boldsymbol{\mu}_1 = \boldsymbol{\alpha}\boldsymbol{\rho}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times 1$ vectors, in the model (1) with $k = 1$ and $n_d = 1$:

$$\Delta \mathbf{X}_t = \mathbf{\Pi} \mathbf{X}_{t-1} + \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \boldsymbol{\varepsilon}_t. \tag{6}$$

The DGP in the simulations has $n = 5$, $\boldsymbol{\alpha} = (a_1, a_2, 0, 0, 0)'$, $\boldsymbol{\beta} = (1, 0, 0, 0, 0)'$, so that $\boldsymbol{\beta}'\boldsymbol{\alpha} = a_1$, $\boldsymbol{\rho} = \mathbf{0}$, $\boldsymbol{\mu}_0 = \mathbf{0}$ and $\mathbf{\Omega} = \mathbf{I}_n$. We consider values of (a_1, a_2), with $a_1 = -0.1, -0.2, -0.4, -0.8$ and $a_2 = 0, -0.1, -0.2, -0.4, -0.8$.

The test for rank one gives size of the tests. The rejection probabilities of a nominal 5% level test are reported in the left columns in Table 2. Johansen (2002) noted that if we are close to an $I(1)$ model with lower rank ($a_1 \to 0$, $a_2 = 0$), then the distribution is shifted to the left, and the rejection probability is less than 5%, and if we are close to an $I(2)$ model ($a_1 \to 0$, $a_2 \neq 0$), the distribution is shifted to the right, and the rejection probability is greater than 5%. The size of the asymptotic test is at least 27% for $T = 50$ and 20% for $T = 100$, since this is the rejection probability we get for $a_1 = -0.1$ and $a_2 = -0.8$. By using the correction factor we get a test with a maximal rejection probability of about 7% for $a_1 = a_2 = -0.8$ for both $T = 50$ and $T = 100$. The maximal rejection probability of the bootstrap test is 7% for $T = 50$ and 6% for $T = 100$ for $a_1 = -0.1$ and $a_2 = -0.8$. The size of the FDB test is slightly closer to the nominal level 5%. Finally we note that the test of rank zero, $\mathbf{\Pi} = \mathbf{0}$ and $\boldsymbol{\mu}_1 = \mathbf{0}$, in the model (1) gives power of the tests and are reported in the right columns in Table 2. The bootstrap test has much higher power than the corrected test for most values of a_1 and a_2, but the difference in power between the bootstrap tests based on restricted and unrestricted residuals is negligible. For example, for $T = 50$ when $a_1 = -0.1$ and $a_2 = -0.4$ the power of the corrected test is about 20% and the power of the bootstrap test based on unrestricted residuals is 48%. For $T = 100$ the power of the corrected test is 81% and the power of the bootstrap test is 94%. Going more into details, we remark that for $T = 50$ and some values of a_1 and a_2 the corrected test has higher power than the bootstrap test, but for $T = 100$ and all values of a_1 and a_2 the bootstrap test has higher power than the corrected test.

The maximal number of simulations with a root of the characteristic polynomial outside the unit circle was 19 for $a_1 = -0.1$ and $a_2 = -0.8$ and $T = 50$ and 1 for $a_1 = -0.1$, $a_2 = -0.2, -0.4$ and $T = 100$.

The performance of the tests with real data is illustrated by an example involving US interest rates in the next section.

Table 2: Test of $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, $\boldsymbol{\mu}_1 = \boldsymbol{\alpha}\boldsymbol{\rho}'$ (left columns) and $\mathbf{\Pi} = \mathbf{0}$, $\boldsymbol{\mu}_1 = \mathbf{0}$ (right columns) in model (6). The nominal significance level is 5%. In the simulations $n = 5$, $\boldsymbol{\alpha} = (a_1, a_2, 0, 0, 0)'$, $\boldsymbol{\beta} = (1, 0, 0, 0, 0)'$, $\boldsymbol{\rho} = \mathbf{0}$, $\boldsymbol{\mu}_0 = \mathbf{0}$, $\mathbf{\Omega} = \mathbf{I}_n$. The number of simulations is $10,000$ and the number of bootstrap replications is 1000. See Table 1 for explanations.

a_2/a_1		-0.1		-0.2		-0.4		-0.8	
					$T = 50$				
0.0	Q	0.014	0.132	0.015	0.143	0.028	0.219	0.078	0.583
	Q_S	0.002	0.032	0.003	0.035	0.005	0.062	0.018	0.278
	Q_C	0.004	0.039	0.005	0.044	0.009	0.084	0.038	0.378
	Q^*	0.005	0.049	0.006	0.053	0.009	0.091	0.035	0.349
	Q^{**}	0.006	0.049	0.007	0.054	0.011	0.093	0.039	0.348
	Q^+	0.005	0.050	0.007	0.053	0.010	0.091	0.038	0.350
	Q^{++}	0.006	0.049	0.008	0.054	0.012	0.093	0.042	0.348
-0.1	Q	0.023	0.163	0.020	0.164	0.032	0.236	0.079	0.591
	Q_S	0.003	0.045	0.003	0.044	0.006	0.070	0.018	0.285
	Q_C	0.005	0.049	0.006	0.052	0.009	0.091	0.038	0.383
	Q^*	0.007	0.064	0.008	0.062	0.009	0.100	0.035	0.357
	Q^{**}	0.008	0.064	0.008	0.063	0.012	0.100	0.038	0.356
	Q^+	0.009	0.064	0.008	0.063	0.010	0.099	0.037	0.358
	Q^{++}	0.010	0.064	0.010	0.063	0.013	0.100	0.042	0.356
-0.2	Q	0.056	0.283	0.038	0.236	0.040	0.281	0.083	0.621
	Q_S	0.013	0.098	0.007	0.072	0.007	0.090	0.019	0.310
	Q_C	0.013	0.077	0.010	0.072	0.012	0.111	0.039	0.412
	Q^*	0.017	0.130	0.012	0.100	0.014	0.126	0.036	0.384
	Q^{**}	0.020	0.130	0.014	0.102	0.016	0.126	0.041	0.385
	Q^+	0.021	0.132	0.014	0.101	0.015	0.125	0.037	0.384
	Q^{++}	0.024	0.130	0.018	0.102	0.018	0.126	0.043	0.385
-0.4	Q	0.168	0.662	0.110	0.530	0.078	0.470	0.090	0.722
	Q_S	0.056	0.419	0.029	0.263	0.016	0.199	0.023	0.409
	Q_C	0.026	0.202	0.024	0.187	0.024	0.210	0.045	0.514
	Q^*	0.053	0.479	0.036	0.329	0.027	0.257	0.041	0.489
	Q^{**}	0.052	0.475	0.038	0.327	0.031	0.257	0.044	0.492
	Q^+	0.061	0.480	0.041	0.329	0.030	0.259	0.042	0.489
	Q^{++}	0.059	0.475	0.044	0.327	0.034	0.257	0.046	0.492
-0.8	Q	0.268	0.980	0.210	0.958	0.145	0.903	0.011	0.934
	Q_S	0.103	0.925	0.073	0.848	0.042	0.707	0.042	0.796
	Q_C	0.020	0.725	0.032	0.721	0.043	0.685	0.065	0.897
	Q^*	0.065	0.945	0.058	0.887	0.047	0.769		
	Q^{**}	0.058	0.943	0.055	0.883	0.050	0.765		
	Q^+	0.070	0.944	0.063	0.885	0.053	0.768		
	Q^{++}	0.060	0.943	0.060	0.883	0.053	0.765		

Table 2 continued

a_2/a_1		-0.1		-0.2		-0.4		-0.8	
					$T = 100$				
0.0	Q	0.011	0.104	0.019	0.160	0.048	0.435	0.076	0.984
	Q_S	0.005	0.053	0.008	0.085	0.022	0.288	0.045	0.963
	Q_C	0.004	0.040	0.008	0.070	0.026	0.281	0.063	0.978
	Q^*	0.005	0.058	0.010	0.094	0.029	0.303		
	Q^{**}	0.008	0.059	0.013	0.092	0.033	0.300		
	Q^+	0.006	0.059	0.011	0.092	0.029	0.299		
	Q^{++}	0.008	0.059	0.013	0.092	0.034	0.300		
-0.1	Q	0.027	0.184	0.028	0.218	0.053	0.481	0.077	0.987
	Q_S	0.012	0.105	0.011	0.124	0.023	0.325	0.047	0.967
	Q_C	0.009	0.070	0.011	0.098	0.029	0.318	0.065	0.981
	Q^*	0.014	0.114	0.015	0.132	0.030	0.340		
	Q^{**}	0.017	0.113	0.018	0.134	0.035	0.340		
	Q^+	0.015	0.113	0.015	0.134	0.031	0.344		
	Q^{++}	0.017	0.113	0.018	0.134	0.035	0.340		
-0.2	Q	0.086	0.501	0.054	0.413	0.065	0.594	0.079	0.992
	Q_S	0.047	0.365	0.027	0.275	0.035	0.469	0.049	0.977
	Q_C	0.025	0.197	0.022	0.204	0.043	0.553	0.067	0.988
	Q^*	0.040	0.378	0.028	0.288				
	Q^{**}	0.041	0.378	0.031	0.285				
	Q^+	0.043	0.379	0.030	0.288				
	Q^{++}	0.044	0.378	0.032	0.285				
-0.4	Q	0.168	0.968	0.118	0.913	0.088	0.899	0.081	0.998
	Q_S	0.103	0.940	0.066	0.840	0.053	0.834	0.051	0.995
	Q_C	0.030	0.839	0.042	0.751	0.054	0.882	0.067	0.997
	Q^*	0.062	0.943	0.052	0.848				
	Q^{**}	0.058	0.940	0.051	0.844				
	Q^+	0.063	0.944	0.053	0.848				
	Q^{++}	0.056	0.940	0.052	0.844				
-0.8	Q	0.198	1.000	0.146	1.000	0.105	1.000	0.086	1.000
	Q_S	0.117	1.000	0.084	1.000	0.067	1.000	0.053	1.000
	Q_C	0.026	1.000	0.043	1.000	0.059	1.000	0.068	1.000
	Q^*	0.060	1.000	0.056	1.000		1.000		1.000
	Q^{**}	0.056	1.000	0.052	1.000		1.000		1.000
	Q^+	0.060	1.000	0.055	1.000		1.000		1.000
	Q^{++}	0.054	1.000	0.052	1.000		1.000		1.000

3.3 The Dickey–Fuller Test for Rank Zero in the Model with k Lags

In the third experiment we study the size of the Dickey–Fuller test for rank zero in the model with k lags, i.e., the test of $\mathbf{\Pi} = \mathbf{0}$ and $\boldsymbol{\mu}_1 = \mathbf{0}$ in the model (1) with $n_d = 1$:

$$\Delta\mathbf{X}_t = \mathbf{\Pi}\mathbf{X}_{t-1} + \sum_{i=1}^{k-1} \mathbf{\Gamma}_i \Delta\mathbf{X}_{t-i} + \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \boldsymbol{\varepsilon}_t. \qquad (7)$$

We consider the simple case of a DGP, where we assume that $\mathbf{\Pi} = \mathbf{\Gamma}_1 = \cdots = \mathbf{\Gamma}_{k-1} = \mathbf{0}$ and $\boldsymbol{\mu}_1 = \mathbf{0}$, in which case $\Delta\mathbf{X}_t = \boldsymbol{\mu}_0 + \boldsymbol{\varepsilon}_t$. In the simulations $n = 5$, $k = 1, \ldots, 4$, $\boldsymbol{\mu}_0 = \mathbf{0}$ and $\mathbf{\Omega} = \mathbf{I}_n$. In this experiment the number of parameters relative to the sample size is of interest. The series lengths are $T = 50$, $T = 100$, $T = 200$ and $T = 500$. The rejection probabilities of a nominal 5% level test are reported in Table 3. The asymptotic test is oversized, in particular in models with large values of k. Johansen (2002) found that as long as the number of parameters per observation, kn/T, is less than 0.2, the correction factor gives a reasonable test. The bootstrap and FDB tests have rejection probabilities close to the nominal level for all values of k and T. For example, for $n = 5$, $k = 4$, $T = 100$, so that $kn/T = 0.2$, and a nominal 5% level test, the rejection probability of the asymptotic test is 48% and the correction factor gives a test with rejection probability about 12%. The bootstrap and FDB tests have rejection probabilities around 5%. The rejection probabilities of all tests are close to the nominal level 5% for $T = 500$, which is what we would expect.

The values of n, k and T considered here are empirically relevant and the size distortion of the asymptotic test is very serious for the series lengths in economic and financial data. The correction factor is not able to correct the size of the asymptotic test if the number of parameters per observation is greater than 0.2, but the size of the bootstrap and FDB tests is close to the nominal level even when the number of parameters per observation is 0.4. This shows that the bootstrap test can be used in overparameterised models and we still get a test that has the correct size. The correction factor does not guard against size distortion in models with long lag lengths. We look at this problem more closely in an application to international stock prices in the next section.

Table 3: Test of $\mathbf{\Pi} = \mathbf{0}$, $\boldsymbol{\mu}_1 = \mathbf{0}$ in model (7). The nominal significance level is 5%. In the simulations $n = 5$, $\mathbf{\Pi} = \mathbf{0}$, $\boldsymbol{\mu}_1 = \mathbf{0}$, $\mathbf{\Gamma}_i = \mathbf{0}$, $i = 1, \ldots, k-1$, $\boldsymbol{\mu}_0 = \mathbf{0}$, $\mathbf{\Omega} = \mathbf{I}_n$. The number of simulations is $10,000$ and the number of bootstrap replications is 1000. See Table 1 for explanations.

k/T		50	100	200	500
1	Q	0.146	0.094	0.075	
	Q_S	0.035	0.050	0.054	
	Q_C	0.069	0.069	0.064	
	Q^*	0.051	0.054	0.050	
	Q^{**}	0.053	0.054	0.053	
	Q^+	0.052	0.053	0.049	
	Q^{++}	0.054	0.054	0.050	
2	Q	0.405	0.183	0.111	
	Q_S	0.034	0.047	0.056	
	Q_C	0.094	0.074	0.068	
	Q^*	0.055	0.052	0.052	
	Q^{**}	0.054	0.052	0.053	
	Q^+	0.055	0.054	0.052	
	Q^{++}	0.055	0.053	0.052	
3	Q	0.757	0.284		
	Q_S	0.026	0.042		
	Q_C	0.193	0.082		
	Q^*	0.055	0.051		
	Q^{**}	0.053	0.051		
	Q^+	0.053	0.052		
	Q^{++}	0.051	0.052		
4	Q	0.967	0.463		
	Q_S	0.025	0.042		
	Q_C	0.425	0.107		
	Q^*	0.059	0.052		
	Q^{**}	0.052	0.051		
	Q^+	0.058	0.051		
	Q^{++}	0.053	0.049		

4 FINANCIAL TIME SERIES

The likelihood ratio test has been widely applied to the empirical analysis of financial time series. We consider two empirical applications to financial time series. The dependence of the small sample distribution of the likelihood ratio test on the parameters is illustrated by an application to US interest rates. The effect of the lag length on the size of the likelihood ratio test is demonstrated through an application to testing for cointegration between international stock prices. To save space we only report the results for the bootstrap tests based on restricted residuals since the difference in size and power between the likelihood ratio test based on restricted and unrestricted residuals is negligible.

For the bootstrap tests with real data it is convenient to report the p-values. The bootstrap and FDB p-values are computed by the formulas

$$p^* = \frac{1}{B}\sum_{j=1}^{B} I(Q_j^* \geq Q) \tag{8}$$

and

$$p^{**} = \frac{1}{B}\sum_{j=1}^{B} I(Q_j^* \geq Q^{**}(1-\alpha^*)). \tag{9}$$

For comparison with the asymptotic quantiles we have also computed the 95% bootstrap and FDB quantiles. Finally, in order to mimic the actual applications we simulate the data and compute the rejection probabilities for the simulated data.

4.1 US Interest Rates

The literature which applies cointegration to the theory of the term structure of interest rates is by now extensive. An influential early paper is Hall et al. (1992), who show that the term structure of US treasury bill yields can be modelled as a cointegrated system. They show that the theory implies that if there are n series, then the cointegration rank is $n-1$, so that there is a single common stochastic trend driving the system.

The data consist of $T = 130$ monthly observations from 1995(1) to 2005(10) on the 1, 3, 6, 9 and 12 month US Interbank Rates. All interest rates are nominal and annualised. The data were obtained from Datastream. We estimate a VAR(3) model with an unrestricted constant and two dummy variables to account for interest rate shocks in January and September 2001 (the second outlier is due to the effect of the terrorist attacks on September 11 2001). The eigenvalues, likelihood ratio statistics, corrected likelihood ratio statistics, bootstrap and FDB p-values are given in Table 4. The number of bootstrap replications is $B = 1,000,000$. Applying the bootstrap and FDB tests to the data on US interest rates we find that the tests accept the hypothesis that the cointegration rank is $n-1$, so in this case the asymptotic test and the bootstrap tests lead to the same conclusion that the cointegration rank is $r = 4$ if we use the nominal significance level 5%. The largest unrestricted root of the companion matrix when $r = 4$ is 0.8284, which is clearly a stationary root. The corrected test leads to the wrong conclusion that $r = 3$ rather than $r = 4$. It is worth mentioning that the hypothesis that the spread between the interest rates span the cointegration space is rejected by the data. Since hypothesis testing on the cointegration vectors is not our concern here we do not go into details, but see Hall et al. (1992) for a discussion and Omtzigt and Fachin (2002) for bootstrap tests of restrictions on the cointegration vectors.

Table 4: The eigenvalues, likelihood ratio statistics, corrected likelihood ratio statistics, bootstrap and FDB p-values. p^* and p^{**} denote the bootstrap and FDB bootstrap p-values.

r	$\widehat{\lambda}_{r+1}$	Q	Q_S	Q_C	p^*	p^{**}
US interest rates data						
0	0.614	248.226	218.908	226.493	0.000	0.000
1	0.382	127.439	112.387	76.256	0.000	0.000
2	0.296	66.413	58.569	40.218	0.000	0.000
3	0.138	21.897	19.311	13.403	0.001	0.000
4	0.023	3.014	2.658	1.971	0.295	0.346
Five largest eigenvalues of the companion matrix: $0.9615, 0.8659, 0.6870 \pm 0.2280i, 0.5011 + 0.5778$						
International stock prices data						
0	0.171	73.085	00.000	65.700	0.375	0.449
1	0.096	34.653				
2	0.047	13.882				
3	0.014	3.987				
4	0.006	1.182				
Five largest eigenvalues of the companion matrix: $0.9862, 0.9839, 0.9442, 0.8924 \pm 0.0847i$						

The bootstrap and FDB 95% quantiles are reported in Table 5. We find that for the hypothesis $r = 0$ the bootstrap and FDB quantiles are larger than the asymptotic quantile, but for $r > 0$ the situation is reversed. This illustrates the fact that for some points in the parameter space the asymptotic test is in fact undersized and then the correction factor corrects the test in the wrong direction. The bootstrap and FDB tests have size much closer to the nominal significance level 5%.

In order to further investigate the properties of the tests we simulate the data for $r = 0, \dots, n-1$ using the estimated parameters $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Gamma}}_1, \dots, \widehat{\boldsymbol{\Gamma}}_{k-1}, \widehat{\boldsymbol{\mu}}_0, \widehat{\boldsymbol{\Omega}})$. We simulate $10,000$ time series with $T = 130$ observations. The processes are started at the actual initial values. In each simulation we estimate the parameters and the 95% simulated and bootstrap quantiles. We compare in Table 5 the asymptotic 95% quantiles with the simulated and bootstrap quantiles. To estimate the bootstrap quantiles we use $B = 1,000,000$ bootstrap replications. The simulated and bootstrap quantiles are close to the asymptotic quantiles, except for the test of $r = 0$ and the last hypothesis $r = 4$. The hypothesis $r = 4$ is uninteresting since a rejection means that all series are stationary. We apply the tests to the simulated data. The number of bootstrap replications is $B = 1000$. The asymptotic test is oversized when $r = 0$. From Table 6 we find that for

Table 5: Asymptotic, simulated and bootstrap 95% quantiles.

k	r	95% as	95% sim	95% B	95% FDB
		US interest rates data			
3	0	68.68	80.421	82.645	83.292
	1	47.21	45.516	45.427	37.075
	2	29.38	27.589	27.619	23.122
	3	15.34	14.025	13.988	11.624
	4	3.84	5.856	5.645	4.984
		International stock prices data			
6	0	68.68	91.878	96.106	101.939

$r = 0$, the rejection probability of the asymptotic test is about 23%, the rejection probability of the corrected test is about 10%, and the rejection probabilities of the bootstrap and FDB tests are about 4.5%. For $r > 0$ the asymptotic test is undersized and then the correction factor corrects in the wrong direction. This corresponds to the case in the simulations in Section 3.2 when we are close to an $I(1)$ model with lower rank and then the size of the asymptotic test is below the nominal significance level 5%. The bootstrap test is also undersized but the size of the FDB test is close to the nominal significance level 5%. Turning to power, we find that the asymptotic test has good power properties, but using the correction factor decreases power considerably. The power of the bootstrap test is always higher than the power of the corrected test. The difference in power between the bootstrap and FDB tests is small.

4.2 International Stock Prices

The likelihood ratio test has been applied is to test for cointegration between international stock markets (Ahlgren and Antell, 2002, and Kasa, 1992). Based on the likelihood ratio test most authors have concluded that international stock prices are cointegrated, although this conclusion seems to contradict the theory of efficient markets. The null hypothesis of no cointegration is not rejected using the bootstrap likelihood ratio test, and the bootstrap quantiles are much larger than the asymptotic quantiles, so the asymptotic test is seriously oversized.

The data consist of month-end stock market index observations for the United States, the United Kingdom, Japan, Germany and France from 1988(1) to 2005(7). The indices are the MSCI total return indices in US dollars and were retrieved from Datastream. The number of monthly observations is $T = 211$. A VAR(6) model with an unrestricted constant and two dummy variables was fitted. In estimating the model it was necessary to include two dummy variables to account for outliers which occurred in 1998(8) and 2002(9). First we apply the tests to the actual data with $B = 1,000,000$ bootstrap replications. The eigenvalues,

Table 6: Rejection probabilities of a nominal 5% level test for the simulated data. The number of simulations is 10,000 and the number of bootstrap replications is 1000. See Table 1 for explanations.

k	r	Q	Q_S	Q_C	Q^*	Q^{**}
			US interest rates data			
				$r=0$		
3	0	0.227	0.069	0.099	0.046	0.043
				$r=1$		
	0	0.224	0.067	0.098	0.059	
	1	0.032	0.001	0.000	0.065	
				$r=2$		
	0					
	1	0.184	0.068	0.001	0.137	
	2	0.029	0.007	0.000	0.055	
				$r=3$		
	0					
	1					
	2	0.151	0.069	0.003	0.150	
	3	0.023	0.010	0.001	0.066	
				$r=4$		
	0					
	1					
	2					
	3	0.176	0.126	0.049	0.069	
	4	0.146	0.074	0.002	0.121	
			International stock prices data			
				$r=0$		
6	0	0.474	0.173	0.278	0.087	0.073

likelihood ratio statistics, correction factors, corrected likelihood ratio statistics, bootstrap and FDB p-values are given in Table 4. If we use the significance level 5% we find that the asymptotic test rejects $r = 0$, and the corrected and bootstrap tests accept $r = 0$. The bootstrap and FDB 95% quantiles for $r = 0$ are 96.106 and 101.939, respectively, which are much larger than the asymptotic quantile 68.68. The bootstrap and FDB p-values for $r = 0$ are 38% and 45%, respectively. Thus there is very little statistical evidence of cointegration in the international stock prices data. This confirms the conjecture in Ahlgren and Antell (2002) that the evidence of cointegration is due to the use of asymptotic rather than small sample critical values.

In order to further investigate the properties of the tests we simulate the data for $r = 0$ using the estimated parameters $(\widehat{\mathbf{\Gamma}}_1, \ldots, \widehat{\mathbf{\Gamma}}_{k-1},\ \widehat{\boldsymbol{\mu}}_0,\ \widehat{\mathbf{\Omega}})$. We simulate $10,000$ time series with $T = 211$ observations. The processes are started at the actual initial values. In each simulation we estimate the parameters, the correction factor and the 95% quantile. We compare in Table 5 the asymptotic 95% quantile with the simulated and bootstrap quantiles. To estimate the bootstrap quantile we use $B = 1,000,000$ bootstrap replications. The simulated and bootstrap quantiles are larger than the asymptotic quantile, and this is what explains the size distortion of the asymptotic test. We apply the tests to the simulated data. The number of bootstrap replications is $B = 1000$. From Table 6 we find that for $r = 0$, the rejection probability of the asymptotic test is 47%, the rejection probability of the corrected test is 28%, and the rejection probabilities of the bootstrap and FDB tests are 8.7% and 7.3%, respectively.

5 CONCLUSIONS

In this article we have investigated the behaviour of two different approaches for bootstrapping the likelihood ratio trace test of cointegration rank based on restricted and unrestricted residuals. We have also evaluated the FDB procedure put forward by Davidson and MacKinnon (2001). The relative performance of bootstrap tests are compared with the performance of small sample corrections.

The bootstrap test works well in most cases and in those cases where the bootstrap test is oversized (undersized) the FDB test reduces the size distortion of the bootstrap test. The FDB test provides a bias correction when the size distortion of the bootstrap test mainly depends on the parameters (e.g. when the model is close to an $I(2)$ model or an $I(1)$ model with a lower rank). If the size distortion of the asymptotic test mainly comes from the lag length then the bootstrap test has the correct size and there is no room for the FDB test to improve the size properties of the test. The difference between the bootstrap tests based on restricted and unrestricted residuals is negligible. There is some evidence that the bootstrap procedure based on restricted residuals (i.e. residuals from models that impose the null hypothesis of cointegration rank) has slightly better size properties than the test based on unrestricted residuals. The difference in power between the two bootstrap tests is small. We recommend the use of the test based on restricted residuals since it is computationally less burdensome.

We have attempted to consider several regions of the parameter space, different lag lengths and series lengths. In almost all cases we find that the bootstrap test has better size properties than the corrected test. In most of our simulations for power we find that the bootstrap and FDB tests have higher power than the corrected test and for some regions in the parameter space the difference in power is large. The difference in power between the bootstrap and FDB tests is very small, so there are no costs involved in terms of loss of power from using the FDB test to correct the bias of the bootstrap test.

A general conclusion from our Monte Carlo experiments is that the rejection probability of the bootstrap test of cointegration rank is close to the nominal 5% level. In terms of size the bootstrap test dominates the correction factor over most areas of the parameter space. Some specific conclusions are the following. As the parameter approaches a boundary point, where the process is almost $I(2)$, the size of a nominal 5% level test tends to one, and the corrected size tends to zero, since the correction factor has a singularity at that point. If the process is close to being $I(2)$, the rejection probability of the bootstrap test is closer to the nominal level. In models with long lag lengths the test based on asymptotic critical values can have very large size. The correction factor does not manage to correct the size distortion. The size of the bootstrap test is very close to the nominal 5% level. This shows that the bootstrap test can be used in overparameterised models and we still get a test that has the correct size.

Our applications to financial time series show that the bootstrap is useful for improving the finite sample properties of the likelihood ratio trace test applied to financial time series.

References

[1] Ahlgren, N. and Antell, J. (2002), 'Testing for Cointegration between International Stock Prices', *Applied Financial Economics*, 12, 851–861.

[2] Ahn, S. K. and Reinsel, C. G. (1990), 'Estimation for Partially Nonstationary Multivariate Autoregressive Models', *Journal of the American Statistical Association*, 85, 813–823.

[3] Beran, R. (1988), 'Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements', *Journal of the American Statistical Association*, 83, 687–697.

[4] Davidson, R. and MacKinnon, J. G. (2001), Improving the Reliability of Bootstrap Tests, Queen's University Institute for Economic Research Discussion Paper No. 995.

[5] Doornik, J. A., Hendry, D. F. and Nielsen, B. (1998), 'Inference in Cointegrated Models: UK M1 Revisited', *Journal of Economic Surveys*, 12, 533–572.

[6] Giersbergen, N. P. A. van (1996), 'Bootstrapping the Trace Statistic in VAR Models: Monte Carlo Results and Applications', *Oxford Bulletin of Economics and Statistics*, 58, 391–408.

[7] Hall, A. D., Anderson, H. M. and Granger, C. W. J. (1992), 'A Cointegration Analysis of Treasury Bill Yields', *Review of Economics and Statistics*, 74, 116–126.

[8] Harris, R. I. D. and Judge, G. (1998), 'Small Sample Testing for Cointegration using the Bootstrap Approach', *Economics Letters*, 58, 31–37.

[9] Johansen, S. (1996), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (Oxford: Oxford University Press).

[10] Johansen, S. (2002), 'A Small Sample Correction for the Test of Cointegrating Rank in the Vector Autoregressive Model', *Econometrica*, 70, 1929–1961.

[11] Johansen, S., Hansen, H. and Fachin, S. (2005), A Simulation Study of some Functionals of Random Walk, Discussion Paper, www.math.ku.dk/˜sjo.

[12] Kasa, K. (1992), 'Common Stochastic Trends in International Stock Markets', *Journal of Monetary Economics*, 29, 95–124.

[13] Mantalos, P. and Shukur, G. (2001), 'Bootstrapped Johansen Tests for Cointegrating Relationships: A Graphical analysis', *Journal Statistical Computation and Simulation* , 68, 351-371.

[14] Nielsen, B. (1997), 'On the Distribution of Tests for Cointegration Rank', Discussion Paper 133, Nuffield College, Oxford.

[15] Omtzigt, P. and Fachin, S. (2002), 'Bootstrapping and Bartlett Corrections in the Cointegrated VAR Model', Manuscript.

[16] Paparoditis, E. and Politis, D. N. (2005), 'Bootstrapping Unit Root Tests for Autoregressive Time Series', *Journal of the American Statistical Association*, 100, 545–553.

[17] Reimers, H.-E. (1992), 'Comparison of Tests for Multivariate Cointegration', *Statistical Papers*, 33, 335–359.

[18] Ruiz, E. and Pascual, L. (2002), 'Bootstrapping Financial Time Series', *Journal of Economic Surveys*, 16, 271–300.

[19] Swensen, A. R. (2004), Bootstrapping Algorithms for Testing and Determining the Cointegration Rank in VAR Models, Department of Mathematics, University of Oslo, Statistical Research Report No. 5.