WP 67

Bootstrap Testing and Approximate Finite Sample Distributions for Tests of Linear Restrictions on Cointegrating Vectors*

Mikael Gredenhoff† Tor Jacobson‡

April 7, 1998

Abstract

This paper considers computer intensive methods to inference on cointegrating vectors in maximum likelihood cointegration analysis. The likelihood ratio test statistics used in the literature are known to have an asymptotic χ^2–distribution. However, previous simulation studies show that the size distortion of the test can be considerable for small samples. Typically the nominal significance level, say 5%, is much smaller than the attained actual level, and as a consequence, too many true null hypotheses will be rejected. It is demonstrated how a parametric bootstrap can be implemented, frequenly resulting in a nearly exact α level test. Furthermore, response surface regression is used to examine small sample properties of the asymptotic likelihood ratio test. The estimated equations can be used as approximate finite-sample corrections, allowing rough, but easily applied, corrections of the LR test.

Key words: Likelihood ratio test, Bootstrap hypothesis testing, Small sample corrections, Response surface regressions, Monte Carlo simulations

JEL Classification: C12; C32.

---

*We are grateful to Katarina Juselius for helping out with data, and to Mickael Andersson, Sune Karlsson, Rolf Larsson, Anders Vredin, and Anders Warne for helpful discussions.

†Department of Economic Statistics, Stockholm School of Economics, 113 83 Stockholm, Sweden. e-mail: stmg@hhs.se

‡Sveriges riksbank, 103 37 Stockholm, Sweden. e-mail: tor.jacobson@riksbank.se

This paper has earlier been presented in Mikael Gredenhoff's PhD thesis "Bootstrap Inference in Time Series Econometrics".

1. Introduction

Estimation of long-run economic relationships by maximum likelihood cointegration analysis has become increasingly popular in applied work.[1] One reason for this is the straightforward treatment of multivariate aspects of the estimation problem, i.e. the simultaneous estimation of two or more long run relations. Another reason is the possibility of inference for the elements of the cointegrating vectors that generate the long-run economic relationships. However, all distributional results within the maximum likelihood cointegration model rely on asymptotic considerations; likelihood ratio testing for cointegrating rank, the number of cointegrating vectors in the system, leads to a non-standard inference situation, whereas conditional likelihood ratio testing, for given cointegrating rank, is standard with test statistics being asymptotically χ^2. Hence, it is important to study the behavior for small to moderate samples of sizes empirical research usually encounters, say 50 to 200 observations. The number of simulation studies evaluating small sample properties is rapidly growing, but the majority concern estimation of cointegrating vectors and testing for cointegrating rank. To the best of our knowledge, only two papers deal with testing of linear restrictions on the cointegrating vectors for given rank; Jacobson (1995) and Podivinsky (1992). Both papers convey rather optimistic pictures regarding the size distortion problem. For a nominal 5%-test using a small sample size of $T = 50$, the two papers report empirical sizes of 0.0826 and 0.0898, respectively. Still, the size of a sample, whether small or large, is not an absolute entity but must be judged in relation to the complexity of the model one proposes to estimate, as well as the appropriateness of the specification. Both papers consider very simple Data Generating Processes (DGP's) with two or three cointegrated series, a minimum number of lags and just one cointegrating vector and, hence, a small number of parameters.

In contrast, Jacobson, Vredin and Warne (1998) consider an empirical labor market model involving four endogenous variables, two stationary exogenous variables, four lags, two cointegrating vectors and a set of seasonal dummies. Resampling from the estimated model based on an original sample of 104 quarterly observations, three tests of null hypotheses involving restrictions on the cointegrating vectors are evaluated in terms of empirical sizes. The results, 0.3170,

[1] For an excellent introduction to the maximum likelihood cointegration method, see Johansen and Juselius (1990). This reference also contains an instructive application on Danish money demand. Theoretical results are found in Johansen (1988, 1991), and a full account of the methods is provided by Johansen (1995).

0.2895 and 0.3481 in comparison with a nominal size of 0.05, indicate that inference based on the asymptotic approximation of a χ^2-distribution can be severely misleading.[2] So what would normally be thought of as a reasonably large sample, $T = 104$, could for inference purposes be quite inadequate due to the many estimated parameters in the empirical model.

One could describe the problem as one of lacking coherence between the test statistic and its reference distribution and there are, in principle, two distinct routes to alleviate the problem; either for given test statistic correct the reference distribution, or, for given reference distribution correct the statistic in use. Bartlett adjustment of likelihood ratio test statistics is one possibility to improve inference that recently has received interest in this context. Consider a test statistic C_T that converges to C_∞, with an asymptotic error of order T^{-1} or smaller. C_∞ has a known distribution which provides the critical values for the asymptotic test. Now, we would like to obtain a transformed test statistic C_T^*, such that C_T^* converges to C_∞, and only with error terms of order T^{-2} or less at play.[3] In other words, we want a correction of C_T which eliminates the influence of error terms of orders T^{-1}. Such a correction could be based on the expectation of C_T, recognizing that $\frac{C_T}{EC_T}$ tends to $\frac{C_\infty}{EC_\infty}$ as $T \to \infty$, and hence $C_T \approx EC_T \frac{C_\infty}{EC_\infty}$. Larsson (1998*a*) and Nielsen (1997) considers Bartlett adjustment for a univariate counterpart of the trace test for cointegrating rank, i.e. a test for the presence of a unit root in a univariate autoregressive process. Whereas Johansen (1998) derives a Bartlett corrected likelihood ratio test for linear restrictions, i.e. the test situation that this paper addresses. Due to the intricate analysis, Johansen treats a special case of the general, cointegrated vector autoregressive model that we consider. The results are promising, but with limited applicability so far.

Alternatively, we could consider a corrected distribution for the test statistic at hand, that is replace the critical values of the limit distribution with such that will generate an actual test size closer to the nominal one. Analytically this amounts to Edgeworth expansions, or related techniques, of the distribution function, see Barndorff-Nielsen and Cox (1989) or Hall (1992) for overviews. Bootstrap hypothesis testing is a plausible numerical alternative, which in fact can be expressed and interpreted in terms of Edgeworth expansions as shown by

[2] Jacobson et al. (1998) calculate the empirical test sizes for one of the null hypotheses reported above for various sample sizes T. Even for $T = 2000$, the empirical and the nominal sizes do not quite coincide. Some results are; $T = 200 \Leftrightarrow 0.187, T = 400 \Leftrightarrow 0.103, T = 1000 \Leftrightarrow 0.068$ and $T = 2000 \Leftrightarrow 0.056$.

[3] Stictly speaking, this correction only applies to the mean, although higher moments and fractiles can also be expected to be closer approximated by the asymptotic distribution.

Hall. Although the consistency of bootstrapping in the unit root context is still unclear, Harris (1992) has evaluated bootstrapping of Dickey-Fuller unit root tests and Giersbergen (1996) has recently presented promising results for the multivariate maximum likelihood trace test for cointegrating rank. Larsson (1998*b*) uses saddlepoint techniques to approximate small sample corrections of the lower tails of the distributions for some unit root test statistics.

This paper proposes use of bootstrap hypothesis testing as a tractable way to improve inference for linear restrictions. The outline is the following. Section 2 briefly introduces the maximum likelihood method and, in particular, the likelihood ratio test of linear restrictions that subsequently will be evaluated. Section 3 discusses aspects of simulation based testing, i.e. the bootstrap hypothesis test. In Section 4 we present the design of the Monte Carlo experiments and the complex data generating processes based on the empirical monetary vector error correction model estimated in Juselius (1997).We will also examine small-sample properties of the asymptotic likelihood ratio test by estimating response surface regressions. The objective is to establish how the complexity of the model, in terms of number of dimensions, lags,and cointegrating vectors, is related to the size of the test conditional on sample size. Results are given in Section 5. Some concluding remarks will end the paper.

2. Maximum likelihood cointegration

The base-line econometric specification for maximum likelihood cointegration is a VAR-representation of an n-dimensional time series x_t according to

$$\Pi(L)\,x_t = \varepsilon_t, \qquad (t = 1, 2, \ldots, T), \tag{2.1}$$

where $\Pi(L)$ is an $n \times n$ matrix polynomial of order k given by $\Pi(\lambda) = I_n - \sum_{j=1}^{k} \Pi_j \lambda^j$, where L is the lag operator and λ a complex number. Since we focus on integrated processes x_t, an assumption regarding the roots of $\Pi(L)$ is necessary, i.e. $|\Pi(\lambda)| = 0$ if and only if $|\lambda| > 1$ or possibly $\lambda = 1$. The error term ε_t is assumed to be $i.i.d.N_n(0, \Sigma)$.

A slight reparameterization of (2.1) yields a vector error correction, VECM, representation for x_t suitable for estimation of the cointegrating relationships. Letting $\Gamma(\lambda) = I_n - \sum_{i=1}^{k-1} \Gamma_i \lambda^i$ where $\Gamma_i = -\sum_{j=i+1}^{k} \Pi_j$ and $\alpha\beta' = \Pi = -\Pi(1)$ we get

$$\Gamma(L)\,\Delta x_t = \alpha\beta' x_{t-1} + \varepsilon_t, \qquad (t = 1, 2, \ldots, T), \tag{2.2}$$

where Δ is the first difference operator. Writing $\alpha\beta' = \Pi$ reflects an assumption of reduced rank $r < n$ for Π, implying that α and β are $n \times r$ matrices. Johansen (1991), in a version of the Granger representation theorem, state conditions such that $\beta' x_t$ and Δx_t are integrated of order zero and x_t is integrated of order one. When $r > 0$, x_t is cointegrated of order (1,1). The cointegrating vectors are found in the r columns of β, whereas the rows of α have an interpretation as "adjustment coefficients" that determine how $\beta' x_t$ enters in the n equations.

Maximum likelihood estimation of (2.2) implies reduced rank regression, and in particular, finding solutions to an eigenvalue problem, see Johansen (1991, 1992) for details. Inference for the cointegrating rank r in (2.2) is carried out by use of a likelihood ratio test, the *trace* test. This test has a non-standard asymptotic distribution and simulated critical values are used in practice. For given rank r, however, the likelihood ratio principle leads to standard inference, i.e. test statistics for linear restrictions on β have asymptotic χ^2-distributions, see Johansen and Juselius (1992). They discuss three classes of hypotheses. In the first class the hypotheses under consideration can be expressed as: $\Pi = \alpha\varphi' H'$, that is $\beta = H\varphi$ where $H\,(n \times s)$, $r \leq s \leq n$, is a known matrix that specifies the restriction that is imposed on *all* cointegrating vectors. The test statistic is given by

$$W_{LR,1} = T \sum_{i=1}^{r} \ln \left[\frac{\left(1 - \widehat{\lambda}_{H,i}\right)}{\left(1 - \widehat{\lambda}_i\right)} \right], \tag{2.3}$$

where $\widehat{\lambda}_{Hi}$ and $\widehat{\lambda}_i$ are the eigenvalues found as solutions to the eigenvalue problem implied by maximum likelihood estimation of the restricted and unrestricted models. $W_{LR,1}$ is asymptotically χ^2 with $r\,(n-s)$ degrees of freedom.

In the second hypothesis class, r_1 of the r cointegrating vectors $\beta = (H, \psi)$, are considered known (typically given by economic theory) and specified by the matrix $H\,(n \times r_1)$, whereas the remaining $r_2 = r - r_1$ relations are estimated without restrictions. In this case the test statistic is

$$W_{LR,2} = T \left[\sum_{i=1}^{r_1} \ln\left(1 - \widehat{\lambda}_{C.H,i}\right) + \sum_{i=1}^{r_2} \ln\left(1 - \widehat{\lambda}_{H,i}\right) - \sum_{i=1}^{r} \ln\left(1 - \widehat{\lambda}_i\right) \right], \tag{2.4}$$

where $\widehat{\lambda}_{C.H,i}$, $\widehat{\lambda}_{H,i}$ and $\widehat{\lambda}_i$ are the eigenvalues found as solutions to the eigenvalue problem implied by maximum likelihood estimation of the concentrated likelihood, the restricted model, and the unrestricted model, respectively. $W_{LR,2}$ is asymptotically χ^2 with $r_1\,(n-r)$ degrees of freedom.

The third class of hypothesis is formulated for some arbitrary restrictions on r_1 of the cointegrating vectors $\beta = (H\varphi, \psi)$, and the remaining $r - r_1$ relations are estimated without restrictions. Thus, $H\,(n \times r_1)$ is known and the maximum likelihood solutiton is found by an iterative algorithm, see Johansen (1995), which gives the test statistic as

$$W_{LR,3} = T\left[\sum_{i=1}^{r_1} \ln\left(1 - \widehat{\lambda}_{C.H,i}\right) + \sum_{i=1}^{r_2} \ln\left(1 - \widehat{\lambda}_{H,i}\right) - \sum_{i=1}^{r} \ln\left(1 - \widehat{\lambda}_{i}\right)\right], \qquad (2.5)$$

where $\widehat{\lambda}_{C.H,i}$ are the eigenvalues when β is concentrated with respect to $H\varphi$,and $\widehat{\lambda}_{H,i}$ are the eigenvalues for the restricted model, and $\widehat{\lambda}_i$ the eigenvalues for the unrestricted model. The test statistic is also in this case asymptotically distributed as χ^2 but with $(n - s - r_2)r_1$ degrees of freedom. The last hypothesis can esily be extended to a more general form, given as $\beta = (H_1\varphi, H_2\psi)$ where H_1 is restrictions of the first r_1 cointegrating relations, and H_2 are the restrictions on the remaining relations.

3. Small sample correction by bootstrapping

The bootstrap approach provides a feasible method for estimation of the small-sample distribution of a statistic.[4] The basic principle is to approximate this distribution by a bootstrap distribution, which can be obtained by simulation. In short, this is done by generating a large number of resamples, based on the original sample, and by computing the statistics of interest in each resample. The collection of bootstrap statistics, suitably ordered, then constitutes the bootstrap distribution.

3.1. The Bootstrap Test

The objective of a general (one-sided) test is to compute the p-value function

$$p\left(\hat{W}_{LR}\right) = p\left(W_{LR} \geq \hat{W}_{LR}\,|\Psi_0, T\right) \qquad (3.1)$$

where Ψ_0 is the DGP under the null hypothesis, and $\hat{W}_{LR}$ is the realized value of a test statistic W_{LR} based on a sample of length T. Since Ψ_0 is unknown this

[4] Efron and Tibshirani (1993) is an accessible introduction, Hall (1992) is more of a theoretical foundation.

p-value function has to be approximated, which is regularly done using asymptotic theory. For asymptotic theory to be valid it is required that $p\left(\hat{W}_{LR}\right)$ should not depend on Ψ_0 and T, which is usually not true in small samples. An alternative to an asymptotic solution is to estimate the finite-sample DGP by the bootstrap DGP $\hat{\Psi}_0$, that is to use a bootstrap test.

If B bootstrap samples, each of size T, are generated in accordance with $\hat{\Psi}_0$ and their respective test statistics W^*_{LR} are calculated using the same test statistic W_{LR} as above, the estimated bootstrap p-value function is defined by the quantity

$$p^*\left(\hat{W}_{LR}\right) = B^{-1}\sum_{i=1}^{B} I\left(W^*_{LR,b} \geq \hat{W}_{LR}\right), \quad i = 1, \ldots B, \tag{3.2}$$

where $I(\cdot)$ equals one if the inequality is satisfied and zero otherwise. The null hypothesis is rejected when the selected significance level exceeds $p^*\left(\hat{W}_{LR}\right)$.

The bootstrap testing procedure is a general tool and can be applied to all tests that allow for the implementation of the null-hypothesis in the bootstrap. Davidson and MacKinnon (1996*a*) conclude that the size distortion of a bootstrap test is of the order $T^{-1/2}$ smaller than that of the corresponding asymptotic test. A further refinement of the order $T^{-1/2}$ can be obtained in the case of an asymptotically pivotal statistic, i.e. a statistic whose limiting distribution is independent of unknown nuisance parameters. Since the test functions considered in this paper are asymptotically χ^2, the predicted refinements are thus of order T^{-1}. For further teoretical considerations, see Davidson and MacKinnon (1996*a*), and for other examples on implementation of the bootstrap test, see Andersson and Gredenhoff (1997, 1998).

3.2. Construction of the Bootstrap Samples

The original non-parametric bootstrap suggested by Efron (1979), is designed for *iid* observations. It usually fails for dependent observations, such that we have in e.g. time series analysis, since the order of the observations is affected. Dependencies in data can be maintained in the bootstrap resamples by using a model-based bootstrap, which is the natural way to proceed in our case since a well-defined statistical model forms the null-hypothesis.

When testing for linear restrictions on cointegrating vectors, the DGP Ψ_0 is characterized by an unknown specification. Since the null model, and consequently Ψ_0, is unknown, the estimated (bootstrap) DGP $\hat{\Psi}_0$ is used to create the bootstrap samples. In our case this means that the estimated error correction model is used as the resampling model,

$$\hat{\Gamma}(L)\,\Delta x_t = \hat{\alpha}\widehat{\beta}' x_{t-1} + \hat{\varepsilon}_t. \tag{3.3}$$

This resampling model clearly obeys the null-hypothesis for e.g. $\hat{\beta} = (H\hat{\varphi})$, i.e. the linear restrictions on the cointegrating relations stated in the null-hypothesis are satisfied. Resampling is done with a simple parametric algorithm which makes use of the normality assumption for the disturbances ε_t in (2.2). This implies that the bootstrap residuals ε_t^* are independent draws from a normal distribution with mean zero and variance $\hat{\Sigma}$. The bootstrap samples $\mathbf{x}_i^*$, $i = 1, ..., B$, are then created recursively, through equation (3.3), using the bootstrap residuals ε_t^*.

4. Design of the Monte Carlo simulation experiments

This section deals with the design of simulation experiments that seek to evaluate the bootstrap test in terms of size accuracy and power. However, before taking on the bootstrap test, we will examine small-sample properties of the asymptotic likelihood ratio test of linear restrictions in (2.3). The purpose is to provide a (very rough) reference guide to the degree of test size distortion for models of varying complexity and also to help interpretation of the simulation results for the bootstrap test.

4.1. Response surface regression

It seems reasonable that in general the size accuracy of the asymptotic likelihood ratio test will deteriorate as the number of dimensions, n, and lags, k, of the model increases. In order to quantify this relationship, we propose fitting of response surface regressions with simulation estimated empirical quantiles as regressands and functions of n, k, and cointegrating rank, r, as well as T, as regressors. MacKinnon (1994) estimates approximate small sample distributions for unit root test statistics using response surface regressions of the following form:

$$q^p(T_i) = \theta_\infty^p + \theta_1^p T_i^{-1} + \theta_2^p T_i^{-2} + \varepsilon_i, \tag{4.1}$$

where θ_∞^p is pth quantile of the asymptotic distribution, which is what MacKinnon is estimating, and $q^p(T_i)$ is the estimated pth quantile in the ith experiment using a sample size T_i.

Since θ_{∞}^{p} is known to be a $\chi_{p}^{2}(\cdot)$ with $r(n-s)$ degrees of freedom for the likelihood ratio statistic in (2.3), we will model the deviation between the small-sample estimate and the asymptotic value of the pth quantile, $q^{p}(T_i)-\chi_{p}^{2}(\cdot)$, as follows:

$$q^{p}(T_i)-\chi_{p}^{2}(\cdot)=\theta_{1}^{p}T_{i}^{-1}+\theta_{2}^{p}T_{i}^{-2}+\theta_{3}^{p}f(k_i)+\theta_{4}^{p}g(n_i)+\theta_{5}^{p}h(r_i)+\varepsilon_i, \qquad (4.2)$$

where $f(\cdot)$, $g(\cdot)$ and $h(\cdot)$ are functions of the number of lags, dimensions, and cointegrating vectors in the model of the ith experiment using a simplified version of the empirical model based data generating process presented below.

4.2. Data generating process

In Monte Carlo evaluations of econometric methods it is common practise to use stylized, simplified data generating processes. In the case of cointegrated VAR-processes this usually means small numbers of dimensions, lags and cointegrating vectors. There are obvious advantages with this approach, a high degree of experimental control since fewer parameters have to be accounted for, and less computing time. The drawback is little scope to gain insights on the behaviour of the methods in realistic situations such that we are likely to encounter in empirical analyses. We will sacrifice control for realism and use a complex data generating process; a Danish monetary VECM estimated in Juselius (1997) and based on a sample that has previously (in parts) been analysed in Johansen and Juselius (1990) and in Juselius (1993, 1994).

Whereas Juselius (1997) analyses both $I(1)$ and $I(2)$-representations, we will make use of the $I(1)$-representation only. The sample covers the period 1974:1-1993:4 for the following variables taken in logarithms: m_t, the money stock measured as M3, y_t, income measured as the real gross domestic product, p_t, prices measured as the implicit gross domestic product price deflator, $R_{d,t}$, the average bank deposit rate, and, finally, $R_{b,t}$, the effective bond rate. Juselius considers the following orders of integration:

$$\begin{matrix} m_t & y_t & p_t & R_{d,t} & R_{b,t} \\ I(2) & I(1) & I(2) & I(1) & I(1) \end{matrix},$$

and formulates her $I(1)$-model in terms of the transformed variables

$$\left[\begin{matrix} (m_t-p_t) & y_t & \Delta p_t & R_{d,t} & R_{b,t} \end{matrix}\right].$$

In what follows we will construct the DGP's with the above $I(1)$-vector evaluated in the VECM in (2.2) for various model specifications. Except for the empirical application reported in the end of the paper, all results throughout are for a test of the linear restriction $(m_t - p_t) = y_t$. That is, that the quantity theory constant of proportionality between the real money stock and income is unity, $\beta' = [1 \quad -1 \quad 0 \quad 0 \quad 0]$.

4.3. Experimental design

The Monte Carlo experiments are designed to evaluate the parametric bootstrap test in terms of size accuracy and power. In particular we want to see how the original and the bootstrap test performs under different specification of the VECM, conditional on sample size. To do this, we base each data generating process on the empirical model from Juselius (1997) presented above. Each DGP is constructed for a combination of system dimension, lag-length, and cointegrating rank. The following combinations are considered in the following experiments:

- Size evaluation: $n \in \{5,4,3,2\}$, $k \in \{4,3,2\}$, $r \in \{3,2,1\}$, and $T \in \{40,60,80,100,200\}$.
- Power evaluation: $n \in \{5,4,3,2\}$, $k \in \{4,2\}$, $r \in \{3,2,1\}$, and $T \in \{60,100,200\}$.
- Response surface regressions: $n \in \{5,4,3,2\}$, $k \in \{4,3,2\}$, $r \in \{3,2,1\}$, and $T \in \{40,45,\dots,95,100,110,\dots,140,150, 175,200,225,250,300,350,400,500,600,700,800,1000\}$

For the smaller models, $n \in \{4,3,2\}$, we have eliminated the variables $R_{b,t}$, $R_{d,t}$, and Δp_t, and in that order.

The size and power evaluations concern the second class of hypothesis; a test for the presence of r_1 known vectors in the cointegrating space, along with r_2 unrestricted vectors, $\beta = (H, \psi)$, see (2.4). In our case $r_1 = 1$, and as noted above $H\,(n \times r_1) = [1 \quad -1 \quad 0 \quad 0 \quad 0]'$.

For the experiments evaluating test power, data has been generated under the following three alternatives

$$\begin{array}{l} H_{A,1} : \beta = [1 \quad -1.1 \quad 0 \quad 0 \quad 0] \\ H_{A,2} : \beta = [1 \quad -1.3 \quad 0 \quad 0 \quad 0] \\ H_{A,3} : \beta = [1 \quad -1.5 \quad 0 \quad 0 \quad 0] \end{array}.$$

Each Monte Carlo experiment is based on 1,000 replicates and all bootstrap distributions are generated from resampling and calculation of the test statistic 1,000 times, i.e. $B = 1,000$. Naturally, the level of accuracy could be improved using a larger number of Monte Carlo replicates, a 95% confidence interval around a 5% nominal size is $[3.6 - 6.4]$ for 1,000 replicates. Even so, the number of replicates, both in the Monte Carlo and in the bootstrap, seem adequate for our purposes. Some pilot experiments were made to examine the sensitivity of the size estimates for $B \in \{500, 1000, 2000, 5000\}$ (not reported), but no distinct patterns were found, perhaps due to an inadequate number of Monte Carlo replicates (2,000).

5. Results

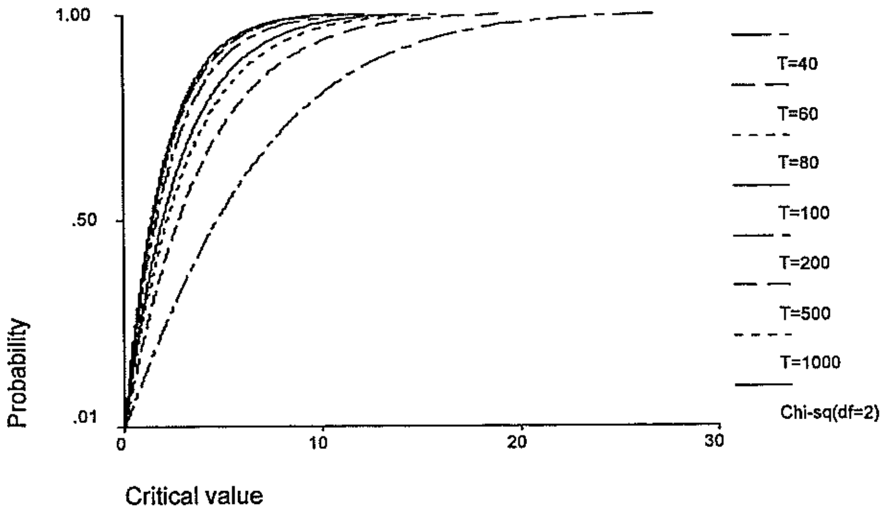
The outline of this section is: response surface regressions results, followed by Monte Carlo simulation results on the size and power properties of the bootstrap test, and finally, in an empirical application, we will present bootstrap test results for a set of hypotheses evaluated in Juselius (1997).

5.1. Response surface regressions

We have followed MacKinnon (1994) very closely in the design of these experiments. The input for the response surface regressions presented below have thus been calculated as follows: for a given combination of dimension, n , lag-order, k, and cointegrating rank, r, 29 sample sizes, ranging from $T = 40$ up to $T = 1000$, have been evaluated in 50 Monte Carlo experiments, with 5000 replicates in each. We have, for a given sample size and specification, estimated 50 sets of 199 percentiles, i.e. $\hat{q}^p_{T,i}, i = 1, \ldots, 50$. This construction of the Monte Carlo experiments is due to the fact that the variances of estimated percentiles are non-constant, for small sample sizes T they tend to be relatively larger. Hence some procedure to account for heteroscedasticity is desirable. MacKinnon (1994) uses a form of generalized method of moments estimation, which has a straightforward implementation for the estimation of (4.1). Using the GMM-procedure, suitably adopted, leads to weighted least squares estimation of the response surfaces. That is, the sample means $\bar{\hat{q}}^p_T$, calculated from 50 Monte Carlo experiments, are regressed according to (4.2), using the inverse of the corresponding standard error $\hat{\sigma}_{pT}$ of each $\bar{\hat{q}}^p_T$ as weights.

Studying the results, we will first consider two sets of simulated cumulative

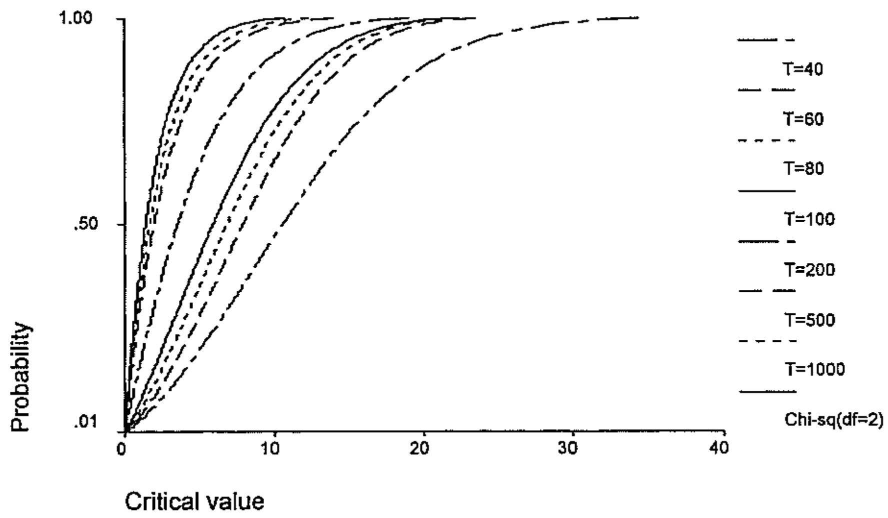
Figure 5.1: Cumulative distributions for varying sample sizes T, using specification $n = 3$, $k = 3$, and $r = 1$.



distributions for model specifications with a common asymptotic distribution, a χ^2-distribution with 2 degrees of freedom. Figure 5.1 and 5.2 present the empirical cumulative distribution functions for a selection of finite-sample sizes and given by 199 percentiles, $\overline{\hat{q}}^p_T$, calculated in the Monte Carlo experiments. These figures are constructed directly from the estimated mean percentiles, so no smothing function is used.The reference curve – a $\chi^2(2)$ – is the solid line to the very left. It is clear that the asymptotic distribution is not a satisfactory approximation for small samples. In Figure 5.1 we can see that even for a VECM with relatively few parameters, the finite-sample distributions are not anywhere close to the asymptotic distribution for sample sizes smaller than 200 observations. Consequently, for a model with richer parametric structure as in Figure 5.2, deviations from the asymptotic distribution are even larger. For the specification $(n = 5, k = 4, r = 3)$, a sample size of at least 500 observations is needed.

The response surface regressions are constructed according to the number of degrees of freedom in the asymptotic distribution, for the different specifications of the VECM. This classification works well for degrees of freedom equal to 2, 3, and 4. For these cases the relationship between the system dimension, n, and

Figure 5.2: Cumulative distributions for varying sample sizes T, using specification $n = 5$, $k = 4$, and $r = 3$.



size distortion is more or less linear, but for d.f. $= 1$ the relationship seems to be non-linear. To improve the fit of the response surface regression for this case, we have estimated one regression for each possible dimension.

In Figure 5.3 we can see that the fitted response surface regression explains size distortion very well, in fact so well that the two series are difficult to discern except for large T. The slope of the curve describes the correction due to sample size and the texture is the correction for lag-order. This correction increases as the sample size decreases. Let us illustrate the effects in two simple examples for a test with 4 degrees of freedom , using the regression given in Table 5.1. For a model with lag-order 3 and sample size $T = 80$, the regression suggests a correction of 7.88, and for a model with the same lag-order, but a sample size of $T = 100$, we get a correction of 5.37. If we instead fix the sample size to $T = 100$ and increase the lag-order to 4, then the regression predicts a correction of 9.21.

5.2. Size

Tables 5.2-5.5 report the Monte Carlo estimated sizes for the likelihood ratio test in (2.4) and its bootstrapped analogue. They are organized according to the

Figure 5.3: Deviation between asymptotic and simulated percentiles for a test with 4 degrees of freedom and the fitted response surface regression. Each cycle represents a given sample size for three lag-orders.

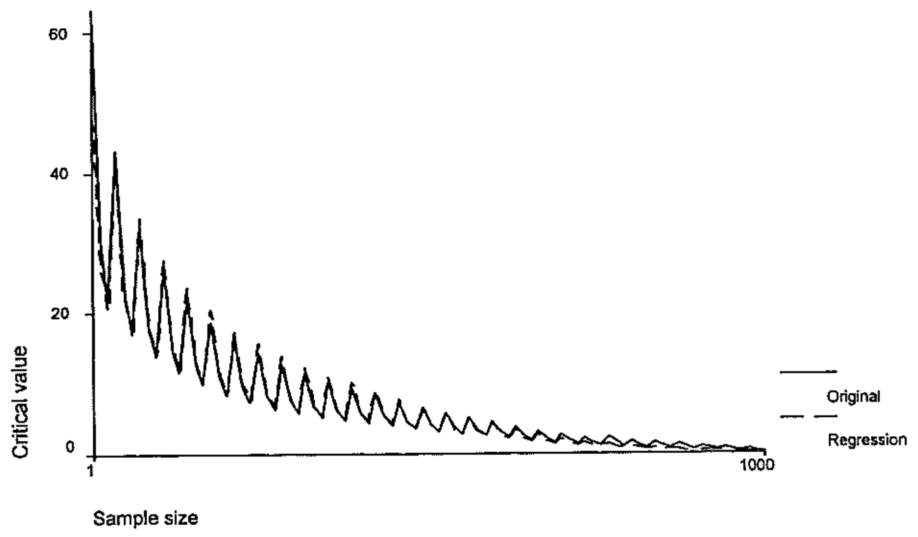


Table 5.1: Response surface regressions, conditional on the number of degrees of freedom.

d.f.	Parameter estimates						R^2
4	$1/t$	$1/t^2$	l_3/t^2	l_4/t^2			
	165.55	26359.05	10840.15	49177.30			98.9
	(6.8)	(15.4)	(7.5)	(29.5)			
3	$1/t$	$1/t^2$	l_3/t^2	l_4/t^2	$n_5/\sqrt{t}$	r_2/t	
	132.44	16346.40	5934.27	21475.69	12.54	-105.09	99.3
	(11.5)	(21.3)	(12.1)	(39.9)	(5.5)	(-4.9)	
2	$1/t$	$1/t^2$	l_3/t	l_4/t	$n_3/\sqrt{t}$	$n_4/\sqrt{t}$	
	242.80	9460.34	25.30	141.39	-12.03	-7.73	
	(4.7)	(5.7)	(2.4)	(13.1)	(-3.4)	(-2.2)	
	$n_5/\sqrt{t}$	r_2/t^2	r_3/t^2				
	36.30	-1804.06	-7167.15				97.9
	(9.7)	(-1.6)	(-6.3)				
1	$1/t$	$1/t^2$	l_3/t	l_4/t	$n_3/\sqrt{t}$	$n_4/\sqrt{t}$	
	80.78	9956.20	36.88	9.66	2.38	3.78	
	(5.5)	(11.3)	(4.8)	(1.3)	(1.8)	(2.8)	
	r_2/t^2	r_3/t^2					
	-4597.61	-3200.90					95.6
	(-5.9)	(-4.2)					
$n = 4$	$1/t$	$1/t^2$	l_3/t	l_4/t	l_3/t^2	l_4/t^2	
	67.70	6189.83	171.83	142.50	-3452.00	-1528.40	99.7
	(11.2)	(16.6)	(18.8)	(15.8)	(-6.4)	(-2.8)	
$n = 3$	$1/t$	$1/t^2$	l_3/t	l_4/t	l_3/t^2	l_4/t^2	
	57.15	8769.59	142.52	62.83	-5101.28	-6068.38	99.9
	(16.6)	(48.7)	(29.6)	(12.7)	(-19.9)	(-23.2)	
$n = 2$	$1/t$	$1/t^2$	l_3/t	l_4/t	l_3/t^2	l_4/t^2	
	165.02	8539.08	-88.93	-96.73	2082.72	270.10	99.2
	(14.6)	(12.8)	(-5.5)	(-5.9)	(2.1)	(0.3)	

Values presented within brackets are t-values of corresponding coefficient

Table 5.2: Estimated size, in percent, for the original and the bootstrap test at a nominal significance level of 5 percent. The asymptotic distribution is a $\chi^2(4)$.

				T				
n	k	r	Test	40	60	80	100	200
5	4	1	Orig.	98.0	84.7	62.3	46.5	17.2
			Boot.	13.7	10.7	7.3	7.4	6.5
5	3	1	Orig.	85.0	58.4	35.7	26.4	12.0
			Boot.	11.2	9.5	6.1	7.1	4.6
5	2	1	Orig.	72.0	34.5	22.7	15.6	10.0
			Boot.	9.9	5.6	4.8	4.8	5.6

degrees of freedom in the asymptotic distributions. In general we find that the asymptotic approximation becomes worse as the degrees of freedom increase. This is also true for the bootstrap test. But, whereas the asymptotic test needs $T = 200$ observations and 1 degree of freedom in order for the estimated size to be anywhere near the nominal test size, it can be seen that the size distortion for the bootstrap test is quite modest even for $T = 40$ and 4 degrees of freedom.[5] In fact, when $T \geq 60$ we frequently find that the bootstrap sizes are not significantly larger than the nominal size.

The overall impression is that the size distortion for the bootstrap test acts in a similar fashion as does the asymptotic test, only to a much lesser extent. Thus, we find that the bootstrap test deteriorates as the number of lags and cointegrating vectors increase. With some caution, we may detect the same effect occurring as the dimension of the system increases. However, for larger sample sizes and smaller degrees of freedom, these patterns disappear and the nominal and estimated sizes coincide.

5.3. Power

Tables 5.6-5.9 report Monte Carlo estimated power for the likelihood ratio test in (2.4) and its bootstrapped analogue. Again we have organized the results according to the degrees of freedom in the asymptotic distributions. The purpose of this set of simulations is to establish the bearing of the theoretical prediction

[5]We intepret the size distortion for the bootstrap test as a reflection of an inadequately estimated (bootstrap) DGP $\hat{\Psi}_0$ in small samples.

Table 5.3: Estimated size, in percent, for the original and the bootstrap test at a nominal significance level of 5 percent. The asymptotic distribution is a $\chi^2(3)$.

				T				
n	k	r	Test	40	60	80	100	200
5	4	2	Orig.	95.6	78.1	57.0	38.6	14.9
			Boot.	12.5	10.4	9.3	8.1	4.5
5	3	2	Orig.	87.5	62.8	39.1	29.5	13.1
			Boot.	10.5	9.6	8.2	8.0	5.2
5	2	2	Orig.	55.5	23.8	15.7	13.3	9.7
			Boot.	8.3	6.0	5.4	5.5	5.8
4	4	1	Orig.	81.4	52.2	31.3	22.8	10.7
			Boot.	11.4	8.9	5.8	5.3	4.4
4	3	1	Orig.	55.6	29.8	20.0	16.4	8.3
			Boot.	9.2	5.3	4.3	5.9	5.0
4	2	1	Orig.	57.4	26.9	16.5	12.1	6.7
			Boot.	8.3	6.3	5.0	5.0	4.8

of Davidson and MacKinnon (1996*b*), namely that power of the bootstrap test will, for practical purposes, not be smaller than the size adjusted power of the asymptotic test.[6] In order to reduce the computational burden we have chosen not to size adjust the bootstrap power estimates, but in view of the modest size distortion reported above, this should not hamper interpretability. Of course, the results will also provide information, albeit limited, on what power we may expect for the likelihood ratio test in a realistic test situation.

We find that the overall outcome supports Davidson and MacKinnons' result, the bootstrap power is almost as good as the asymptotic power on most occasions. It is sometimes, and for unknown reason, dramatically worse. For instance, when $p = 5$ and $r = 3$, we see that the bootstrap tests performs poorly for both lag-orders. Since the asymptotic power also behaves strangely for these cases (e.g. a smaller power for $T = 200$ than for $T = 60$ when $k = 4$), this may be a reflection of the somewhat limited experimental control implied by use of an empirical DGP.

Unlike the experiments regarding test size, it is difficult to detect how the

[6]The statement of Davidson and MacKinnon (1996*b*) is that the power of the bootstrap test is predicted to differ from the power of the size adjusted asymptotic test by an amount of the same order in T , as the size distortion of the bootstrap test itself.

Table 5.4: Estimated size, in percent, for the original and the bootstrap test at a nominal significance level of 5 percent. The asymptotic distribution is a $\chi^2(2)$.

				T				
n	k	r	Test	40	60	80	100	200
5	4	3	Orig.	84.9	74.7	65.6	56.2	27.3
			Boot.	3.5	8.5	8.3	8.3	5.2
5	3	3	Orig.	73.7	53.7	38.5	28.1	12.4
			Boot.	7.5	7.4	6.6	6.8	5.2
5	2	3	Orig.	72.4	53.7	41.8	34.5	16.0
			Boot.	7.0	5.1	7.0	6.2	5.2
4	4	2	Orig.	62.6	33.1	21.1	16.4	7.8
			Boot.	9.1	5.5	4.0	6.5	5.1
4	3	2	Orig.	63.8	40.5	24.9	19.0	8.6
			Boot.	9.4	8.1	5.6	5.7	4.5
4	2	2	Orig.	39.3	19.9	12.1	11.0	5.9
			Boot.	7.6	6.6	4.5	5.8	3.6
3	4	1	Orig.	56.0	53.7	23.3	18.6	8.5
			Boot.	8.8	7.3	7.0	6.9	4.5
3	3	1	Orig.	35.9	18.7	12.8	10.7	6.9
			Boot.	6.8	6.3	4.7	5.0	4.8
3	2	1	Orig.	45.9	24.1	13.3	10.5	7.5
			Boot.	8.0	6.7	4.5	4.9	5.9

Table 5.5: Estimated size, in percent, for the original and the bootstrap test at a nominal significance level of 5 percent. The asymptotic distribution is a $\chi^2(1)$.

				T				
n	k	r	Test	40	60	80	100	200
4	4	3	Orig.	44.7	25.2	16.9	14.4	8.3
			Boot.	7.4	5.9	6.1	4.5	4.9
4	3	3	Orig.	33.8	22.6	14.9	13.4	8.4
			Boot.	5.3	7.5	5.2	6.3	5.8
4	2	3	Orig.	28.2	14.6	9.8	10.7	7.0
			Boot.	5.4	4.9	5.0	5.5	5.1
3	4	2	Orig.	24.8	15.8	14.8	11.1	7.0
			Boot.	6.6	6.3	6.3	5.1	5.0
3	3	2	Orig.	27.7	18.8	13.5	12.5	7.9
			Boot.	7.3	6.6	5.4	5.6	6.4
3	2	2	Orig.	34.0	18.4	13.2	10.4	6.3
			Boot.	8.4	6.6	6.3	6.0	4.5
2	4	1	Orig.	31.0	15.8	13.6	9.9	7.5
			Boot.	7.2	5.5	5.4	4.9	4.3
2	3	1	Orig.	34.9	18.6	12.3	9.5	5.5
			Boot.	5.9	6.2	5.9	4.5	3.7
2	2	1	Orig.	40.1	30.8	16.2	13.4	6.5
			Boot.	7.7	7.4	5.0	5.7	4.5

Table 5.6: Power, in percent, for the original and the bootstrap test at a nominal level of 5 percent, presented for 4 df.

				Alternative hypothesis								
				−1.1			−1.3			−1.5		
n	k	r		60	100	200	60	100	200	60	100	200
5	4	1	Orig.	10.0	10.7	88.4	9.7	40.2	100	12.9	83.6	100
			Boot.	10.3	8.5	86.8	10.2	29.8	100	11.4	58.5	100
5	2	1	Orig.	3.9	4.8	10.6	4.3	6.3	29.9	8.5	32.0	32.0
			Boot.	5.1	4.8	10.7	5.6	6.2	26.6	6.7	28.0	28.0

power is related to the size and complexity of the system. However, in general, and as expected, the power increases with sample size and distance between the null and the alternative. The power for the larger sample size, $T = 200$, is reasonable, irrespective of which alternative we use. For the sample size which is frequently at hand in empirical applications, $T = 100$, the results are not very reassuring. For the smaller sample size, $T = 60$, the power estimates are only occassionaly significantly larger than the nominal test size.

5.4. Empirical application

The purpose of the following empirical application is to demonstrate how inference about long-run economic relationships may shift when asymptotic tests are substituted for bootstrap analogues. We have re-evaluated the asymptotic tests of the hypotheses labeled $\mathcal{H}_1 - \mathcal{H}_{12}$ in Table 6.1 in Juselius (1997), moreover corresponding bootstrap tests have been calculated, see Table 5.11. These hypotheses are examples of the third class, $\beta = (H\varphi, \psi)$, i.e. r_1 vectors are restricted and remaining $r - r_1$ vectors are estimated unrestrictedly.

Hypotheses $\mathcal{H}_3, \mathcal{H}_9, \mathcal{H}_{11}$, and $\mathcal{H}_{12}$ cannot be rejected by either test, likewise both tests reject hypotheses $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_6$, and $\mathcal{H}_{10}$, although $\mathcal{H}_2$ and $\mathcal{H}_{10}$ are borderline cases using the bootstrap test. Hypothesis $\mathcal{H}_4$ concerns the stationarity of the real bond rate and is not rejected by the bootstrap test. Consequently $\mathcal{H}_2$ - the stationarity of a linear combination of inflation and the nominal bond rate - is also insignificant. Bootstrap testing of hypothesis $\mathcal{H}_7$ indicates that the real deposit rate is also stationary, and so is a linear combination of inflation and the nominal deposit rate implied by $\mathcal{H}_8$.

Table 5.7: Power, in percent, for the original and the bootstrap test at a nominal level of 5 percent, presented for 3 df.

				Alternative hypothesis								
				−1.1			−1.3			−1.5		
n	k	r		60	100	200	60	100	200	60	100	200
5	4	2	Orig.	25.8	15.4	70.5	27.0	33.7	100	31.2	67.0	100
			Boot.	10.1	8.9	53.8	9.8	18.4	99.8	10.4	34.2	100
5	2	2	Orig.	5.6	6.4	33.1	5.4	12.9	100	6.7	19.1	100
			Boot.	5.9	6.1	31.9	6.1	13.1	100	6.6	18.3	100
4	4	1	Orig.	5.4	5.4	84.9	7.2	19.3	100	9.0	18.5	100
			Boot.	9.2	7.1	84.9	8.5	17.8	100	9.6	16.9	100
4	2	1	Orig.	5.5	5.3	15.9	9.1	16.0	96.1	21.0	42.6	96.8
			Boot.	5.5	5.7	15.9	8.5	14.5	95.5	11.3	22.7	93.8

Table 5.8: Power, in percent, for the original and the bootstrap test at a nominal level of 5 percent, presented for 2 df.

				Alternative hypothesis								
				−1.1			−1.3			−1.5		
n	k	r		60	100	200	60	100	200	60	100	200
5	4	3	Orig.	17.0	8.8	16.1	16.1	15.5	41.1	17.9	9.5	10.6
			Boot.	8.3	8.9	13.9	9.3	11.9	21.3	8.3	8.5	11.0
5	2	3	Orig.	5.4	7.0	40.8	7.9	36.1	82.2	12.5	33.0	75.9
			Boot.	5.7	6.7	29.1	9.6	17.5	29.9	10.2	13.1	22.4
4	4	2	Orig.	7.1	6.2	62.8	8.2	10.8	100	12.5	9.6	78.7
			Boot.	5.3	6.1	62.4	6.2	9.9	100	6.7	7.4	77.2
4	2	2	Orig.	5.3	5.1	5.3	6.0	5.3	6.2	8.7	13.8	74.2
			Boot.	5.8	4.9	5.4	5.9	5.1	5.9	7.3	12.1	72.5
3	4	1	Orig.	7.3	7.3	71.7	8.6	16.9	100	11.7	23.9	100
			Boot.	7.2	6.2	69.2	8.0	14.0	100	8.7	17.2	100
3	2	1	Orig.	5.4	4.6	8.2	7.5	9.9	67.4	13.5	29.7	96.2
			Boot.	5.5	4.7	8.0	8.1	9.3	66.1	9.3	21.0	92.1

Table 5.9: Power, in percent, for the original and the bootstrap test at a nominal level of 5 percent, presented for 1 df.

				Alternative hypothesis								
				−1.1			−1.3			−1.5		
n	k	r		60	100	200	60	100	200	60	100	200
4	4	3	Orig.	5.5	4.3	17.2	7.2	10.9	61.5	9.8	19.0	81.7
			Boot.	5.0	6.5	17.5	7.0	11.6	60.0	7.5	16.8	75.8
4	2	3	Orig.	4.8	4.9	21.3	3.7	8.6	99.9	3.5	41.9	100
			Boot.	4.2	4.7	21.5	3.2	8.6	99.9	3.1	41.0	100
3	4	2	Orig.	5.7	6.5	88.9	5.5	24.4	100	5.2	46.0	100
			Boot.	5.5	6.0	88.3	5.0	22.1	100	4.4	42.4	100
3	2	2	Orig.	5.8	4.7	5.8	5.7	4.1	21.2	9.1	6.7	9.3
			Boot.	5.0	4.8	5.8	5.5	4.7	21.0	6.9	5.3	8.5

Table 5.10: Original and bootstrap test of liner restriction in the cointegrating space, the restriction on β are defined as $\beta = (\beta_1 H, \psi)$.

							original test		Boot-test	
	$m_t - p_t$	y_t	Δp_t	$R_{d,t}$	$R_{b,t}$	$D83$	$\chi^2(\nu)$	*pval*	*per*	*pval*
$\mathcal{H}_1$	0	1	0	14.1	0	-.08	13.0(1)	0.00	10.7	0.03
$\mathcal{H}_2$	0	1	0	0	3.9	-.09	13.5(1)	0.00	12.4	0.04
$\mathcal{H}_3$	0	1	14.5	0	0	.08	0.53(1)	0.47	6.11	0.58
$\mathcal{H}_4$	0	0	-1	0	1	.001	14.8(2)	0.00	14.6	0.05
$\mathcal{H}_5$	0	0	1	0	-0.2	.014	7.23(1)	0.01	7.32	0.05
$\mathcal{H}_6$	0	0	0	1	-1	-.011	11.5(2)	0.00	14.7	0.01
$\mathcal{H}_7$	0	0	-1	1	0	-.011	5.4(2)	0.07	10.9	0.24
$\mathcal{H}_8$	0	0	-1.4	1	0	-.017	5.2(1)	0.02	7.53	0.11
$\mathcal{H}_9$	0	0	0	1	-0.5	-.003	0.07(1)	0.80	7.64	0.84
$\mathcal{H}_{10}$	0	0	1	-0.5	0.05	.017	7.8(1)	0.01	7.06	0.04
$\mathcal{H}_{11}$	1	-1	0	0	7.2	-.18	0.00(1)	0.96	6.66	0.97
$\mathcal{H}_{12}$	1	-1	0	-14.1	14.1	-1.80	0.02(1)	0.89	5.98	0.90

per denots the 5% percentile of the bootstrap distribution. Compare with $\chi^2_{0.05}(1) = 3.84$, $\chi^2_{0.05}(2) = 5.99$. The bootstrap-tests are based one 5000 replicates.

Table 5.11: Estimated sizes for testing an hypothesis of type $\beta = (H\varphi, \psi)$, in percent, at a nominal level of 5 percent.

$T = 80$				
			Test	
n	k	r	Orig.	Boot.
4	4	3	16.1	4.6
4	4	2	24.5	4.7
4	3	3	17.1	5.3
4	3	2	24.8	6.4
4	2	3	11.8	6.0
4	2	2	13.6	6.2

Finally, in a minor Monte Carlo experiment we examine the size properties of the bootstrap tests applied in Table 5.11. For this experiment the restriction matrix H is set to

$$H\,(n \times s) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}' .$$

Due to the extensive computational effort involved (the maximum likelihood solution has to be iterated), the experiment is restricted to the following cases: $n = 4$, $k \in \{4, 3, 2\}$, $r \in \{3, 2\}$ and $T = 80$. The results in Table 5.12 show that there is no significant size distortion for the bootstrap tests.

6. Conclusions

The likelihood ratio test statistics that are used for checking linear restrictions on cointegrating vectors, are not χ^2 distributed in small samples. Depending on the complexity of the empirical model, convergence towards the asymptotic distribution is attained for various small sample sizes, but rarely for such that e.g. quarterly macro data imply.

This paper demonstrates that a parametrically bootstrapped likelihood ratio test is, more or less, unaffected by size distortions. Moreover, the power of the bootstrap test turns out to be almost as good, or bad, as size adjusted power for the asymptotic test. These results are based on Monte Carlo simulations using an empirical model as data generating process. Hence, we believe that they have bearing for the test behaviour in empirical models.

Extensive simulation experiments have provided input for response surface regressions that seek to explain the size distortion of the asymptotic test in terms of system dimension, lag order, cointegrating rank, and sample size. The fit for the regressions are extremely good, and suggests that they could be used for inference purposes, albeit being based on one particular empirical model.

The general conclusion is that bootstrap hypothesis testing is a useful device for robust inference in this context. Obviously, further work is needed to check sensitivity against model mis-specification such as incorrect lag order and deviation from the normality assumption.

References

[1] ANDERSSON, M.K. and M.P. GREDENHOFF (1997). "Bootstrap Testing for Fractional Integration". *Working Paper Series in Economics and Finance*, No. 188, Stockholm School of Economics.

[2] ANDERSSON, M.K. and M.P. GREDENHOFF (1998) "Robust Testing for Fractional Integration Using the Bootstrap". *Working Paper Series in Economics and Finance*, No. 218, Stockholm School of Economics.

[3] BARNDORFF-NIELSEN, O.E. AND COX, D.R. (1989). "*Asymptotic Techniques for Use in Statistics*", London, Chapman and Hall.

[4] DAVIDSON, R. and J. G. MACKINNON (1996*a*). "The Size Distortion of Bootstrap Tests". *GREQAM Document de Travail No.* 96A15 and *Queens Institute for Economic Research Discussion Paper No.* 936.

[5] DAVIDSON, R. and J. G. MACKINNON (1996*b*). "The Power of Bootstrap Tests". *Queens Institute for Economic Research Discussion Paper No.* 937.

[6] EFRON, B. (1979). "Bootstrap Methods: Another Look at the Jacknife". *Annals of Statistics*, 7, 1-26.

[7] EFRON, B. and R. J. TIBSHIRANI (1993). "*An Introduction to the Bootstrap*", New York, Chapman and Hall.

[8] GIERSBERGEN, N.P.A. VAN (1996). "Bootstrapping the trace statistics in VAR models: Monte Carlo results and applications", *Oxford Bulletin of Economics and Statistics*, 58, 391-408.

[9] HALL, P. (1992). "*The Bootstrap and Edgeworth Expansion*". Springer Verlag. New York.

[10] HARRIS, R.I.D. (1992). "Small Sample Testing For Unit Roots", *Oxford Bulletin of Economics and Statistics*, 54, 615-625.

[11] JACOBSON, T. (1995). "Simulating Small Sample Properties of the Maximum Likelihood Cointegration Method: Estimation and Testing". *Finnish Economic Papers*, 8, 96-107.

[12] JACOBSON, T.,VREDIN, A. AND WARNE, A. (1998). "Are Real Wages and Unemployment Related?" , *Economica*, 65, 69-96.

[13] JOHANSEN, S. (1988). "Statistical analysis of cointegration vectors". *Journal of Economic Dynamics and Control*, 12, 231-254.

[14] JOHANSEN, S. (1989). "*Likelihood based inference on cointegration: theory and applications*". Lecture Notes, Institute of Mathematical Statistics, University of Copenhagen.

[15] JOHANSEN, S. (1991). "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models". *Econometrica*, 59, 1551-1580.

[16] JOHANSEN, S. (1992). "Determination of cointegration rank in the presence of a linear trend". *Oxford Bulletin of Economics and Statistics*, 52, 169-210.

[17] JOHANSEN, S. (1995). "*Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*". Oxford University Press.

[18] JOHANSEN, S. (1997). "*A small sample correction for the test for hypotheses on the cointegrating vectors*".Manuscript, European University Institute, Florence.

[19] JOHANSEN, S. AND JUSELIUS, K. (1990). "Maximum likelihood estimation and inference on cointegration - with applications to the demand for money". *Oxford Bulletin of Economics and Statistics*, 52, 169-210.

[20] JOHANSEN, S. AND JUSELIUS, K. (1992). "Testing structuralhypotheses in a multivariate cointegration analysis of the PPP and UIP for UK". *Journal of Econometrics*, 53, 211-244.

[21] JUSELIUS, K. (1993). "VAR models and Haavelmo's probability approach to macroeconomic modelling". *Empirical Economics*, 18, 595-622.

[22] JUSELIUS, K. (1994). "On the duality between long-run relations and common trends in the I(1) and the I(2) case. An application to aggregate money holdings". *Econometric Reviews* ,13, 151-178.

[23] JUSELIUS, K. (1997). "A Structured VAR in Denmark under Changing Monetary Regimes". Manuscript, University of Copenhagen.

[24] LARSSON, R. (1998*a*). "Bartlett corrections for a unit root test statistic". *Journal of Time Series Analysis*, forthcoming.

[25] LARSSON, R. (1998*b*). "Distribution approximation of unit root tests in autoregressive models". *The Econometric Journal* , forthcoming.

[26] MACKINNON, J.H. (1994). "Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests". *Journal of Business and Economic Statistics*, 12, 167-176.

[27] NIELSEN, B. (1997). "Bartlett correction of the unit root test in autoregressive models". *Biometrika*, 84, 500-504.

[28] PODIVINSKY, J.M. (1992). "Small sample properties of tests of linear restrictions on cointegrating vectors and their weights". *Economic Letters*, 39, 13-18.