BOOTSTRAP INFERENCE IN TIME SERIES ECONOMETRICS

Mikael Gredenhoff

AKADEMISK AVHANDLING
som för avläggande av ekonomie doktorsexamen
vid Handelshögskolan I Stockholm framlägges
till offentlig granskning fredagen
den 29 maj, 1998 kl. 13.15 I sal Torsten.
Handelshögskolan, Sveavägen 65

STOCKHOLM 1998

Bootstrap Inference in Time Series Econometrics

STOCKHOLM SCHOOL OF ECONOMICS
EFI, THE ECONOMIC RESEARCH INSTITUTE

EFI Mission

EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution which works independently of economic, political and sectional interests.

EFI conducts theoretical and empirical research in management and economic sciences, including selected related disciplines. EFI also encourages and assists in the publication and distribution of its research findings, and is involved in the doctoral education at the Stockholm School of Economics.

EFI selects its projects based on the need for theoretical or practical development of a research domain, on methodological interests, and on the generality of a problem.

Research Centres		**Directors**
A	Management and Organisation Theory	Prof. Sven-Erik Sjöstrand
B	Accounting and Managerial Finance	Prof. Lars Östman
C	Mangerial Economics	Prof. Peter Jennergren
CEE	Centre for Ethics and Economics	Adjunct Prof. Hans de Geer
CFR	Centre for Risk Research	Prof. Lennart Sjöberg
CHE	Centre for Health Economics	Prof. Bengt Jönsson
D	Marketing, Distribution and Industry Dynamics	Prof. Lars-Gunnar Mattsson
ES	Economic Statistics	Prof. Anders Westlund
F	Public Management	Prof. Nils Brunsson
FDR	The Foundation for Distribution Research	Acting Prof. Richard Wahlund
FI	Finance	Prof . Clas Bergström
I	Information Management	Prof.. Mats Lundeberg
IEG	International Economics and Geography	Prof.. Mats Lundahl
P	Economic Psychology	Prof. Lennart Sjöberg
PMO	Man and Organisation	Prof. Bengt Stymne
PSC	Policy Science	Adjunct Prof . Brita Schwarz
RV	Law	Prof. Bertil Wiman
S	Economics	Prof. Lars Bergman
T	Industrial Production	Prof . Christer Karlsson

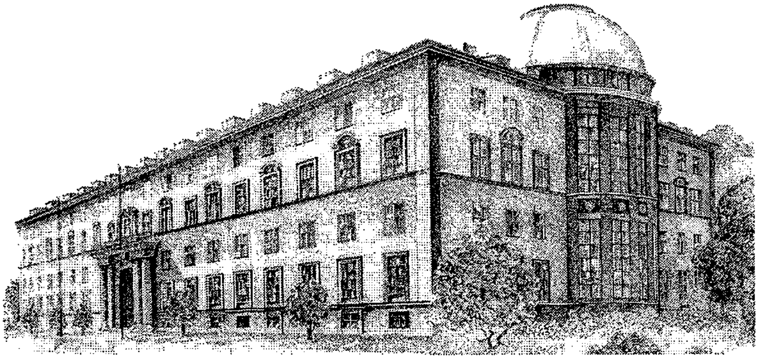
Chairperson: Prof. Sven-Erik Sjöstrand
Director: Ekon Lic. Rune Castenäs

EFI, Box 6501, S-113 83 Stockholm, Sweden • Internet: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se

BOOTSTRAP INFERENCE IN TIME SERIES ECONOMETRICS

Mikael Gredenhoff

Stockholm 1998



STOCKHOLM SCHOOL OF ECONOMICS
EFI, THE ECONOMIC RESEARCH INSTITUTE

A Dissertation for the Degree of Doctor of
Philosophy in Econometrics, presented at
Stockholm School of Economics, 1998.

Abstract:

This volume contains five essays in the field of time series econometrics. The main issue discussed is the lack of coherence between small sample and asymptotic inference. Frequently, in modern econometrics distributional results are strictly only valid for a hypothetical infinite sample. Studies show that the attained actual level of a test may be considerable different from the nominal significance level, and as a consequence, too many true null hypotheses will falsely be rejected. This leads, in the extension, to applied users that too often reject evidence in data for teoretical predictions.

In large, the thesis discusses how computer intensive methods may be used to adjust the test distribution, such that the actual significance level will coincide with the desired nominal level.

The first two essays focus on how to improve testing for persistence in data, through a bootstrap procedure within a univariate framework.

The remaining three essays are studies of multivariate time series models. The third essay considers the identification problem of the basic stationary vector autoregressive model, which is also the base-line econometric specification for maximum likelihood cointegration analysis.

In the fourth essay the multivariate framework is expanded to allow for components of different integrating order and in this setting the paper discusses how fractional cointegration affects the inference in maximum likelihood cointegration analysis.

The fifth essay consider once again the bootstrap testing approach, now in a multivariate application, to correct inference on long-run relations in maximum likelihood cointegration analysis.

ISBN 91-7258-476-9

Keywords: Bootstrap Testing, Small Sample Corrections, Cointegration, Fractional Integration, Monte Carlo Simulations, Lag Order Determination

Printed by
Gotab, Stockholm 1998

Distributed by
EFI, Stockholm School of Economics, Box 6501, S-113 83 Stockholm, Sweden

Contents

Preface

Acknowledgements

The process of completing my Doctor's degree at Stockholm School of economics has been a lengthy one involving everything from pleasant tasks as to meet new friends to some more difficult accomplishments as writing this thesis. Under this time I has become greatly indepted to a number of people. My advisor, Tor Jacobson provided encouragement, as well as insightful and constructive advise. His support has helped me a great deal to reach the end. Sune Karlsson my secondary advisor, but also coauthor, whose comments greatly improved the contents in the thesis. One person that has contributed with a specially large share in this work is Michael Andersson, not only because he is my coauthor to a large part, but also because he is a great room mate, friend and a smooth and easy snowboarder.

I would like to thank the faculty members and Ph.D. students at the Department of Economic Statistics at the Stockholm School of Economics. Furthermore, I wish to express my sincere gratitude to a number of colleagues and friends, listed below, who have helped and encouraged me along the way; Anders Westlund who is the head of the department, Monica Peijne and Carina Morton-Fincham who are the department secretaries, Mickael Löthgen, Magnus Tambour

and Nicklas Zethraeus that started the same year as Ph.D. students, Stefan Åsbrink who has been a good friend, Stefan Lundbergh that has contributed especially to my fitness, my old friends that probably didn't think this would have an end, Magnus Anseklev, Roger Hanning, Gunnar Möller and Henrik Sorin, finally I am deeply grateful to my parents, Kerstin and Leif, who have always been there when needed.

Stockholm, May 1998
Mikael Gredenhoff

List of Papers

1. BOOTSTRAP TESTING FOR FRACTIONAL INTEGRATION, *Working Paper Series in Economics and Finance, No. 188*, Stockholm School of Economics, 1997. With M. Andersson.

2. ROBUST TESTING FOR FRACTIONAL INTEGRATION USING THE BOOTSTRAP, *Working Paper Series in Economics and Finance, No. 218*, Stockholm School of Economics, 1998. With M. Andersson.

3. LAG-LENGTH SELECTION IN VAR-MODELS USING EQUAL AND UNEQUAL LAG-LENGTH PROCEDURES, *Working Paper Series in Economics and Finance, No. 177*, Stockholm School of Economics, 1997. With S. Karlsson.

4. POWER AND BIAS OF LIKELIHOOD BASED INFERENCE IN THE COINTEGRATION MODEL UNDER FRACTIONAL COINTEGRATION, *Working Paper Series in Economics and Finance, No. 221*, Stockholm School of Economics, 1998. With M. Andersson.

5. BOOTSTRAP TESTING AND APPROXIMATE FINITE SAMPLE DISTRIBUTIONS FOR TESTS OF LINEAR RESTRICTIONS ON COINTEGRATING VECTORS, *Unpublished manuscript*, Stockholm School of Economics, 1998.

Part I

Thesis Summary

1

Introduction

This thesis consists of five essays in the field of time series econometrics, presented in Part 2. The chapters in Part 2 are self-contained entities and can be read independently of one another. The purpose of the first part of the thesis is to give an orientation and a introduction to the main topics covered, whereas Chapter 2 to provides an overview of the thesis.

1.1 Computer intensive statistical methods

In the last decade the tools of econometrics has become more complex, in the sense that inference in economic models no longer is associated with simple analytical solutions. Consequently, new methods to solve these complex problem has also been developed. On way to account for more complex numerical difficulties, that recently has shown high potential, is to use simulation based inference. The topic of simulation base inference is very broad, i.e. from numerical estimation of moments to construction of empirical distributions. Hence, the bootstrap techniques applied and discussed in this thesis is just a very small part of a topic that in the future probably will increase substantially in importance.

1.1.1 The Bootstrap

The main problem discussed in this thesis could be described as the lack of coherence between the test statistic and its reference distribution. 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. In the latter case analytical correction, as Bartlett adjustment, is commonly used. In the former situation, a corrected distribution for the test statistic could be considered, 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 Barndorrf-Nielsen and Cox (1989), Field and Ronchetti (1990) or Hall (1992) for overviews.

The bootstrap is a plausible numerical alternative to analytical calculations, to give a approximation or when the asymptotic distribution is known, a more accurate one for small samples. The simple idea is to treat the known sample as if this was the population for which the distribution of interest is to be evaluated. This can account for an approximation of the distribution of interest that is at least as accurate as a first order asymptotic approximation. The Bootstrap hypothesis testing can in fact be expressed and interpreted in terms of Edgeworth expansions as shown by Hall.

The simple idea can be exemplified with a data set consisting of independent and identically distributed observations, $X_1, \ldots, X_n$, having the realized sample $X_1 = x_1, \ldots, X_n = x_n$. The bootstrap resample are a random sample drawn with replacement from the observations $x_1, \ldots, x_n$, realized into $X_1^*, \ldots, X_n^*$. From the original sample the estimator of the sample mean, $\hat{\theta}(X_1, \ldots, X_n)$, and the the estimator of the standard deviation $\hat{\sigma} = \left[Var\,\hat{\theta}(X_1, \ldots, X_n)\right]^{1/2}$ can be calculated, using the bootstrap one resampling yields, similar as above, the estimators $\hat{\theta}(X_1^*, \ldots, X_n^*)$ and $\hat{\sigma}_{boot} = \left[Var\,\hat{\theta}(X_1^*, \ldots, X_n^*)\right]^{1/2}$. Hence, by repeating the bootstrap resampling and the calculations of estimators, an empirical bootstrap distribution can be generated.

The resampling is usually the critical step in the implementation of the bootstrap. The original non-parametric bootstrap suggested by Efron, see for instance Efron and Tibshirani (1993), is designed for *iid* observations. Such resampling can in itself be done in several ways, for example as above, probably the most common; a non-parametric resampling which is based directly on the observations, and the observations are random draws with replacement. A *second* example is a simple parametric algorithm which makes use of a distributional assumption for the observations. This implies that the bootstrap observations are independent draws from a prespecified distribution, where the moments of the distribution are estimates from the sample at hand.

These simple resampling procedures usually fails for dependent observations, such that we have in e.g. time series analysis, since the order of the observations is affected. However, there are several approaches to generalize the bootstrap for dependent data. One can allowing for dependences in the observations and use a procedure that account for the dependences, e.g. a blockbootstrap see for instance Künsch (1989) or a bootstrap that resample the spectral densities see Hjorth (1994) for an overwieve. Another way, common in econometric applications, is to implement a model structure, and resample the *iid* residuals non-parametrically or parametrically. In such situations the specified model may not always capture all the structure in the process, and thus the residuals are not independent. Moreover, in this case the residuals may exhibit some structure such as ARCH effects, kurtosis or skewness and the resampling must be carried out in a way that suitably captures these features, depending on the purpose of the bootstrap.

An accessible introduction to the bootstrap is Efron and Tibshirani (1993), and in need of a more theoretical foundation and in a econometric framework see Hall (1992), Horowitz (1995) and Davidson and MacKinnon (1996*a*,1996*b*).

1.1.2 Hypothesis testing

Traditionally, the bootstrap techniques has frequently been considered for measures of statistical errors such as standard deviation, bias, prediction error and confidence intervals. The relation between confidence intervals and hypothesis testing in standard inference is well established. Also in the context of bootstrap similar issues are of interest, such as the issue of pivotalness relating to the order of accuracy in the adjustment. However, there is a conceptual difference in how to implement the bootstrap in a hypothesis testing situation. Most important, is to base the test on a resample that satisfy the null-hypothesis, even if the population fails to satisfy the null-hypothesis. If this is satisfied the power of the original test will not be distorted and in generally close to the power of the size adjusted asymptotic test, see Davidson and MacKinnon (1996*b*). The trade off between power and a correct level of a test does not need to be a problem when the bootstrap test is considered. Since the bootstrap procedure only corrects the level, and the power of the test depends on the power of the asymptotic test.

The objective of the test is to compute the *p*-value function,

$$p(\hat{\tau}) = p(\tau \geq \hat{\tau} \,|\Psi_0, T), \tag{1.1}$$

where Ψ_0 is the true data generating process (*DGP*) under the null hypothesis, T is the sample size and $\hat{\tau}$ is a realized value of the test statistic τ based on the original sample $\mathbf{x} = [x_1, ..., x_T]'$. The DGP Ψ_0 is characterized by an unknown specification. Since the null model, and hence Ψ_0, is unknown the estimated (bootstrap) DGP $\hat{\Psi}_0$ is employed to create the bootstrap samples. The basic idea is to create a large number of such samples which all obey the null-hypothesis and, as far as possible, resemble the original sample.

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

$$p^*(\hat{\tau}) = B^{-1} \sum_{i=1}^{B} I(\hat{\tau}^* \geq \hat{\tau}), \quad i = 1, \ldots B, \tag{1.2}$$

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

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. One caution should be noted, that the bootstrap should not be expected to work if a first order expansion does not, see Hall (1992). Otherwise, Davidson and MacKinnon (1996*a*) show 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.

When testing in econometrics it is common that a well-defined statistical model forms the null-hypothesis thus the dependencies in data can be maintained in the bootstrap resamples by using a model-based bootstrap. In this case the specified statistical model works as the bootstrap DGP $\hat{\Psi}_0$, and the resampling can be done through a simple residual resampling as discussed in previous section.

1.2 Time series processes

1.2.1 Fractional integrated models

Although most analysis of non-stationarity in univariate time series is devoted to unit roots, fractional integrated models introduce a more diverse method for analysis. The fractionally integrated process permits representations between I(0) and I(1), which imply series with different long run predictions and effects of shocks compared with conventional macroeconomics. This model is a generalization of the ARIMA class of models, and allows the order of differencing to take non-integer values. As an introduction, consider the simplest case of fractional integrated white noise, disregarding the AR(p) and the MA(q) parts that allow for the low frequencies,

$$(1 - B)^d x_t = \varepsilon_t$$

where ε_t is white noise and B the backshift operator. The differencing parameter, d, can be any real value and consequently the differencing polynomial is defined by an infinite binomial expansion,

$$(1 - B)^d = 1 - dB - \tfrac{1}{2}d(1-d)B^2 - \tfrac{1}{6}d(1-d)(2-d)B^3 - \dots .$$

The effect is that the coefficients diminish very slowly, and thus a autoregressive approximation includes a large number of lags. Granger and Joyeux (1980) and Hosking (1981) show that if $|d| < \frac{1}{2}$ the process is stationary and invertible with finite variance. Hence, if $d < \frac{1}{2}$ the process is stationary with an infinite moving average representation and if $d > -\frac{1}{2}$ the process is invertible with an infinite autoregressive representation. The autocorrelation for a fractional noise series is then defined as

$$\rho_i = \frac{\gamma_i}{\gamma_0} = \frac{\Gamma(1-d)\,\Gamma(i+d)}{\Gamma(d)\,\Gamma(i+1-d)},$$

where $\Gamma(\cdot)$ is the gamma function. Consequently, if the differencing parameter is positive, the autocorrelations are positive and the ACF declines at a slow hyperbolical rate; compare with the exponential decay of an AR(1). When $d \geq \frac{1}{2}$ the process has infinite variance and is thus non-stationary but it will nevertheless be mean reverting. If $d = 0$ the process collapses to white noise.

The fractional integrated model has been used in various applications, as a small sample; in macroeconomics to model persistence in GNP see Diebold and Rudebusch (1989) and Sowell (1992), modelling of long term volatility in the financial market see Ding, Granger and Engle (1993), testing for long memory in stock market see Lo (1991). For a broader survey see for instance Baillie (1996).

1.2.2 Cointegrated vector autoregressive models

Multivariate time series models are important tools for economic analysis to day, mainly due to innovations in multivariate analysis of long-run economic relationships. Since Sims (1980), the basic model used in multivariate econometrics has been a stationary vector autoregressive model with independent Gaussian errors. This model is in itself rather restricted for economic analysis but it yields straightforward statistical analysis, thanks to the relatively easy likelihood function. Furthermore, the statistical analysis can be done in similar ways for a variety of restrictions and generalizations. One of the most important generalizations is the possibility to analyze stationary and non-stationary variables in the same model, in order to describe the long-run relations as well as the short-run dynamics. For this purpose maximum likelihood cointegration analysis has become increasingly popular in applied work. 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.

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

$$\Pi(L)\, x_t = \varepsilon_t, \qquad (t = 1, 2, \ldots, T), \tag{1.3}$$

where $\Pi(L)$ is an $n \times n$ matrix polynomial of order p given by $\Pi(\lambda) = I_n - \sum_{j=1}^{p} \Pi_j \lambda^j$, 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 *iid* $N_n(0, \Sigma)$.

A slight reparameterization of (1.3) yields a vector error correction, VECM, representation for x_t suitable for estimation of the cointegrating relationships. Letting $\Gamma(\lambda) = I_n - \sum_{i=1}^{p-1} \Gamma_i \lambda^i$ where $\Gamma_i = -\sum_{j=i+1}^{p} \Pi_j$ and $\alpha\beta' = \Pi = -\Pi(1)$ we get

$$\Gamma(L)\Delta x_t = \alpha\beta' x_{t-1} + \varepsilon_t, \qquad (t = 1, 2, \ldots, T), \tag{1.4}$$

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 (1.4) 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 (1.4) 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

$$W_{LR,1} = T \sum_{i=1}^{r} \ln \left[\frac{\left(1 - \widehat{\lambda}_{H,i}\right)}{\left(1 - \widehat{\lambda}_i\right)} \right], \tag{1.5}$$

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

$$\begin{aligned} W_{LR,2} &= T \left[\sum_{i=1}^{r_1} \ln\left(1 - \widehat{\lambda}_{C.H,i}\right) \right. \\ &\qquad \left. + \sum_{i=1}^{r_2} \ln\left(1 - \widehat{\lambda}_{H,i}\right) - \sum_{i=1}^{r} \ln\left(1 - \widehat{\lambda}_i\right) \right], \qquad (1.6) \end{aligned}$$

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 solution is found by an iterative algorithm, see Johansen (1995), which gives the test statistic as

$$\begin{aligned} W_{LR,3} &= T \left[\sum_{i=1}^{r_1} \ln\left(1 - \widehat{\lambda}_{C.H,i}\right) \right. \\ &\qquad \left. + \sum_{i=1}^{r_2} \ln\left(1 - \widehat{\lambda}_{H,i}\right) - \sum_{i=1}^{r} \ln\left(1 - \widehat{\lambda}_i\right) \right], \qquad (1.7) \end{aligned}$$

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 easily 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.

Empirical applications of cointegrated relations are presently increasing in numbers, the Swedish contributions in this topic can be exemplified with this sample of studies; Vredin and Warne (1991) studied the correlation between balance of current account and savings/investments, Mellander, Vredin and Warne (1992) examined if the growth rate in consumption, investment, output and terms of trade can be explained by technology innovations, Jacobson, Vredin and Warne (1997) discuss the hysteresis effect in Scandinavian unemployment, Jacobson, Vredin and Warne (1998) analyzed the relation between Swedish real wage and unemployment in different time frames, Jacobson and Ohlsson (1994) studied private and public wages setting in Sweden, Jacobson and Ohlsson (1996) examined the potential of work-sharing as a policy tool with respect to unemployment, Nessén (1996) tests long-run purchasing power parity, Becker (1997) treats Ricardian equivalence, and investigates how expected changes in the budget of the US public sector influence the private consumption, with cointegrated vectors based on intertemporal budget restrictions for the public and the private sector, Becker and Palzov (1997) analyzes similar problems on Swedish data and Bergman (1995), finally, implement a comparative analysis with respect to output and inflation for five countries, Sweden, Great Britain, USA, Japan and Germany.

1.2.3 Fractional cointegration

In the common cointegrated setting, discussed in previous section, a time series process $\mathbf{x}_t$ is said to be cointegrated of order $CI\left(\delta, b\right)$ if the variables of $\mathbf{x}_t$ individually are integrated of order $I\left(\delta\right)$, while a linear combination of the variables, denoted $\beta'\mathbf{x}_t$, is $I\left(\delta - b\right)$. The variables are in equilibrium if $\beta'\mathbf{x}_t$ equals some constant $\boldsymbol{\mu}$, but in most time periods $\mathbf{x}_t$ is not in equilibrium and the quantity $\mathbf{z}_t = \beta'\mathbf{x}_t$ may be called the equilibrium error. Commonly in applied work δ and b are both equal to unity, but there are examples of analysis of $I\left(2\right)$ processes.

The $\delta = b = 1$ case is appealing to empirical economists since it provides a framework to estimate long-run steady states, given by economic theory, using stationary linear combinations of nonstationary variables. However, the notion of cointegration may be generalized to real values, that is allowing for fractional δ and b. The theory of fractional integration was introduced by Granger and

Joyeux (1980) and Hosking (1981), and is considered when modelling persistence in time series, se introduction in section 1.2.1. The distinction between $I(0)$ and $I(1)$ is rather arbitrary; the relevant concept is mean-reversion in the equilibrium error. Mean-reversion does not require a strictly $I(0)$ process; the effect of a shock also dies out, although at a slow hyperbolic rate, for an $I(d)$ process with $d < 1$. Moreover, a similar interpretation as in the $\delta = b = 1$ case is possible within the fractional framework. If $\delta > 1/2$ and $\delta \geq b > \delta - 1/2$ the variables are non-stationary (their variances are infinite) but there exists a stationary linear combination (a long-run relationship) of the variables.

Fractional cointegration is presently a rather unknown topic, hence empirical applications are rare. Still, one example is Cheung and Lai (1993), who applied the theory of fractional cointegration on the long-run purchasing power parity hypothesis, in an attempt to capture a wider range of mean-reverting behavior than standard cointegration analysis can analyze.

2

Summary of Papers

The thesis begins by discussing time series econometrics in a univariate setting and further on extends the framework to a multivariate. The common topic in both frameworks is the analyzis of long-run properties in economic data. In the univariate framework Paper 1 and 2 focus on how to improve testing for persistence in data, through a bootstrap procedure. Paper 3-5 are studies of multivariate time series models. Paper 3 consider the identification problem of the basic stationary vector autoregressive model, which is the baseline econometric specification for maximum likelihood cointegration analysis. In paper 4 the multivariate framework is expanded to allow for components of different integrating order and in this setting the paper discusses how fractional cointegration affects the inference in maximum likelihood cointegration analysis. The last paper considers once again the bootstrap testing approach, now in a multivariate application, to correct inference on long-run relations in maximum likelihood cointegration analysis.

Paper 1-2: The fractionally integrated autoregressive moving average processes generalize linear ARIMA models by allowing for non-integer differencing powers and thereby provide a more flexible framework for analyzing persistence in time series data. This flexibility enables fractional processes to model stronger data dependence than that allowed in stationary ARMA models without resorting to non-stationary unit-root processes. However, estimators of the fractional model exhibit larger bias and are computationally more demanding. It is therefore beneficial to discriminate fractionally integrated processes from ARIMA specifications in a first modelling step, that is to test the null-hypothesis of an integer differencing power against a fractional alternative. For this purpose the literature frequently utilizes the Geweke and Porter-Hudak (1983) test, the modified rescaled range test of Lo (1991) and Lagrange multiplier tests, see e.g. Agiakloglou and Newbold (1994). The size and power of these asymptotic tests are investigated by Cheung (1993) and Agiakloglou and Newbold. One finding in their studies is the existence of non-negligible small-sample size distortions.

To improve inference, classical statistical theory employs expansions to provide analytical corrections. By numerical means, similar corrections can be given by bootstrap methods. While analytical corrections modify the test statistic to approach the asymptotic distribution more rapidly, the bootstrap adjusts the critical values so that the true size of the test converges to its nominal value.

The first paper applies a parametric bootstrap testing procedure. In the paper it is shown that for normally distributed autoregressions and moving averages processes the bootstrap testing procedure has better size properties than the original tests, that is the bootstrap corrects existing size distortions without introducing new ones. All bootstrap tests are close to exact, with an exception for MA(1) processes with a large positive root.

In the second paper the investigation of improving tests for fractional integration is extended to consideration of characteristic features such as non-normality (i.e. with excess skewness and kurtosis) and conditional heteroskedasticity. Empirical evidence shows that many financial time series display such characteristic features, for instance Ding, Granger and Engle (1993) report evidence of autocorrelations between distant lags for long lags in the absolute returns of the Standard and Poor 500, S&P500, composite stock index. Furthermore, Granger and Ding (1995) show that the absolute value of the rate of return for a variety of stock prices, commodity prices and exchange rates exhibit excess skewness and kurtosis.

The previous study implemented a bootstrap method, in order to size-adjust fractional integration tests. Again the bootstrap is used to correct for size distortions. Furthermore, a comparison of parametric, non-parametric and heteroscedasticity invariant resampling algorithms is also done. The aim of the paper is to find tests that are robust to non-normalities and ARCH effects in data, and thus are suitable when testing for long-memory in financial and economic time series. The results suggest that the performance of a bootstrap testing procedure depends to some extent on the chosen resampling algorithm. However, all (but one) bootstrap tests are superior to the original version of the tests, in the sense that the bootstrap tests have better size properties. The main conclusions are that the bootstrap tests are remarkably well-sized (whereas the originals are not) and robust to non-normalities and ARCH effects, and that reliable testing for fractional integration in many cases requires a bootstrap test, or some other refinement.

Paper 3: It is well known that inference in vector autoregressive models depends crucially on the choice of lag-length. Various lag-length selection procedures have been suggested and evaluated in the literature. In these evaluations the possibility that the true model may have unequal lag-length has, however, received little attention. In practice it is not known if the true model has equal lag-length or not and there is little evidence in the literature of the losses made by falsely assuming an equal lag-length or using an "unnecessarily" general procedure when the assumption holds. Using an equal lag-length procedure and consistent information criteria will, of course, asymptotically select the true or maximum lag-length. This is, however, less clear with the sample sizes common in empirical work. This paper extends the existing literature by investigating the performance of equal and unequal lag-length selection procedures when the true model has equal, as well as unequal, lag-length. In doing this we consider a variety of data generating processes designed to replicate salient features of macroeconomic data.

Choosing the lag-length in VAR models is not an easy task and the cost of using a mis-specified model depends on the intended use of the model. If inference and hypothesis testing is the primary concern it is important to avoid under-parameterization. In a forecasting application, on the other hand, the bias introduced by under-parameterization may well be offset by the reduction in the variance of the estimates.

When choosing a lag-length selection strategy it is wise to keep these issues in mind. The BIC criterion is highly parsimonious and underestimates the true lag-length in almost all cases considered here, even for large sample sizes. The BIC criteria may thus be useful in forecasting applications, but should be avoided when inference is the primary concern. The AIC and HQ criteria are less parsimonious and estimate the correct lag-length better than BIC. AIC underestimates the lag-length less frequently than HQ and may thus be the preferred criteria when performing inference.

The main concern of this paper is the distinction between lag-length selection strategies which assume equal lag-lengths and procedures which allow for different lag-lengths.

In smaller Models, in terms of the number of parameters, the difference in performance between the equal and unequal lag-length procedures is small.

For the larger models, the equal lag-length procedure performs reasonably well and selects the maximum lag-length correctly in a majority of the replicates when AIC or HQ is used. The difference in performance between the equal and unequal lag-length models is very small for the equal lag-length procedure.

It is thus difficult to give clear guidelines for the choice between equal and unequal lag-length procedures. The performance seems to be highly model dependent. A tentative conclusion is that the equal lag-length procedure works well with equal lag length models as well as unequal-lag length models with a relatively simple lag structure. For models with more complicated lag structures, with holes in the lag polynomials, the unequal lag-length procedure may be a better choice.

Paper 4: This paper investigates the maximum likelihood cointegration procedure when the process is fractionally cointegrated. For the sake of comparison, the common case of cointegrated series is also included in the study. The results suggest that the likelihood ratio test for cointegration also has high power against fractional alternatives, and hence possesses the ability to detect slow mean-reversion in the equilibrium error. However, if fractional cointegration is present, the usual maximum likelihood procedure may lead to incorrect inference since persistence in the equilibrium error will then be modelled by an $I(0)$ instead of an $I(d)$ specification.

The high power of the LR test against fractional alternatives and the severely biased ML estimates under fractional cointegration suggest that the standard likelihood based approach should be used with caution. In particular, if the equilibrium error is likely to be ruled by persistence we recommend that a secondary test should be used to separate fractionally cointegrated series from series that are cointegrated of an integer order. This may be conducted by the Engle-Granger procedure combined with the bootstrap tests as in paper 1-2, which are robust to AR and MA components.

Paper 5: 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 in the sense that test statistics are 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 in small sample. Possibly due to the very simple Data Generating Processes (DGP's) considered 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. Their results indicate that inference based on the asymptotic approximation of a χ^2-distribution, for a reasonably large empirical sample, can be severely misleading.

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.

This paper proposes use of bootstrap hypothesis testing as a tractable way to improve inference for linear restrictions. A numerical method to correct the limit distribution, that in fact can be expressed and interpreted in terms of Edgeworth expansions as shown by 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.

The Monte Carlo experiments is based on a complex data generating processes as given by the empirical monetary vector error correction model estimated in Juselius (1997). One purpose of the analyses 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.

From a response surface regression analysis it is clear that the asymptotic distribution is not a satisfactory approximation for finite samples. 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 deviations from the asymptotic distribution are even larger.

The results in the 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.

To further show the implication of size distortions an empirical application demonstrates how inference about long-run economic relationships may shift when asymptotic tests are substituted for bootstrap analogues. This is done through a re-evaluation of the asymptotic tests of the hypotheses of linear restrictions on cointegrating relations in Juselius (1997).

References

AGIAKLOGLOU, C. and P. NEWBOLD (1994) Lagrange Multiplier Tests for Fractional Difference. *Journal of Time Series Analysis* 15 (3), 253-262.

BAILLIE, R.T. (1996) Long Memory Processes and Fractional Integration in Econometrics. *Journal of Econometrics* 73, 5-59.

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

BECKER T. (1997) Investigating Ricardian Equivalence in a Common Trends Model. forthcoming, *Journal of Monetary Economics.*

BECKER T. and A. PALZOW (1997) Real Effects of Budget Deficts - Theory and Evidence. forthcoming, *Swedish Economic Policy Review.*

BERGMAN M. (1995) International Evidence on the Source of Macroeconomic Fluctuations. *European Economic Review* 40, 1237-1258.

CHEUNG, Y.-W. (1993) Tests for Fractional Integration: A Monte Carlo Investigation. *Journal of Time Series Analysis* 14 (4), 331-345.

CHEUNG, Y.-W. and K.S. LAI (1993) A fractional Cointegration Analysis of Purchasing Power Parity, *Journal of Business and Economic Statistics* 11, 103-112.

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.

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

DIEBOLD, F.X. and G.D. RUDEBUSCH (1989) Long Memory and Persistence in Aggregate Output. *Journal of Monetary Economics 24*, 189-209.

DING, Z.,C.W.J. GRANGER and R.F. ENGLE (1993) A Long Memory Property of Stock Return and a New Model. *Journal of Empirical Finance* 1, 83-106.

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

FIELD, C. and E. RONCHETTI (1990) *Small Sample Asymptotics.* Institute of Mathematical Statistics, Lecture Notes-Monograph series 13.

GEWEKE, J. and S. PORTER-HUDAK (1983) The Estimation and Application of Long Memory Time Series Models. *Journal of Time Series Analysis* 4 (4), 221-238.

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.

GRANGER, C.W.J. and R. JOYEUX (1980) An Introduction to Long-Memory Time Series Models and Fractional Integration. *Journal of Time Series Analysis* 1 (1), 15-29.

GRANGER, C.W.J. and Z. DING (1995) Stylized Facts on the Temporal and Distributional Properties of Daily Data from Speculative Markets, *Unpublished manuscript,* Department of Economics, University of California, San Diego.

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

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

HJORTH, U.S.J. (1994) *Computer Intensive Statistical Methods.* Chapman and Hall, London.

HOROWITZ, J.L. (1995) Bootstrap Methods in Econometrics: Theory and Numerical Performence. *Paper presented at the 7th World Congress of Econometric Society,* Tokyo.

HOSKING, J.R.M. (1981) Fractional Differencing. *Biometrica* 68, 165-176.

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

JACOBSON T. and R. LARSSON (1996) Bartlett Correction for a Likelihood Ratio Cointegration Test. Mimeo, Department of Economic Statistic, Stockholm School of Economics.

JACOBSON T., A. VREDIN and A. WARNE (1997) Common Trends and Hysteresis in Scandinavian Unemployment. forthcoming *European Economic Review.*

JACOBSON, T.,VREDIN, A. and WARNE, A. (1998). Are Real Wages and Unemployment Related?. *Economica,* forthcomimg.

JACOBSON T. and H.OHLSSON (1994) Long-Run Relations Between Private and Public Sector Wages in Sweden. *Empirical Econometrics* 19, 343-360.

JACOBSON T. and H.OHLSSON (1996) Working Time, Employment, and Worksharing: Evidence from Sweden. Working Paper No. 135, Stockholm School of Economics.

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

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

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

JOHANSEN, S. and JUSELIUS, K. (1992) Testing structural hypotheses in a multivariate cointegration analysis of the PPP and UIP for UK. *Journal of Econometrics* 53, 211-244.

KUNSCH, H.R. (1989) The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics* 17, 1217-1241.

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

LO, A.W. (1991) Long Term Memory in Stock Market Prices. *Econometrica* 59 (5), 1279-1313.

MELLANDER, E., A. VREDIN and A. WARNE (1992) Stochastic Trend and Economic Fluctuations in a Small Open Economy. *Journal of Applied Econometrics* 7, 369-394.

NESSEN, M. (1996) Trends in Prices and Exchange Rates: Test of Long-Run Purchasing Power Parity. *Empirical Econometrics* 21, 381-400.

VREDIN, A. and A. WARNE, (1991) Current Account and Macroeconomic Fluctuations. *Scandinavian Journal of Econometrics* 93, 511-530.

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

SIMS, C.A. (1980) Macroeconomics and Reality. *Econometrica* 48, 1-48.

SOWELL, F.B. (1992) Modelling Long Run Behavior with a Fractional ARIMA Model. *Journal of Monetary Economics* 29, 277-302.

Part II

Included papers

3

Bootstrap Testing for Fractional Integration

ABSTRACT: Asymptotic tests for fractional integration, such as the Geweke-Porter-Hudak test, the modified rescaled range test and Lagrange multiplier type tests, exhibit size distortions in small samples. This paper investigates a parametric bootstrap testing procedure, for size correction, by means of a computer simulation study. The bootstrap provides a practical method to eliminate size distortions in the case of an asymptotic pivotal statistic while the power, in general, is close to the corresponding size-adjusted asymptotic test. The results are very encouraging and suggest that a bootstrap testing procedure does correct for size distortions.

KEY WORDS: Long-memory; ARFIMA; Parametric resampling; Small-sample; Monte Carlo simulation; Size correction
JEL-CLASSIFICATION: C15; C22; C52

3.1 Introduction

The fractionally integrated autoregressive moving average (ARFIMA) model has recently received considerable attention in economics, and also in other research areas. ARFIMA processes generalize linear ARIMA models by allowing for non-integer differencing powers and thereby provide a more flexible framework for analyzing time series data. This flexibility enables fractional processes to model stronger data dependence than that allowed in stationary ARMA models without resorting to non-stationary unit-root processes. However, estimators of the fractional model exhibit larger bias and are computationally more demanding. It is therefore beneficial to discriminate fractionally integrated processes from ARIMA specifications in a first modelling step, that is to test the null-hypothesis of an integer differencing power against a fractional alternative. For this purpose the literature frequently utilizes the Geweke and Porter-Hudak (1983) test, the modified rescaled range test of Lo (1991) and Lagrange multiplier tests, see e.g. Agiakloglou and Newbold (1994). The size and power of these asymptotic tests are investigated by Cheung (1993) and Agiakloglou and Newbold. One finding in their studies is the existence of non-negligible small-sample size distortions.

To improve inference, classical statistical theory employs expansions to provide analytical corrections. By numerical means, similar corrections can be given by bootstrap methods. While analytical corrections modify the test statistic to approach the asymptotic distribution more rapidly, the bootstrap adjusts the critical values so that the true size of the test converges to its nominal value. This paper applies a parametric bootstrap testing procedure. According to Davidson and MacKinnon (1996a), the size distortion of such a procedure, based on parameter estimates under the null, will be at least one full order, $O(T^{-1})$, smaller than the distortion of the asymptotic test. Thus, the bootstrap is said to provide a trustworthy technique to perform inference in small samples and yields, under regularity conditions, exact or close to exact tests. The purpose of this paper is to examine this claim.

The paper is organized as follows. Section 2 briefly describes the tests and Section 3 introduces the bootstrap testing procedure. Section 4 contains the Monte Carlo simulation study, where the size and power of the tests are compared with their bootstrap analogues. Section 5 concludes the paper.

3.2 Testing for Fractional Integration

A time series $\{x_t\}$ follows an ARFIMA(p, d, q)[1] process if

$$\phi(B)(1-B)^d x_t = \theta(B) a_t, \tag{3.1}$$

where $\{a_t\}$ is a series of independently and identically distributed disturbances with mean zero and variance $\sigma_a^2 < \infty$, and $\phi(B)$ and $\theta(B)$ are the autoregressive and moving average polynomials in the backshift operator B. If the roots of $\phi(B)$ and $\theta(B)$ are outside the unit circle and $d < |0.5|$, x_t is both stationary and invertible. When $d > 0$, x_t is persistent in the sense that the autocorrelations are not absolutely summable. Thus there exists a region $(0 < d < 0.5)$ where the ARFIMA model is capable of generating stationary series which are persistent.[2] If $d \neq 0$ the process displays long-memory characteristics, such as a hyperbolic autocorrelation decay, while the stationary ARMA model exhibits a faster geometrical decay (given the existence of AR parameters).

If d is integer-valued, the ARFIMA process reduces to an ARIMA process. The tests are applicable on stationary and invertible series, and the series are subsequently differenced or summed until this is satisfied. $d = 0$ is thus a natural null-hypothesis when testing for fractional integration.

3.2.1 The Periodogram Regression Test of Geweke and Porter-Hudak

Geweke and Porter-Hudak (1983), henceforth referred to as *GPH*, proposed the following non-parametric periodogram regression:

$$\ln\{I_x(\omega_j)\} = \alpha - d\ln\{4\sin^2(\omega_j/2)\} + v_j, \tag{3.2}$$

for the estimation of the fractional difference parameter. $I_x(\omega_j)$ is the periodogram at the harmonic frequencies $\omega_j = 2\pi j/T$, where $j = 1, ..., g(T)$. With a proper choice of $g(T)$, the ordinary least squares (OLS) estimator of d is consistent and the distribution of

---

[1] The properties of the fractionally integrated ARMA model are presented by Granger and Joyeux (1980) and Hosking (1981).

[2] Persistence is commonly found in Economic time series, i.e. real exchange rates and unemployment.

$\left(\hat{d}_{OLS} - d\right) / SE\left(\hat{d}_{OLS}\right)$ is asymptotically normal. The known variance of v, $\pi^2/6$, is used to increase the efficiency of the test and $g\left(T\right)$ is commonly selected as $T^{1/2}$.

3.2.2 The Modified Rescaled Range Test

The rescaled range statistic was proposed by Hurst (1951) and has been refined by Mandelbrot (1972) and MacLeod and Hipel (1978). A version of the statistic, which is robust to short-range dependence in data, was suggested by Lo (1991). This modified rescaled range (*MRR*) statistic is defined by the ratio

$$\tilde{Q}_T = \frac{R_T}{\hat{\sigma}_T\left(k\right)}, \tag{3.3}$$

where the range and standard error are calculated by

$$\begin{aligned} R_T &= \max_{0<i\leq T}\sum_{t=1}^{i}\left(x_t - \overline{x}\right) - \min_{0<i\leq T}\sum_{t=1}^{i}\left(x_t - \overline{x}\right) && (3.4) \\ \hat{\sigma}_T^2\left(k\right) &= \hat{\sigma}^2 + 2\sum_{j=1}^{k}\sum_{i=j+1}^{T}\left(1 - \frac{j}{k+1}\right)\left(x_i - \overline{x}\right)\left(x_{i-j} - \overline{x}\right) && (3.5) \end{aligned}$$

The truncation lag k depends on the short-term correlation structure of the series and is set, according to Andrews' (1991) data dependent formula, to the integer part of $\left(3T/2\right)^{\frac{1}{3}}\left\{2\hat{\rho}/\left(1-\hat{\rho}^2\right)\right\}^{\frac{2}{3}}$, where $\hat{\rho}$ denotes the sample first-order autocorrelation coefficient and $\hat{\sigma}^2$ the maximum likelihood variance estimate. Asymptotic critical values of the MRR test are given by Lo (1991).

3.2.3 A Lagrange Multiplier Test

The LM test, denoted *REG*, of Agiakloglou and Newbold (1994) is carried out through the likelihood based auxiliary regression

$$\hat{a}_t = \sum_{i=1}^{p}\beta_i W_{t-i} + \sum_{j=1}^{q}\gamma_j Z_{t-j} + \delta K_m + u_t, \tag{3.6}$$

where

$K_m = \sum_{j=1}^{m} j^{-1}\hat{a}_{t-j}$, $\hat{\theta}\left(B\right)W_t = x_t$, $\hat{\theta}\left(B\right)Z_t = \hat{a}_t$ and u_t is *iid* normal.

$\hat{a}_t$ and $\hat{\theta}(B)$ are the estimated residual and MA polynomial from the ARFIMA specification (3.1) under the null-hypothesis.

The autoregressive and moving average orders p and q are estimated by the Bayesian information criterion (BIC) of Schwartz (1978). According to Agiakloglou and Newbold a small value of the truncation lag m is preferable, therefore m is set equal to five. The equation (4.12) is fitted by non-linear least squares (the $IMSL$ routine DNSLSE) over the time period $t = m+1, ..., T$. The usual t-test of the hypothesis $\delta = 0$ together with asymptotically normal critical values constitutes the LM test.

3.2.4 The Size and Power of the GPH, MRR and REG tests

Cheung (1993) presents size and power for the MRR and GPH tests. This is done for a variety of AR(1), MA(1) and ARFIMA(0,d,0) processes with positive and negative parameter values. The MRR test is conservative for autoregressions, that is the empirical size is smaller than the nominal, for almost every parameter value and serial length. For large positive AR parameters, the GPH test is severely over-sized, whereas it is well-sized for the remaining parameter values. Rejection frequencies of both the MRR and GPH are notably larger than the nominal significance level when the MA parameter is close to -0.9.

The empirical size of the REG test is similar to the asymptotic size according to Agiakloglou and Newbold (1994). In contrast to the thorough investigation of the MRR and GPH tests, the size of the REG test is only computed for $\phi = 0.5$ and 0.9, and $T = 100$. Under the unrealistic assumption of a known AR order, the REG test exhibits high rejection frequencies when the true process is fractionally integrated. A lower power is expected when the lag-order is unknown. The MRR test has difficulties in detecting positive fractional integration, especially in moderate sample sizes. Independently of the serial length, the GPH test displays a low rejection frequency for weakly persistent processes. Our study confirms these conclusions and extends them for the REG test.

3.3 The Bootstrap Test

The finite-sample distribution of a test statistic may not always coincide with its asymptotic distribution. One feasible way to estimate the small-sample distribution is through a bootstrap procedure, see for instance Horowitz (1995) for an introduction and overview. The size distortion of a bootstrap test is of an order $T^{-1/2}$ smaller than that of the corresponding asymptotic test. A further refinement of an 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. This is achieved without the complex derivations of analytical higher order expansions. If the significance level of a test is calculated using a bootstrap procedure, an exact or close to exact test is often the result, which enables more reliable inference in finite samples. Following Davidson and MacKinnon (1996a), such a procedure will be referred to as a bootstrap test.

The objective of the test is to compute the p-value function,

$$p(\hat{\tau}) = p(\tau \geq \hat{\tau} \,|\Psi_0, T), \tag{3.7}$$

where Ψ_0 is the true data generating process (*DGP*) under the null hypothesis, T is the sample size and $\hat{\tau}$ is a realized value of the test statistic τ based on the original sample $\mathbf{x} = [x_1, ..., x_T]'$. The DGP Ψ_0 is characterized by an unknown ARMA(p, q) specification. Since the null model, and hence Ψ_0, is unknown the estimated (bootstrap) DGP $\hat{\Psi}_0$ is employed to create the bootstrap samples. The basic idea is to create a large number of such samples which all obey the null-hypothesis and, as far as possible, resemble the original sample.

In this paper we use a parametric bootstrap algorithm[3], for which the DGP $\hat{\Psi}_0$ is based on parameter estimates under the null, that is retrieving $\hat{\Psi}_0$ from the estimated ARMA$(\hat{p}, \hat{q})$ model,

$$\left(1 - \hat{\phi}_1 B - ... - \hat{\phi}_{\hat{p}} B^{\hat{p}}\right) x_t = \left(1 + \hat{\theta}_1 B + ... + \hat{\theta}_{\hat{q}} B^{\hat{q}}\right) \hat{a}_t, \tag{3.8}$$

where $\hat{a}_t$ is the residual at time t. Alternatively, the re-sampling model may be estimated by

$$\left(1 - \tilde{\phi} B - ... - \tilde{\phi}_{\tilde{p}} B^{\tilde{p}}\right) x_t = \tilde{a}_t, \tag{3.9}$$

---

[3] The use of a parametric bootstrap is motivated by the assumed normality of the data. Further resampling procedures are evaluated by Andersson and Gredenhoff (1998).

which can be regarded as the estimated AR representation of the bootstrap DGP. The models (3.8) and (3.9) are estimated, conditional on stationarity and invertibility conditions, by the BIC and non-linear least squares and OLS (the *IMSL* routines DNSLSE and DRLSE) respectively. The orders p and q are allowed to a maximum lag of five for the ARMA model, whereas a maximum lag p of 30 is allowed for the AR specification.[4]

The bootstrap samples, each denoted $\mathbf{x}_r^*, r = 1, ..., R$, are created recursively using

$$x_{r,t}^* = \phi^* \left(B\right)^{-1} \theta^* \left(B\right) a_t^*, \tag{3.10}$$

where $\phi^* \left(B\right)$ and $\theta^* \left(B\right)$ are the estimated polynomials of $\hat{\Psi}_0$. The values for $\{a_t^*\}$ are independent draws from a normal distribution with mean zero and variance $s_{\hat{a}}^2$ or $s_{\tilde{a}}^2$.

If R bootstrap re-samples, each of size T, and their respective test statistics τ_r^* are generated, the estimated bootstrap p-value function, for a two-sided test, is defined by

$$p^* \left(\hat{\tau}\right) = R^{-1} \sum_{r=1}^{R} I\left(\left|\tau_r^*\right| \geq \left|\hat{\tau}\right|\right), \tag{3.11}$$

where $I\left(\cdot\right)$ equals one if the inequality is satisfied and zero otherwise, and the number of bootstrap replicates R is chosen as 1000. The null hypothesis is rejected when the selected significance level exceeds $p^* \left(\hat{\tau}\right)$.

Davidson and MacKinnon (1996b) show that the power of a bootstrap test, based on a pivotal statistic, is generally close to the size-adjusted asymptotic test. Even if the statistic is only close to pivotal this is generally true.

3.4 The Monte Carlo Study

The experiment covers first order autoregressions and moving averages, and fractional noise series of lengths $T = 50$, 100, 300 and 500. We generate $T + 100$ normally distributed pseudo random numbers, using the *IMSL* routine DRNNOA, and discard the first 100 observations to reduce the effect of initial values. The AR and MA series

---

[4] Preliminary results suggest that no significant AR parameters enter the estimated polynomial after the $30th$ lag when the true process is an MA(1) with $\theta = 0.9$.

are then constructed recursively and the fractional noise series are generated using the algorithm of Diebold and Rudebusch (1991).

The Monte Carlo study involves 1000 replicates (series), where each series is tested for fractional integration using the tests described in Sections 2 and 3.

The bootstrap resamples are created by the ARMA (3.8) and AR (3.9) specifications. Reported results are based on the AR resampling model, due to its better performance. The AR specification works better than a pure MA resampling model even when the true process is a moving average, regardless of parameter values.

Estimated size and power of the different processes in the study are computed as the rejection frequencies of the non-fractional null hypothesis.

3.4.1 AR and MA Processes

The empirical sizes of the tests are examined for the specifications

$$x_t = \phi x_{t-1} + a_t \tag{3.12}$$

and

$$x_t = a_t + \theta a_{t-1}, \tag{3.13}$$

where the members of $\{a_t\}$ are *iid* $N(0,1)$. The AR and MA parameters ϕ and θ are set equal to ±0.1, ±0.5 and ±0.9. Table 3.1[5] presents the sensitivity, at a nominal 5% level of significance, of the empirical size with respect to AR and MA parameters.

The estimated size of the MRR test for both AR and MA processes differs, in general significantly at the 5% level, from the nominal size. Significant differences, based on a 95% acceptance interval, are obtained when the rejection frequencies lie outside $(3.6, 6.4)$. The MRR test is in general over-sized for both AR and MA processes, for large negative parameters, and conservative for large positive parameters. Exactly as in Cheung (1993), AR series with parameter $\phi = -0.5$ lead to very low rejection frequencies. The MRR test is always conservative for autoregressions near the unit circle for larger sample sizes ($T = 300$ and 500), whereas the rejection frequencies increase

---

[5] All results are approximately valid for the 1% and 10% nominal significance level and for T=300 and 500.

TABLE 3.1. Rejection percentage of the nominal 5 percent fractional integration test when the data follow an AR(1) or MA(1) process of length T.

	MRR		GPH		REG	
ϕ/θ	Orig.	Boot.	Orig.	Boot.	Orig.	Boot.
T=50			*AR(1) Processes*			
-0.9	**18.9**	**2.2**	4.7	5.4	**6.9**	4.7
-0.5	**0.9**	**3.1**	4.7	5.3	6.2	4.1
-0.1	**6.7**	4.5	5.3	5.1	5.8	3.9
0.1	5.0	3.8	5.5	4.8	5.7	4.1
0.5	**2.5**	**3.1**	**8.2**	5.1	**6.8**	4.7
0.9	**1.2**	4.1	**63.6**	3.6	**7.8**	4.8
T=100						
-0.9	6.4	**2.1**	6.4	5.6	5.6	4.8
-0.5	**0.8**	5.0	5.7	5.0	5.1	4.6
-0.1	**6.6**	4.8	4.9	4.5	6.2	3.6
0.1	**6.8**	5.9	4.9	5.2	**6.9**	5.1
0.5	**2.3**	4.8	**8.3**	4.7	6.4	4.7
0.9	**0.8**	3.9	**71.8**	3.7	5.1	4.7
T=50			*MA(1) Processes*			
-0.9	4.8	5.4	**41.2**	**25.8**	**9.3**	**9.5**
-0.5	5.5	5.0	**7.7**	5.0	**3.3**	**1.6**
-0.1	**7.3**	4.5	5.0	4.5	**6.6**	5.4
0.1	**6.7**	4.9	4.3	4.0	**7.4**	5.3
0.5	3.7	4.3	4.4	4.8	4.0	4.4
0.9	**2.0**	**3.2**	5.2	3.7	**10.0**	6.3
T=100						
-0.9	**9.9**	**10.5**	**50.1**	**36.3**	**8.0**	**7.2**
-0.5	4.2	4.8	**7.9**	5.6	**2.8**	**2.8**
-0.1	**6.6**	5.1	4.9	5.5	5.6	4.3
0.1	5.4	4.7	5.0	5.4	5.2	4.1
0.5	**3.2**	5.7	5.4	5.7	**2.7**	4.0
0.9	**2.3**	4.8	6.0	4.4	**6.8**	4.4

The number reported in the table is the rejection percentage of the two-sided 5% test. Numbers in bold face denote significant deviations from the nominal size. Under the null hypothesis of no fractional integration, the 95% acceptance interval of the rejection percentage equals (3.6, 6.4). In the table head, Orig. denotes the original test and Boot. the corresponding bootstrap test.

with T for moving averages with $\theta = -0.9$. The GPH test is well-sized, except for highly short-term AR/MA dependent series with positive roots[6]. Agiakloglou *et al.* (1993) show that large positive AR and MA roots bias the periodogram (4.8), resulting in biased estimates of d and hence large test statistics. These results are close to those of Cheung.

Extending the results of Agiakloglou and Newbold (1994), we find the REG test well-sized, compared with the other tests, for the entire AR parameter space when $T = 100$. However, the test is over-sized for $T = 50$ and $|\phi|$ close to unity. This over-sizing tendency, close to the unit circle, is enhanced for moving average processes. This is most pronounced for series of length $T = 50$, where large empirical sizes also occur for small parameters. The performance of the REG test improves with the serial length (considering also $T = 300$ and 500).

The simulation results suggest, in general, that the bootstrap testing procedure is able to improve the tests. Moreover, every bootstrap test has better size properties than any of the original tests. In more detail, the bootstrap MRR test is found conservative when $\phi = -0.9$, whereas $\theta = -0.9$ leads to higher rejection frequencies than the nominal significance level. Over the parameter space, the dispersion of the sizes for the bootstrap test is smaller than that for the original test. The bootstrap procedure improves the MRR test, that is only two out of twelve (AR and MA) empirical sizes differ significantly from the nominal size at sample size 100, compared to nine for the original test.

The size problems encountered by the GPH test for autoregressions are adjusted by the bootstrap procedure. In particular, the bootstrap correction is remarkable for $\phi = 0.9$ processes. The bootstrap is also able to correct for size distortions due to intermediate positive MA roots and can partly adjust the size for large positive roots. The empirical size for $\theta = -0.9$ is unfortunately still very large for the bootstrap GPH test. One might think that this is due to the AR resampling, but the size adjustment is even smaller when using a pure MA resampling. Furthermore, the bootstrap procedure does not impose distortions where the original GPH test is well-sized. In

---

[6] Positive values of ϕ imply positive roots and positive values of θ negative roots.

TABLE 3.2. Rejection percentage of the nominal 5 percent fractional integration test when the data follow an ARFIMA(0,d,0) process of length T=100.

	MRR		GPH		REG	
d	Orig.	Boot.	Orig.	Boot.	Orig.	Boot.
-0.45	14.6	16.5	21.8	21.0	43.9	41.0
-0.25	11.3	13.6	8.3	11.1	27.8	37.6
-0.05	3.7	6.1	3.9	5.3	3.8	6.7
0	5.0	5.3	5.0	5.0	5.0	6.0
0.05	5.8	6.9	4.4	5.3	10.8	8.0
0.25	14.7	14.3	17.0	16.5	35.3	25.9
0.45	4.0	18.6	40.9	11.8	22.5	19.6

See note to Table 3.1. The original tests are size adjusted.

general, the bootstrap GPH test works considerably better than the original test.

The bootstrap procedure corrects the size distortions of the REG test for autoregressive processes, that is the rejection frequencies of the bootstrap REG test are always within the acceptance bounds. The bootstrap also corrects when the process is an MA with $\theta > -0.5$, whereas $\theta \leq -0.5$ processes lead to significant size distortions for all serial lengths. Except for these cases, the bootstrap REG test is correctly (on the 95% level) sized and more robust than its original version, in particular for MA series.

3.4.2 Fractional Processes

The power of the tests against ARFIMA$(0, d, 0)$ is studied using data constructed by

$$(1 - B)^d x_t = a_t, \tag{3.14}$$

where the fractional differencing parameter, d, is set equal to ± 0.05, ± 0.25 and ± 0.45. Table 3.2 presents the power of the tests as a function of d.The simulation results verify that the power of the bootstrap tests are close to the power of the size-adjusted asymptotic tests. We find the dispersion of the different power functions least pronounced for the REG test, which relates to the small size improvements of the bootstrap. The REG tests and the original MRR test reduce in power when the true d is close to 0.5 compared to a slightly lower d value, which is not the case for the bootstrap MRR.

When specifying the auxiliary regression (4.12) for the REG test, a large true fractional differencing power is interpreted as a large autoregressive order, yielding decreased rejection frequencies for the test. For the MRR test, the truncation lag k in (4.9) increases with d, i.e. too many autocorrelations are included in the variance correction term (4.11), resulting in a negatively biased estimate of the d parameter which lowers the power of the original test.

A substantially lower power is found for the bootstrap GPH compared to its original version when $d = 0.45$. A large differencing power results in a rich parameter structure of the ARMA resampling model in the bootstrap procedure. The rich parametrization implies that the resample periodograms resemble the periodogram of the original, highly persistent process. Thus, the bootstrap GPH test will have difficulties in distinguishing fractional processes from ARMA specification, which can be seen in Table 3.2.

The power properties suggest that the REG test is superior when testing for fractional integration in small samples. In larger samples the MRR test is more powerful when $d < 0$, and the GPH test is more powerful when $d > 0$.

3.5 Conclusions

The bootstrap testing procedure has better size properties than the original tests, that is the bootstrap corrects existing size distortions without introducing new ones. All bootstrap tests are close to exact on the 95% acceptance level, with an exception for MA(1) processes with a large positive root.

In general, the power of the bootstrap tests are close to the power of the corresponding size-adjusted asymptotic tests. The REG test is the most powerful test in small samples and by using the bootstrap version we get a test which is robust to ARMA components and has power properties similar to those of the original test. In larger samples the bootstrap MRR and GPH have higher power, when the alternative hypothesis is one-sided.

We conclude that a bootstrap testing procedure provides a practical and effective method to improve existing tests for fractional integration.

Acknowledgments

Valuable discussions with and comments by Tor Jacobson, Sune Karlsson and Rolf Tschernig are gratefully acknowledged. We also thank participants at the International Symposium on Forecasting (ISF '97), Bridgetown meeting. The usual disclaimer applies.

References

AGIAKLOGLOU, C., P. NEWBOLD and M. WOHAR (1993) Bias in an Estimator of the Fractional Difference Parameter. *Journal of Time Series Analysis* 14 (3), 235-246.

AGIAKLOGLOU, C. and P. NEWBOLD (1994) Lagrange Multiplier Tests for Fractional Difference. *Journal of Time Series Analysis* 15 (3), 253-262.

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, Dept. of Economic Statistics, Stockholm School of Economics, Sweden.

ANDREWS, D. (1991) Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* 59, 817-858.

CHEUNG, Y.-W. (1993) Tests for Fractional Integration: A Monte Carlo Investigation. *Journal of Time Series Analysis* 14 (4), 331-345.

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

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

DIEBOLD, F. X. and G. D. RUDEBUSCH (1991) On the Power of Dickey-Fuller Tests against Fractional Alternatives. *Economics Letters* 35, 155-160.

GEWEKE, J. and S. PORTER-HUDAK (1983) The Estimation and Application of Long Memory Time Series Models. *Journal of Time Series Analysis* 4 (4), 221-238.

GRANGER, C. W. J. and R. JOYEUX (1980) An Introduction to Long-Memory Time Series Models and Fractional Integration. *Journal of Time Series Analysis* 1 (1), 15-29.

HOROWITZ, J. L. (1995) Bootstrap Methods in Econometrics: Theory and Numerical Performance. Paper presented at the 7th World Congress of the Econometric Society, Tokyo.

HOSKING, J. R. M. (1981) Fractional Differencing. *Biometrica* 68, 165-176.

HURST, H. E. (1951) Long Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers* 116, 770-799.

LO, A. W. (1991) Long Term Memory in Stock Market Prices. *Econometrica* 59 (5), 1279-1313.

MACLEOD, A. I. and K. W. HIPEL (1978) Preservation of the Rescaled Adjusted Range, Part I, A Reassessment of the Hurst Phenomenon. *Water Resources Research* 14 (3), 491-518.

MANDELBROT, B. B. (1972) Statistical Methodology for Non-Periodic Cycles: From the Covariance to R/S Analysis. *Annals of Economic and Social Measurements* 1, 259-290

SCHWARTZ, G. (1978) Estimating the order of a model. *The Annals of Statistics* 2, 461-464

4

Robust Testing for Fractional Integration using the Bootstrap

ABSTRACT: Asymptotic tests for fractional integration are usually badly sized in small samples, even for normally distributed processes. Furthermore, tests that are well-sized under normality may be severely distorted by non-normalities and ARCH errors. This paper demonstrates how the bootstrap can be implemented to correct for such size distortions. It is shown that a well-designed bootstrap test based on the MRR and GPH tests is exact, and a procedure based on the REG test is nearly exact.

KEY WORDS: Long-memory; Resampling; Skewness and kurtosis; ARCH; Monte Carlo; Size correction.
JEL-CLASSIFICATION: C12; C15; C22; C52.

4.1 Introduction

Many financial time series display characteristic features such as observations that are non-normally distributed (i.e. with excess skewness and kurtosis), conditionally heteroskedastic and ruled by long-memory. For instance Ding, Granger and Engle (1993) report evidence of autocorrelations between distant lags for long lags in the absolute returns of the Standard and Poor 500, S&P500, composite stock index. Furthermore, Granger and Ding (1995) show that the absolute value of the rate of return for a variety of stock prices, commodity prices and exchange rates exhibit excess skewness and kurtosis.

Long-memory is usually described by fractionally integrated specifications, hence testing for long-memory may be performed via a test for a fractional differencing power. For this purpose, several tests have been proposed and some of the most popular are thoroughly investigated by Cheung (1993), who also studies the influence of ARCH disturbances. Andersson and Gredenhoff (1997) implement a bootstrap method, in order to size-adjust fractional integration tests. The bootstrap provides a trustworthy technique for estimation of the small-sample distribution of a statistic. When using a bootstrap test the null-distribution is retrieved by bootstrap methods and hence the critical values are adjusted to give exact tests.

This paper investigates some fractional integration tests when the data are non-normal or the residuals are heteroskedastic. Again, the bootstrap is used to correct for size distortions. Another extension is a comparison of parametric, non-parametric and heteroskedasticity invariant resampling algorithms. The aim of the paper is to find tests that are robust to non-normalities and ARCH effects in data, and thus are well-suited when testing for long-memory in financial and economic time series.

The results suggest that the performance of a bootstrap testing procedure depends to some extent on the chosen resampling algorithm. However, all (but one) bootstrap tests are superior to the original version of the tests, in the sense that the bootstrap tests have better size properties.

The paper is organized as follows. Section 2 describes the bootstrap testing procedure and Section 3 contains a Monte Carlo simulation study where the sizes of the tests are presented for normal and non-normal data and processes with ARCH errors. Section 4 concludes the paper.

4.2 The Bootstrap Testing Procedure

The bootstrap, see for instance Efron and Tibshirani (1993), provides a feasible method for estimation of the small-sample distribution of a statistic. The basic principle is to approximate this distribution by a bootstrap distribution, which can be retrieved 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.

4.2.1 The Bootstrap Test

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

$$p(\hat{\tau}) = p(|\tau| \geq |\hat{\tau}| \, |\Psi_0, T) \tag{4.1}$$

where Ψ_0 is the data generating process (*DGP*) under the null hypothesis, and $\hat{\tau}$ is the realized value of a test statistic τ based on a sample of length T. Since Ψ_0 is unknown this *p*-value function has to be approximated, which is regularly done using asymptotic theory. For asymptotic theory to be valid it is required that $p(\hat{\tau})$ 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. According to Davidson and MacKinnon (1996*a*), a bootstrap test is understood as a test for which the significance level is calculated using a bootstrap procedure.

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

$$p^*(\hat{\tau}) = R^{-1} \sum_{r=1}^{R} I(|\tau_r^*| \geq |\hat{\tau}|), \tag{4.2}$$

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

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 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.

This paper employs the bootstrap technique on fractional integration tests. In order to handle non-normal or conditionally heteroskedastic data, we refine the bootstrap testing procedure of Andersson and Gredenhoff (1997) to include these cases. The bootstrap tests are based on the periodogram regression test of Geweke and Porter-Hudak, GPH, (1983), the modified rescaled range, MRR, test (Lo, 1991) and the Lagrange multiplier REG test of Agiakloglou and Newbold (1993).[1]

A fractionally integrated autoregressive moving average (ARFIMA) time series process is described by

$$\phi(B)(1-B)^d x_t = \theta(B) a_t, \qquad t = 1, ..., T \tag{4.3}$$

where the roots of $\phi(B)$ and $\theta(B)$ have all roots outside the unit circle and a_t is *iid* with mean zero and variance $\sigma_a^2 < \infty$. The differencing parameter d is allowed to take any real number, but if d is restricted to the set of integers the specification (4.3) reduces to an ARIMA process. The sample autocorrelation function of a long-memory process may be approximated by a fractionally integrated model, hence testing for long-memory can be done by a test on d. Such tests are applied to stationary and invertible series and $d = 0$ is thus a natural null-hypothesis.[2]

When testing for fractional integration, the DGP Ψ_0 is characterized by an unknown ARMA(p, q) 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.

4.2.2 Construction of the Bootstrap Samples

The original non-parametric bootstrap of Efron (1979), designed for *iid* observations, usually fails for dependent observations, e.g. time series, since the order of the observations is affected. Dependencies in data can be maintained in the bootstrap resample by using a model-based bootstrap, which is the natural way to proceed in our case since a well-defined model forms the null-hypothesis. A model free procedure, such as a moving blocks bootstrap or a spectral resampling scheme, may also preserve dependencies. However, model free

---

[1] These tests are briefly described in *Appendix* A.

[2] Stationarity and invertibility require that $d < |1/2|$. The ARFIMA model is presented in greater detail by Granger and Joyeux (1980) and Hosking (1981).

approaches deviate from the bootstrap testing idea of Davidson and MacKinnon (1996*a*, *b*), in the sense that the resemblance between the bootstrap samples and the original sample is sacrificed. This is due to the implementation of the null-hypothesis, which in this situation is done by filtering the series through the long-memory filter $(1-B)^{\hat{d}}$, where $\hat{d}$ is an estimate of the differencing parameter. A further drawback is that the bootstrap test would then in general be sensitive to the estimate of d.

For the bootstrap fractional integration tests we use the resampling model,

$$\left(1-\hat{\phi}_0-\hat{\phi}_1 B-\ldots-\hat{\phi}_{\hat{p}} B^{\hat{p}}\right) x_t=\hat{a}_t, \tag{4.4}$$

which clearly obeys the null-hypothesis and can be regarded as the estimated AR representation of the process. The autoregressive order $\hat{p}$ is selected from the values $(0, 1 \ldots, 5)$ for the size evaluation and up to 25 for the power, by the Bayesian information criterion (BIC) of Schwartz (1978) , and the parameters are estimated by ordinary least squares (OLS). The use of the BIC is motivated by comparisons, not reported in the paper, with the AIC of Akaike (1974). Furthermore Andersson and Gredenhoff (1997) use the AR approximation as well as an ARMA resampling model, and find that the former performs better.

The bootstrap samples $\mathbf{x}_r^*$, $r=1,...,R$, are created recursively by the equation

$$x_{r,t}^*=\hat{\phi}(B)^{-1} a_t^*, \tag{4.5}$$

where $\hat{\phi}(B)$ is the polynomial of (4.4) and a_t^* are the bootstrap residuals. In this study the number of bootstrap replicates is $R=1,000$.

Four resampling algorithms are utilized to generate the bootstrap residuals a_t^*. The *first* algorithm, b_1, makes use of a normality assumption for the disturbances a_t in (4.3), and is denoted the simple parametric bootstrap. In this resampling the residuals a_t^* are independent draws from a normal distribution with mean zero and variance $s_{\hat{a}}^2$.

A *second* similar but non-parametric resampling scheme (denoted b_2) does not impose distributional assumptions but is directly based on the estimated residuals $\hat{a}_t$. The bootstrap residuals are drawn, with replacement, from the recentered and degrees of freedom corrected residual vector. One typical bootstrap residual is constructed as

$$a_t^* = \sqrt{\frac{T}{T-\hat{p}-1}} \times \hat{a}_s,$$

where s is $U\left(\hat{p}+1, T\right)$ distributed.

The *third* and *fourth* resampling algorithms are constructed to preserve ARCH(1) dependence in the residuals. ARCH is introduced to the autoregression, $\phi\left(B\right)x_t = a_t$, by the equation $a_t = \sqrt{\omega_t}\varepsilon_t$, where the conditional variance is given by $\omega_t = \beta_0 + \beta_1 a_{t-1}^2$. The assumed normality of ε_t allows joint estimation of the parameters through maximization of the log-likelihood function

$$l\left(\phi_0, \ldots, \phi_p, \beta_0, \beta_1 \left|\mathbf{x}\right.\right) = -\frac{1}{2T}\sum_{t=1}^{T}\left(\log\omega_t + \frac{a_t^2}{\omega_t}\right).$$

For the optimization, we use the numerical method of Davidon, Fletcher and Powell, see for instance Press *et al.* (1992). The resamplings are based on a parametric or a non-parametric algorithm, similar to those above. In the parametric case (denoted b_3), a residual series $\tilde{\varepsilon}_t$ is created by independent draws from a $N\left(0, s_{\hat{\varepsilon}}^2\right)$ distribution. For the non-parametric (b_4) scheme the members of $\left\{\tilde{\varepsilon}_t\right\}$ are drawn from the degrees of freedom adjusted elements of $\left\{\hat{\varepsilon}_t\right\}$. The bootstrap residuals are then built by imposing the estimated conditional dependency, according to the equations

$$\tilde{\omega}_t = \hat{\beta}_0 + \hat{\beta}_1 a_{t-1}^{*2}$$

and

$$a_t^* = \tilde{\varepsilon}_t\sqrt{\tilde{\omega}_t}.$$

This implies that a_t^* has an unconditional variance of $\hat{\beta}_0/\left(1-\hat{\beta}_1\right)$.

4.3 The Monte Carlo Study

The Monte Carlo study involves 1000 replicates (series), where each series is tested for fractional integration using the original tests and the different bootstrap tests described in Section 4. The rejection frequencies of the non-fractional null-hypothesis, i.e. the empirical sizes, are evaluated and compared. The power of a bootstrap test is in general close to that of the size adjusted asymptotic test (see Davidson and MacKinnon 1996*b*). In particular, for the asymptotic tests in this study and normal processes, Andersson and Gredenhoff (1997) demonstrate this for a bootstrap test with the simple parametric resampling scheme.

To evaluate the size of the tests first order autoregressions,

$$(1 - \phi B)\, x_t = a_t, \tag{4.6}$$

of length $T = 100$ are generated and the parameter ϕ is set equal to the values $\{0, 0.1, 0.5, 0.7, 0.9\}$. To reduce the initial-value effect, an additional 100 observations are generated. We construct the data in order to display three different characteristics: normality, non-normality (skewness and excess kurtosis) and ARCH errors. The characteristics are introduced via the disturbances a_t.

4.3.1 Normal Processes

The experiment examining the empirical size of the tests under normality is based on the process (4.6) where the disturbances $\{a_t\}$ are *iid* normally distributed with mean zero and variance equal to unity. Table 4.1 presents the sensitivity of the empirical size with respect to the investigated AR parameters for a nominal 5% level of significance. Significant differences from the nominal size are obtained when the rejection frequencies lie outside the 95% acceptance interval $(3.6, 6.4)$.

The estimated size of the original MRR test always differs significantly from the 5% nominal level. In particular, the MRR test is strongly conservative for large positive parameters. The GPH test is severely over-sized for highly short-term dependent series, which is explained by a biased periodogram regression estimate due to large positive AR roots, see Agiakloglou *et al.* (1993). Compared with the other original tests the REG test is well-sized; only one significant size-distortion can be found. A more detailed presentation of the tests is given in Andersson and Gredenhoff (1997).

The results suggest that the MRR and GPH bootstrap tests, regardless of resampling, give exact tests in the sense that the estimated sizes of the tests coincide with the nominal. The bootstrap REG test based on the simple parametric resampling is almost exact, whereas the non-parametric resampling (which does not incorporate

TABLE 4.1. Rejection percentage of the nominal 5 percent fractional integration test when the data follow an AR(1), of length 100, with normal errors.

Test		ϕ				
		0.0	0.1	0.5	0.7	0.9
MRR	o	**7.6**	**6.8**	**2.3**	**1.3**	**0.8**
	b_1	5.3	5.9	4.8	4.7	3.9
	b_2	4.5	4.2	4.9	5.0	4.4
	b_3	6.2	6.4	6.0	5.2	3.9
	b_4	6.0	6.4	5.5	5.3	4.3
GPH	o	4.9	4.9	**8.3**	**17.9**	**71.8**
	b_1	5.0	5.2	4.7	4.3	3.7
	b_2	5.0	5.1	4.6	4.2	4.0
	b_3	5.3	5.8	5.0	4.4	3.6
	b_4	5.4	5.5	5.0	4.4	3.6
REG	o	5.9	**6.9**	6.4	5.1	5.1
	b_1	6.0	5.1	4.7	4.2	4.7
	b_2	4.5	3.7	5.6	**7.9**	**19.2**
	b_3	5.0	4.6	3.7	**3.3**	**3.4**
	b_4	4.8	4.8	3.6	**3.5**	**3.5**

The number reported is the rejection percentage of the two-sided 5% test. Bold face denotes a significant deviation from the nominal size. Under the null-hypothesis of no fractional integration, the 95% acceptance interval of the rejection percentage equals (3.6, 6.4). o denotes the original test and $b_1 - b_4$ the bootstrap testing procedures described in Section 4.

the normality) produces notably large sizes for strongly dependent AR processes. The resamplings that account for the (non-existing) ARCH effects have resonable estimated sizes, however conservative for ϕ equal to 0.7 and 0.9.

4.3.2 Non-Normal Processes

In the non-normal case, the disturbances a_t are distributed with mean, variance, skewness and kurtosis equal to $0, 1, \gamma_s$ and γ_k respectively. The members of $\{a_t\}$ are generated by the transformation,

$$a_t = c_0 + c_1\alpha_t + c_2\alpha_t^2 + c_3\alpha_t^3 \qquad \alpha_t \sim N(0,1), \qquad (4.7)$$

proposed by Fleichmann (1978). In this situation, the sequence $\{a_t\}$ will have a distribution dependent upon the constants c_i, which can be solved for using a non-linear equation system specified as a function of selected skewness and kurtosis. γ_s and γ_k are chosen to gen-

TABLE 4.2. Rejection percentage of the nominal 5 percent fractional integration test when the data follow an AR(1), of length 100, with non-normal errors.

Test		ϕ				
		0.0	0.1	0.5	0.7	0.9
MRR	o	**7.7**	6.0	**2.2**	**2.0**	**0.5**
	b_1	5.0	5.5	5.3	5.4	4.8
	b_2	4.7	4.8	4.4	4.8	4.6
	b_3	4.9	6.0	5.5	4.2	4.2
	b_4	4.7	6.0	5.2	4.5	3.8
GPH	o	4.9	5.6	**8.0**	**16.3**	**72.2**
	b_1	4.2	4.2	3.7	4.5	3.9
	b_2	4.4	4.3	5.0	4.7	3.6
	b_3	4.2	5.4	5.3	5.0	3.7
	b_4	4.4	5.4	5.1	5.2	4.1
REG	o	5.5	**7.3**	**7.7**	**6.6**	**8.7**
	b_1	4.9	5.0	3.8	4.7	**2.4**
	b_2	5.1	3.9	4.5	**6.9**	**13.8**
	b_3	3.9	5.8	4.5	**3.1**	**1.5**
	b_4	3.8	4.7	4.2	**1.8**	**1.1**

See note to Table 4.1. The skewness and kurtosis of the disturbances are selected in order to give a skewness of 2.0 and a kurtosis of 9.0, for all processes.

erate series x_t with a skewness and kurtosis of 2 and 9 respectively.[3] The empirical size of the tests under non-normality is reported in Table 4.2.

The original MRR and GPH tests are robust to excess skewness and kurtosis, in the sense that the results are similar to the normal case. This does not imply that the tests are well-sized, since the distortions of the MRR test are slightly more articulated and the GPH test is still severely over-sized for large parameters. In contrast to the other tests, the original REG test is sensitive when the data do not fulfill the normality assumption. The difference between the empirical and nominal size is in general significant.

The estimated sizes of all bootstrap MRR and GPH tests never differ significantly from the nominal 5%. The bootstrap REG tests behave as in the normal case. That is, the parametric b_1 works well, the non-parametric b_2 is over-sized for large parameters and b_3 and b_4 are conservative for the same parameters.

---

[3] The expressions for the determination of $c_i, i = 0, ..., 3$ and how the residual skewness and kurtosis depend on those of the time series process are given in *Appendix* B.

4.3.3 Processes with ARCH Errors

For the final set of processes the assumption of identically and independently distributed errors is relaxed. Instead we consider the effect of heteroskedasticity of ARCH(1) type, which implies that the disturbances are conditionally distributed as $a_{t|t-1} \sim N(0, \omega_t)$, where $\omega_t = 1 - \beta + \beta a_{t-1}^2$ and $\beta < 1$. The parametrization implies that the unconditional variance of a_t equals unity, and the parameter β is selected as 0.5 and 0.9. The 0.9 parameter imposes a strong conditional dependence in the disturbances, but the fourth moment of the disturbance process does not exist. As a complement, the weaker ARCH dependence of $\beta = 0.5$ is also investigated.

Results in Table 4.3 show that the MRR test is quite robust also against conditional heteroskedasticity. However, compared with the case of uncorrelated data, cf. $\beta = 0$, the test tends to be more conservative as the ARCH parameter increases. The GPH test is unaffected by ARCH in the disturbances. In short, these tests have the same size problem as with uncorrelated disturbances. On the other hand, the usually well-sized REG test is very sensitive to ARCH effects and exhibits a seriously distorted size for $\beta = 0.5$ and in particular for $\beta = 0.9$.

The robustness of the MRR and GPH tests against ARCH effects can be detected in the bootstrap tests. As a result all bootstrap MRR and GPH tests are exact for all generated combinations of β and ϕ.

The disappointing size of the original REG test is partly inherited by b_1 and b_2. Furthermore, the increasing pattern with the AR parameter for b_2 is still present. However, the size distortions of b_1 and b_2 are smaller for the lower value of β. The REG test, overall, requires that the resampling scheme allows for ARCH effects. This is exactly what bootstraps b_3 and b_4 do, and despite a few conservative values these bootstraps are not only superior to the original test, but also much better than b_1 and b_2.

4.3.4 Size Comparisons and Power

Table 4.4 supplies an overview of the tests that exhibit the best size properties, judged by the number of significant results based on the 95% acceptance region, for the respective processes. All bootstrap MRR and GPH tests work well and have equivalent size properties, for all processes investigated. The simple non-parametric bootstrap

TABLE 4.3. Rejection percentage of the nominal 5 percent fractional integration test when the data follow an AR(1), of length 100, with ARCH errors.

		ϕ				
Test		0.0	0.1	0.5	0.7	0.9
				$\beta = 0.5$		
MRR	o	5.5	4.0	**2.4**	**1.8**	**0.6**
	b_1	4.4	4.8	5.1	3.9	3.8
	b_2	4.3	4.8	4.7	4.3	4.1
	b_3	5.1	5.1	4.8	5.4	3.9
	b_4	4.7	4.9	4.1	5.2	3.8
GPH	o	4.5	4.3	5.7	**17.6**	**71.6**
	b_1	4.0	4.1	4.0	3.9	3.6
	b_2	3.8	4.2	4.0	3.6	4.1
	b_3	3.8	3.5	4.0	4.5	4.2
	b_4	4.0	3.8	3.9	4.8	4.5
REG	o	**9.8**	**9.7**	**9.0**	**7.4**	**10.4**
	b_1	**7.6**	**8.5**	5.4	4.6	**7.3**
	b_2	**8.7**	4.6	6.6	4.6	**25.2**
	b_3	4.3	3.8	3.9	3.9	3.7
	b_4	4.2	4.0	3.9	3.7	4.0
				$\beta = 0.9$		
MRR	o	**3.4**	**3.1**	**1.4**	**0.8**	**0.7**
	b_1	3.7	3.8	4.0	4.0	5.3
	b_2	4.0	3.6	4.1	3.7	3.7
	b_3	5.2	5.4	4.8	4.3	4.2
	b_4	5.1	5.3	4.8	4.0	4.3
GPH	o	4.7	5.0	**6.5**	**16.8**	**70.9**
	b_1	4.7	5.0	4.1	4.2	4.1
	b_2	5.4	5.3	4.7	3.8	3.8
	b_3	4.0	4.7	4.2	6.2	4.7
	b_4	4.0	4.7	4.4	6.4	4.6
REG	o	**28.6**	**29.3**	**29.5**	**29.2**	**33.4**
	b_1	**7.9**	**8.4**	**7.3**	**7.1**	**7.3**
	b_2	**9.2**	**6.8**	**8.8**	**14.4**	**24.9**
	b_3	**3.1**	3.8	**3.5**	3.8	4.3
	b_4	3.6	3.8	**3.3**	4.1	4.4

See note to Table 4.1. The error processes follow an ARCH process with parameter β.

TABLE 4.4. Best test based on size properties.

	Normal	*Non − normal*	*ARCH*	*Overall*
MRR	$b_1 - b_4$	$b_1 - b_4$	$b_1 - b_4$	$b_1 - b_4$
GPH	$b_1 - b_4$	$b_1 - b_4$	$b_1 - b_4$	$b_1 - b_4$
REG	$b_1, (o, b_3, b_4)$	$b_1, (b_3, b_4)$	$b_4\,(b_3)$	b_3, b_4

The Table reports the best test based on the smallest number of significant size distortions. The test within brackets is regarded as almost equivalent to the ones without.

REG test is badly sized when the generated process has an AR parameter close to the unit circle, regardless of the characteristics of the disturbance process. Otherwise, all REG tests, even the original, perform well under normality, whereas for non-normality the simple parametric bootstrap is best. When ARCH errors are introduced, the bootstraps that account for the heteroskedasticity clearly adjust the size of the REG test better. Since these resamplings also work for uncorrelated errors, b_3 and b_4 exhibit the best REG performance overall.

A bootstrap MRR or GPH test is shown to be exact, and a well-designed bootstrap test based on the REG test nearly exact. Furthermore, the stylized DGPs of this study are quite well-behaved, whereas in empirical situations they are not. Consequently, the asymptotic tests are likely to be more distorted and the gain from a bootstrap test to be even larger.

Davidson and MacKinnon (1996*b*) show that the power of a bootstrap test, based on a pivotal statistic, is generally close to the size-adjusted asymptotic test. Table 4.5 presents the power of the tests for fractionally integrated white noise, $(1 - B)^d x_t = a_t$, where the members of $\{a_t\}$ have a normal, non-normal[4] or heteroskedastic distribution. Only the parametric bootstraps are reported, because of the similar power properties of the corresponding non-parametric resamplings. However, combined with the REG test the simple parametric bootstrap b_1 exhibits, at positive differencing parameters, notably better power properties than the simple non-parametric resampling b_2.

---

[4] The skewness and kurtosis of the residuals as functions of the time series moments are given in *Appendix* B.

TABLE 4.5. Rejection percentage of the nominal 5 percent fractional integration test when the data follow fractional noise of length 100.

Test		d						
		-0.45	-0.25	-0.05	0.0	0.05	0.25	0.45
		Normal processes						
MRR	o	14.6	11.3	3.7	5.0	5.8	14.7	4.0
	b_1	16.5	13.6	6.1	5.3	6.9	14.3	18.6
	b_3	13.6	12.4	4.4	6.2	6.6	11.9	12.3
GPH	o	21.8	8.3	3.9	5.0	4.4	17.0	40.9
	b_1	21.0	11.1	5.3	5.0	5.3	16.5	11.8
	b_3	17.6	9.3	4.8	5.3	5.3	7.1	12.2
REG	o	43.9	27.8	3.8	5.0	10.8	35.3	22.5
	b_1	41.0	37.6	6.7	6.0	8.0	25.9	19.6
	b_3	19.9	23.9	3.1	5.0	5.6	16.1	14.1
		Non-normal processes						
MRR	o	6.9	8.7	5.0	5.0	6.8	15.8	6.8
	b_1	10.8	8.2	4.7	5.0	5.4	11.8	18.6
	b_3	8.0	8.7	5.3	4.9	6.0	11.6	18.4
GPH	o	25.9	10.8	6.1	5.0	7.0	15.4	40.8
	b_1	20.4	10.5	4.5	4.2	5.3	8.4	10.0
	b_3	19.7	9.2	5.6	4.2	6.6	7.9	8.9
REG	o	42.9	32.6	4.2	5.0	10.8	37.9	22.5
	b_1	42.2	37.7	5.5	4.9	6.1	27.1	15.8
	b_3	23.1	30.2	4.2	3.9	5.3	20.4	16.7
		ARCH processes						
MRR	o	10.8	8.0	4.9	5.0	6.0	19.1	17.1
	b_1	8.3	5.7	3.1	3.7	4.8	11.2	20.3
	b_3	9.4	7.3	4.3	5.2	5.2	13.0	13.7
GPH	o	21.1	10.1	5.9	5.0	6.0	14.8	40.7
	b_1	19.4	11.4	5.6	4.7	4.9	8.9	12.7
	b_3	17.0	9.2	5.5	4.0	5.8	7.7	10.6
REG	o	31.7	16.6	4.5	5.0	7.0	21.2	10.9
	b_1	35.9	25.1	9.5	7.9	9.5	21.6	18.7
	b_3	11.6	9.3	2.8	3.1	2.4	9.3	7.5

The number reported is the rejection percentage of the two-sided 5% test. o denotes the original test, and b_1 and b_3 the bootstrap testing procedures described in Section 4.

The power of the MRR and GPH tests are preserved by all bootstrap procedures, except for processes with $d = 0.45$. In this case too many autocorrelations are included in the variance correction term of the MRR test, resulting in a negatively biased estimate of the fractional differencing parameter which lowers the power of the original test. This phenomenon is not experienced by the bootstrap tests, which have well-behaved power curves. For the GPH test a large differencing parameter results in a rich parameter structure of the resampling model, which implies that the resample periodograms resemble the periodogram of the highly persistent original process. Thus, the bootstrap GPH test has difficulties in distinguishing fractional processes from AR specifications. The power of the REG test is preserved by the simple parametric bootstrap, whereas the ARCH resamplings have a lower power throughout.

On the basis of the estimated power, two major situations are detected. If we cannot rule out ARCH effects in the disturbances, the highest power is given by a simple bootstrap MRR or GPH test. However, if there are no ARCH effects (in theory or data) then the simple parametric REG test clearly outperforms all other testing procedures.

4.4 Conclusions

The concept of bootstrap testing for fractional integration works extraordinarily well. If the significance level is calculated by a bootstrap procedure an exact test is almost always the result. However, the choice of resampling algorithm may affect the degree of size adjustment. For instance, if the original test is sensitive to distributional assumptions, in particular ARCH effects, this should be accounted for when specifying the resampling model. However, if the test is robust to ARCH errors, the choice of resampling is not very important for the size properties of that test.

Since economic and financial data are often heteroskedastic we recommend the use of the parametric ARCH resampling scheme for the REG test. However, if prior information suggests that the investigated series does not have ARCH effects, the simple parametric bootstrap has equivalent size properties and a higher power, and should thus be used.

The MRR and GPH tests, which are robust to deviations from the iid normality of the disturbances, have nice size properties for all bootstrap procedures. Due to the simplicity and the slightly higher power of the simple algorithms, they are preferred when bootstrapping the MRR and GPH tests.

The main conclusions are that the bootstrap tests are remarkably well-sized (whereas the originals are not) and robust to non-normalities and ARCH effects, and that reliable testing for fractional integration in many cases requires a bootstrap test.

Acknowledgments

We are greatly indebted to Tor Jacobson and Sune Karlsson for useful discussions and comments. Financial support from the Tore Browaldh Foundation is gratefully acknowledged. The usual disclaimer applies.

References

AGIAKLOGLOU, C., P. NEWBOLD and M. WOHAR (1993) Bias in an Estimator of the Fractional Difference Parameter. *Journal of Time Series Analysis* 14 (3), 235-246.

AGIAKLOGLOU, C. and P. NEWBOLD (1994) Lagrange Multiplier Tests for Fractional Difference. *Journal of Time Series Analysis* 15 (3), 253-262.

AKAIKE, H. (1974) A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* 19(6), 716-723

ANDERSSON, M.K. and M.P. GREDENHOFF (1997) Bootstrap Testing for Fractional Integration. *Working Paper Series in Economics and Finance*, No. 188, Dept. of Economic Statistics, Stockholm School of Economics, Sweden.

CHEUNG, Y.-W. (1993) Tests for Fractional Integration: A Monte Carlo Investigation. *Journal of Time Series Analysis* 14 (4), 331-345.

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.

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

DING, Z., C. W.J. GRANGER, and R. F. ENGLE (1993) A Long Memory Property of Stock Market Returns and a New Model, *Journal of Empirical Finance* 1, 83-106.

EFRON, B. (1979) Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics* 7, 1-26.

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

FLEICHMANN, A. I. (1978) A Method for Simulating Non-Normal Distributions. *Psychometrika* 43 (4), 521-532.

GEWEKE, J. and S. PORTER-HUDAK (1983) The Estimation and Application of Long Memory Time Series Models. *Journal of Time Series Analysis* 4 (4), 221-238.

GRANGER, C. W.J. and Z. DING (1995) Stylized Facts on the Temporal and Distributional Properties of Daily Data from Speculative Markets, *Unpublished manuscript,* Department of Economics, University of California, San Diego.

GRANGER, C. W. J. and R. JOYEUX (1980) An Introduction to Long-Memory Time Series Models and Fractional integration. *Journal of Time Series Analysis* 1 (1), 15-29.

HOROWITZ, J. L. (1995) Bootstrap Methods in Econometrics: Theory and Numerical Performance. Paper presented at the 7th World Congress of the Econometric Society, Tokyo.

HOSKING, J. R. M. (1981) Fractional Differencing. *Biometrika* 68, 165-176.

LO, A. W. (1991) Long Term Memory in Stock Market Prices. *Econometrica* 59 (5), 1279-1313.

PRESS, H., TEUKOLISKY, S., VETTERLING, W. AND B. FLANNERY (1992) *Numerical Recipes in Fortran: The Art of Scientific Computing,* second edition. Cambridge University Press, New York, USA.

SCHWARTZ, G. (1978) Estimating the order of a model. *The Annals of Statistics* 2, 461-464

Appendix A: Tests for Fractional integration

Consider the regression equation

$$\ln\left\{I_x\left(\omega_j\right)\right\} = \alpha - d\ln\left\{4\sin^2\left(\omega_j/2\right) + v_j\right\}, \tag{4.8}$$

where $I_x\left(\omega_j\right)$ is the periodogram at the harmonic frequencies $\omega_j = 2\pi j/T$, and $j = 1, ..., g\left(T\right) = T^{1/2}$. The ordinary least squares (OLS) estimator of d is then consistent and the distribution of $\left(\hat{d}_{OLS} - d\right)/SE\left(\hat{d}_{OLS}\right)$ is asymptotically normal. This is the periodogram regression estimation/testing procedure of Geweke and Porter-Hudak (1983).

Lo (1991) proposes a modified rescaled range (MRR) statistic when testing for fractional integration. This modified rescaled range is defined by the ratio

$$\tilde{Q}_T = \frac{R_T}{\hat{\sigma}_T\left(k\right)}, \tag{4.9}$$

where the range and the standard error are calculated by

$$R_T = \max_{0<i\leq T}\sum_{t=1}^{i}\left(x_t - \overline{x}\right) - \min_{0<i\leq T}\sum_{t=1}^{i}\left(x_t - \overline{x}\right) \tag{4.10}$$

$$\hat{\sigma}_T^2\left(k\right) = \hat{\sigma}^2 + 2\sum_{j=1}^{k}\sum_{i=j+1}^{T}\left(1 - \frac{j}{k+1}\right)\left(x_i - \overline{x}\right)\left(x_{i-j} - \overline{x}\right). \tag{4.11}$$

The truncation lag, k, is set, to the integer part of

$$\left(3T/2\right)^{\frac{1}{3}}\cdot\left\{2\hat{\rho}/\left(1 - \hat{\rho}^2\right)\right\}^{\frac{2}{3}}$$

, where $\hat{\rho}$ denotes the sample first-order autocorrelation coefficient and $\hat{\sigma}^2$ the maximum likelihood variance estimate. Asymptotic critical values of the MRR test are given by Lo (1991).

The LM type test, denoted REG, of Agiakloglou and Newbold (1994) is carried out through the likelihood based auxiliary regression

$$\hat{a}_t = \sum_{i=1}^{p}\beta_i W_{t-i} + \sum_{j=1}^{q}\gamma_j Z_{t-j} + \delta K_m + u_t, \tag{4.12}$$

where

$$K_m = \sum_{j=1}^{m} j^{-1} \hat{a}_{t-j}, \hat{\theta}(B) W_t = x_t, \hat{\theta}(B) Z_t = \hat{a}_t \text{ and } u_t \text{ is } iid \text{ normal.}$$

$\hat{a}_t$ and $\hat{\theta}(B)$ are the estimated residuals and MA polynomial under the null and m is a prespecified truncation lag. The equation (4.12) is fitted by OLS over the time period $t = m+1, ..., T$ and the usual t-statistic for the null hypothesis $\delta = 0$ follows an asymptotic $N(0,1)$ distribution.

Appendix B: Generation of Non-normal Data

Appendix B.1:The Skewness and Kurtosis Relationship

Under the assumptions in Section 4 the skewness and (raw) kurtosis for the disturbance process are given by

$$\gamma_s = \Gamma_s \frac{(1-\phi^3)}{(1-\phi^2)^{3/2}}$$

and

$$\gamma_k = \frac{\Gamma_k(\phi^2+1) - 6\phi^2}{1-\phi^2},$$

where Γ_1 and Γ_2 are the corresponding moments of the AR(1) process.

In the fractionally integrated case the disturbance skewness and kurtosis are given as

$$\gamma_s = \Gamma_s \frac{\sum_{i=0}^{\infty} \delta_i^3}{var^{3/2}(x)}$$

$$\gamma_k = \left\{ (\Gamma_k - 3) \sum_{i=0}^{\infty} \delta_i^4 + 3 \sum_{j=0}^{\infty} \sum_{k \neq j} \delta_j^2 \delta_k^2 \right\} / var^2(x),$$

where δ_i is the *ith* weight in the moving average representation,

$$x_t = \sum_{i=0}^{\infty} \delta_i a_{t-i},$$

for the fractionally integrated process. The weights are given by

$$\begin{aligned}\delta_0 &= 1\\ \delta_1 &= d\\ \delta_i &= \frac{1}{i}\delta_{i-1}\left(i-1-d\right), \qquad \text{for } i>1.\end{aligned}$$

Appendix B.2: The Fleichmann Algorithm

The constants in (4.7) are given as the solutions to the following system of equations,

$$\begin{aligned}c_0 &= -c_2\\ \gamma_k &= 24\left[c_1c_3 + c_2^2\left(1+c_1^2+28c_1c_3\right)+\right.\\ &\quad \left.c_3^2\left(12+48c_1c_3+141c_2^2+225c_3^2\right)\right]\\ c_2 &= \frac{\gamma_s}{2\left(c_1^2+24c_1c_3+105c_3^2+2\right)}\end{aligned}$$

and

$$2 = 2c_1^2 + 12c_1c_3 + \frac{\gamma_s^2}{\left(c_1^2+24c_1c_3+105c_3^2+2\right)^2} + 30c_3^2.$$

5

Lag-length Selection in VAR-models Using Equal and Unequal Lag-Length Procedures

ABSTRACT: It is well known that inference in vector autoregressive models depends crucially on the choice of lag-length. Various lag-length selection procedures have been suggested and evaluated in the literature. In these evaluations the possibility that the true model may have unequal lag-length has, however, received little attention. In this paper we investigate how sensitive lag-length estimation procedures, based on assumptions of equal or unequal lag-lengths, are to the true model structure. The procedures are based on information criteria, we give results for AIC, HQ and BIC. In the Monte Carlo study we generate data from a variety of VAR models with properties similar to macro-economic time series. In summary, we find that the commonly used procedure based on equal lag-length together with AIC and HQ performs well in VAR models up to a lag-length of 3. The procedure (due to Hsiao) based on unequal lag-lengths produces reasonable results when the true model has unequal lag-lengths. The Hsiao procedure also tends to do better in models with a more complicated lag structure.

KEYWORDS: Vector autoregression, Order selection, Information Criteria, Monte Carlo simulation.
JEL CODES: C32, C51, C53

5.1 Introduction

Consider a time series $y_t = (y_{1t}, \ldots, y_{kt})'$ generated by a stationary VAR(p) model

$$\mathbf{y}_t = \mathbf{v} + \sum_{j=1}^{p} \mathbf{A}_j \, \mathbf{y}_{t-j} + \mathbf{u}_t \tag{5.1}$$

with constant parameters $(\mathbf{v}, \mathbf{A}_1, \ldots, \mathbf{A}_p, \mathbf{\Sigma})$ and $\mathbf{u}_t \sim iid \; \mathbf{N}[\mathbf{0}, \mathbf{\Sigma}]$. In practice, the order, p, of the VAR model is unknown and must be chosen on the basis of the available data. Mis-specification of the order will result in either a model with false zero restrictions and inconsistent estimates, or an over-parameterized model resulting in a loss of efficiency.

The choise of lag-length is thus a critical step in the estimation procedure and affects the inference in the VAR model: Hsiao (Hsiao 1979, Hsiao 1981) demonstrates that the acceptance or rejection of a null hypothesis in many cases depends on the selected lag-length. Boswijk and Franses (1992) show that for cointegration tests, an underestimated lag-length will lead to a substantial increase in size, and an over-parameterization results in a loss of power. Cheung and Lai (1993) find similar results in a study of the Johansen trace test. DeSerres and Guay (1995) find that a too short lag-length in a structural VAR model can lead to a significant bias in the estimates of the permanent and transitory components. Kunitomo and Yamamoto (1985) show that the prediction mean square error of a fitted model is always larger than that of the true model when the fitted order overstates the true lag-length.

The performance of various lag-length selection strategies and information criteria has been studied extensively. Unfortunately, the vast majority of these studies have dealt with strategies based on the assumption of equal lag-length and evaluated the procedures on equal lag-length models. In addition, previous studies have mainly consider small, 2 variable, models. See, for example, Nickelsburg (1985), Lütkephol (1985) and Jacobson (1995). Nickelsburg considers true models with unequal lag-length and a maximum lag of 4 and finds that no information criteria performs well in "reasonable" sample sizes. Specifically, all the criteria, including AIC, tend to underestimate the true (maximum) lag-length. Lütkepohl differs in that he only considers models of lag-length 1 and 2 but varies the parameter values of the models extensively. The conclusions drawn by Lütkepohl also differ; he finds that BIC and HQ perform best, with little tendency to underestimate the lag-length. This result may be due in part to the short true lag-lengths and the correspondingly smaller scope for underfitting. Jacobson studies 2 variable Cointegrated systems with lag order 2 and finds that the HQ criterion performs well in conjunction with the multivariate Box-Pierce portmanteau test to protect against underestimation of the lag-length.

Due to the simplicity of the equal lag-length procedures, the choice of lag-length is usually made under this assumption. If the assumption of equal lag-length is invalid, the assumption of equal lag-length will lead to either the imposition of false zero restrictions or an over-parameterized model. That is, even if the true lag-length (the maximum lag-length for the entire model) can be estimated, this model will be over-parameterized since at least on of the parameter matrix $(\mathbf{A}_p, \mathbf{A}_{p-1}, ...)$ will contain zeros. Alternatively, the lag-length is underestimated and at least one variable with lag p is falsely excluded from the model when the $\mathbf{A}_p$ matrix is left out.

Hsiao (1979) suggested a lag selection procedure that allows for different lag-lengths. The Hsiao procedure can be expected to perform better than procedure based on the equal lag-length assumption when the true model have unequal lag-lengths. On the other hand, it is reasonable to expect the equal lag-length procedures to have an edge when the assumption of equal lag-length holds.

In practice it is not known if the true model has equal lag-length or not and there is little evidence in the literature of the losses made by falsely assuming an equal lag-length or using an "unnecessarily" general procedure when the assumption holds. Using an equal lag-length procedure and consistent information criteria will, of course, asymptotically select the true or maximum lag-length. This is, however, less clear with the sample sizes common in empirical work. This paper extends the existing literature by investigating the performance of equal and unequal lag-length selection procedures when the true model has equal as well as unequal lag-length. In doing this we consider a variety of data generating processes designed to replicate salient features of macroeconomic data. To our knowledge, this is also the first systematic study of the performance of the Hsiao procedure.

In Section 2 we briefly discuss the lag-length selection strategies and the information criteria. Section 3 describes the Monte Carlo design. Section 4 presents the results and Section 5 concludes the paper.

5.2 Lag-length selection strategies

Under the assumption of equal lag-length the identification procedure is straightforward, using a multivariate information criteria or a sequential test procedure. The common lag-length is simply chosen as the one wich minimizes the information criteria. Identification procedures allowing for different lag-lengths are more complicated and are discussed in Section 2.2.

5.2.1 Information criteria

The Akaike (Akaike 1969, Akaike 1974) Information Criterion (AIC), is given by

$$AIC(p) = \ln \left| \hat{\Sigma}_p \right| + \frac{2k^2 p}{T}, \tag{5.2}$$

where T is the number of observations, k the dimension of the time series, p the estimated number of lags and $\hat{\Sigma}_p$ the estimated white noise covariance matrix. Shibata (1976) proves that the AIC criterion, in the univariate AR(p) representation, is inconsistent in the sense that asymptotically it overestimates the true order with a nonzero probability.

In contrast to AIC, HQ and BIC behave more parsimoniously and are known to be consistent, that is the probability of choosing the true lag-length goes to one asymptotically. The Schwarz (1978) criterion (BIC),

$$BIC(p) = \ln \left| \hat{\Sigma}_p \right| + \frac{k^2 p \ln T}{T}, \tag{5.3}$$

is strongly consistent for a standard VAR model, see Geweke and Meese (1981). The Hannan and Quinn (1979) criterion (HQ),

$$HQ(p) = \ln \left| \hat{\Sigma}_p \right| + \frac{2pk^2 ln \ln T}{T}, \tag{5.4}$$

is consistent for AR processes in the univariate case, and strongly consistent if $k > 1$. The multivariate representations, assuming equal lag-length, of AIC, HQ and BIC are as in equations (5.2) to (5.4), and for the equal lag-length procedure the order, $0 \leq p \leq P$, is simply chosen to minimize one of these criteria.

Some small sample properties of BIC, HQ and AIC were established by Lütkephol (1985). BIC is the most parsimonious criterion, $\hat{p}(BIC) \leq \hat{p}(HQ)$ for all T and $\hat{p}(BIC) \leq \hat{p}(AIC)$ for $T \geq 8$. For $T \geq 16$, HQ is more restrictive than AIC. Moreover Lütkepohl finds that the BIC and the HQ criteria underestimate, and the AIC criterion over-estimates the order in finite samples.

As an alternative to information criteria, a sequential test-procedure may be used. Given the assumed maximum lag-length, P, the successively more restrictive hypotheses, $A_P = 0$, $A_{P-1} = 0, \ldots$, are tested until the null is rejected, or $A_1 = 0$ is accepted. If the significance levels are allowed to vary for the individual tests it is possible to set the significance level in such a way that the outcome is identical to the one obtained from the use of an arbitrary information criterion, see Teräsvirta and Mellin (1986) for univariate models. In practice, sequential test-procedures are, in general, used with a common significance level. Note that the overall significance level will be considerably larger than the level for the individual tests.

5.2.2 Lag-length selection procedures without assumptions of equal lag-length

Relaxing the equal lag-length restriction and allowing for more flexibility in the model specification complicates matters since the number of possible combinations of lag-lengths increases dramatically. Consider one equation in a possibly unequal lag-length VAR model,

$$\begin{aligned} y_{i,t} &= v_i + a_{i\,1,1} y_{1,t-1} + \cdots + a_{i\,1,p_1} y_{1,t-p_1} + \cdots \\ &\quad + a_{i\,k,1} y_{k,t-1} + \cdots + a_{ik,p_k} y_{K,t-p_k} + u_i, \end{aligned} \tag{5.5}$$

and the task of selecting the k lag-lengths, $p_1, ..., p_k$. If the largest lag-length considered is P there are $(P+1)^k$ different combinations of lag-length in one equation to investigate. For the whole system there are $(P+1)^{k^2}$ possible specifications to investigate. If every possible combination of lag-lengths is to be investigated, we have to restrict ourselves to small models. This is rather unsatisfactory,

TABLE 5.1. Modified Hsiao procedure.

Step 1:	Calculate the information criterion for lags 0 to P of the dependent variable. Choose the lag that minimizes the criterion.
Step 2:	Calculate the information criterion for lags 1 to P for each of the variables not yet included in the model. Choose the combination of variable and lag which gives the lowest value for the criterion, and add the variable to the model if the criterion is lower than in the previous step. Repeat until no more variables are added.

and in order to simplify the computations, variations of the Hsiao (1979) procedure together with univariate information criteria are frequently used.

The first step in the Hsiao procedure is to rank the variables according to their potential predictive power for each dependent variable. Then we consider one equation at a time and investigate the lags of the variables in the order implied by the ranking. This approach reduces the number of regressions to be estimated to at most $(2k-1)(P+1)$. To rank the variables according to their potential predictive power can sometimes be difficult, and Edlund and Karlsson (1993) avoid the ranking of variables at the cost of increasing the number of regressions to be estimated. In the modified procedure the first step selects the number of lags of the dependent variable. In the second step the variables not yet included in the equation are evaluated up to P lags. The combination of variable and lag-length with the smallest information criteria is added to the equation if this leads to a reduction in the information criteria for the equation. This is repeated until no more variables are added. With this modification, the maximum number of regressions to be estimated is $k(k+1)P/2$. The procedure is summarized in Table 5.1.

The Hsiao procedure does not investigate all combinations of lag-length. Consequently there is no guarantee that the correct specification is tried. Therefore the method is not consistent, even if consistent information criteria are used.

TABLE 5.2. Properties of the generated VAR models.

Model	Lag-Length	Variables	Dynamics	Lag Structure
1	1	3	Real roots	Unequal lag-length
2	2	2	Compex roots: period 7.5	Unequal lag-length
3	3	3	Complex roots: period 5 & 2.5	Equal lag-length
4	3	3	Complex roots: period 20 & 4	Equal lag-length
5	3	3	Same as 3	Unequal, Triangular
6	3	3	Same as 4	Unequal, Triangular
7	4	3	Real roots	Unequal, zero coefficients in all parameter matrices

5.3 Model selection and design of the Monte Carlo simulation

The Monte Carlo simulation is designed to analyze and compare the properties of the different lag-length selection strategies and information criteria in finite samples. We use seven VAR models, specified to mimic various aspects of the dynamics typically displayed by macro-economic time series. The true lag-length varies from one to four lags, with both equal and unequal lag-length structure. The properties of the models are summarized in Table 5.2 and the full specification is given in Table 5.3. Note that models 3 and 5 and models 4 and 6, respectively, have the same dynamic properties, i.e. the same dominating roots.

The first two models are provided as a reference to the smaller models with lag-lengths 1 and 2 previously investigated by Lütkephol (1985). Model 2 has complex roots corresponding to a cyclical component with a cycle length of 7.5 observations. Models 3 to 6 are a group of lag-length models that are specified to mimic the salient features of (stationary) macroeconomic time series. Models 3 and 5 have been specified with cycle components of lengths 5 and 2.5 observations, and models 4 and 6 have cyclical components with lengths of 20 and 4 observations. That is, models 3 and 5 correspond to yearly data with a business cycle component and models 4 and 6 correspond to quarterly data with business cycle and seasonal components. Of these, Models 3 and 4 have equal lag-length and Models 5 and6 have "triangular", unequal lag-length, lag structures. Model 7 has a maximum lag-length of 4 and a more complex lag structure. The lag structure is not triangular and contains zero elements in various positions. The estimation of lag-length for this model should be more challenging task for the procedures.

For each VAR model, two sets of 1000 samples are generated. One set is obtained using a diagonal variance-covariance matrix (the identity matrix, I) in the innovations u_t, and the second from a non-diagonal variance-covariance matrix. The non-diagonal variance-covariance matrices are

$$\Sigma_2 = \begin{bmatrix} 0.09 & 0.03 \\ 0.03 & 0.04 \end{bmatrix} \text{ and } \Sigma_3 = \begin{bmatrix} 2.25 & 0.2 & 0.8 \\ 0.2 & 1.0 & 0.5 \\ 0.8 & 0.5 & 0.74 \end{bmatrix}$$

for the two and three variable models, respectively.

The RAN1 subroutine of Press, Teukolsky, Vetterling and Flannery (1985) and a transformation routine (GASDEV) are used to generate the standard normal pseudo random numbers. The first 100 observations are discarded, and the samples of size 50, 100 and 200 used in identification and estimation are obtained from the remaining observations.

5.4 Simulation results

Tables of results are presented in Appendices A, B and C. In Appendix A, Tables 5.4-5.17, present numbers of correct, under- and overestimated lag-lengths for the equal and the unequal procedures, estimated for a sample size of 50 observations. Corresponding tables for sample sizes 100 and 200 are given in appendix B and C. For the Hsiao (unequal lag-length) procedure we report the number of correct, under and overestimated lag-lengths for each equation and variable. With the equal lag-length procedure, the selected lag-length is common to all equations and variables and the correct lag-length is taken to be the largest lag in the VAR model.

TABLE 5.3. Parameter values of Model of the generated VAR-models.

A_1			A_2			A_3		
Model	1							
0.5	0	0						
0.1	0.1	0.3						
0	0.3	0.3						
Model	2							
0.5	0.1		0	0				
0.4	0.5		0.25	0				
Model	3							
-0.90	0.60	1.00	-0.90	0.60	1.00	0.05	-0.09	0.16
-0.27	0.56	1.74	-0.27	0.56	1.74	0.26	-0.32	-0.40
0.30	0.16	0.68	0.30	0.16	0.68	-0.08	0.11	-0.12
Model	4							
-0.09	0.24	0.91	-0.30	-0.09	0.045	0.03	-0.10	0.21
0.23	0.25	1.86	0.18	-0.62	0.18	0.30	-0.40	0.56
0.07	0.29	0.65	-0.07	0.063	-0.50	-0.09	0.14	-0.16
Model	5							
-0.11	0	0	-0.34	0	0	-0.38	0	0
2.23	-0.85	0	-0.16	-0.29	0	-1.06	-0.03	0
-1.82	1.07	1.30	0.13	-0.08	-0.44	0.32	0.03	0.03
Model	6							
-0.29	0	0	-0.62	0	0	-0.49	0	0
-0.26	-0.20	0	-0.77	-0.36	0	-1.24	-0.07	0
-0.66	0.75	1.30	0.30	-0.4	-0.44	0.36	0.05	0.03
Model	7							
0	-1.00	0	1.13	-0.10	0	0	0.30	0
0	0.70	0	0	0.38	0	0	-0.24	0
0.20	0.90	0.8	1.20	-0.50	-0.25	-0.30	0	0
						A_4		
						-0.31	0	0
						0	0	0
						0	0	0

In addition to the AIC, BIC and HQ information criteria we also used the Final Prediction Error (FPE) criterion (Akaike 1969) to select the lag-lengths. We do not report the results for FPE since they are very similar to the ones for AIC. For small sample sizes the AIC criterion overestimates the lag-length slightly more than FPE, while for larger sample sizes the results are almost identical. The latter is in agreement with Lütkephol (1985), who found that the two criteria are asymptotically equivalent, i.e. in a Taylor expansion it can be seen that they only differ by a term of order T^{-2}. Finally, we only report results for VAR models generated using the identity residual variance-covariance matrix, since the results using non-diagonal variance-covariance matrixes are very similar.

5.4.1Result for $T = 50$

When discussing the results of the simulation study we concentrate on the results for $T = 50$, which is the most challenging sample size in the Monte Carlo study. In Section 5 we discuss the effects of increasing the sample size.

5.4.1.1 Simulation results for Model 1 and Model 2

Model 1 and Model 2 are small, unequal lag-length models. In Model 1 we have three variables and a maximum lag-length of 1,while Model 2 has 2 variables and lag-length 2.

For Model 1 and the equal lag-length procedure (Table 5.4) the BIC and HQ criteria estimate the correct order in almost all replicates. The AIC overestimates the order in 10% of the replicates.

If an unequal lag-length procedure is used with Model 1 (Table 5.5 in Appendix A), AIC and HQ give the best overall performance. However, the BIC criterion estimates the zero lag-lengths better than AIC and HQ. In Equation 2 all the variables enter with lag-length 1 (the maximum lag-length in the model). This poses a challenge to the unequal lag-length procedure and all three information criteria frequently underestimate the lag-lengths in this equation.

When the equal lag-length procedure is used with Model 2 (Table 5.6) all criteria underestimate the lag-lengths to a greater extent than in Model 1. The AIC criterion estimates the correct maximum lag-length most frequently and underestimates less often than either BIC or HQ. AIC is also the criterion that overestimates the lag-length most frequently. The BIC criterion underestimates the lag-length in almost all replications and HQ in 63% of the replications.

Using the unequal lag-length procedure, AIC and HQ perform best with Model 2 (Table 5.7). Again, BIC underestimates the lag-length most frequently and the AIC overestimates most frequently. With Model 2 the first equation is the most challenging and all criteria tend to underestimate the lag-length of the second variable. This can be explained by the second variable having a small absolute parameter value on the last (and only) lag. The effect of dropping the variable from the equation is thus relatively small and the unequal lag-length procedure has difficulties picking up this effect.

5.4.1.2 Simulation results for the large models

Models 3 and 4 are equal lag-length VAR models with a lag-length of 3 and Models 5 and 6 are their unequal lag-length counterparts, also with a maximum lag-length of 3. Model 7, finally, is a more complicated unequal lag-length model with a maximum lag-length of 4.

Turning first to the equal lag-length models and the equal lag-length procedure, the HQ and AIC criteria give the best results. HQ has a tendency to underestimate the lag-length and AIC tends to overestimate. For Model 3 (Table 5.8) they select the correct lag-length equally often. With Model 4 (Table 5.10) HQ selects the correct lag-length in 85% and AIC in 68% of the replicates. BIC consistently underestimates the lag-length and selects the correct lag-length in less than 1% of the replicates.

Using the unequal lag-length procedure (Tables 5.9 and 5.11) all criteria underestimate the true lag-length to a large extent. With the exception of a few combinations of equation and variable, BIC underestimates the lag-length in more than 90% of the replicates. AIC and HQ estimate the correct lag-length in equally often, 3%-40% of the replicates depending on variable and equation. In line with their general tendency, AIC overestimates the lag-length to a greater extent and HQ underestimates more frequently than AIC. As with the equal lag-length procedure, the performence is slightly better with Model 4 than with Model 3.

For the lag-length 3, unequal lag-length, Models 5 and 6 the performance of the equal lag-length procedure (Tables 5.12 and 5.14) is very similar to the equal lag-length Models 3 and 4. For Model 6 all criteria perform better than with Model 5, AIC and HQ underestimate the lag-length in 2% of the replicates and BIC estimates the correct lag-length in 29% of the replicates. This is similar to Model 4 which also has a cyclical component with long period which leads to larger lag-lengths being selected.

Turning to the unequal lag-length procedures it is worth noting that Models 5 and 6 are triangular systems. For equation 1, where only the dependent variable enters, both AIC and HQ perform very well. They identify the zero coefficients and estimate the true lag-length for the dependent variable in over 60% of the replicates. As can be expected, BIC identifies the zero coefficients but also tends to underestimate the lag-length of the dependent variable. The underestimation is most severe in Model 5; for Model 6 BIC underestimates the lag-length in only 19% of the replicates. For Equation 2, where two variables enter the equation with lag-length 3, the results differ markedly between the models. In Model 5 AIC and BIC overestimate the lag-length of the dependent variable in over 90% of the replicates. For Model 6 they estimate the lag-length correctly in about 50% of the replicates and underestimate more frequently than they overestimate. The lag-length of variable 1 is estimated correctly in a majority of the replicates and more frequently in Model 6, where the lag-length is never underestimated, than in Model 5. In the third equation all three variables enter with 3 lags and the results are very similar to the results for the equal lag-length models.

For the relatively complicated, unequal lag-length, Model 7 the equal lag-length procedures fails, estimating the correct lag-length in less than 10% of the replicates. There is a clear tendency to underestimate the lag-length even for AIC. With the unequal lag-length procedure the variables not entering into the equations (variable 3 in Equation 1 and variables 1 and 3 in Equation 2) are identified in the majority of the replicates even for AIC. Variable 1 in Equation 3 is the only case where we have a lag-length of 4 and this is identified correctly in a majority of the replicates, except for BIC which underestimates the lag-length. For the remaining combinations of variable and equation the true lag-length is 2 or 3 and these are estimated less well. The lag-length of the dependent variable is overestimated and the lag-length of other variables is underestimated.

5.4.2 Increasing the sample size

As can be expected, increasing the sample size to 100 and 200 observations improves the performance of both the equal and the unequal procedure. Consequently, we will restrict ourselves to a discussion of the differences between the procedures and criteria as the sample size increases.

Using equal lag-length procedures, the performance of AIC and HQ improves for all models although AIC overestimates the lag-length in about 10% of the replicates even for large samples. BIC consistently underestimates the lag-length even for $T = 200$ when the true lag-length is greater than 1. Exceptions to this are Models 4 and 6. For Model 4 there is a dramatic improvement for BIC when going from $T = 100$ to $T = 200$ and in Model 6 BIC estimates the lag-length correctly in 94% of the replicates even for $T = 100$. For Model 7 there is little improvement with the increase in sample size; only AIC shows a marked improvement, selecting the correct lag-length in 42% of the replicates when $T = 200$.

In the unequal lag-length procedure the frequency of underestimated lag-lengths decreases whereas the frequency of overestimates changes little when the sample size increases for Models 1, 2, 3 and 6. The pattern with "difficult" equations where all criteria tend to underestimate the lag-length of one or more variables remains, even if the frequency of underestimation decreases compared to $T = 50$. HQ and AIC tend to overestimate the lag-length of the dependent variable to a greater extent as the sample size increases. This is especially so in Models 4 and 5, where even BIC tends to overestimate the lag-length of the dependent variable for large T. BIC, on the other hand, tends to underestimate the lag-length of variables other than the dependent even for $T = 200$. The results for the third equation in Models 5 and 6 show a markedly different pattern. The lag-length of the second variable is underestimated by all three criteria in almost all replicates and the lag-length of the first variable is overestimated in over 60% of the replicates. For Model 5 the lag-length of the dependent variable tends to be underestimated by HQ and BIC, whereas AIC and HQ overestimate in Model 6. With Model 7 there is a gradual improvement with the increase in sample size, but the same general pattern as for $T = 50$ emerges. In particular we note that the lag-length of the dependent variable in the third equation is always overestimated by all three criteria.

5.5 Conclusions

Choosing the lag-length in VAR models is not an easy task and the cost of using a mis-specified model depends on the intended use of the model. If inference and hypothesis testing is the primary concern it is important to avoid under-parameterization. In a forecasting application, on the other hand, the bias introduced by under-parameterization may well be offset by the reduction in the variance of the estimates.

When choosing a lag-length selection strategy it is wise to keep these issues in mind. The BIC criterion is highly parsimonious and underestimates the true lag-length in almost all cases considered here, even for large sample sizes. This is contrary to the finding of Lütkephol (1985) that BIC estimates the lag-length correctly more frequently than other criteria. This discrepancy is due to Lütkepohl only considering models with lag-lengths 1 and 2, where the scope for underestimating the lag-length is small. The BIC criteria may thus be useful in forecasting applications, but should be avoided when inference is the primary concern.

The AIC and HQ criteria are less parsimonious and estimate the correct lag-length better than BIC. For larger models we can see a tendency for HQ to select the correct lag-length more frequently than AIC. The performance of AIC and especially HQ also improves faster with increasing sample size than BIC. AIC underestimates the lag-length less frequently than HQ and may thus be the preferred criteria when performing inference.

The main concern of this paper is the distinction between lag-length selection strategies which assume equal lag-lengths and procedures which allow for different lag-lengths. In the small Models 1 and 2, the difference in performance between the equal and unequal lag-length procedures is small. These are unequal lag-length models and the equal lag-length procedure tends to underestimate the maximum lag-length in the model. The unequal lag-length procedure estimates most of the individual lag-lengths correctly in a majority of the cases, but there are also one or more combinations of equation and variable where the lag-length is rarely estimated correctly.

For the larger 3 variable, 3 lag, Models 3 to 6, the equal lag-length procedure performs reasonably well and selects the maximum lag-length correctly in a majority of the replicates when AIC or HQ is used. The difference in performance between the equal and unequal lag-length models is very small for the equal lag-length procedure. The unequal lag-length procedure identifies the zero lag-lengths in the unequal lag-length models and in general performs well in equations where one or more variables are excluded. In equations where all variables enter and in the equal lag-length models the unequal lag-length procedure performs less well.

With the more complicated Model 7, the equal lag-length procedure underestimates the lag-length. The unequal lag-length procedure identifies the zero lag-lengths and estimates the correct lag-length with a high frequency for about half of the variables. In most of the remaining combinations of equation and variable there is a clear tendency to underestimate the lag-length.

It is thus difficult to give clear guidelines for the choice between equal and unequal lag-length procedures. The performance seems to be highly model dependent. A tentative conclusion is that the equal lag-length procedure works well with equal lag length models as well as unequal-lag length models with a relatively simple lag structure, such as Models 5 and 6. For models with more complicated lag structures, such as Model 7 where there are holes in the lag polynomials, the unequal lag-length procedure may be a better choice.

In practice, it is not certain that the data generating process is a VAR model. The issue then chances from one of finding the true model to one of finding the best approximation within the class of VAR models. It is unclear how the results in this paper apply to the latter issue. Although we have made general comments about desirable properties of the lag-length selection procedure when performing inference and when forecasting, it is unclear how the actual choice of lag-length selection procedure affects inference and the forecasting performance. These and other issues are left for future research.

Acknowledgments

We thank Tor Jacobson for helpful suggestions. Comments from participants at the International Symposium on Forecasting in Toronto, 1995, are gratefully acknowledged. Finanical support from the Swedish Research Council for the Humanities and the Social Sciences (HSFR) is gratefully acknowledged.

References

Akaike, H. (1969), 'Fitting autoregressive models for prediction', *Annals of the Institute of Mathematical Statistics* **21**, 243-247.

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**,716-723.

Boswijk, P. & Franses, P. (1992), 'Dynamic specification and cointegration', *Oxford Bulletin of Economics and Statistics* **54**, 369-380.

Cheung, Y. W. & Lai, K. (1993), 'Finite-sample size of Johansen's likelihood ratio test for cointegration', *Oxford Bulletin of Economics and Statistics* **55**, 313-328.

DeSerres, A. & Guay, A. (1995), Selection of truncation lag in structural VAR's (or VECM's) with long-run restriction, 95-9, Bank of Canada.

Edlund, P. O. & Karlsson, S. (1993), 'Forecasting the swedish unemployment rate. VAR vs. transfer function modelling', *International Journal of Forecasting* **9**, 61-76.

Geweke, J. F. & Meese, R. A. (1981), 'Estimating regression models of finite but unknown order', *International Economic Review* **22**, 55-70.

Hannan, E. J. & Quinn, B. G. (1979), 'The determination of the order of an autoregression', *Journal of the Royal Statistical Society* pp. 190-195.

Hsiao, C. (1979),' Autoregressive modelling of canadian money and income data', *Journal of the American Statistical Association* **74**, 553-560.

Hsiao, C. (1981),' Autoregressive modelling and money-income causality detection', *Journal of Monetary Economics* **7**, 85-106.

Jacobson, T. (1995), 'On the determination of lag order in vector autoregressions of cointegrated systems', *Computational Statistics* **10**, 177-192.

Kunitomo, N. & Yamamoto, T. (1985), 'Properties of predictors in misspecified autoregressive time series models', *American Statistical Association* **80**, 941-950.

Lütkepohl, H. (1985), 'Comparisons of criteria for estimating the order of a vector autoregressive process', *Journal of Time Series Analysis* **6**, 35-52.

Nickelsburg, G. (1985), 'Small-sample properties of dimensionally statistics for fitting VAR models to aggregate economic data', *Journal of Econometrics* **2**, 183-192.

Press, Teukolsky, s., Vetterling, W. & Flannery, B. (1985), *Numerical Recipes in Fortran: The Art of Sientific Computing*, Cambridge university press.

Schwarz, G. (1978), 'Estimating the dimensio of a model', *The Annals of Statistics* **6**, 461-464.

Shibata, R. (1976), 'Selection of the order of an autoregressive model by Akaike's information criterion', *Biometrica* **63**, 117-126.

Teräsvirta, T & Mellin, I. (1986), 'Model selection criteria and model selection test in regression models', *Scandinavian Journal of Statistics* **13**, 159-171.

Appendix A: Selection of lag-lengths, T=50

Tables in Appendix A give the frequencies (out of 1000 replicates) of correct-, under and over-estimation of the lag lengths for T=50. Results are given for both the equal and the unequal lag length procedures, using the AIC, HQ and BIC information criteria.

TABLE 5.4. Selected lag-lengths: Model 1, equal procedure,$T = 50$.

	1 lag, Correct estimation	Overestimation
AIC	895	105
HQ	987	13
BIC	1000	0

True model has unequal lag-lengths with maximum lag equal to one.

TABLE 5.5. Selected lag-lengths: Model 1, unequal procedure, $T = 50$.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	24	0	0	718	696	718	258	304	282
	2	507	296	433	282	482	361	211	222	206
	3	0	463	64	686	372	696	314	165	240
HQ	1	50	0	0	796	833	850	154	167	150
	2	630	410	541	263	471	361	107	119	98
	3	0	559	99	824	351	780	176	90	121
BIC	1	276	0	0	716	991	991	8	9	9
	2	931	834	810	67	166	189	2	0	1
	3	0	823	425	990	176	572	10	1	3

True model has unequal lag-lengths with maximum lag equal to one.

TABLE 5.6. Selected lag-lengths: Model 2, equal procedure, $T = 50$.

	Under-estimation	Correct estimation (2 lags)	Over-estimation
AIC	391	470	249
HQ	631	332	47
BIC	996	4	0

True model has unequal lag-lengths with maximum lag equal to two.

TABLE 5.7. Selected lag-lengths: Model 2, unequal procedure, $T = 50$.

		Under-estimation		Correct estimation		Over-estimation	
Crit.	Eq.	Variable		Variable		Variable	
		1	2	1	2	1	2
AIC	1	25	735	727	111	248	154
	2	154	0	608	666	248	334
HQ	1	44	830	829	98	127	72
	2	226	0	630	786	144	214
BIC	1	221	984	776	15	3	1
	2	654	4	334	978	12	18

True model has unequal lag-lengths with maximum lag equal to two.

TABLE 5.8. Selected lag-lengths: Model 3, equal procedure, $T = 50$.

	Under-estimation	Correct estimation	Over-estimation
AIC	124	599	277
HQ	368	598	144
BIC	1000	0	0

True model has equal lag-length of three.

TABLE 5.9. Selected lag-lengths: Model 3, unequal procedure, $T = 50$.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	489	669	295	88	87	186	432	244	519
	2	484	515	171	174	172	345	342	313	484
	3	779	816	532	129	51	216	92	133	252
HQ	1	673	829	508	73	70	148	254	101	344
	2	665	707	263	151	145	339	184	148	398
	3	824	916	625	117	31	195	59	53	180
BIC	1	998	994	999	1	5	1	1	1	0
	2	834	983	874	166	15	52	0	2	74
	3	987	1000	986	12	0	14	1	0	0

True model has equal lag-length of three.

TABLE 5.10. Selected lag-lengths: Model 4, equal procedure, T=50.

	Under-estimation	Correct estimation	Over-estimation
AIC	19	682	299
HQ	114	848	38
BIC	991	9	0

True model has equal lag-lengths with maximum lag equal to three.

TABLE 5.11. Selected lag-lengths: Model 4, unequal procedure, T=50.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	537	693	182	227	192	126	236	115	692
	2	342	88	169	305	101	368	356	811	463
	3	750	596	345	75	105	87	175	299	568
HQ	1	741	635	318	190	199	99	69	166	583
	2	534	177	240	294	140	393	172	683	367
	3	920	691	548	38	85	84	42	224	368
BIC	1	992	938	972	8	61	5	0	1	23
	2	975	729	645	24	186	126	1	85	229
	3	997	982	968	1	3	21	2	15	11

True model has equal lag-lengths with maximum lag equal to three.

TABLE 5.12. Selected lag-lengths: Model 5, equal procedure, $T = 50$.

	Under-estimation (1 lag)	Under-estimation (2 lags)	Correct estimation (3 lags)	Over-estimation ($\geq$ 4 lags)
AIC	0	73	630	363
HQ	0	227	734	39
BIC	234	764	2	0

True model has unequal lag-lengths with maximum lag equal to three.

TABLE 5.13. Selected lag-lengths: Model 5, unequal procedure, $T = 50$.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	95	0	0	622	729	670	283	271	330
	2	119	3	0	532	43	546	349	996	454
	3	42	651	600	229	107	126	729	242	274
HQ	1	179	0	0	673	837	807	148	162	193
	2	187	6	0	628	91	712	185	903	288
	3	57	767	760	317	108	112	625	125	128
BIC	1	825	0	0	172	961	993	3	39	7
	2	501	342	0	406	281	983	93	377	17
	3	340	968	979	528	23	15	132	9	6

True model has unequal lag-lengths with maximum lag equal to three.

TABLE 5.14. Selected lag-lengths: Model 6, equal procedure, T=50.

	Under-estimation (1 lags)	Under-estimation (2 lags)	Correct estimation (3 lags)	Over-estimation (≥4 lags)
AIC	0	1	699	300
HQ	0	2	943	55
BIC	192	521	287	0

True model has unequal lag-lengths with maximum lag equal to three.

TABLE 5.15. Selected lag-lengths: Model 6, unequal procedure, T=50.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	0	0	0	683	719	643	317	281	357
	2	0	267	0	654	449	575	346	284	425
	3	86	532	204	118	214	236	796	254	560
HQ	1	23	0	0	802	863	801	175	136	199
	2	0	383	0	800	477	750	200	140	250
	3	135	663	322	138	199	299	727	138	379
BIC	1	189	0	0	704	981	985	107	19	15
	2	3	811	0	977	185	987	30	4	13
	3	758	951	914	60	44	72	182	5	14

True model has unequal lag-lengths with maximum lag equal to three.

TABLE 5.16. Selected lag-lengths: Model 7, equal procedure, $T = 50$.

	Under-estimation (2 lags)	Undert-esimation (3 lags)	Correct estimation (4 lags)	Over-estimation (≥ 5 lags)
AIC	574	134	81	211
HQ	913	64	10	13
BIC	1000	0	0	0

True model has unequal lag-lengths with maximum lag equal to four.

TABLE 5.17. Selected lag-lengths: Model 7, unequal procedure, $T = 50$.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	65	635	0	586	101	604	349	264	396
	2	0	374	0	664	367	702	336	259	298
	3	594	271	5	125	305	21	281	424	974
HQ	1	122	809	0	668	83	785	210	108	225
	2	0	438	0	807	330	814	193	232	186
	3	748	409	17	106	344	29	146	247	954
BIC	1	509	998	0	481	2	970	10	0	30
	2	0	957	0	980	42	961	20	1	39
	3	980	886	249	14	91	60	6	23	691

True model has unequal lag-lengths with maximum lag equal to four.

Appendix B: Selection of lag-lengths, T=100

Tables in Appendix B give the frequencies (out of 1000 replicates) of correct-, under and over-estimation of the lag lengths for T=100. Results are given for both the equal and the unequal lag length procedures, using the AIC, HQ and BIC information criteria.

TABLE 5.18. Selected lag-lengths: Model 1, equal procedure, T=100.

	1 lag, Correct estimation,	Overestimation
AIC	936	64
HQ	999	1
BIC	1000	0

True model has unequal lag-lengths with maximum lag equal to one.

TABLE 5.19. Selected lag-lengths: Model 1, unequal procedure, T=100.

		Underestimation			Correct estimation			Overestimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	0	0	0	764	739	760	236	361	240
	2	366	91	256	421	639	543	213	270	201
	3	0	266	6	711	531	771	289	203	223
HQ	1	3	0	0	884	884	899	113	116	101
	2	510	182	369	395	693	544	95	125	87
	3	0	417	11	867	492	882	133	91	107
BIC	1	37	0	0	961	997	999	2	3	1
	2	917	635	688	82	362	302	1	3	10
	3	0	828	97	998	172	902	2	0	1

True model has unequal lag-lengths with maximum lag equal to one.

TABLE 5.20. Selected lag-lengths: Model 2, equal procedure, T=100.

	Under-estimation	Correct estimation (2 lags)	Over-estimation
AIC	144	716	140
HQ	405	577	18
BIC	995	5	0

True model has unequal lag-lengths with maximum lag equal to two.

TABLE 5.21. Selected lag-lengths: Model 2, unequal procedure, T=100.

		Under-estimation		Correct estimation		Over-estimation	
Crit.	Eq.	Variable		Variable		Variable	
		1	2	1	2	1	2
AIC	1	1	661	693	187	306	152
	2	34	0	721	635	245	365
HQ	1	2	802	844	150	154	48
	2	76	0	810	796	114	204
BIC	1	22	974	975	26	3	0
	2	422	0	574	990	4	10

True model has unequal lag-lengths with maximum lag equal to two.

TABLE 5.22. Selected lag-lengths: Model 3, equal procedure, T=100.

	Under-estimation	Correct estimation	Over-estimation
AIC	9	908	85
HQ	87	911	2
BIC	998	2	0

True model has equal lag-lengths with maximum lag equal to three.

TABLE 5.23. Selected lag-lengths: Model 3, unequal procedure, T=100.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	416	677	91	71	131	197	513	192	712
	2	294	332	41	364	197	492	352	471	467
	3	698	748	355	155	127	311	147	125	334
HQ	1	630	872	215	65	75	205	305	53	580
	2	503	563	79	293	198	561	204	139	360
	3	794	903	581	154	47	305	52	49	114
BIC	1	992	984	987	7	14	4	1	2	9
	2	774	984	582	215	12	174	11	4	244
	3	969	1000	974	31	0	25	0	0	1

True model has equal lag-lengths with maximum lag equal to three.

TABLE 5.24. Selected lag-length:s Model 4, equal procedure, T=100.

	Under-estimation	Correct estimation	Over-estimation
AIC	0	897	103
HQ	2	991	7
BIC	855	145	0

True model has equal lag-lengths with maximum lag equal to three.

TABLE 5.25. Selected lag-lengths: Model 4, unequal procedure, T=100.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	262	469	48	426	284	156	312	247	798
	2	120	3	43	535	14	419	245	983	538
	3	751	438	102	125	198	47	224	364	851
HQ	1	496	570	116	395	315	128	109	115	756
	2	278	26	105	589	57	528	133	917	367
	3	907	669	288	59	125	97	33	206	615
BIC	1	975	817	867	25	174	5	0	9	128
	2	935	360	370	63	294	357	2	346	253
	3	999	922	946	1	5	33	0	73	21

True model has equal lag-lengths with maximum lag equal to three.

TABLE 5.26. Selected lag-lengths Model 5, equal procedure, T=100.

	Under-estimation (1 lag)	Under estimation (2 lags)	Correct estimation (3 lags)	Over-estimation (≥4 lag)
AIC	0	2	907	91
HQ	0	33	961	6
BIC	0	974	26	0

True model has unequal lag-lengths with maximum lag equal to three.

TABLE 5.27. Selected lag-lengths Model 5, unequal procedure, T=100.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	6	0	0	726	782	733	268	218	267
	2	12	0	0	683	0	640	305	1000	360
	3	1	757	471	88	102	210	911	141	319
HQ	1	25	0	0	859	900	882	116	100	118
	2	32	0	0	842	10	843	126	990	157
	3	1	894	718	172	73	166	827	43	116
BIC	1	412	0	0	584	982	999	4	18	1
	2	299	15	0	697	214	993	4	761	7
	3	35	982	987	657	17	11	308	1	2

True model has unequal lag-lengths with maximum lag equal to three.

TABLE 5.28. Selected lag-lengths: Model 6, equal procedure, T=100.

	Under-estimation (1 lag)	Under-estimation (2 lags)	Correct estimation (3 lags)	Over-estimation (≥4 lags)
AIC	0	0	906	94
HQ	0	0	991	9
BIC	0	60	940	0

True model has unequal lag-lengths with maximum lag equal to three.

TABLE 5.29. Selected lag-lengths: Model 6, unequal procedure, T=100.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	0	0	0	720	795	709	280	205	291
	2	0	111	0	705	609	641	295	280	359
	3	15	575	41	117	239	207	868	186	752
HQ	1	1	0	0	880	913	878	119	87	122
	2	0	197	0	878	686	857	122	117	143
	3	50	723	85	143	216	331	807	61	574
BIC	1	35	0	0	961	999	996	4	1	4
	2	0	670	0	996	327	998	4	3	2
	3	461	937	728	158	58	232	381	5	40

True model has unequal lag-lengths with maximum lag equal to three.

TABLE 5.30. Selected lag-lengths: Model 7, equal procedure, T=100.

	Under-estimation (2 lags)	Under-estimation (3 lags)	Correct estimation (4 lags)	Over-estimation ($\geq$5 lags)
AIC	468	337	157	38
HQ	878	111	11	0
BIC	1000	0	0	0

True model has unequal lag-lengths with maximum lag equal to four. The Correct estimation is conditional on the lag -length restriction.

TABLE 5.31. Selected lag-lengths: Model 7, unequal procedure, T=100.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	4	493	0	544	284	711	452	323	289
	2	0	144	0	732	578	788	268	278	212
	3	584	97	0	202	552	1	214	351	999
HQ	1	19	732	0	706	212	888	275	56	112
	2	0	291	0	882	589	895	118	120	105
	3	769	195	0	154	647	1	77	158	999
BIC	1	187	0	0	795	0	983	15	1000	17
	2	0	853	0	992	147	986	8	0	14
	3	980	680	35	17	312	12	3	8	953

True model has unequal lag-lengths with maximum lag equal to four.

Appendix C: Selection of lag-lengths, T=200

Tables in Appendix C give the frequencies (out of 1000 replicates) of correct-, under and over-estimation of the lag lengths for T=200. Results are given for both the equal and the unequal lag length procedures, using the AIC, HQ and BIC information criteria.

TABLE 5.32. Selected lag-lengths: Model 1, equal procedure, T=200.

	1 lag, Correct estimation,	Overestimation
AIC	946	54
HQ	1000	0
BIC	1000	0

True model has unequal lag-lengths with maximum lag equal to one.

TABLE 5.33. Selected lag-lengths: Model 1, unequal procedure, T=200.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	0	0	0	760	763	748	240	237	252
	2	156	7	59	618	611	736	226	382	205
	3	0	95	0	725	698	753	275	207	247
HQ	1	0	0	0	910	916	900	90	84	100
	2	285	27	129	628	788	795	87	185	76
	3	0	185	1	893	745	906	107	70	93
BIC	1	0	0	0	1000	1000	1000	0	0	0
	2	823	264	540	172	730	459	1	6	1
	3	0	691	5	998	309	995	2	0	0

True model has unequal lag-lengths with maximum lag equal to one.

TABLE 5.34. Selected lag-lengths: Model 2, equal procedure, T=200.

	Under-estimation	Correct estimation, (2 lags)	Over-estimation
AIC	14	854	142
HQ	137	853	10
BIC	962	38	0

True model has unequal lag-lengths with maximum lag equal to two.

TABLE 5.35. Selected lag-lengths: Model 2, unequal procedure, T=200.

		Under-estimation		Correct estimation		Over-estimation	
Crit.	Eq.	Variable		Variable		Variable	
		1	2	1	2	1	2
AIC	1	0	547	563	271	437	172
	2	0	0	758	552	232	448
HQ	1	0	678	770	267	230	55
	2	2	0	902	752	96	248
BIC	1	1	950	989	50	10	0
	2	141	0	859	989	0	11

True model has unequal lag-lengths with maximum lag equal to two.

TABLE 5.36. Selected lag-lengths: Model 3, equal procedure, T=200.

	Under-estimation	Correct estimation	Over-estimation
AIC	0	953	47
HQ	0	999	1
BIC	829	171	0

True model has equal lag-lengths with maximum lag equal to three.

TABLE 5.37. Selected lag-lengths: Model 3, unequal procedure, T=200.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	231	583	9	33	250	176	736	167	815
	2	61	137	1	613	164	566	236	699	423
	3	675	552	130	186	284	378	139	164	492
HQ	1	540	818	36	38	151	217	422	31	747
	2	217	332	6	555	211	685	228	457	409
	3	782	813	340	183	121	449	35	66	211
BIC	1	987	994	755	8	6	27	5	0	218
	2	889	950	172	65	40	474	46	10	454
	3	921	1000	933	78	0	67	1	0	0

True model has equal lag-lengths with maximum lag equal to three.

TABLE 5.38. Selected lag-lengths: Model 4, equal procedure, T=200.

	Under-estimation	Correct estimation	Over-estimation
AIC	0	946	54
HQ	0	999	1
BIC	131	869	0

True model has equal lag-lengths with maximum lag equal to three.

TABLE 5.39. Selected lag-lengths: Model 4, unequal procedure, T=200.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	41	413	6	493	364	88	466	223	906
	2	5	0	1	696	0	343	296	1000	656
	3	642	416	3	186	302	6	178	282	991
HQ	1	150	572	18	621	343	134	229	85	848
	2	31	0	3	852	1	590	135	999	407
	3	887	647	49	82	237	39	31	116	912
BIC	1	894	698	481	105	284	42	1	18	477
	2	597	59	249	402	139	677	1	802	74
	3	999	747	816	1	13	80	0	240	104

True model has equal lag-lengths with maximum lag equal to three.

TABLE 5.40. Selected lag-lengths: Model 5, equal procedure, T=200.

	Under-estimation (2 lags)	Correct estimation (3 lags)	Over-estimation ($\geq$4 lags)
AIC	0	937	63
HQ	0	998	2
BIC	550	450	0

True model has unequal lag-lengths with maximum lag equal to three.

TABLE 5.41. Selected lag-lengths: Model 5, unequal procedure, T=200.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	0	0	0	710	798	766	290	202	234
	2	0	0	0	713	0	684	287	1000	316
	3	0	787	276	19	99	240	981	134	484
HQ	1	0	0	0	897	919	921	103	81	79
	2	0	0	0	912	0	875	88	1000	125
	3	1	943	579	31	39	241	968	18	180
BIC	1	0	0	0	964	994	999	36	6	1
	2	45	0	0	954	15	997	1	985	3
	3	1	998	975	384	1	25	615	1	0

True model has unequal lag-lengths with maximum lag equal to three.

TABLE 5.42. Selected lag-lengths: Model 6, equal procedure, T=200.

	Under-estimation (2 lags)	Correct estimation (3 lags)	Over-estimation (≥4 lags)
AIC	0	942	58
HQ	0	997	3
BIC	0	1000	0

True model has unequal lag-lengths with maximum lag equal to three.

TABLE 5.43. Selected lag-lengths: Model 6, unequal procedure, T=200.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	0	0	0	718	815	715	282	185	285
	2	0	14	0	724	676	700	276	310	300
	3	0	655	0	51	190	79	949	155	921
HQ	1	0	0	0	897	938	918	103	62	82
	2	0	40	0	905	833	875	95	127	125
	3	1	792	5	84	169	190	925	39805	
BIC	1	0	0	0	997	999	1000	3	1	0
	2	0	348	0	996	651	997	4	1	3
	3	73	864	208	220	135	589	707	1	203

True model has unequal lag-lengths with maximum lag equal to three.

TABLE 5.44. Selected lag-lengths: Model 7, equal procedure, T=200.

	Under-estimation (2 lags)	Under-stimation (3 lags)	Correct estimation (4 lags)	Over-estimation (≥5 lags)
AIC	88	444	419	49
HQ	571	382	47	0
BIC	1000	0	0	0

True model has unequal lag-lengths with maximum lag equal to four. The Correct estimation is conditional on the lag -length restriction.

TABLE 5.45. Selected lag-lengths: Model 7, unequal procedure, T=200.

		Under-estimation			Correct estimation			Over-estimation		
Crit.	Eq.	Variable			Variable			Variable		
		1	2	3	1	2	3	1	2	3
AIC	1	0	218	0	347	539	747	653	243	253
	2	0	17	0	766	718	813	234	265	187
	3	487	3	0	280	667	0	233	330	1000
HQ	1	1	441	0	533	488	902	466	71	98
	2	0	49	0	941	869	934	59	86	66
	3	693	15	0	249	850	0	58	135	1000
BIC	1	23	949	0	920	51	997	57	0	3
	2	0	539	0	994	461	994	6	0	6
	3	968	329	0	32	670	0	0	21	1000

True model has unequal lag-lengths with maximum lag equal to four.

6

Power and Bias of Likelihood Based Inference in the Cointegration Model under Fractional Cointegration

ABSTRACT: This paper investigates how fractional cointegration affects the common maximum likelihood cointegration procedure. It is shown that the likelihood ratio test of no cointegration has considerable power against fractional alternatives. In contrast to the case of a cointegrated system, the usual maximum likelihood estimator gives severely biased estimates of the long-run relation under fractional cointegration. This suggests that the standard likelihood approach should be used with caution and that a test to separate fractionally cointegrated series from series that are cointegrated of an integer order should be executed prior to estimation.

KEY WORDS: Error correction; Likelihood ratio test; Maximum likelihood; Fractional integration.
JEL-CLASSIFICATION: C12; C32

6.1 Cointegration and Fractional Integration

A time series process $\mathbf{x}_t$ is said to be cointegrated of order $CI\left(\delta, b\right)$ if the variables of $\mathbf{x}_t$ individually are integrated of order $I\left(\delta\right)$, while a linear combination of the variables, denoted $\beta'\mathbf{x}_t$, is $I\left(\delta - b\right)$. The variables are in equilibrium if $\beta'\mathbf{x}_t$ equals some constant $\boldsymbol{\mu}$, but in most time periods $\mathbf{x}_t$ is not in equilibrium and the quantity $\mathbf{z}_t = \beta'\mathbf{x}_t$

may be called the equilibrium error. Commonly in applied work δ and b are both equal to unity, but there are examples of analysis of $I(2)$ processes.

The $\delta = b = 1$ case is appealing to empirical economists since it provides a framework to estimate long-run steady states, given by economic theory, using stationary linear combinations of non-stationary variables. However, the notion of cointegration may be generalized to real values, that is allowing for fractional δ and b. The theory of fractional integration was introduced by Granger and Joyeux (1980) and Hosking (1981), and is considered when modelling persistence in time series. The distinction between $I(0)$ and $I(1)$ is rather arbitrary; the relevant concept is mean-reversion in the equilibrium error. Mean-reversion does not require a strictly $I(0)$ process; the effect of a shock also dies out, although at a slow hyperbolic rate, for an $I(d)$ process with $d < 1$. Moreover, a similar interpretation as in the $\delta = b = 1$ case is possible within the fractional framework. If $\delta > 1/2$ and $\delta \geq b > \delta - 1/2$ the variables are non-stationary (their variances are infinite) but there exists a stationary linear combination (a long-run relationship) of the variables.

Cointegrated systems, where δ and b are integer-valued, are usually analyzed using (consistent and efficient) likelihood based inference, see Johansen (1995). The nice properties of the maximum likelihood estimation (*MLE*) procedure are likely to be sacrificed when the convergence to equilibrium follows the slow rate of a fractionally integrated specification, due to mis-specification of the likelihood function. For such circumstances, this paper demonstrates that the ML estimates of the long-run relationship are severely biased. Furthermore, it is shown that the likelihood ratio test for no cointegration has considerable power against both cointegrated and fractionally cointegrated alternatives. Thus it is not possible, using the LR test, to discriminate between cointegrated and fractionally cointegrated systems.

The large bias and high power under fractional cointegration implies that the use of the standard likelihood based approach requires careful consideration. It is therefore necessary to determine if the system is fractionally cointegrated prior to estimation. This can be done by testing for fractional integration in the equilibrium error.

For power comparisons the Engle-Granger (1987) procedure with the Dickey-Fuller (1979) and the Geweke and Porter-Hudak (1983) tests are included in the study.

6.2 Testing and Estimation

Two common approaches in the literature when testing for cointegration and estimation of cointegrated models are the Engle and Granger (1987) two-step procedure and the maximum likelihood procedure. The testing procedure of Engle and Granger first estimates the cointegrating relation, that is the linear combination

$$x_{1t} = \varphi x_{2t} + \eta_t.$$

The residual series η_t is interpreted as the equilibrium error. Cheung and Lai (1993) show that the usual least squares estimator $\hat{\varphi}$ converges in probability to φ also for fractionally integrated η_t. If the series are not cointegrated, η_t contains a unit root. In the second step we test the null-hypothesis of a unit root by the augmented Dickey-Fuller, ADF, test or the test for fractional integration of Geweke and Porter-Hudak, GPH. In this study the GPH regression is based on T^v ordinates, where v equals 0.5 (the recommendation of GPH) and 0.9, the value that maximizes the power when testing for cointegration according to Andersson and Lyhagen (1997).

The maximum likelihood approach is based on the p-dimensional vector error correction model $(VECM)$,

$$\Delta \mathbf{x}_t = \Pi \mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \quad t = 1, ..., T, \tag{6.1}$$

where Π and Γ_i are $p \times p$ parameter matrices, $\boldsymbol{\mu}$ is an intercept and $\boldsymbol{\varepsilon}_t$ an *iid* $N(\mathbf{0}, \Omega)$ error term. Using the maximum likelihood estimation procedure of Johansen (1988), the long run properties and short run dynamics are analyzed jointly and efficient inference is possible. Following Johansen, the first step is to determine the rank r of the cointegration space spanned by β. If $r = 0$, the series are $I(1)$ but not cointegrated and if $r = p$ then $\mathbf{x}_t$ is weakly stationary and there are no unit roots. When $0 < r < p$, there is reduced rank and the parameter matrix may be decomposed as $\Pi = \alpha\beta'$, where α

and β are $p \times r$ matrices. Then $\beta' \mathbf{x}_t$ is $I(0)$ and hence mean-reverting. Given the estimated rank r, the model is estimated by maximization of the likelihood function or equivalently by solving an eigenvalue problem. Details of the estimation procedure are given by Johansen (1995).

Testing for a reduced rank is performed by the *trace* statistic, which is the likelihood ratio test statistic for the hypothesis of at most r cointegrated vectors. The LR statistic is given by

$$-2\log\left(H\left(r\right)|H\left(p\right)\right) = -T\sum_{i=r+1}^{p}\log\left(1-\hat{\lambda}_i\right), \tag{6.2}$$

where $1 > \hat{\lambda}_1 > ... > \hat{\lambda}_p > 0$ are the eigenvalues that maximize the likelihood function. The statistic has an asymptotic distribution which can be expressed as functions of Brownian motions. The percentiles of the asymptotic distribution are tabulated by Johansen (1995).

6.3 The Power of the Tests

The experiment examining the power of the LR test is based, exactly as in Chung and Lai (1993), on the bivariate system

$$\begin{aligned} x_{1t} + x_{2t} &= u_{1t} \\ x_{1t} + \alpha x_{2t} &= u_{2t}, \end{aligned} \tag{6.3}$$

where α equals two and $(1-B)\,u_{1t} = \varepsilon_{1t}$. u_{2t} is generated as an autoregression, $(1-\phi B)\,u_{2t} = \varepsilon_{2t}$, or as the fractional specification $(1-B)^d u_{2t} = \varepsilon_{2t}$. The system is constructed so that $\delta = 1$ and $b = 1 - d$. The members of $\{\varepsilon_{1t}\}$ and $\{\varepsilon_{2t}\}$ are *iid* $N(0,1)$. When $\phi = 1$ or $d = 1$ the processes are not cointegrated, whereas $\phi < 1$ implies cointegration and $d < 1$ fractional cointegration.

Table 6.1 reports the power of the LR test against autoregressive alternatives at an exact[1] 5% level. The likelihood ratio test is very powerful against AR alternatives when $\phi \leq 0.75$ and $T = 100$, whereas $T = 50$ requires a parameter of 0.55 to give a high power.

---

[1] Empirical null distributions of the tests are given in *Appendix A*.

TABLE 6.1. Power, in percent, of the cointegration tests against autoregressive alternatives. The size of the tests equals five percent.

	ϕ									
Test	.95	.85	.75	.65	.55	.45	.35	.25	.15	.05
	T=50									
L1	5.9	13.5	30.3	55.6	79.5	93.8	98.8	99.8	100	100
L2	5.9	11.8	23.1	39.1	55.9	71.0	82.2	90.0	94.8	97.1
G5	6.1	11.9	21.6	32.1	41.1	47.5	51.6	53.3	54.3	53.5
G9	6.4	14.2	30.8	52.3	72.5	86.4	93.6	97.3	98.8	99.5
A	7.2	18.4	39.2	62.1	79.9	90.3	95.8	98.2	99.2	99.6
	T=100									
L1	9.0	42.0	86.1	99.3	100	100	100	100	100	100
L2	8.6	35.8	73.7	94.5	99.2	99.9	100	100	100	100
G5	8.2	26.6	46.5	59.9	65.5	67.2	67.0	65.8	63.9	61.7
G9	8.9	36.7	75.1	94.6	99.2	99.9	100	100	100	100
A	11.8	55.0	91.1	99.2	99.9	100	100	100	100	100

In the Table margin, L1 and L2 denote the LR test where the number of lags in (6.1) are 1 and 2 respectively. For the ADF (A) test, l is the number of augmentation lags and T^v the number of ordinates in the GPH regression (G5 has v=0.5 and G9 v=0.9). All tests are size adjusted and the power is obtained through 50,000 replicates.

The results also suggest that the test is sensitive to mis-specification; a power loss is found when $k = 2$. This power loss decreases with the serial length and significance level. Boswijk and Franses (1992) show that an under-specified error correction model leads to considerable size distortions and over-specification to a severe power loss. The ADF test with augmentation lag one and GPH test with $v = 0.9$ each has a power similar to the LR test with $k = 1$. Moreover, at serial length $T = 200$ all tests except the GPH with $v = 0.5$ are very powerful already at $\phi = 0.85$.[2]

Results in Table 6.2 show that the LR test also has quite high power against fractional alternatives when $d \leq 0.55$ and $T = 100$. However, for the shorter serial length a similar high power is not experienced until $d \leq 0.35$. The power loss, due to over-specification of the lag-length in the VECM, is more pronounced than in the autoregressive case. For fractional alternatives, the GPH test with $v = 0.9$ has higher power than the LR and ADF tests.

The results suggest that the likelihood ratio test also has high power for fractional alternatives, which implies that we cannot dis-

---

[2] Results for $T = 200$ and the 1% and 10% levels of significance are given in *Appendix B.*

TABLE 6.2. Power, in percent, of the cointegration tests against fractional alternatives. The size of the tests equals five percent.

	d									
Test	.95	.85	.75	.65	.55	.45	.35	.25	.15	.05
					T=50					
L1	5.3	8.8	16.8	31.5	51.8	67.6	92.7	99.2	100	100
L2	5.0	6.2	9.0	14.0	22.1	30.6	53.6	74.8	89.0	96.3
G5	6.2	8.9	12.1	16.6	21.8	26.1	34.4	42.1	47.4	51.6
G9	7.4	15.0	27.7	45.9	64.8	78.0	91.7	97.0	98.8	99.5
A	5.8	9.3	14.9	24.3	37.3	49.0	74.4	90.5	97.3	99.4
					T=100					
L1	5.7	12.9	30.4	57.4	81.5	91.0	99.8	100	100	100
L2	5.1	8.4	15.6	29.1	49.7	66.8	94.1	99.7	100	100
G5	6.3	9.8	14.0	20.5	27.9	34.0	45.7	53.0	57.5	59.6
G9	9.2	24.7	52.6	79.6	94.1	98.0	100	100	100	100
A	6.2	12.4	24.8	43.6	65.3	78.3	97.4	99.9	100	100

See note to Table 6.1.

criminate between cointegration and fractional cointegration using the LR or any of the other tests.

6.4 Properties of the ML Estimator

As seen in the previous section, we are likely to find evidence of (say $CI\,(1,1)$) cointegration when the true process is fractionally cointegrated, $CI\,(1,1-d)$. To examine the estimation procedure we again generate data according to (6.3). The bias and root mean squared error ($RMSE$) are used to evaluate the maximum likelihood estimator of the Π matrix, i.e. the long-run relationship.

A general error correction representation for the system (6.3) is expressed as

$$\Delta \mathbf{x}_t = \Pi D\,(d,B)\,\mathbf{x}_t + \varepsilon_t,$$

where $D\,(d,B) = (1-B)^d - (1-B)$ is a scalar operation on $\mathbf{x}_t$. The common CI model is obtained when $d=0$ since $D\,(d,B)$ then collapses to just the backshift operator B. In the autoregressive case the true Π matrix is given by

$$\Pi = (\alpha-1)^{-1}\begin{bmatrix} (1-\phi) & \alpha\,(1-\phi) \\ -\,(1-\phi) & -\alpha\,(1-\phi) \end{bmatrix},$$

where ϕ is the AR parameter in the disturbance process u_{2t} and α is equal to two.

When u_{2t} is generated as a fractionally integrated process, Π reduces to

$$\Pi = (\alpha - 1)^{-1} \begin{bmatrix} 1 & \alpha \\ -1 & -\alpha \end{bmatrix}.$$

Under both specifications the variance-covariance matrix Ω is given by

$$\begin{aligned} E\left[\varepsilon_t \varepsilon_t'\right] &= (\alpha - 1)^{-2} \begin{bmatrix} \alpha & -1 \\ -1 & 1 \end{bmatrix} E\left[\boldsymbol{\xi}_t \boldsymbol{\xi}_t'\right] \begin{bmatrix} \alpha & -1 \\ -1 & 1 \end{bmatrix}' \\ &= \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}, \end{aligned} \tag{6.4}$$

where the members of $\{\boldsymbol{\xi}_t\}$ are *iid* $N(\mathbf{0}, \mathbf{I})$.

Larger parameter values than those presented in the tables give more biased estimates and a higher RMSE. Moreover, the opposite is valid for parameters smaller than those reported. Setting k equal to two does not alter the findings, we merely experience a higher estimation bias and RMSE for all specifications.

In the autoregressive case, see Table 6.3, the bias of the estimated Π matrix is quite small for all specifications and decreases with the sample size. To illuminate further, the relative bias (bias/π_{ij}) is almost identical for all estimates of the elements of Π, and it decreases rapidly with the distance between ϕ and the unit-circle.

According to the definition of the mean square error, the MSE may be decomposed into the variance of the parameter estimate and the square of the bias. In the autoregressive case, the variance dominates the MSE and thus the RMSE. Table 6.3 shows that the root of the mean squared error is fairly stable for each element of $\hat{\Pi}$ (that is $\hat{\pi}_{ij}$, $i, j = 1, 2$), but increases slightly as ϕ decreases. However, if the RMSE is adjusted for the size of the true parameter, yielding the ratio RMSE/π_{ij}, we notice that this quantity decreases with ϕ. Moreover, the RMSE/π_{ij} ratio for $\hat{\pi}_{11}$ and $\hat{\pi}_{12}$ is notably larger than those for $\hat{\pi}_{21}$ and $\hat{\pi}_{22}$; this is an effect of the design of the disturbance covariance matrix (6.4).

Table 6.4 presents the bias and RMSE when the series are fractionally cointegrated according to a fractionally differenced white noise process with parameter d. The absolute value of the bias of $\hat{\Pi}$ is very large for the specifications presented, for instance when $T = 100$ and

TABLE 6.3. Bias and RMSE for the maximum likelihood estimation procedure when the equilibrium error follows an AR(1) process. The lag-length k=1.

	ϕ				
	$T = 50$				
Test	.85	.65	.55	.45	.25
π_{11}					
True	.150	.350	.450	.550	.750
Bias	.041	.048	.045	.041	.032
RMSE	.253	.291	.305	.314	.327
π_{12}					
True	.300	.700	.900	1.10	1.50
Bias	.104	.103	.095	.085	.065
RMSE	.490	.579	.609	.627	.655
π_{21}					
True	-.150	-.350	-.450	-.550	-.750
Bias	-.051	-.051	-.048	-.043	-.033
RMSE	.167	.187	.193	.199	.206
π_{22}					
True	-.300	-.700	-.900	-1.10	-1.50
Bias	-.113	-.106	-.098	-.087	-.066
RMSE	.324	.371	.385	.396	.410
	$T = 100$				
	.85	.65	.55	.45	.25
π_{11}					
True	.150	.350	.450	.550	.750
Bias	.027	.029	.024	.021	.017
RMSE	.149	.188	.200	.210	.223
π_{12}					
True	.300	.700	.900	1.10	1.50
Bias	.060	.060	.050	.043	.034
RMSE	.297	.376	.401	.420	.447
π_{21}					
True	-.150	-.350	-.450	-.550	-.750
Bias	-.030	-.029	-.025	-.022	-.017
RMSE	.098	.120	.127	.133	.141
π_{22}					
True	-.300	-.700	-.900	-1.10	-1.50
Bias	-.063	-.059	-.050	-.044	-.033
RMSE	.193	.239	.254	.265	.281

The simulation results are based on 50,000 replicates.

TABLE 6.4. Bias and RMSE for the maximum likelihood estimation procedure when the equilibrium error follows a fractionally integrated process.

	d				
	$T = 50$				
Test	.75	.55	.25	.15	.05
π_{11}					
True	1.00	1.00	1.00	1.00	1.00
Bias	-.832	-.643	-.243	-.126	-.027
RMSE	.879	.723	.417	.361	.335
π_{12}					
True	2.00	2.00	2.00	2.00	2.00
Bias	-1.62	-1.26	-.481	-.249	-.054
RMSE	1.71	1.41	.832	.721	.670
π_{21}					
True	-1.00	-1.00	-1.00	-1.00	-1.00
Bias	.809	.628	.240	.125	.027
RMSE	.832	.669	.326	.249	.211
π_{22}					
True	-2.00	-2.00	-2.00	-2.00	-2.00
Bias	1.60	1.24	.478	.249	.054
RMSE	1.64	1.32	.647	.494	.419
	$T = 100$				
	.75	.55	.25	.15	.05
π_{11}					
True	1.00	1.00	1.00	1.00	1.00
Bias	-.881	-.703	-.274	-.145	-.038
RMSE	.897	.735	.359	.272	.233
π_{12}					
True	2.00	2.00	2.00	2.00	2.00
Bias	-1.74	-1.39	-.547	-.289	-.077
RMSE	1.77	1.45	.717	.544	.466
π_{21}					
True	-1.00	-1.00	-1.00	-1.00	-1.00
Bias	.868	.695	.273	.145	.038
RMSE	.876	.713	.313	.208	.150
π_{22}					
True	-2.00	-2.00	-2.00	-2.00	-2.00
Bias	1.72	1.38	.545	.172	.090
RMSE	1.74	1.42	.625	.415	.299

See note to Table 6.3.

$d = 0.75$ we notice a bias of $86 - 88\%$ of the true parameter. Furthermore, the relative bias is constant over true Π, decreases with d and is not negligible until d equals 0.05, which is almost a white noise process. Unlike in the AR case, the estimates become more biased as the serial length increases (see also $T = 200$ in *Appendix B*). Thus we find no evidence of consistency of the ML estimation procedure under fractional cointegration.

Contrary to the AR case, the squared bias dominates the RMSE under fractional cointegration. Thus the effect of the variance of the parameter estimates on the RMSE is quite small, but increases as d tends to zero. The impact of the bias to the RMSE implies that the RMSE$/\pi_{ij}$ ratios are similar for large d. For small d, the RMSE$/\pi_{ij}$ ratios are slightly larger for $\hat{\pi}_{11}$ and $\hat{\pi}_{12}$, because of the smaller bias and the design of the disturbance covariance matrix. Moreover, the RMSE decreases with d, which is mainly an effect of the reduced bias.

6.5 Conclusions

This paper investigates the maximum likelihood cointegration procedure when the process is fractionally cointegrated. For the sake of comparison, the common case of cointegrated series is also included in the study. The results suggest that the likelihood ratio test for cointegration also has high power against fractional alternatives, and hence possesses the ability to detect slow mean-reversion in the equilibrium error. However, if fractional cointegration is present, the usual maximum likelihood procedure may lead to incorrect inference since persistence in the equilibrium error will then be modelled by an $I(0)$ instead of an $I(d)$ specification.

The maximum likelihood procedure works well for cointegrated systems, that is the estimates exhibit a small bias and seem to be consistent. Unfortunately this appears not to be the case for fractionally cointegrated series, where the likelihood function is incorrectly specified. Consequently, the ML estimation technique produces strongly biased estimates.

The high power of the LR test against fractional alternatives and the severely biased ML estimates under fractional cointegration suggest that the standard likelihood based approach should be used with caution. In particular, if the equilibrium error is likely to be ruled by persistence we recommend that a secondary test should be used to separate fractionally cointegrated series from series that are cointegrated of an integer order. This may be conducted by the Engle-Granger procedure combined with the bootstrap tests of Andersson and Gredenhoff (1997, 1998), which are robust to AR and MA components.

Acknowledgments

The authors would like to thank Tor Jacobson and Sune Karlsson for valuable comments. Financial support from the Tore Browaldh Foundation is gratefully acknowledged. The usual disclaimer applies.

References

Andersson, J. and J. Lyhagen, 1997, A note on the power of the GPH test for cointegration, Research Report 97:7, Department of Statistics, Uppsala University, Sweden.

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, Sweden.

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, Sweden.

Boswijk, P. and P.H. Franses, 1992, Dynamic specification and cointegration, Oxford Bulletin of Economics and Statistics 54, 369-380

Cheung, Y.-W. and K.S. Lai, 1993, A fractional cointegration analysis of purchasing power parity, Journal of Business and Economic Statistics 11, 103-112.

Dickey, D.A. and W.A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74, 427-431.

Engle, R.F. and C.W.J. Granger, 1987, Co-integration and error correction: representation, estimation, and testing, Econometrica 55, 251-276.

Geweke, J. and S. Porter-Hudak, 1983, The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221-238.

Granger, C.W.J. and R. Joyeux, 1980, An introduction to long-memory time series models and fractional integration, Journal of Time Series Analysis 1, 15-29.

Hosking, J.R.M., 1981, Fractional differencing, Biometrica 68, 165-176.

Johansen, S., 1988, Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control 12, 231-254

Johansen, S., 1995, Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, Oxford University Press, Oxford, UK.

Appendix A: Empirical Null Distribution of the Tests

TABLE 6.5. Emprical distribution of the tests under no cointegration. Sample size T=50.

Perc.	LR,k=1	LR,k=2	GPH,v=0.5	GPH,v=0.9	ADF,l=1
0.5	1.0497	1.1254	-3.1229	-3.6328	-3.8556
1	1.2671	1.3542	-2.6974	-3.2466	-3.5844
2.5	1.6386	1.7715	-2.1527	-2.6889	-3.2217
5	2.0379	2.1935	-1.7367	-2.2022	-2.9124
10	2.5981	2.7883	-1.2983	-1.7007	-2.5858
20	3.4689	3.7039	-0.8144	-1.1292	-2.1982
30	4.2246	4.5121	-0.4948	-0.7314	-1.9338
40	4.9608	5.2966	-0.2339	-0.4104	-1.7135
50	5.7266	6.1059	0.0066	-0.1231	-1.5054
60	6.5816	7.0162	0.2402	0.1575	-1.2913
70	7.5854	8.0929	0.4875	0.4578	-1.0548
80	8.9151	9.4888	0.7753	0.7933	-0.7532
90	11.0113	11.6747	1.1821	1.2326	-0.2438
95	12.9535	13.7821	1.5203	1.5918	0.2107
97.5	14.8347	15.7109	1.8279	1.8942	0.6463
99	17.1305	18.3777	2.2011	2.2404	1.1485
99.5	18.8596	20.2265	2.4752	2.4776	1.5009
Mean	6.3772	6.7967	-0.0362	-0.1887	6.3772
Std	3.4448	3.6611	1.0015	1.1631	3.4448
Skew	1.1437	1.1698	-0.3298	-0.3749	1.1437
Kurt	4.8727	5.0755	3.8341	3.3923	4.8727

In the table head, k is the number of lags in (6.1) for the LR test. For the ADF test, l is the number of augmentation lags and T^v the number of ordinates in the GPH regression. The distribution is obtained through 100,000 replicates.

TABLE 6.6. Emprical distribution of the tests under no cointegration. Sample size T=100.

Perc.	LR,k=1	LR,k=2	GPH,v=0.5	GPH,v=0.9	ADF,l=1
0.5	1.0209	1.0784	-3.1039	-3.5263	-3.6912
1	1.2337	1.2931	-2.7189	-3.1368	-3.4582
2.5	1.6019	1.6643	-2.1737	-2.6116	-3.1258
5	2.0033	2.0684	-1.7613	-2.1801	-2.8424
10	2.5569	2.6914	-1.3299	-1.6957	-2.5257
20	3.4051	3.5181	-0.8452	-1.1320	-2.1575
30	4.1351	4.2742	-0.5204	-0.7410	-1.8964
40	4.8553	5.0158	-0.2525	-0.4197	-1.6802
50	5.3955	5.7761	-0.0091	-0.1230	-1.4762
60	6.4466	6.6378	0.2273	0.1657	-1.2688
70	7.4243	7.6487	0.4772	0.4678	-1.0386
80	8.7218	8.9677	0.7670	0.8062	-0.7411
90	10.7265	11.0434	1.1691	1.2627	-0.2433
95	12.6053	13.0021	1.5019	1.6372	0.2235
97.5	14.3833	14.8694	1.8048	1.9502	0.6292
99	16.7144	17.2541	2.1749	2.3138	1.1133
99.5	18.3801	19.1332	2.4207	2.5459	1.4558
Mean	6.2341	6.4304	-0.0549	-0.1782	-1.4239
Std	3.3579	3.4623	1.0045	1.1644	0.9217
Skew	1.1663	1.1896	-0.3376	-0.2906	0.4035
Kurt	5.1207	5.2392	3.7438	3.2402	3.7767

See note to table (6.5).

TABLE 6.7. Emprical distribution of the tests under no cointegration. Sample size T=200.

Perc.	LR,k=1	LR,k=2	GPH,v=0.5	GPH,v=0.9	ADF,l=1
0.5	1.0262	1.0352	-3.0609	-3.4473	-3.6608
1	1.2293	1.2613	-2.7011	-3.0866	-3.4108
2.5	1.6157	1.6472	-2.1863	-2.5678	-3.0801
5	1.9919	2.0324	-1.7877	-2.1556	-2.8062
10	2.5496	2.5817	-1.3593	-1.6846	-2.5037
20	3.3823	3.4343	-0.8731	-1.1241	-2.1335
30	4.1015	4.1762	-0.5426	-0.7388	-1.873
40	4.8019	4.8848	-0.2746	-0.4216	-1.6547
50	5.5405	5.6261	-0.0276	-0.1315	-1.4503
60	6.3614	6.4599	0.2133	0.1544	-1.2481
70	7.3321	7.4579	0.4681	0.4561	-1.0211
80	8.6036	8.7285	0.7571	0.8044	-0.7301
90	10.5889	10.7491	1.1601	1.2727	-0.2503
95	12.4654	12.6758	1.4956	1.6587	0.2029
97.5	14.2643	14.5313	1.7906	1.9744	0.6102
99	16.5892	16.8683	2.1472	2.3426	1.0963
99.5	18.2103	18.5489	2.3728	2.5939	1.4125
Mean	6.1687	6.2674	-0.0715	-0.1744	-1.4063
Std	3.3158	3.3702	1.0066	1.1596	0.9092
Skew	1.1776	1.1836	-0.3098	-0.2438	0.3893
Kurt	2.1635	2.1728	0.5581	0.2064	3.7846

See note to table (6.5).

Appendix B: Additional Power, Bias and RMSE Tables

TABLE 6.8. Power, in percent, of the cointegration tests against autoregressive alternatives. Sample size 50.

	ϕ									
Test	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05
	Size 1%									
L1	1.3	3.3	9.0	23.0	45.7	70.6	88.8	96.9	99.5	99.9
L2	1.2	2.6	6.2	13.3	23.4	35.7	49.3	62.6	73.9	82.3
G5	1.3	2.5	5.4	9.0	12.7	15.9	18.6	20.1	20.8	20.8
G9	1.4	3.3	9.1	20.1	36.4	55.1	70.7	81.9	89.5	93.8
A	1.6	4.5	12.3	26.2	43.6	60.8	74.8	85.0	91.3	94.8
	Size 5%									
L1	5.9	13.5	30.3	55.6	79.5	93.8	98.8	99.8	100	100
L2	5.9	11.8	23.1	39.1	55.9	71.0	82.2	90.0	94.8	97.1
G5	6.1	11.9	21.6	32.1	41.1	47.5	51.6	53.3	54.3	53.5
G9	6.4	14.2	30.8	52.3	72.5	86.4	93.6	97.3	98.8	99.5
A	7.2	18.4	39.2	62.1	79.9	90.3	95.8	98.2	99.2	99.6
	Size 10%									
L1	11.7	24.1	47.1	73.5	91.1	98.1	99.7	100	100	100
L2	11.4	21.6	38.0	57.2	73.3	85.3	92.6	96.3	98.3	99.3
G5	12.1	22.0	36.0	49.4	59.0	65.2	68.8	70.4	70.8	69.9
G9	12.5	25.3	47.3	70.1	85.9	94.4	97.8	99.2	99.7	99.9
A	13.8	32.4	58.2	79.6	91.4	96.7	98.8	99.6	99.9	99.9

See note to table (6.1).

TABLE 6.9. Power, in percent, of the cointegration tests against autoregressive alternatives. Sample size 100.

	ϕ									
Test	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05
					Size 1%					
L1	1.9	14.6	53.5	90.8	99.4	100	100	100	100	100
L2	1.8	11.3	38.7	72.2	91.3	98.1	99.7	99.9	100	100
G5	1.7	6.7	15.8	24.2	29.8	32.1	32.3	32.0	30.8	29.6
G9	2.0	13.0	43.1	76.5	93.4	98.5	99.7	99.9	100	100
A	2.6	21.1	61.6	90.5	98.5	99.9	100	100	100	100
					Size 5%					
L1	9.0	42.0	86.1	99.3	100	100	100	100	100	100
L2	8.6	35.8	73.7	94.5	99.2	99.9	100	100	100	100
G5	8.2	26.6	46.5	59.9	65.5	67.2	67.0	65.8	63.9	61.7
G9	8.9	36.7	75.1	94.6	99.2	99.9	100	100	100	100
A	11.8	55.0	91.1	99.2	99.9	100	100	100	100	100
					Size 10%					
L1	17.0	60.1	94.8	99.9	100	100	100	100	100	100
L2	16.5	53.7	87.4	98.3	99.9	100	100	100	100	100
G5	15.9	42.4	64.3	75.4	79.6	80.8	80.4	79.2	77.5	75.5
G9	16.6	53.4	87.2	98.1	99.8	100	100	100	100	100
A	22.1	74.4	97.4	99.9	100	100	100	100	100	100

See note to table (6.1).

TABLE 6.10. Power, in percent, of the cointegration tests against autoregressive alternatives. Sample size 200.

	ϕ									
Test	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05
					Size 1%					
L1	5.6	72.7	99.9	100	100	100	100	100	100	100
L2	5.3	62.4	98.4	100	100	100	100	100	100	100
G5	3.9	28.5	52.1	62.1	65.5	65.1	63.0	60.1	57.1	53.6
G9	4.6	52.5	94.7	99.8	100	100	100	100	100	100
A	8.3	83.5	99.8	100	100	100	100	100	100	100
					Size 5%					
L1	21.1	95.4	100	100	100	100	100	100	100	100
L2	20.1	90.9	99.9	100	100	100	100	100	100	100
G5	16.2	62.9	82.5	87.6	88.6	87.6	86.2	83.9	81.2	78.1
G9	16.8	81.2	99.4	100	100	100	100	100	100	100
A	30.0	98.4	100	100	100	100	100	100	100	100
					Size 10%					
L1	35.5	99.0	100	100	100	100	100	100	100	100
L2	33.9	97.3	100	100	100	100	100	100	100	100
G5	27.7	78.1	91.2	94.0	94.4	93.9	92.8	91.4	89.4	87.1
G9	28.5	90.6	99.8	100	100	100	100	100	100	100
A	47.7	99.7	100	100	100	100	100	100	100	100

See note to table (6.1).

TABLE 6.11. Power, in percent, of the cointegration tests against fractional alternatives. Sample size 50.

	d									
Test	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05
	Size 1%									
L1	1.1	2.2	5.1	12.4	26.3	42.5	75.5	93.8	99.1	99.9
L2	0.9	1.1	1.9	3.6	6.6	10.9	24.3	42.2	62.0	79.3
G5	1.2	1.8	2.6	3.9	5.3	6.7	10.0	13.1	16.1	19.3
G9	1.6	3.6	8.1	16.9	29.7	44.2	65.3	80.3	88.7	93.6
A	1.1	2.1	4.0	7.7	14.3	22.7	43.6	65.8	83.1	93.4
	Size 5%									
L1	5.3	8.8	16.8	31.5	51.8	67.6	92.7	99.2	100	100
L2	5.0	6.2	9.0	14.0	22.1	30.6	53.6	74.8	89.0	96.3
G5	6.2	8.9	12.1	16.6	21.8	26.1	34.4	42.1	47.4	51.6
G9	7.4	15.0	27.7	45.9	64.8	78.0	91.7	97.0	98.8	99.5
A	5.8	9.3	14.9	24.3	37.3	49.0	74.4	90.5	97.3	99.4
	Size 10%									
L1	10.5	16.0	27.5	45.1	65.7	78.3	96.7	99.8	100	100
L2	9.9	12.2	16.6	24.1	35.0	45.5	70.0	87.0	95.7	99.0
G5	12.1	16.7	22.2	29.0	36.5	42.0	52.0	59.8	65.0	68.7
G9	14.2	26.0	43.2	63.2	79.8	88.5	96.9	99.1	99.7	99.9
A	11.2	16.6	24.7	37.0	51.8	62.9	85.6	96.2	99.2	99.9

See note to table (6.1).

TABLE 6.12. Power, in percent, of the cointegration tests against fractional alternatives. Sample size 100.

	d									
Test	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05
	Size 1%									
L1	1.2	3.7	12.6	33.5	63.1	81.2	99.1	100	100	100
L2	1.1	1.9	4.5	11.1	24.8	41.8	79.2	96.9	99.8	100
G5	1.2	2.1	3.1	4.9	7.3	9.8	15.2	19.9	24.0	27.1
G9	2.1	7.7	23.6	51.3	77.8	91.6	99.1	99.9	100	100
A	1.4	3.4	8.7	20.3	40.2	58.5	90.0	99.2	100	100
	Size 5%									
L1	5.7	12.9	30.4	57.4	81.5	91.0	99.8	100	100	100
L2	5.1	8.4	15.6	29.1	49.7	66.8	94.1	99.7	100	100
G5	6.3	9.8	14.0	20.5	27.9	34.0	45.7	53.0	57.5	59.6
G9	9.2	24.7	52.6	79.6	94.1	98.0	100	100	100	100
A	6.2	12.4	24.8	43.6	65.3	78.3	97.4	99.9	100	100
	Size 10%									
L1	11.1	21.7	43.1	69.3	88.2	94.1	100	100	100	100
L2	10.3	15.2	25.7	42.4	63.5	77.4	97.4	99.9	100	100
G5	12.1	18.1	24.9	34.1	43.7	51.0	63.2	69.6	72.7	74.1
G9	16.9	38.6	68.2	89.0	97.1	99.0	100	100	100	100
A	12.0	21.4	37.3	57.6	76.1	85.3	98.8	100	100	100

See note to table (6.1).

TABLE 6.13. Power, in percent, of the cointegration tests against fractional alternatives. Sample size 200.

	d									
Test	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05
					Size 1%					
L1	1.4	6.5	26.4	62.1	87.2	94.8	100	100	100	100
L2	1.2	3.1	10.4	29.5	59.4	79.1	99.3	100	100	100
G5	1.5	2.6	5.1	9.1	15.7	23.0	36.0	45.0	50.5	52.3
G9	2.9	17.9	56.9	89.8	98.4	99.6	100	100	100	100
A	1.6	5.5	17.9	43.3	71.3	85.1	99.6	100	100	100
					Size 5%					
L1	6.3	18.7	48.4	79.6	94.0	97.9	100	100	100	100
L2	5.6	11.4	27.3	53.3	78.4	89.7	99.9	100	100	100
G5	6.7	11.8	20.0	31.5	44.5	55.6	70.3	76.4	78.3	77.4
G9	11.4	42.4	81.4	96.8	99.5	99.9	100	100	100	100
A	7.2	17.3	39.4	66.6	85.2	92.4	100	100	100	100
					Size 10%					
L1	12.2	28.7	60.3	86.3	96.2	98.8	100	100	100	100
L2	11.2	19.7	39.9	66.0	85.8	93.3	100	100	100	100
G5	13.1	21.2	33.4	47.8	61.5	71.6	83.2	86.7	87.6	86.7
G9	20.1	57.2	89.1	98.2	99.7	99.9	100	100	100	100
A	13.3	27.4	52.1	76.3	89.9	95.0	100	100	100	100

See note to table (6.1).

TABLE 6.14. Bias and RMSE for the maximum likelihood estimation procedure when the equilibrium error follows an AR(1) or a fractionally integrated process. The lag-length k=1.

	$T = 200$				
	ϕ				
Test	.85	.65	.55	.45	.25
π_{11}					
True	.150	.350	.450	.550	.750
Bias	.015	.014	.014	.011	.008
RMSE	.094	.127	.137	.145	.156
π_{12}					
True	.300	.700	.900	1.10	1.50
Bias	.032	.029	.027	.023	.015
RMSE	.189	.254	.274	.291	.311
π_{21}					
True	-.150	-.350	-.450	-.550	-.750
Bias	-.016	-.014	-.013	-.012	-.008
RMSE	.061	.081	.087	.092	.098
π_{22}					
True	-.300	-.700	-.900	-1.10	-1.50
Bias	-.033	-.029	-.027	-.023	-.016
RMSE	.122	.161	.173	.184	.197
	d				
	.75	.55	.25	.15	.05
π_{22}					
True	1.00	1.00	1.00	1.00	1.00
Bias	-.914	-.751	-.295	-.160	-.045
RMSE	.920	.765	.336	.227	.168
π_{22}					
True	2.00	2.00	2.00	2.00	2.00
Bias	-1.81	-1.49	-.589	-.320	-.090
RMSE	1.83	1.52	.671	.455	.335
π_{22}					
True	-1.00	-1.00	-1.00	-1.00	-1.00
Bias	.907	.747	.294	.159	.045
RMSE	.910	.765	.314	.191	.112
π_{22}					
True	-2.00	-2.00	-2.00	-2.00	-2.00
Bias	1.81	1.49	.587	.319	.090
RMSE	1.81	1.51	.627	.381	.223

See note to Table 6.3.

7

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

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.

7.1 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

---

[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).

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, 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, coin-

---

[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.

tegrated 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 Barndorrf-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 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.

7.2 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

$$\Pi(L)\, x_t = \varepsilon_t, \qquad (t = 1, 2, \ldots, T), \tag{7.1}$$

where $\Pi(L)$ is an $n \times n$ matrix polynomial of order p given by $\Pi(\lambda) = I_n - \sum_{j=1}^{p} \Pi_j \lambda^j$, 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 (7.1) yields a vector error correction, VECM, representation for x_t suitable for estimation of the cointegrating relationships. Letting $\Gamma(\lambda) = I_n - \sum_{i=1}^{p-1} \Gamma_i \lambda^i$ where $\Gamma_i = -\sum_{j=i+1}^{p} \Pi_j$ and $\alpha\beta' = \Pi = -\Pi(1)$ we get

$$\Gamma(L)\, \Delta x_t = \alpha\beta' x_{t-1} + \varepsilon_t, \qquad (t = 1, 2, \ldots, T), \tag{7.2}$$

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 (7.2) 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 (7.2) 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

$$W_{LR,1} = T \sum_{i=1}^{r} \ln \left[\frac{\left(1 - \widehat{\lambda}_{H,i}\right)}{\left(1 - \widehat{\lambda}_i\right)} \right], \tag{7.3}$$

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 $r\left(n-s\right)$ 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 $H\left(n \times r_1\right)$, whereas the remaining $r_2 = r - r_1$ relations are estimated without restrictions. In this case the test statistic is

$$W_{LR,2} = T\left[\sum_{i=1}^{r_1} \ln\left(1 - \widehat{\lambda}_{C.H,i}\right) + \sum_{i=1}^{r_2} \ln\left(1 - \widehat{\lambda}_{H,i}\right) - \sum_{i=1}^{r} \ln\left(1 - \widehat{\lambda}_i\right)\right], \tag{7.4}$$

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 $r_1\left(n-r\right)$ 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, $H\left(n \times r_1\right)$ is known and the maximum likelihood solutiton is found by an iterative algorithm, see Johansen (1995), which gives the test statistic as

$$W_{LR,3} = T\left[\sum_{i=1}^{r_1} \ln\left(1 - \widehat{\lambda}_{C.H,i}\right) + \sum_{i=1}^{r_2} \ln\left(1 - \widehat{\lambda}_{H,i}\right) - \sum_{i=1}^{r} \ln\left(1 - \widehat{\lambda}_i\right)\right], \tag{7.5}$$

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.

7.3 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.

7.3.1 The Bootstrap Test

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

$$p\left(\hat{W}_{LR}\right) = p\left(W_{LR} \geq \hat{W}_{LR} \,|\Psi_0, T\right) \tag{7.6}$$

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 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

$$p^*\left(\hat{W}_{LR}\right) = B^{-1} \sum_{b=1}^{B} I\left(W_{LR}^* \geq \hat{W}_{LR}\right), \tag{7.7}$$

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

---

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

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 theoretical considerations, see Davidson and MacKinnon (1996*a*), and for other examples on implementation of the bootstrap test, see Andersson and Gredenhoff (1997, 1998).

7.3.2 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,

$$\hat{\Gamma}(L)\,\Delta x_t = \widehat{\alpha\beta'}x_{t-1} + \hat{\varepsilon}_t. \tag{7.8}$$

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 (7.2). This implies that the bootstrap residuals ε_t^* are independent draws from a normal distribution with mean zero and variance $\widehat{\Sigma}$. The bootstrap samples $\mathbf{x}_i^*$,

$i = 1, ..., B$, are then created recursively, through equation (7.8), using the bootstrap residuals ε_t^*.

7.4 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 (7.3). 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.

7.4.1 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:

$$q^p(T_i) = \theta_\infty^p + \theta_1^p T_i^{-1} + \theta_2^p T_i^{-2} + \varepsilon_i, \tag{7.9}$$

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 (7.3), 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:

$$q^p(T_i) - \chi_p^2(\cdot) = \theta_1^p T_i^{-1} + \theta_2^p T_i^{-2} + \theta_3^p f(k_i) + \theta_4^p g(n_i) + \theta_5^p h(r_i) + \varepsilon_i, \tag{7.10}$$

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.

7.4.2 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:

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

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

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

In what follows we will construct the DGP's with the above $I(1)$-vector evaluated in the VECM in (7.2) 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, $\beta' = [1 \quad -1 \quad 0 \quad 0 \quad 0]$.

7.4.3 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 $T \in \{40,45,\ldots,95,100,110,\ldots,140,150, 175,200,225,250,300,350,400,500,600,700,800,1000\}$

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 (7.4). 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).

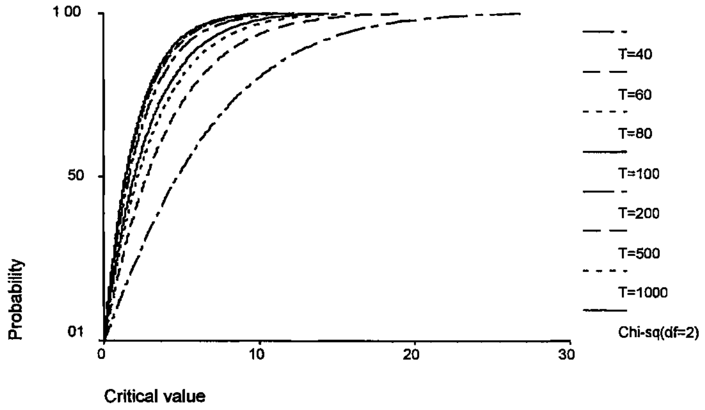
7.5 Results

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).

7.5.1 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 (7.9). 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 (7.10), using the inverse of the corresponding standard error $\hat{\sigma}_{pT}$ of each $\bar{\hat{q}}^p_T$ as weights.

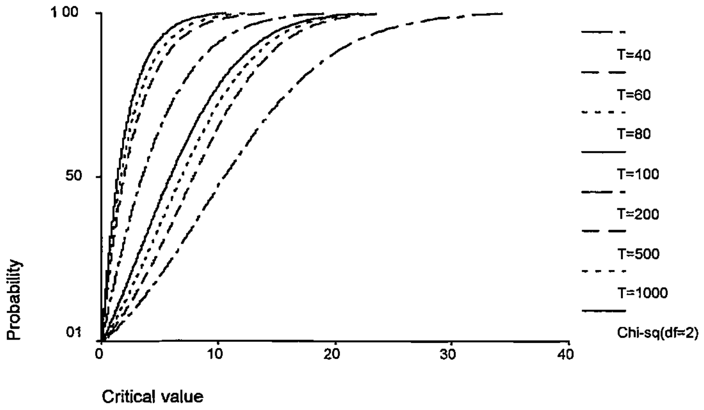
FIGURE 7.1. Cumulative distributions for varying sample sizes T, using specification $n = 3$, $k = 3$, and $r = 1$.



Studying the results, we will first consider two sets of simulated cumulative 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, $\overline{q}_T^p$, 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 size distortion is more or less linear, but for d.f. $= 1$ the relationship seems to be non-

FIGURE 7.2. Cumulative distributions for varying sample sizes T, using specification $n = 5$, $k = 4$, and $r = 3$.



linear. To improve the fit of the response surface regression for this case, we have estimated one regression for each possible dimension.

In Figure 7.3 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.

7.5.2 Size

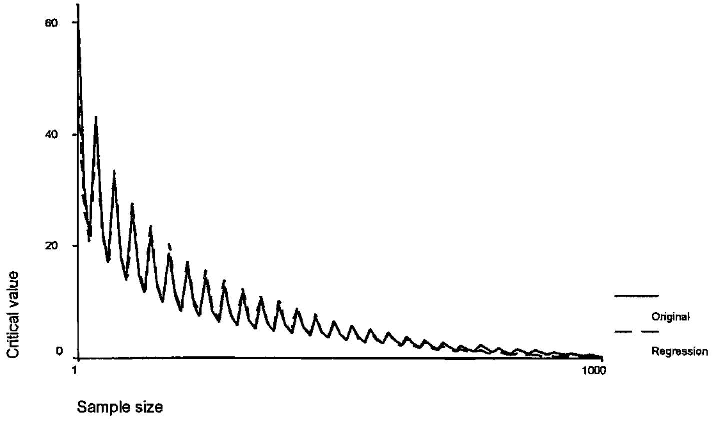
Tables 5.2-5.5 report the Monte Carlo estimated sizes for the likelihood ratio test in (7.4) and its bootstrapped analogue. They are organized according to the degrees of freedom in the asymptotic distributions. In general we find that the asymptotic approximation becomes worse as the degrees of freedom 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

TABLE 7.1. Response surface regressions, conditional on the number of degrees of freedom.

Parameter estimates					
$d.f. = 4$	$R^2 = 98.9$				
$1/t$	$1/t^2$	l_3/t^2	l_4/t^2		
165.55	26359.05	10840.15	49177.30		
(6.8)	(15.4)	(7.5)	(29.5)		
$d.f. = 3$	$R^2 = 99.3$				
$1/t$	$1/t^2$	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
(11.5)	(21.3)	(12.1)	(39.9)	(5.5)	(-4.9)
$d.f. = 2$	$R^2 = 97.9$				
$1/t$	$1/t^2$	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			
(9.7)	(-1.6)	(-6.3)			
$d.f. = 1$	$R^2 = 95.6$				
$1/t$	$1/t^2$	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				
(-5.9)	(-4.2)				
$n = 4$	$R^2 = 99.7$				
$1/t$	$1/t^2$	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
(11.2)	(16.6)	(18.8)	(15.8)	(-6.4)	(-2.8)
$n = 3$	$R^2 = 99.9$				
$1/t$	$1/t^2$	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
(16.6)	(48.7)	(29.6)	(12.7)	(-19.9)	(-23.2)
$n = 2$	$R^2 = 99.2$				
$1/t$	$1/t^2$	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
(14.6)	(12.8)	(-5.5)	(-5.9)	(2.1)	(0.3)

Values presented within brackets are t-values of corresponding coefficient

FIGURE 7.3. Original series of 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.



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.

7.5.3 Power

Tables 5.6-5.9 report Monte Carlo estimated power for the likelihood ratio test in (7.4) 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

---

[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 7.2. 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

TABLE 7.3. 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

TABLE 7.4. Estimated size, in percent, for the original and the bootstrap test at a nominal significance level of 5 percent. The asymptotic distribution is a χ^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 7.5. 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

establish the bearing of the theoretical prediction 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 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.

7.5.4 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

---

[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 7.6. 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	10.0	10.7	88.4	9.7	40.2	100	12.9	83.6	100
Orig.									
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

TABLE 7.7. 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 7.8. 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 7.9. 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

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 $\mathcal{H}_8$.

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

TABLE 7.10. Original and bootstrap test of liner restriction in the cointegrating space, the restriction on β are defined as $\beta = (\beta_1 H, \psi)$.

	m_t-						original test		Boot-test	
	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.

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.

7.6 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

TABLE 7.11. 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

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.

Acknowledgments

I'am grateful to Katarina Juselius for helping out with data, and to Michael Andersson, Sune Karlsson, Rolf Larsson, Anders Vredin, and Anders Warne for helpful discussions.

References

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.

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.

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

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.

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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.

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.

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

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.

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

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

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

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

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

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

EFI
The Economic Research Institute

Reports since 1994

Published in the language indicated by the title

1994

Andersson, H., Ett industriföretags omvandling. En studie av Hägglunds förändringsprocess 1922-81 med bas i företagets produkter, relationer och resurser.
Andersson, H., En produkthistoria.
(separat publicerad bilaga till ovanstående)
Cowan, R., Hultén, S., Escaping Lock-in: The Case of the Electric Vehicle. Research Report.
Dahlfors, G., Jansson, P., Essays in Financial Guarantees and Risky Debt.
Företag och marknader i förändring - dynamik i nätverk,
red. Mattsson, L-G., Hultén, S.
EFI-/Nerenius & Santérus Förlag.
Helgesson, C-F., Coordination and Change in Telecommunications. Research Report.
Lundström, T., Wijkström, F., The Swedish Nonprofit Sector. Research Report.
Nessén, M., Essays on Exchange Rates and International Finance.
Niemeyer, J., Essays on Market Microstructure -Empirical Evidence from Some Nordic Exchanges.
Nilsson, K., Rapportering i svenska storföretag 1993. Delrapport 1. Faktiska tider och påverkande faktorer för månatlig resultatrapportering. Research Report.
Nilsson, K., Rapportering i svenska storföretag 1993. Delrapport 2. Rapporteringstider för kvartals/tertials-rapporter. Research Report.
Nordenlöw, L., Entreprenörskap i industriella nätverk. Research Report.
Norén, N-E., Om Flexibilitetspåverkande beslut. Research Report.
Normark, P., Medlemsägda företag. Organisering av strategiska förändringar.
Nyman, K., Studier av inkomstför-delningar med hjälp av mikrosimule-ringsmodeller. Research Report.
Thodenius, B., Användningen av ledningsinformationssystem i Sverige: Lägesbild 1993. Research Report.
Sjöholm, G., Redovisningsmått vid bolagisering. Utformning och effekter.
Wijkström, F., Det organiserade ideella arbetet i Sverige. Research Report.
Wirsäll, N-E., Holmlöv, PG., Den nyckfulle Svensson. Reflektioner efter ett seminarium om strategier för handeln. Fonden för Handels- och Distributionsforskning, EFI.

1995

Becker, T., Essays on Stochastic Fiscal Policy, Public Debt and Private Consumption.
Blomberg, J., Ordning och kaos i projektsamarbete - en socialfenomenologisk upplösning av en organisationsteoretisk paradox.
Brodin, B., Lundkvist, L., Sjöstrand, S-E., Östman, L., Styrelsearbete i koncerner
Brännström, T., Bias Approximation and Reduction in Vector Autoregressive Models. Research Report.
Ekonomisk politik i omvandling. red. Jonung, L.
Gunnarsson, J., Wahlund, R.,
Hushållens finansiella strategier.
En explorativ studie.
Fonden för Handels- och Distributionsforskning, EFI.
Höök, P., Chefsutveckling ur könsperspektiv - Mentorskap och nätverk på Vattenfall - Research Report.
Levin, J., Olsson, P., Looking Beyond the Horizon and Other Issues in Company Valuation. Research Report.
Mägi, A., Customer Statisfaction in a Store Performance Framework. Research Report.
Persson, P-G., Modeling the Impact of Sales Promotion on Store Profits.
Roman, L., Institutions in Transition. A Study of Vietnamese Banking.
Sandberg, J., Statistisk metod - dess vetenskapliga hemvist, grundläggande principer och möjligheter inom samhällsvetenskapen. Research Report.
Sandberg, J., How Do We Justify Knowledge Produced by Interpretative Approaches? Research Report.

Schuster, W., Redovisning av konvertibla skuldebrev och konvertibla vinstandels-bevis - klassificering och värdering.

Söderberg, K., Farmartjänsten -Ny kooperation inom lantbruket. Research Report.

Söderqvist, T., Benefit Estimation in the Case of Nonmarket Goods. Four Essays on Reductions of Health Risks Due to Residential Radon Radiation.

Thorén, B., Användning av information vid ekonomisk styrning - månadsrapporter och andra informationskällor.

1996

Advancing your Business. People and Information Systems in Concert. red. Lundeberg, M., Sundgren, B.

Att föra verksamheten framåt. Människor och informationssystem i samverkan. red. Lundeberg, M., Sundgren, B. EFI/Studentlitteratur

Andersson, P., Concurrence, Transition and Evolution - Perspectives of Industrial Marketing Change Processes.

Andersson, P., The Emergence and Change of Pharmacia Biotech 1959-1995. The Power of the Slow Flow and the Drama of Great Events.

Asplund, M., Essays in Industrial Economics.

Delmar, F., Entrepreneurial Behavior & Business Performance.

Edlund, L., The Marriage Market: How Do You Compare?

Gunnarsson, J., Three Studies of Financial Behavior. Research Report.

Hedborg, A., Studies of Framing, Judgment and Choice.

Holgersson, C., Höök, P., Ledarutveck-ling för kvinnor - Uppföljning av en sats-ning på Volvo - Research Report.

Holmberg, C., Stores and Consumers - Two Perspectives on Food Purchasing.

Håkansson, P., Wahlund, R., Varumärken. Från teori till praktik. Fonden för Handels- och Distributionsforskning, Ekonomiska Forskningsinstitutet, EFI.

Karlsson, A., The Family Business as an Heirloom. Research Report.

Linghag, S., Man är handelsstudent. Research Report.

Molin, J., Essays on Corporate Finance and Governance.

Mägi, A., The French Food Retailing Industry - A Descriptive Study. The Foundation for Distribution Research.

Mölleryd, B., Så byggdes en världsindustri - Entreprenörskapets betydelse för svensk mobiltelefoni. Research Report.

Nielsen, S., Omkostningskalkulation for avancerede produktionsomgivelser - en sammenligning af stokastiske og deterministiske omkostningskalkulationsmodeller.

Normark, P., Danielsson, L., Larsson, A., Lundblad, P., Kooperativa nyckeltal. Research Report.

Sandin, R., Heterogeneity in Oligopoly: Theories and Tests.

Sandén-Håkansson, U., Från kampanjmål till mediemix - en studie av samarbete mellan annonsörer, reklambyråer och mediebyråer. Research Report.

Stein, J., Söderlund, M., Framgång i arbetet, strategier för att utföra arbetsuppgifterna, arbetsuppgifternas karaktär och utbildningskvalitet. En empirisk studie av civilekonomer. Research Report.

Strömberg, P., Thorburn, K., An Empirical Investigation of Swedish Corporations in Liquidation Bankruptcy. Research Report.

Söderlund, M., Och ge oss den nöjda kunden. En studie av kundtillfredsställelse och dess orsaker och effekter. Research Report.

Thodenius, B., Användningen av ledningsinformationssystem i Sverige: Lägesbild 1995. Research Report.

Ulfsdotter, U., Internationalisering för expansion eller hemmamarknadsförsvar? De nordiska marknaderna för fruktyoghurt 1982-1994.

Westelius, A., A Study of Patterns of Communication in Management Accounting and Control Projects.

Wijkström, F., Den svenska ideella sektorn och pengarna. Research Report.

Örtendahl, M., Health and Time - A Problem of Discounting.

1997

Alexius, A., Essays on Exchange Rates, Prices and Interest Rates.

Andersson, B., Essays on the Swedish Electricity Market.

Berggren, N., Essays in Constitutional Economics.

Ericsson, J., Credit Risk in Corporate Securities and Derivatives. Valuation and Optimal Capital Structure Choice.

Changli He, Statistical Properties of Garch Processes.

Charpentier, C., Budgeteringens roller, aktörer och effekter. En studie av budgetprocesserna i en offentlig organisa-tion.

De Geer, H., Silfverberg, G., Citizens'Trust and Authorities' Choices, A Report from The Fourth International Conference on Ethics in the Public Service, Stockholm June 15-18, 1994.

Friberg, R., Prices, Profits and Exchange Rates.

Från optionsprissättning till konkurslagstiftning. red. Bergström, C., Björk, T.

Hagerud, G.E., A New Non-Linear GARCH Model.

Haksar, A., Environmental Effects of Economywide Policies: Case Studies of Costa Rica and Sri Lanka.

Holmgren, M., Datorbaserat kontrollrum inom processindustrin; erfarenheter i ett tidsperspektiv.

Jennergren, P., Tutorial on the McKinsey Model for Valuation of Companies.

Lagerlöf, J., Essays on Political Economy, Information, and Welfare.

Lange, F., Wahlund, R., Planerade och oplanerade köp - Konsumenternas plane-ring och köp av dagligvaror.

Löthgren, M., Essays on Efficiency and Productivity; Contributions on Bootstrap, DEA and Stochastic Frontier Models.

Nilsson, H.E., Using Conceptual Data Modelling in Management Accounting: A Case Study.

Sjöberg, L., Ramsberg, J., En analys av en samhällsekonomisk bedömning av änd-rade säkerhetsföreskrifter rörande heta arbeten.

Säfvenblad, P., Price Formation in Multi-Asset Securities Markets.

Sällström, S., On the Dynamics of Price Quality.

Södergren, B., På väg mot en horisontell organisation? Erfarenheter från närings-livet av decentralisering och därefter.

Tambour, M., Essays on Performance Measurement in Health Care.

Thorén, B., Berg-Surwee, U., Områdesarbete i Östra Hökarängen - ett försök att studera effekter av decentralisering. **Zhang Gang.**, Chinese Rural Enterprises Between Plan and Market.

Åhlström, P., Sequences in the Profess of Adopting Lean Production.

Åkesson, G., Företagsledning i strategiskt vakuum. Om aktörer och förändrings-processer.

Åsbrink, S., Nonlinearities and Regime Shifts in Financial Time Series.

1998

Andersson, M.K., On Testing and Forecasting in Fractionally Integrated Time Series Models.

Gredenhoff, M., Bootstrap Inference in Time Series Econometrics.

Johansson, S., Savings Investment, and Economic Reforms in Developing Countries.

Segendorff, B., Essays on Bargaining and Delegation.

Wijkström, F., The Non-Profit Sector in Sweden.

Zethraeus, N., Essays on Economic Evaluation in Health Care. Evaluation of Hormone Replacement Therapy and Uncertainty in Economic Evaluations.