

# On Testing and Forecasting in Fractionally Integrated Time Series Models



# 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

A	Management and Organisation Theory
B	Accounting and Managerial Finance
C	Managerial Economics
CEE	Centre for Ethics and Economics
CFR	Centre for Risk Research
CHE	Centre for Health Economics
D	Marketing, Distribution and Industry Dynamics
ES	Economic Statistics
F	Public Management
FDR	The Foundation for Distribution Research
FI	Finance
I	Information Management
IEG	International Economics and Geography
P	Economic Psychology
PMO	Man and Organisation
PSC	Policy Science
RV	Law
S	Economics
T	Industrial Production

### Directors

Prof. Sven-Erik Sjöstrand
Prof. Lars Östman
Prof. Peter Jennergren
Adjunct Prof. Hans de Geer
Prof. Lennart Sjöberg
Prof. Bengt Jönsson
Prof. Lars-Gunnar Mattsson
Prof. Anders Westlund
Prof. Nils Brunsson
Acting Prof. Richard Wahlund
Prof. Clas Bergström
Prof. Mats Lundeberg
Prof. Mats Lundahl
Prof. Lennart Sjöberg
Prof. Bengt Stymne
Adjunct Prof. Brita Schwarz
Prof. Bertil Wiman
Prof. Lars Bergman
Prof. Christer Karlsson

Chairperson: Prof. Sven-Erik Sjöstrand

Director: Ekon Lic. Rune Castenäs

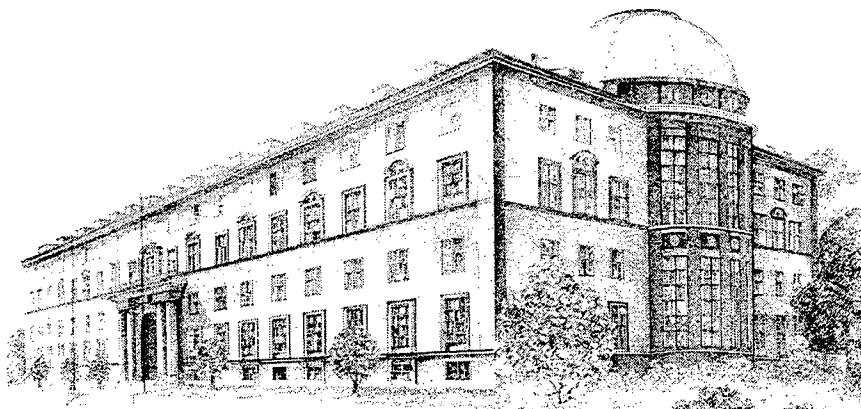
---

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

# On Testing and Forecasting in Fractionally Integrated Time Series Models

by

Michael K. Andersson



STOCKHOLM SCHOOL OF ECONOMICS  
EFI, THE ECONOMIC RESEARCH INSTITUTE



DISSERTATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY  
IN ECONOMETRICS, PRESENTED AT THE STOCKHOLM SCHOOL  
OF ECONOMICS 1998

### **Abstract.**

Andersson, M.K., 1998. On Testing and Forecasting in Fractionally Integrated Time Series Models. Stockholm ISBN 91-7258-467-X.

This volume contains five essays in the field of time series econometrics. All five discuss properties of fractionally integrated processes and models.

The first essay, entitled **Do Long-Memory Models have Long Memory?**, demonstrates that fractional integration can enhance the memory of ARMA processes enormously. This is however not true for all combinations of differencing, autoregressive and moving average parameters.

The second essay, with the title **On the Effects of Imposing or Ignoring Long-Memory when Forecasting**, investigates how the choice between modelling stationary time series as ARMA or ARFIMA processes affect the accuracy of forecasts. The results suggest that ignoring long-memory is worse than imposing it and that the maximum likelihood estimator for the ARFIMA model is to prefer.

The third essay, **Power and Bias of Likelihood Based Inference in the Cointegration Model under Fractional Cointegration**, investigates the performance of the usual cointegration approach when the processes are fractionally cointegrated. Under these circumstances, it is shown that the maximum likelihood estimates of the long-run relationship are severely biased.

The fourth and fifth essay, entitled respectively **Bootstrap Testing for Fractional Integration** and **Robust Testing for Fractional Integration using the Bootstrap**, propose and investigate the performance of some bootstrap testing procedures for fractional integration. The results suggest that the empirical size of a bootstrap test is (almost) always close to the nominal, and that a well-designed bootstrap test is quite robust to deviations from standard assumptions.

**Key words:** Long Memory, Bootstrap Testing, Fractional Cointegration

*Michael K. Andersson, Department of Economic Statistics, Stockholm School of Economics, Stockholm, Sweden*

© By the author 1998  
ISBN 91-7258-467-X

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

*for Anna*



# Contents

<b>Preface</b>	<b>xi</b>
Acknowledgements . . . . .	xi
Summary in Swedish for Non-Statisticians . . . . .	xiii
List of Papers . . . . .	xiv
 <b>I Thesis Summary</b>	 <b>1</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Long Memory and Fractional Integration . . . . .	3
1.1.1 Background . . . . .	3
1.1.2 Fractional Integration and The ARFIMA Model	4
1.1.3 Fractional Cointegration . . . . .	6
1.1.4 Other Fractionally Integrated Models . . . . .	7
1.1.5 Empirical Work using Fractional Integration .	8
1.2 The Bootstrap . . . . .	10
1.2.1 The Bootstrap Method . . . . .	11
1.2.2 Resampling Schemes . . . . .	11
1.2.3 The Bootstrap Testing Procedure . . . . .	12
 <b>2 Summary and Main Results of the Papers</b>	 <b>15</b>

2.1	Consequences of Assuming or Ignoring a Fractional Difference . . . . .	16
2.2	On Testing for a Fractional Difference . . . . .	18
	<b>References</b>	<b>21</b>
<b>II</b>	<b>Included Papers</b>	<b>27</b>
1	Do Long-Memory Models have Long Memory?	29
2	On the Effects of Imposing or Ignoring Long Memory when Forecasting	41
3	Power and Bias of Likelihood Based Inference in the Cointegration Model under Fractional Cointegration	57
4	Bootstrap Testing for Fractional Integration	77
5	Robust Testing for Fractional Integration using the Bootstrap	91



# Preface

## Acknowledgements

As far as I recall it. Fall 1990. I was living in the huge student ghetto in Uppsala while taking my first course in Economics. In the corridor opposite to mine lived a man from the Ivory Coast (it took me three months to understand that it was the name of a country) who seemed quite interested in the girl living next to me. While drinking my J.D. he often told me how much a major in Statistics would do for my career, that is a good job with a very high salary. This man, Tim Oké, is still a good friend of mine. Maybe I was feeble-minded back then, but I did take the first course in Statistics, and a second... .

Anders Ågren was the lecturer on my first course in time series analysis and he really got me interested in autocorrelations and the other fascinating stuff. He also supervised my thesis for the BSc. I had by then completed three straight semesters of Statistics. Later, Dr Ågren was advisor also for my MSc thesis, about persistence in Economic time series. Thanks for everything Anders. During the Uppsala years I also met Johan Lyhagen who not only stole my model, he completed his thesis one year ahead of me. Dispite that, I thank him for all suggestions while I was writing this thesis.

After my Masters degree in Uppsala I was accepted for the Ph.D. programme at the department of Economic Statistics at the Stockholm School of Economics. The first two years I was thinking about the ARFIMA model, which crossed my path during the Uppsala period, but I never really thought of it as a topic for my dissertation. So it almost surprised me when Sune Karlsson answered my question about a suitable project accordingly: why not write about fractional integration? Suddenly, it was obvious.

The single most important person for the thesis (next to myself I hope!) is Dr Karlsson, my supervisor. Over the years he has dismissed a few of my ideas, encouraged some and always given me perspicacious comments and suggestions. My debts to Sune are enormous and I'm still very grateful that he agreed on supervising me. Besides Sune Karlsson, my co-author Mikael Gredenhoff has also been important to my work. In October 1996 he entered my room, quite excited, and asked if I had a test for fractional integration. Of course, I replied. Lets bootstrap it, he suggested. And we did. He has not only been my co-author, he is also my snowboarding companion. He needs to put up some more speed though.

There are some other persons at the department, from Carina and Monica to Professor Westlund and Professor Teräsvirta, I would like to thank. Especially, I would like to thank the former room-mate Stefan Erik Åsbrink. In particular I remember him saying, give up the long-nose (slang for being cheated in Swedish) model and he was always laughing when I told him about the future importance of fractional integration. Furthermore, I am grateful to my co-advisors Lars-Erik Öller (for giving me ideas, reading my papers and supplying articles) and Tor Jacobson (for scrutinizing the bootstrap papers, Mikael sort of got the bootstrap idea from him). Mickael Löthgren deserves a piece of acknowledgement. He is certainly a very devoted researcher (watch out for the frontier) and a good friend. Over the years he has supplied tons of valuable information.

Not being a member of the department (not even a statistician), I would like to thank Stefan Nydahl. He co-authored a memo in Uppsala and is still always willing to write a paper.

Financial support from the Tore Browaldh foundation is gratefully acknowledged.

Finally my gratitude and love goes to my pride and joy Anna and my baby-daughter Alicia.

## Summary in Swedish for Non-Statisticians

Denna avhandling i ämnet statistik består av fem uppsatser. Alla fem uppsatserna behandlar långminnesprocesser och modeller. Med långt minne förstås att chocker är varaktiga, dvs att de påverkar seriens värden långt in i framtiden. Beaktande av långt minne kan medföra tillförlitligare analys av data och säkrare prognoser.

Den första uppsatsen visar att långminnesmodellen ofta, men inte alltid, har ett längre minne än den vanligen använda tidsseriemodellen. Ett långt minne innebär (vanligen) att variablen i fråga kan prognosticeras långt fram i tiden.

Den andra artikeln undersöker vad som händer om man ignorerar eventuellt långt minne eller om långt minne felaktigt påtvingas modellen. Studien visar att användandet av långminnesmodeller på kortminnesprocesser inte leder till några avsevärda problem, emedan det omvända kan innebära markant större prognosfel.

Den tredje artikeln behandlar samvariation mellan två variabler. I de flesta ekonomiska tillämpningarna antas samvariationen vara av kortminnestyp. Vi visar att de etablerade test- och skattningsmetoderna fungerar dåligt om samvariationen karaktäriseras av varaktiga chocker.

De två sista uppsatserna föreslår en testprocedur för långt minne. Resultaten av dessa artiklar tyder på att den föreslagna proceduren fungerar mycket väl, även om den sanna processen innehåller vissa störfenomen.

## List of Papers

This dissertation consists of the summary and the following papers, which are referred to by the numerals 1-5:

1. DO LONG-MEMORY MODELS HAVE LONG MEMORY? *Working Paper Series in Economics and Finance No. 227*, Stockholm School of Economics, 1998.
2. ON THE EFFECTS OF IMPOSING OR IGNORING LONG-MEMORY WHEN FORECASTING. *Working Paper Series in Economics and Finance No. 225*, Stockholm School of Economics, 1998.
3. 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.P Gredenhoff.
4. BOOTSTRAP TESTING FOR FRACTIONAL INTEGRATION. *Working Paper Series in Economics and Finance No. 188*, Stockholm School of Economics, 1997. With M.P Gredenhoff.
5. ROBUST TESTING FOR FRACTIONAL INTEGRATION USING THE BOOTSTRAP. *Working Paper Series in Economics and Finance No. 218*, Stockholm School of Economics, 1998. With M.P Gredenhoff.

# **Part I**

## **Thesis Summary**



# 1

## Introduction

*When I was younger I could remember anything, whether it happened or not. –Mark Twain*

This chapter gives an introduction to and explanation of some key concepts. The purpose is to facilitate the understanding of the thesis.

### 1.1 Long Memory and Fractional Integration

#### 1.1.1 Background

What is Memory? Memory is information from the past that we can use in the present. The memory definition of psychologists can easily be adopted to statistical time series analysis. Statistical modelling of time series often involves seeking dependencies between the time series and previous values of that particular series (if the series is self/auto-correlated) and other series. For instance if only lag one autocorrelation is found (that is a short memory), we use todays observation to forecast tomorrow. However, if several previous observations affect todays' the memory is longer. The concept of long-memory is an old one; the Old Testament prophets foretold of the seven years of famine followed by seven years of plenty that Egypt

was to experience. This is called the "Joseph effect", see Mandelbrot (1972), Mandelbrot and Wallis (1968) and Beran (1994), and can be explained by the fact that the floods of the River Nile fertilize the soil implying great abundance for the following years.

The interest in long-memory processes appears to originate in the physical sciences. The most famous example of long memory is (from hydrology) the tidal flows and inflows into reservoirs originally documented by Hurst in 1951. Later, evidence of long memory was also found in economic data. Granger (1966) investigates the estimated power spectra for economic variables and concludes that after removal of an exponential trend in mean the spectrum typically still has considerable power at low frequencies. This is a characteristic (and maybe evidence) of a long-memory process.

### 1.1.2 Fractional Integration and The ARFIMA Model

It is standard practice to (if necessary) difference time series to achieve stationarity. A series that is non-stationary in the levels but stationary in the first differences is said to be integrated of order one, denoted  $I(1)$ , and a series that is stationary after differencing  $d$  times is  $I(d)$ . According to the Box-Jenkins modelling procedure,  $d$  is restricted to the set of integers. However, in the early 80's Granger and Joyeux (1980) and Hosking (1981) suggested that the notion of differencing may be generalized to include real values of  $d$ .

**Definition 1** For any real-valued  $d$ , a fractionally differenced white noise (FDWN) process  $\{x_t\}$  is defined by

$$\Delta^d x_t = (1 - B)^d x_t = a_t, \quad (1.1)$$

where  $\Delta$  and  $B$  denotes the differencing and backshift operators respectively and the sequence  $\{a_t\}$  is a white noise process.

The differencing filter (called the long-memory filter, *LMF*) can be expanded as

$$(1 - B)^d = 1 - dB - \frac{d(1-d)B^2}{2!} - \frac{d(1-d)(2-d)B^3}{6!} - \dots \quad (1.2)$$

When  $d < 1/2$ ,  $\{x_t\}$  is a stationary process with the infinite moving average representation (using the notation of Granger and Joyeux)

$$x_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}, \quad \psi_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}. \quad (1.3)$$



The coefficients of (1.3) dies out at a slow hyperbolical rate. When  $d > -1/2$ ,  $\{x_t\}$  is invertible with the infinite autoregressive representation

$$\sum_{j=0}^{\infty} \pi_j x_{t-j} = a_t, \quad \pi_k = \frac{\Gamma(k-d)}{\Gamma(1-d)\Gamma(k-1)}. \quad (1.4)$$

If we assume that  $x_t$  is a stationary and invertible process with  $\sigma_a^2 = 1.0$ , the following properties can be derived. The covariance function of  $x_t$  is given by

$$\gamma_k = \frac{(-1)^k \Gamma(1-2d)}{\Gamma(1-d+k)\Gamma(1-d-k)}, \quad (1.5)$$

and the spectral density by

$$s(\omega) = \begin{cases} (2 \sin \frac{1}{2}\omega)^{-2d} & \text{for } 0 < \omega \leq \pi \\ \sim \omega^{-2d} & \text{as } \omega \rightarrow \infty. \end{cases} \quad (1.6)$$

The autocorrelations and partial autocorrelations for the process is given by

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \frac{\Gamma(d+k)}{\Gamma(1-d+k)} \quad (1.7)$$

and

$$\phi_{kk} = \frac{d}{k-d}. \quad (1.8)$$

Proofs of the properties are given by Granger and Joyeux (1980) and Hosking (1981). Furthermore, the following is noted for some values (or ranges) of  $d$ .

- $d \notin (-1/2, 1/2)$ . The process is either stationary or invertible, but may be summed or differenced a finite integral number of times until  $d$  is contained in the desired interval.
- $1/2 \leq d < 1$ . The process is not covariance stationary but mean-reverting in the sense the effect of a shock dies out, albeit very slowly.
- $0 < d < 1/2$ . The process is stationary with long-memory and is useful in modelling long-range persistence. The autocorrelations and impulse responses are all positive and decay at a slow hyperbolical rate.

- $d = 0$ . The process collapses to an ARMA process.
- $-1/2 < d < 0$ . The autocorrelations are all negative and decay hyperbolically, and the process is said to be anti-persistent.

The FDWN model may capture the long-run behaviour of a time series process, but it is not flexible enough to model the short-range dynamics of economic time series. But if fractional differencing is introduced to the common ARMA model, yielding the fractionally integrated autoregressive moving average (*ARFIMA*) model, we obtain a model that takes care of the low-lag correlation structure while retaining the long-term hyperbolic decay of the autocorrelation function.

**Definition 2** *An ARFIMA( $p, d, q$ ) process is defined by the equation*

$$\phi_p(B) (1 - B)^d x_t = \theta_q(B) a_t, \quad (1.9)$$

where  $d$  is allowed to assume any real value and the members of the sequence  $\{a_t\}$  are iid with finite variance.

If the roots of  $\phi_p(B)$  and  $\theta_q(B)$ , the autoregressive and moving average polynomials in equation (1.9), lie outside the unit circle and  $|d| < 0.5$ ,  $x_t$  is stationary and invertible. The LMF describes the long-term dependence in the series and the short-term structure is captured by the autoregressive and moving average parameters. This enables separate modelling of the long and short-run dynamics.

### 1.1.3 Fractional Cointegration

A time series process  $\mathbf{x}_t$  is said to be cointegrated of order  $CI(\delta, b)$  if the variables of  $\mathbf{x}_t$  individually are integrated of order  $I(\delta)$ , while a linear combination of the variables, denoted  $\beta' \mathbf{x}_t$ , is  $I(\delta - b)$ . The variables are in equilibrium if  $\beta' \mathbf{x}_t$  equals some constant  $\mu$ , 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  $\delta$  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  $\delta$  and  $b$ . Furthermore, 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, however 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.

Two common approaches when testing and estimating in cointegration models are provided in the literature. The *first* is a two-step procedure, introduced by Engle and Granger (1987), which involves testing for unit roots. A *second* approach is the maximum likelihood procedure of Johansen (1988). The maximum likelihood procedure enables joint analysis of the long-run properties and short-run dynamics, and makes efficient inference possible.

#### 1.1.4 Other Fractionally Integrated Models

A model that can be regarded as an extension of the ARFIMA model is the Gegenbauer ARMA (*GARMA*) model,

$$\phi(B)(1 - 2\xi B + B^2)^\lambda x_t = \theta(B)a_t.$$

A stationary GARMA process, that is when  $|\xi| < 1$  and  $0 < \lambda < 0.5$  or  $|\xi| = 1$  and  $\lambda < 0.25$ , exhibits long-memory in the sense that the spectral density is unbounded as the frequency tends to zero. The GARMA process is described by Gray *et al.* (1989).

Fractional integration can also be combined with seasonality, for instance Porter-Hudak (1990) considers the specification

$$(1 - B^s)^d x_t = a_t,$$

where  $s$  is the seasonal period. A periodic fractionally integrated model, where the integration parameter is allowed to vary among the seasons, is suggested by Franses and Ooms (1995) and a flexible seasonal long-memory model is described by Ooms (1995).

The concept of long-memory is not only valid for the mean of a time series, also the variance may be modelled as a persistent process. For this purpose, Harvey (1993) and Breidt *et al.* (1994, 1995) use

a long-memory stochastic volatility model. Baillie *et al.* (1996a) combine fractional differencing and the GARCH model in order to model a hyperbolic rate of decay for the autocorrelation function and persistent impulse response weights. A similar model, but based on the EGARCH model, is developed by Bollerslev and Mikkelsen (1996).

### 1.1.5 Empirical Work using Fractional Integration

Several decades before fractional integration appeared in the literature, Hurst (1956) analyzed 900 geophysical data series using the range over standard deviation (R/S) statistic. The results (if recalculated) suggest an estimated mean value of the fractional difference parameter (henceforth denoted by  $d$ ) of 0.23 across the 900 series. Baillie (1996) uses a typical series of Hurst (1951, 1956), the tree ring data of Mount Campito, to illustrate the fractionally integrated model. When fitting a fractional white noise process (see equation (1.1))  $d$  is estimated to 0.449 (with a standard error of 0.010). Baillie points out that this very simple specification is able to model all dynamics in the conditional mean of the process. The conditional variance may be modelled by a GARCH(1,1) specification. Hosking (1984) investigates annual flows of the River Nile and retrieves estimates of  $d$  lying between 0.16 and 0.28 (depending on the specified model and definition of the year, July-June or Jan-Dec), and Haslett and Raftery (1989) study Ireland's wind power resource ( $\hat{d} = 0.328$ ).

The initial work on long memory concerned applications in geophysical sciences, but there is quite a large amount of (more recent) applications in economics, notably financial, monetary and macroeconomics. Besides the paper of Granger (1966), studies on persistence in macroeconomic variables dates back to the work of Beveridge and Nelson (1981) and Nelson and Plosser (1982).

Particular interest has been focused on the existence of a unit root in aggregate output (GDP), for instance Campbell and Mankiw (1987) find that the GPD series (for some countries) are consistent with (simulated) unit root processes. In contrast, Cochrane (1988) suggests that the US GDP series is best characterized as trend stationary. Instead of the knife-sharp distinction between stationary and unit root processes, Diebold and Rudebusch (1989) consider fractional integration and utilize the Geweke and Porter-Hudak (1983) procedure to estimate the differencing parameter. The estimates, for

some measures of US real GNP, range from 0.49 to 0.92. The unit root hypothesis is rejected (at a 10% level of significance) in 13 out of 21 cases and the hypothesis of  $d > 0.5$  (non-stationarity) is never rejected. In particular, Diebold and Rudebusch strongly reject a time trend model for quarterly post World War II data, but find the series consistent with a unit root process (even though  $d$  is estimated to 0.50). For the same data, Sowell (1992a) reports a smaller value for the differencing power ( $\hat{d} = 0.41$ ) and confidence intervals that include  $d = 0$  as well as  $d = 1$ .

Diebold and Rudebusch (1991) and Haubrich (1993) consider the so-called Deaton's excessive smoothness paradox, where consumption (under the usual ARIMA representation) appears too smooth for the permanent income hypothesis (PIH) to hold. Haubrich's results suggest that if income follows a fractionally integrated ARMA process, then the observed variance of consumption and income is consistent with the PIH.

Hassler and Wolters (1995) investigate monthly inflation rates of five countries. In addition to testing for unit roots, they fit long-memory models. Their results are against unit roots but for long-memory ( $\hat{d}_{US} \approx 0.41$  and  $0.42$ ,  $\hat{d}_{UK} \approx 0.39$  and  $0.51$ ,  $\hat{d}_{Fra} \approx 0.48$  and  $0.54$ ,  $\hat{d}_{Ger} \approx 0.35$  and  $0.4$ ,  $\hat{d}_{Ita} \approx 0.50$  and  $0.57$ ). Baillie *et al.* (1996b) use an ARFIMA-GARCH model, i.e. the conditional mean is modelled as a long-memory processes and the conditional variance as GARCH(1,1), to examine monthly post World War II inflation for 10 countries. The estimates of the differencing parameter fall within the following ranges: 0 – 0.1: Japan, 0.1 – 0.2: Germany, 0.2 – 0.3: U.K., 0.3 – 0.4: Canada, 0.4 – 0.5: France, Italy and USA, 0.5 – 0.6: Argentina, Brazil and Israel. Thus support is found for long-memory and mean-reversion is found for all countries but Japan, which appears stationary (and thus also mean-reverting).

Foreign exchange rate data have been investigated by Cheung (1993a) and Tschernig (1994). Cheung finds evidence of long memory in the exchange rate change and suggests that the empirical evidence of unit roots are not robust to long-memory alternatives. Tschernig claims that the evidence of Cheung may be weaker, since Cheung's estimation results may be biased due to non-accounted conditional heteroskedasticity and caused by a specific short time period.

In stock market data, evidence of long-memory is usually not found. Cheung and Lai (1995) employ the modified rescaled range

and Geweke-Porter-Hudak tests to investigate long-memory properties of stock index data for 18 countries. Their results suggest that long-memory can seldom be found in nominal and real stock returns. The results are in accordance with those of Lo (1991) who find no evidence of long-memory using the modified rescaled range. However, the original rescaled range ( $R/S$ ) test supports long-memory in stock returns according to Lo. This can be attributed the sensitivity of the  $R/S$  statistic to short-term dynamics. Barkoulas and Baum (1996) find long-memory in some individual returns series but not in stock indices. Evidence for long memory in individual stock returns series have previously been found by Greene and Fielitz (1977).

There are also some applications on fractional cointegration, for instance Cheung and Lai (1993) show that PPP reversion exists and may be characterized by a fractionally integrated specification for three (France, Italy and Canada) out of five countries studied. No evidence of a long-run PPP relation is found for United Kingdom and Japan. The intercountry relations are considered between United States (as home country) and the countries above as foreign above.

Empirical work using fractional integration on Swedish data is quite rare.

## 1.2 The Bootstrap

Bradley Efron thought of naming his procedure the *shotgun* because it has the ability of blow away any statistical problem. Instead Efrons' procedure was given the name bootstrap after the expression 'pulling oneself up by the bootstraps', which means doing the impossible. This reminds us of the stories of Baron von Münchhausen, who manages to pull not only himself but also his horse out of a swamp by his own wigtail. The bootstrap as a statistical technique was introduced in 1979 by Efron and can be used to solve a variety of problems, such as estimation of bias, standard errors, construction of confidence intervals and for size-adjusting tests.

Statistical inference often involves estimating a probability distribution  $F$  given a random sample, drawn from  $F$ . Since  $F$  is usually unknown, the empirical distribution  $\hat{F}$  is often used as an estimate of the entire distribution  $F$ . However, more conveniently the probability distribution may be characterized by some aspects of  $F$ , such as the mean, median, variance, skewness and kurtosis *et cetera*. Now,

the corresponding estimated aspects can serve as approximations to the corresponding unknown quantities. This is called the 'plug-in principle', of which the bootstrap is a direct application.

### 1.2.1 The Bootstrap Method

Consider an independently and identically distributed (*iid*) random sample

$$\mathbf{x}_T = (x_1, \dots, x_T) \quad (1.10)$$

of  $T$  observations on a random variable  $X$  with unknown probability distribution  $F$ . Suppose that we want to estimate the parameter  $\theta = \theta(F)$  on the basis of  $\mathbf{x}$ , which can be done by calculating  $\hat{\theta} = \hat{\theta}(\mathbf{x}_T)$ . The properties of the estimator  $\hat{\theta}$  can be derived from asymptotic theory, and may exhibit poor small-sample properties, it may be a mathematically complicated function or maybe we want to know something about the statistical behaviour of  $\hat{\theta}$ . In these situations, it is beneficial to take advantage of the simplicity and accuracy of a bootstrap procedure.

Let

$$\mathbf{x}_T^* = (x_1^*, \dots, x_T^*) \quad (1.11)$$

denote a so-called bootstrap (re)sample. Each  $x_i^*$  is a random variable with distribution  $\hat{F}$ .  $\hat{F}$  denotes the empirical distribution of the original sample  $\mathbf{x}_T$  and is given by  $\hat{F}(\kappa) = \sum_{i=1}^T I(x_i \leq \kappa)$ , where  $I(\cdot)$  is the usual indicator function. For the bootstrap sample the same statistic as above may be employed to compute the bootstrap estimate  $\hat{\theta}^* = \hat{\theta}(\mathbf{x}_T^*)$ . Albeit there is only one value of  $\hat{\theta}$ , we can create satisfactorily (or affordably) many bootstrap estimates and hence retrieve the empirical distribution (and subsequently the mean, variance, *et c.*) of the statistic of interest.

For further readings on the bootstrap procedure consult Efron and Tibshirani (1993).

### 1.2.2 Resampling Schemes

The crucial step in the bootstrap procedure is the creation of the resamples or equivalently, the construction of  $\hat{F}$  from  $\mathbf{x}_T$ . Following the original bootstrap of Efron (1979), the members of each resample can be drawn with replacement from the collection of original observations. This is a non-parametric (distribution free) bootstrap,

but one may also use a parametric resampling, that is imposing a distributional assumption and construct the resamples as random draws from that particular distribution.

A simple iid resampling (as in the original Efron bootstrap) fails in some circumstances, for instance it is not able to preserve dependencies in data. When dealing with time series, where the observations usually are correlated, we need a resampling procedure that maintain the dependencies. One way of doing this is to use a model-based bootstrap approach. In this case the resamples are constructed by an estimated (or theoretical) model, e.g.

$$\mathbf{x}_T^* = f(\hat{\varphi}, \mathbf{X}_{lag}^*, \mathbf{A}_{t,lag}^*), \quad (1.12)$$

where  $\hat{\varphi}$  is an estimated parameter vector. The matrix  $\mathbf{X}_{lag}^*$  consists of lagged bootstrap observations and the matrix  $\mathbf{A}_{t,lag}^*$  of contemporaneous and lagged residuals, which are drawn according to a resampling algorithm.

Since the outcome may be sensitive to model mis-specification, model-free bootstrap approaches provide competitive alternatives. One model-free (time domain) procedure, similar to the iid bootstrap, is the moving blocks bootstrap of Künsch (1989). However, Lahiri (1993) shows that the block bootstrap fails to approximate the distribution of long range dependent observations. An alternative to the block bootstrap is a, parametric or non-parametric, frequency domain (spectral) bootstrap (see for instance Franke and Härdle (1992) and Nordgaard (1996)).

### 1.2.3 The Bootstrap Testing Procedure

First-order asymptotic theory often gives poor approximations to the distribution of test statistics with the sample sizes common in empirical situations. Thus the true size using asymptotic critical values can differ enourmously from the nominal. One solution to this problem may be the use of higher-order expansions. Another feasible and sometimes more tractable solution is to calculate the significance level by simulation methods. If, for this purpose, a bootstrap procedure is used, we will refer to this test as a bootstrap test, see for instance Horowitz (1995) for an introduction and overview. The theory of bootstrap testing is a recent addition to the bootstrap literature. According to Davidson and MacKinnon (1996a), the size-distortion



of a model-based bootstrap procedure on an asymptotically pivotal statistic, with parameter estimates under the null, can be at least one full order,  $O(T^{-1})$ , smaller than the distortion of the asymptotic test. Furthermore, Davidson and MacKinnon (1996b) show that the power of a bootstrap test, based on a pivotal statistic, is generally close to that of the size-adjusted asymptotic test. Even if the statistic is only close to pivotal this is generally true.

The  $p$ -value function for a general one-sided test is defined by

$$p(\hat{\tau}) = p(\tau \geq \hat{\tau} | \Psi_0, T), \quad (1.13)$$

where  $\Psi_0$  is the data generating process (DGP) under the null hypothesis,  $T$  is the sample size and  $\hat{\tau}$  is a realized value of the test statistic  $\tau$  based on the original sample  $\mathbf{x} = [x_1, \dots, x_T]'$ . Since  $\Psi_0$  is unknown this  $p$ -value function has to be approximated, which is regularly done according to asymptotic theory. For asymptotic theory to be valid it is required that  $p(\hat{\tau})$  should not depend on  $\Psi_0$  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. The quantity (1.13) is then estimated by the bootstrap  $p$ -value

$$p^*(\hat{\tau}) = p(\tau \geq \hat{\tau} | \hat{\Psi}_0, T). \quad (1.14)$$

The DGP  $\hat{\Psi}_0$  may be retrieved by estimation of the null-model to the original sample, in order to obey the null-hypothesis. In practice the bootstrap  $p$ -value (1.14) is approximated by a Monte Carlo simulation from the bootstrap DGP  $\hat{\Psi}_0$ ,

$$p^*(\hat{\tau}) = R^{-1} \sum_{r=1}^R I(\tau_r^* \geq \hat{\tau}), \quad (1.15)$$

where  $R$  is the number of resamples and  $\tau_r^*$  is the calculated bootstrap value of the same test statistic  $\tau$  as above for each resamples series. The indicator function  $I(\cdot)$  equals one if the inequality is satisfied and zero otherwise. Furthermore, in addition to obeying the null-hypothesis the resamples shall (as far as possible) resemble the original sample.



## 2

# Summary and Main Results of the Papers

*An estimator may be consistent, but who lives forever.*

-M.A. (freely after J.M. Keynes)

*Prediction is extremely difficult. Especially about the future.* -Niels Bohr

This thesis investigates the consequences of modelling time series data as fractionally integrated processes. Since the true nature of a time series process is unknown, several models may (potentially) be employed to describe the process. For example, the fractionally integrated ARFIMA model may be an alternative to the common ARMA model.

The issues of the first two papers (referred to as 1 and 2, see the List of Papers) are process memory and forecasting accuracy. The performance of the ARFIMA model is compared with that of the non-fractional ARMA model.

The third article (paper 3) puts fractional integration in a multivariate context; we investigate the usual cointegration maximum likelihood estimation procedure when the series are fractionally cointegrated.

The results of papers 1 - 3 suggest that model mis-specification (of both ARMA and ARFIMA type) may lead to bad inference and

inaccurate forecasts. Consequently, testing for a fractional difference is often an important step prior to estimation. Papers 4 and 5 investigate three common tests for fractional integration, and propose some bootstrap algorithms in order to improve the tests.

Besides studying the properties of the maximum likelihood estimator, paper 3 also investigates if the likelihood ratio test for cointegration also has power against fractional cointegration.

The remaining part of this chapter highlights the most important discoveries of the five articles. The chapter is subdivided into two sections; the first section is concerned with the consequences of model mis-specification and the second sums up our findings on testing for fractional integration.

## 2.1 Consequences of Assuming or Ignoring a Fractional Difference

In a univariate context, paper 1 calculates the memory of (fractionally integrated) ARFIMA(1,  $d$ , 1) processes. By setting the differencing parameter  $d$  equal to zero, the non-fractional ARMA specification is obtained as a special case. We establish the memory using a rule based on the predictability measure of Granger and Newbold (1986). The rule implies that the memory is given by the horizon where the prediction error variance amounts to a pre-specified proportion of the process variance. The memory is closely related to the maximum prediction horizon of a process, but the two quantities may not always coincide.

Paper 1 shows that ARMA processes, that is when  $d = 0$ , display short memory. On the other hand, ARFIMA processes generate long memory under some circumstances, but not always, and it is demonstrated that fractional integration often extends the memory of ARMA processes. In particular, ARFIMA models with large positive differencing powers and positive parameters in the autoregressive polynomial have an extremely long memory. However, the long memory generated by a positive differencing parameter vanishes if a large negative autoregressive parameter is introduced in the process. Furthermore (and more surprisingly), negative moving average parameters (which alone make no memory contribution) absorb a substantial amount of memory induced by fractional integration.

In a previous study, Lyhagen (1997) demonstrates theoretically that ignoring long memory may lead to very high relative mean squared errors of prediction. However, the results of Ray (1993) show that high-order AR models forecast fractional noise series well, even in the long term. Also for fractional noise, Smith and Yadav (1994) point out that there is a potential loss from incorrectly fitting an AR model when the differencing parameter is positive. On the other hand, negative fractional differencing will bring about a performance loss of the AR model only at a horizon of one step. Paper 2 analyzes the forecasting performance of the ARMA and ARFIMA models in terms of their mean squared prediction errors. Three popular estimation procedures for the fractional model are used, namely the one of Geweke and Porter-Hudak (1983), the modified rescaled range of Lo (1991) and maximum likelihood (Sowell, 1992b), in order to compare the performance of the estimation techniques.

The results of paper 2 suggest that the modified rescaled range and maximum likelihood estimators for the fractionally integrated model generate predictions that are almost as accurate as those of the ARMA model, when the true process is a first-order autoregression or moving average. The predictions are also quite similar for negatively fractionally integrated processes. The estimator of Geweke and Porter-Hudak performs badly when the true process is an AR(1) with a large positive parameter. For persistent processes, the GPH and MRR produce short-term predictions that are worse than the ARMA forecasts. However, the opposite is found for intermediate and long-range forecasts. The MLE is better than the ARMA model for all horizons. The GPH technique experiences problems when combining fractional integration with an autoregression, especially for intermediate to large AR parameters. In the case of simultaneous short and long memory, the MRR and in particular the MLE procedures generate better predictions than the ARMA model.

In general, it is worse to ignore than to impose long memory when forecasting. Overall, the MLE is the best of the fractional estimators and the ARFIMA model is better than the ARMA.

Paper 3 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. It is concluded that 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.

To sum up paper 1, 2 and 3; long memory is associated with positively fractionally integrated processes. Pursuing this issue one step further, if an ARFIMA process is modelled by an ARMA specification we are likely to under-estimate the maximum prediction horizon. On the other hand, the prediction horizon may be over-estimated when modelling non-fractional processes with the ARFIMA model. This is of course not a problem if our fractional estimation procedure works well, but the estimators are often severely biased. The estimation bias of the fractionally integrated procedure may also imply a less accurate forecasting performance, compared to that of the ARMA model. We also notice prediction errors that are even greater for the non-fractional model when the true process is fractionally integrated. Finally, if the equilibrium error in the cointegration model is ruled by persistence, that is a fractionally integrated specification, the usual maximum likelihood estimator of the long-run relation is severely biased and displays very large root mean squared errors. The results of paper 1-3 point out the importance of testing for fractional integration.

## 2.2 On Testing for a Fractional Difference

"The design of appropriate test statistics to distinguish between  $I(0)$ ,  $I(d)$  and  $I(1)$  is still at an early stage, and it is likely that this will be an intensively worked area in the next few years." The words of Baillie (1996) and the results of paper 1-3 in this thesis illustrate the importance of improved tests for fractional integration. There exist several tests for fractional integration and paper 4 and 5 investigate the performance of the testing procedure of Geweke and Porter-Hudak, GPH (1983), the modified rescaled range (range over standard deviation, MRR) test of Lo (1991) and a Lagrange multiplier (LM) test (denoted REG) developed by Agiakloglou and Newbold (1994). The LM test is investigated by Agiakloglou and Newbold and the GPH and MRR tests by Cheung (1993). Besides studying the original tests under normality (paper 4 and 5) and under non-normality and heteroskedastic errors (paper 5), paper 4 pro-

pose a model-based parametric bootstrap algorithm in order to size-adjust tests. Paper 5 also considers a non-parametric algorithm and two (one parametric and one non-parametric) resampling schemes that account for heteroskedasticity of ARCH type.

For normally distributed autoregressions and moving averages, the estimated size of the MRR test usually differs from the nominal. The GPH test is usually well-sized, except for AR processes with large positive roots and moving averages with large negative roots. The REG test has the best size-properties, but a tendency to reject the true null-hypothesis is found for short ( $T \leq 50$ ) AR and all MA processes, when the true process has a root close to the unit-circle.

The original MRR and GPH tests are robust to excess skewness and kurtosis and ARCH errors 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 and GPH tests remain. In contrast to the other tests, the original REG test is sensitive to data that do not fulfill the normality assumption. The difference between the empirical and nominal size is then in general significant.

The results for normally distributed autoregressive processes suggest that the MRR and GPH bootstrap tests, regardless of resampling, give exact tests in the sense that the estimated size of the tests coincide with the nominal. The bootstrap REG test based on the simple parametric resampling is very well-sized, whereas the non-parametric resampling (which do not incorporate the normality) produce notably large sizes for strongly dependent AR processes. The resamplings that account for the (non-existing) ARCH effects have reasonable estimated sizes.

The robustness of the MRR and GPH tests to excess skewness and kurtosis and ARCH effects can be detected in the bootstrap tests. As a result all bootstrap MRR and GPH tests are exact for all generated autoregressions. The bootstrap REG tests behave similarly for processes with excess skewness and kurtosis as they do in the normal case. For ARCH processes, the disappointing size of the original REG test is partly inherited by the simple algorithms. The REG test, overall, requires that the resampling scheme allows for ARCH effects.

The simulation results of paper 4 demonstrate that the power of the bootstrap tests are close to the power of the size-adjusted asymptotic tests, when the true process is normally distributed. Paper

5 shows that the power of the MRR and GPH tests are preserved by all bootstrap procedures for fractionally integrated processes with excess skewness and kurtosis or conditionally heteroskedastic error terms. The power of the REG test is preserved by a simple bootstrap, whereas an ARCH resampling has a lower power throughout. Furthermore, the REG test should preferably be combined with a parametric bootstrap. For the MRR and GPH tests, a non-parametric resampling scheme performs equally well. 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 procedures.

The results of paper 3 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, together with 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 described in paper 4 and 5, which are robust to AR and MA components.



# References

*A journey of a thousand miles begins with a single step.*  
-Chinese Proverb

This chapter gives the references used in the Introduction and the Summary.

- AGIAKLOGLOU, C. AND P. NEWBOLD (1994) Lagrange Multiplier Tests for Fractional Difference. *Journal of Time Series Analysis* 15, 253-262.
- BAILLIE, R.T. (1996) Long Memory and Fractional Integration in Econometrics. *Journal of Econometrics* 73, 5-59.
- BAILLIE, R.T., T. BOLLERSLEV AND H.-O. MIKKELSEN (1996a) Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*. Forthcoming.
- BAILLIE, R.T., C.-F. CHUNG AND M.A. TIESLAU (1996b) Analyzing Inflation by the Fractionally Integrated ARFIMA-GARCH Model. *Journal of Applied Econometrics* 11, 23-40.
- BARKOULAS, J.T. AND C.F. BAUM (1996) Long Term Dependence in Stock Returns. *Working Paper 314*, Department of Economics, Boston College, USA.

- BERAN, J. (1994) *Statistics for Long-Memory Processes*. Chapman & Hall, New York, USA.
- BEVERIDGE S. AND C.R. NELSON (1981) A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurements of the 'Business Cycle'. *Journal of Monetary Economics* 7, 151-174.
- BOLLERSLEV, T. AND H.-O. MIKKELSEN (1996) Modeling and Pricing Long-Memory in Stock Market Volatility. *Journal of Econometrics* 73, 151-184.
- BREIDT, F.J., P. DE LIMA AND N. CRATO (1994) Modelling Long Memory Stochastic Volatility. *Working Paper No. 134*, Department of Economics, the Johns Hopkins University.
- BREIDT, F.J., P. DE LIMA AND N. CRATO (1995) On Detection and Estimation of Long Memory in Stochastic Volatility. *Unpublished manuscript*, Department of Economics, the Johns Hopkins University.
- CAMPBELL AND MANKIW (1987) Are Output Fluctuations Transitory? *The Quarterly Journal of Economics* 102, 857-880.
- CHEUNG, Y.-W. (1993a) Long Memory in Foreign-Exchange Rates. *Journal of Business & Economic Statistics* 11, 93-101.
- CHEUNG, Y.-W. (1993b) Tests for Fractional Integration: A Monte Carlo Investigation. *Journal of Time Series Analysis* 14, 331-345.
- CHEUNG, Y.-W. AND K.S. LAI (1993) A Fractional Cointegration Analysis of Purchasing Power Parity. *Journal of Business & Economic Statistics* 11, 103-112.
- CHEUNG, Y.-W. AND K.S. LAI (1995) A search for Long Memory in International Stock Market Returns. *Journal of International Money and Finance* 14, 597-615.
- COCHRANE J.H. (1988) How Big is the Random Walk in GNP. *Journal of Political Economy* 96, 893-920.

- 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 (1989) Long Memory and Persistence in Aggregate Output. *Journal of Monetary Economics* 24, 189-209.
- DIEBOLD, F.X. AND G.D. RUDEBUSCH (1991) Is Consumption too Smooth? Long Memory and the Deaton Paradox. *The Review of Economics and Statistics* 73, 1-9.
- 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*. Chapman & Hall, New York.
- ENGLE, R.F. AND C.W.R. GRANGER (1987) Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* 55, 251-276.
- FRANKE, J. AND W. HARDLE (1992) On Bootstrapping Kernel Spectral Estimates. *Annals of Statistics* 29, 121-145.
- FRANSES, P.H. AND M. OOMS (1995) A Periodic Long Memory ARFIMA(0,  $D(s)$ , 0) Model for Quarterly UK Inflation. *Econometric Institute Report*, Erasmus University Rotterdam, The Netherlands.
- GEWEKE, J.F. 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. (1966) The Typical Spectral Shape of an Economic Variable. *Econometrica* 34, 150-161.

- 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.
- GRANGER, C.W.J. AND P. NEWBOLD (1986). *Forecasting Economic Time Series*, 2nd ed., Academic Press, San Diego, CA, USA.
- GRAY, H.L., N.-F. ZANG AND W.A. WOODWARD (1989) On Generalized Fractional Processes. *Journal of Time Series Analysis* 10, 233-257.
- GREENE, M.T. AND B.D. FIELITZ (1977) Long-Term Dependence in Common Stock Returns. *Journal of Financial Economics* 4, 339-349.
- HARVEY, A.C. (1993) Long Memory in Stochastic Volatility. *Unpublished manuscript*, London School of Economics.
- HASSLER, U. AND J. WOLTERS (1995) Long Memory in Inflation Rates: International Evidence. *Journal of Business & Economic Statistics* 13, 37-45.
- HASLETT, J. AND A.E. RAFTERY (1989) Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource, *Journal of the Royal Statistical Society C* 38, 1-50.
- HAUBRICH, J.G., (1993) Consumption and Fractional Differencing: Old and New Anomalies. *The Review of Economics and Statistics* ?, 767-772.
- 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.
- HOSKING, J.R.M. (1984) Modeling Persistence in Hydrological Time Series using Fractional Differencing. *Water Resources Research* 20, 1898-1908.

- HURST H.E. (1951) Long-term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers* 116, 770-799.
- HURST H.E. (1956) Methods of using Long-term Storage into Reservoirs. *Proceedings of the Institute of Civil Engineers* 1, 519-543.
- JOHANSEN, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231-254
- KUNSCH, H.R. (1989) The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics* 17, 1217-1241.
- LAHIRI, S.N., (1993) On the Moving Blocks Bootstrap under Long Range Dependence. *Statistics and Probability Letters* 18, 405-413.
- LO, A. W. (1991) Long Term Memory in Stock Market Prices. *Econometrica* 59 (5), 1279-1313.
- LYHAGEN J., (1997) Essays on Univariate Long-Memory Models. *Ph. D. Thesis*, Department of Statistics, Uppsala University, Sweden.
- MANDELBROT, B. AND J. WALLIS (1968) Noah, Joseph and Operational Hydrology. *Water Resources Research* 4, 909-918.
- MANDELBROT, B.B. (1972) Statistical Methodology for Nonperiodic Cycles. *Annals of Economic and Social Measurements* 1 (3), 259-290.
- NELSON, C.R. AND C.I. PLOSSER (1982) Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications. *Journal of Monetary Economics* 10, 139-162.
- NORDGAARD, A., (1996) Computer-Intensive Methods for Dependent Observations. *T.D. Thesis (Dissertations No. 409)*, Department of Mathematics, Linköping University, Sweden.
- OOMS, M. (1995) Flexible Seasonal Long Memory and Economic Time Series. *Unpublished manuscript*, Econometric Institute, Erasmus University Rotterdam, The Netherlands.

- PORTER-HUDAK, S. (1990) An Application of the Seasonally Fractionally Differenced Model to the Monetary Aggregates. *Journal of the American Statistical Association* 85, 338-344.
- RAY, B.K. (1993) Modelling Long-Memory Processes for Optimal Long-Range Predictions. *Journal of Time Series Analysis*, 14, 511-525.
- SMITH, J. AND S. YADAV (1994) Forecasting Costs Incurred from Unit Differencing Fractionally Integrated Processes. *International Journal of Forecasting* 10, 507-514.
- SOWELL (1992a) Modeling Long-Run Behavior with the Fractional ARIMA Model. *Journal of Monetary Economics* 29, 277-302.
- SOWELL (1992b) Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models. *Journal of Econometrics* 53, 165-188.
- TSCHERNIG, R. (1994) Long Memory in Foreign Exchange Rates Revisited, *Unpublished manuscript*, Institute of Statistics and Econometrics, Humboldt University of Berlin, Germany.

# **Part II**

## **Included Papers**





# 1

## Do Long-Memory Models have Long Memory?

**Abstract.** This paper calculates the memory of fractionally integrated ARMA(1,  $d$ , 1) processes. It is shown that fractional integration can enhance the memory of common ARMA processes enormously, but also that this is not necessarily the case. Very long memory is found for positively fractionally integrated processes with large positive autoregressive parameters. On the contrary, large negative AR parameters absorb the memory generated by a positive differencing parameter to a great extent. A moving average parameter may also reduce the memory substantially, even if the parameter alone causes virtually no memory.

**Key words:** ARMA, Fractional integration, Wold representation.

**JEL classification:** C22, C52, C53.

### 1.1 Introduction

This paper investigates the memory of fractionally integrated time series processes, with emphasis on how much fractional integration enhances the memory of autoregressive moving average (*ARMA*) processes. Fractionally integrated ARMA (*ARFIMA*) processes are said to be long-memory, due to their slow hyperbolic autocorrela-

tion and impulse response decay, compared to the faster geometrical decay of ARMA processes.

The model memory may be defined by how long (in time) a shock influences the process, which in this study is investigated through the Wold representation. The idea is similar to the decision rule of Öller (1985), which is based on Grangers' and Newbolds' (1986) predictability measure. Öllers' rule has been used by Oké and Öller (1997) on vector ARMA processes.

The memory of fractionally integrated ARFIMA(1, $d$ ,1) processes is investigated in the paper. The fractional cases are compared to processes where the differencing parameter  $d$  is set equal to zero. The results suggest that long-memory processes often have quite a long memory and that positive fractional integration extends the memory of a common ARMA process. However, this is not true for all parameter constellations examined.

## 1.2 A Method to Compare Model Memory

A time series process  $\{x_t\}$  follows an ARFIMA( $p, d, q$ ) process<sup>1</sup> if

$$\phi_p(B) (1 - B)^d x_t = \theta_q(B) a_t, \quad (1.1)$$

where the parameter polynomials are defined by

$$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad (1.2)$$

$$\theta_q(B) = 1 + \theta_1 B + \dots + \theta_q B^q \quad (1.3)$$

$$(1 - B)^d = 1 - dB - \frac{d}{2!} (1 - d) B^2 \dots \quad (1.4)$$

The differencing parameter  $d$ , describing the long-term dependence in the series, is allowed to assume any real value and the members of  $\{a_t\}$  are *i.i.d.* disturbances with mean zero and variance  $\sigma_a^2 < \infty$ . By restricting the differencing power  $d$  to zero, the common ARMA( $p, q$ ) specification is retrieved as a special case of (1.1).

If the polynomials (1.2) and (1.3), jointly describing the short-term behavior, have all roots outside the unit circle and  $d$  lies in the interval  $(-1/2, 1/2)$ , the process is both invertible and stationary.

---

<sup>1</sup> See Granger and Joyeux (1980) and Hosking (1981) for an introduction to fractional integration and an overview of the ARFIMA properties.

Stationarity implies the existence of a Wold representation (sometimes called the memory function),

$$x_t = \phi_p^{-1}(B)\theta_q(B)(1-B)^{-d}a_t = (1 + \psi_1 B + \psi_2 B^2 + \dots)a_t, \quad (1.5)$$

where the psi-weights  $\psi_j$  may be calculated as in e.g. Lyhagen (1997). Equation (1.5) shows how past disturbances influence the series and through this representation, the  $l$ -step forecast error variance of (1.1) can conveniently be calculated as

$$\xi(l) = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2. \quad (1.6)$$

Using the same tools, the variance of  $x_t$  is defined by

$$\sigma_x^2 = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2. \quad (1.7)$$

Details on ARFIMA variance and covariance calculations are given by Sowell (1992) and Chung (1994).

The memory-structure of (1.5), and thus (1.1), is often characterized by the mean (or average) lag,  $\overline{Lag} = \sum_{j=0}^{\infty} j\psi_j / \sum_{i=0}^{\infty} \psi_i$ . The mean lag provides a measure of how slowly the psi-weights die out and thereby enables model memory comparisons. However, the mean lag does not exist in the case of a fractionally integrated process and therefore the quantity cannot be utilized in our study.

Instead, we use a measure based on the prediction error variance. From equations (1.6) and (1.7) we see that  $\xi(l)$  approaches  $\sigma_x^2$  with increasing  $l$  and when they are sufficiently close, that is when

$$\frac{\xi(L+1)}{\sigma_x^2} \geq c \quad c \in (0, 1), \quad (1.8)$$

the memory is given by  $L$ . We employ the following decision rule: the memory of a particular process is  $l_c = L$  if

$$\xi(L) < c\sigma_x^2 \leq \xi(L+1). \quad (1.9)$$

The choice of  $c$  is of course arbitrary and will have an impact on the measured memory (which has no practical interpretation), but the decision rule (1.9) allows comparisons between processes or models.

The notion of process memory is of course closely related to the theoretical maximum prediction horizon of that process. Granger and Newbold (1986) suggest a measure of predictability of a stationary series  $z_t$ , having a finite variance, under squared error loss as the quantity

$$R_{h,z}^2 = 1 - \frac{\text{var} [\hat{a}_{t+j|t}]}{\text{var} [z_{t+h}]}, \quad (1.10)$$

where  $\hat{a}_{t+j|t} = z_{t+h} - \hat{z}_{t+h|t}$ .  $z_{t+h}$  is the actual value of process  $z$  at  $t+h$  and  $\hat{z}_{t+h|t}$  the (optimal) predicted value of  $z$ ,  $h$  steps ahead from  $t$ . If  $j$  equals  $l$  in equation (1.6), and  $z_t$  and  $x_t$  are the same covariance stationary process, then using equation (1.10) the memory definition (1.6) may be written

$$1 - R_{h,z}^2 \geq c \quad c \in (0, 1).$$

The predictability measure of Granger and Newbold has been generalized by Diebold and Kilian (1997) to allow for nonstationary series, general loss functions and multivariate information sets.

The predictability measure of Granger and Newbold and a test similar to that for the coefficient of determination of ordinary regressions may be utilized to estimate the memory. Using the decision rule of Öller (1985), the memory of a model estimated on the process (1.1), given *i.i.d.* normality of  $a_t$ , is  $l_f = L$  if

$$\hat{R}_{L+1,z}^2 < \delta_v(T, k) \leq \hat{R}_{L,z}^2. \quad (1.11)$$

$\hat{R}^2$  is the estimated value of  $R^2$ ,  $v$  the significance level,  $T$  the number of observations and  $k$  the number of parameters in the model. The critical value is given by

$$\delta_v(T, k) = \frac{F(v, k, T - k - 1)}{F(v, k, T - k - 1) - \frac{T - k - 1}{k}}, \quad (1.12)$$

where the quantity  $F(\cdot)$  is the tabulated value of the  $F$ -distribution. The critical value (1.12) depends on  $T$  and  $k$ , and is thus subject to estimation matters, which are beyond the scope of this paper. However, there exists a relationship between (1.9) and (1.11), that is  $l_c = l_f = L$  if  $c = 1 - \delta_v(T, k)$ .

The memory of a process may not always coincide with the practical prediction horizon of the process. For instance, a random walk

in the levels has infinite memory but cannot fruitfully be predicted even at horizon one. This is due to nonstationarity, and it is suggested that such series should be differenced to produce a finite variance series. In the first differences the random walk then has no memory or predictability. For our purposes, the case where the memory and practical prediction horizon diverge is when the autoregressive or fractional difference parameter is very close to unity and  $1/2$  respectively. The memory (and prediction horizon) is then driven by the large variances, which may lead to extremely uncertain long-term forecasts.

## 1.3 Results

In this study the ARFIMA(1,  $d$ , 1) specification,

$$(1 - \phi B)(1 - B)^d x_t = (1 + \theta B) a_t \quad a_t \sim N(0, 1),$$

is employed to illustrate the memory  $L$ .<sup>2</sup> The parameters are chosen as  $d = \{0, 0.10, 0.20, \pm 0.30\}$ , and  $\phi$  and  $\theta = \{0, \pm 0.1 \pm 0.5 \pm 0.9\}$ . Figure 1.1 presents the memory sensitivity of these values of  $\phi$ ,  $d$  and  $\theta$ .

### 1.3.1 Short-Memory Processes

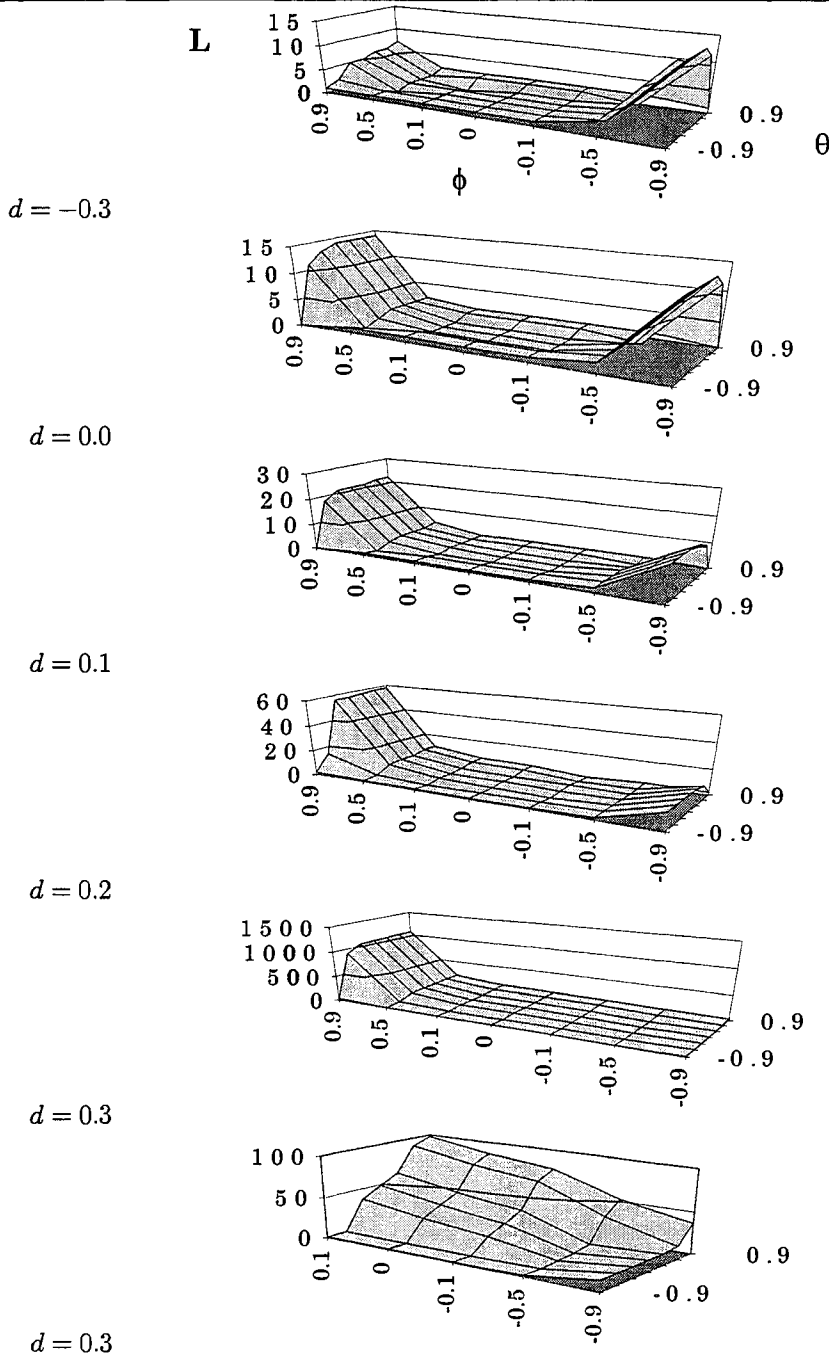
The ARFIMA(1, 0, 0) process, commonly known as AR(1), has geometrically decaying psi-weights  $\psi_j = \phi^j, \forall j \geq 1$ . Consequently, a small  $\phi$  implies that just a few squared terms of (1.6) approximate the variance (1.7) well, resulting in a short memory structure, see the  $d = 0$  picture in Figure 1.1 and the line corresponding to  $\theta = 0$ . However, almost unit root processes exhibit a longer memory. For the autoregressions investigated, the memory is not driven by large process variances, and hence the results may be interpreted in terms of prediction horizons.

A first order moving average has memory one in terms of the Wold representation, i.e.  $\psi_j = 0$  for  $j > 1$ , but if the parameter  $\theta$  is close to zero there is for practical purposes no memory in the model, i.e. the process is almost white noise.

---

<sup>2</sup>The coefficients  $\psi_j$  of (1.5) for  $p$  and  $q$  equal to one are shown in the *Appendix*.

FIGURE 1.1. Memory of ARFIMA(1,d,1) processes.



When combining AR and MA parameters, while  $d = 0$ , we notice that the autoregressive parameter governs the memory, and that a moving average parameter may alter the memory only slightly, e.g. the memory decreases when an MA parameter of opposite sign is introduced to AR processes with a large  $\phi$ .

### 1.3.2 Long-Memory Processes

An introduction of fractional integration to white noise, resulting in fractional noise or ARFIMA(0,  $d$ , 0), increases the memory from  $l_{0.95} = 0$ , for  $d = 0$ , to  $l_{0.95} = 91,226$  for  $d = 0.4$  (not reported in the figure). Thus, long-memory processes may indeed have a long memory-structure. It is worth noticing that the large prediction horizon for the process is not a result of a large process variance; this variance is only twice the disturbance variance.

However, this is not always the case that fractional integration causes long memory, e.g. anti-persistence ( $d < 0$ ) increases the memory only marginally, for instance  $l_{0.95} = 1$  for  $d = -0.3$ . When a  $d < 0$  process also contains a positive autoregressive parameter, the memory is reduced compared to the corresponding ARMA model. Even when the parameters have the same sign, the memory is similar to that of the ARMA model. The results for anti-persistent processes can explain why non-fractional models suffer only slightly, compared to fractional ones, when predicting over-differenced processes. This fact is noticed by Smith and Yadav (1994).

Weakly persistent ( $0 < d \leq 0.2$ ) ARFIMA processes (represented by the  $d = 0.1$  and  $0.2$  pictures in Figure 1.1) behave similarly to ARMA processes, but if the AR polynomial has a large positive parameter the memory is longer, and sometimes much longer (in particular when  $d = 0.2$ ), than that of pure autoregressions. Not only do positive AR parameters have an enhancing effect on the memory, but this is also the case for processes with a positive moving average parameter.

If the degree of differencing equals 0.3, the model really has the capability of generating long memory structures, but yet again the results show that negative AR and MA parameters restrict the process memory. However, a positive MA parameter in addition to the 0.3 differencing power yields a maximum memory of 86 (with  $c = 0.95$ ), while a maximum of 1,051 is found for ARFIMA(1, 0.3, 0) processes.

Some interesting features of the ARFIMA(1,0.3,1) specification are worth pointing out. If the process contains large negative AR and MA parameters the memory is determined by those, implying the same memory as the corresponding ARMA(1,1) process. Furthermore, we see that positive autoregressive parameters trigger the memory, but the memory can still be quite long for some negative values. However, this requires a positive MA parameter. The moving average parameter may also reduce the memory, in particular a negative value can absorb the effect of an AR parameter of a value up to 0.5. AR and MA parameters of different signs cancel out any memory that may occur from these parameters. It is also quite interesting to note that the MA parameter in a full ARFIMA model affects the memory to a great extent, even though the parameter alone generates virtually no memory.

## 1.4 Conclusions

This paper shows that the long-memory ARFIMA model does have long memory under some circumstances, but not always, and that fractional integration often extends the memory of ARMA processes. In particular, ARFIMA models with large positive differencing powers and positive parameters in the autoregressive polynomial do have an extremely long memory. However, the long memory generated by a positive differencing parameter vanishes if a large negative autoregressive parameter is introduced in the process. Furthermore (and more surprisingly), negative moving average parameters (which alone make no memory contribution) absorb a substantial amount of memory induced by fractional integration.

## Acknowledgements

Financial support from the Browaldh Foundation is gratefully acknowledged. I also thank Sune Karlsson and Lars-Erik Öller for discussions and suggestions, and Johan Lyhagen Mickael Löthgren and Timo Teräsvirta for comments on previous drafts. The usual disclaimer applies.



## References

- Chung, C.-H., 'A note on calculating the autocovariances of the fractionally integrated ARMA models', *Economics Letters*, **45**, (1994), 293-297.
- Diebold, F.X. and L. Kilian, 'Measuring Predictability: Theory and Macroeconomic Applications', *Unpublished manuscript*, Department of Economics, University of Pennsylvania, USA, (1997).
- Granger, C. W. J. and Joyeux, R., 'An introduction to long-memory time series models and fractional differencing', *Journal of Time Series Analysis*, **1**, (1980), 15-29.
- Granger, C. W. J. and P. Newbold, '*Forecasting economic time series*', 2nd ed., Academic Press, San Diego, CA, USA, (1986).
- Hosking, J. R. M., 'Fractional differencing', *Biometrika*, **68**, (1981), 165-176.
- Lyhagen, J., 'A matrix evaluation of the moving average representation', *forthcoming in Economics Letters*, (1997).
- Nelson, C. R., 'The interpretation of  $R^2$  in autoregressive-moving average time series models', *The American Statistician*, **30**, (1976), 175-180.
- Oké T. and Öller L.-E., 'Testing for short memory in a VARMA process', in Oké, T., *Pre-testing and specification search in an autoregressive moving average process with extensions to unit root cases*, Unpublished doctoral dissertation, Uppsala University, 1997.
- Parzen, E., 'ARARMA models for time series analysis and forecasting', *Journal of Forecasting*, **1**, (1982), 67-82.
- Smith, J. and Yadav, S., 'Forecasting costs incurred from unit differencing fractionally integrated processes', *International Journal of Forecasting*, **10**, (1994), 507-514.
- Sowell, F., 'Maximum likelihood estimation of stationary univariate fractionally integrated time series models', *Journal of Econometrics*, **53**, (1992), 165-188.

Öller, L.-E., 'How far can changes in general business activity be forecasted', *International Journal of Forecasting*, 1, (1985), 135-141.

# Appendix

## The Wold Representation for the ARFIMA(1,d,1) Model

For tractable specifications, the  $\psi$ -coefficients of the Wold representation can conveniently be calculated in the following way. Equation (1.5) requires that

$$\phi_p^{-1}(B)\theta_q(B)(1-B)^{-d} = \psi(B),$$

where the polynomials, for the ARFIMA(1,  $d$ , 1) model, are

$$\begin{aligned}\phi_p(B) &= 1 - \phi B \\ \theta_q(B) &= 1 + \theta B \\ (1-B)^d &= 1 - d_1 B - d_2 B^2 - \dots \\ \psi(B) &= 1 + \psi_1 B + \psi_2 B^2 + \dots\end{aligned}$$

$d_i$  is given recursively as a function of the differencing parameter  $d$ ,

$$\begin{aligned}d_1 &= d \\ d_i &= \frac{1}{i}d_{i-1}(i-1-d), \quad \text{for } i > 1.\end{aligned}$$

Finally, the  $\psi$ -weights can be calculated by equating coefficients in

$$(1 - \phi B)(1 - d_1 B - d_2 B^2 - \dots)(1 + \psi_1 B + \psi_2 B^2 + \dots) = (1 + \theta B).$$

For this equality to hold, the coefficients of  $B^j$  on each side have to be the same, that is

$$B^1: \quad \psi_1 - \phi - d_1 = \theta \Rightarrow$$

$$\psi_1 = \phi + d_1 + \theta$$

$$B^2: \quad \psi_2 - d_2 - \psi_1\phi - \psi_1d_1 + \phi d_1 = 0 \Rightarrow$$

$$\psi_2 = d_2 + \psi_1\phi + \psi_1d_1 - \phi d_1$$

$$B_{j \geq 3}^j: \quad \psi_j = \sum_{i=0}^{j-1} (\psi_i d_{j-i}) - \phi \sum_{k=1}^{j-2} (\psi_k d_{j-k-1}) + \phi (\psi_{j-1} - d_{j-1}).$$

## 2

# On the Effects of Imposing or Ignoring Long Memory when Forecasting

**Abstract.** Since the true nature of a time series process is often unknown it is important to understand the effects of model choice. This paper examines how the choice between modelling stationary time series as ARMA or ARFIMA processes affects the accuracy of forecasts. This is done, for first-order autoregressions and moving averages and for ARFIMA(1, $d$ ,0) processes, by means of a Monte Carlo simulation study. The fractional models are estimated using the technique of Geweke and Porter-Hudak, the modified rescaled range and the maximum likelihood procedure. We conclude that ignoring long memory is worse than imposing it, when forecasting, and that the ML estimator is preferred.

**Key words:** ARFIMA; Fractional integration; Periodogram regression; Rescaled range; Maximum likelihood; Forecast error

**JEL Classification Code:** C15; C22; C53

## 2.1 Introduction

The use of erroneous models when analyzing time series may have a great impact on, for instance, the accuracy of forecasts and policy decisions. This paper investigates the forecasting performance of the

usual autoregressive moving average (*ARMA*) model when the true process is fractionally integrated and the performance of the fractionally integrated ARMA (*ARFIMA*) model when the true process is non-integrated. The faulty specification is compared to the correct one.

The ARFIMA model generalizes the well-known ARIMA model by allowing for non-integer differencing powers and thereby provides a more flexible framework when examining time series. For example, fractional specifications can model data dependencies which are stronger than those allowed in stationary ARMA models, but weaker than those implied by unit root processes. This is an attractive feature when investigating economic and financial time series, which often exhibit a strong dependence between distant observations. Although the ARMA model is a special case of the ARFIMA model, its use is motivated by its simplicity and smaller bias when the differencing parameter is integer-valued.

In a previous study, Lyhagen (1997) demonstrates theoretically that ignoring long memory (i.e. a fractional differencing power) may lead to very high relative mean squared errors of prediction. However, the results of Ray (1993) show that high-order AR models forecast fractional noise series well, even in the long term. Also for fractional noise, Smith and Yadav (1994) point out that there is a potential loss from incorrectly fitting an AR model when the differencing parameter is positive. On the other hand, negative fractional differencing will bring about a performance loss of the AR model only at a horizon of one step.

This paper analyzes the forecasting performance of the respective models in terms of their mean squared prediction errors. Three popular estimation procedures for the fractional model are used, namely the one of Geweke and Porter-Hudak (1983), the modified rescaled range of Lo (1991) and maximum likelihood (Sowell, 1992), in order to compare the performance of the estimation techniques. As a consequence, we incorporate the effects of estimation bias and possible model selection mistakes. Furthermore, a test for significant differences in prediction performance is utilized.

The results suggest that the forecast errors from ARMA models are larger than those from ARFIMA models and that the maximum likelihood procedure is best.

The paper is organized as follows: the estimation techniques for the ARFIMA model are introduced in Section 2. Section 3 contains the simulation study where the model predictions are compared and Section 4 concludes the paper.

## 2.2 Estimation of the ARFIMA Model

A time series process is said to be integrated of order  $d$ , denoted  $I(d)$ , if it has a stationary and invertible autoregressive moving average (ARMA) representation with uncorrelated disturbances after differencing  $d$  times, that is after applying the filter  $(1 - B)^d$ . When  $d$  is not integer-valued (as required for ARIMA processes) the series is fractionally integrated. An ARFIMA( $p, d, q$ ) process  $\{x_t\}$  is generated by

$$\phi_p(B)(1 - B)^d x_t = \theta_q(B) a_t, \quad (2.1)$$

where the members of the sequence  $\{a_t\}$  are identically and independently distributed (*iid*) with finite variance. If the roots of  $\phi_p(B)$  and  $\theta_q(B)$ , the autoregressive and moving average polynomials, lie outside the unit circle and  $d < 0.5$ ,  $x_t$  is stationary. When  $d > 0$   $x_t$  is persistent, implying that there exists a region,  $d \in (0, 0.5)$ , where the ARFIMA model generates stationary series ruled by long memory. This behavior cannot be mimicked with ARMA models.

The differencing filter, denoted the long-memory filter (*LMF*), describes the long-term dependence in the series and may be expanded as

$$(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k = 1 - dB - \frac{d}{2!} (1 - d) B^2 \dots \quad (2.2)$$

The short-term structure is captured by the autoregressive and moving average parameters, which enables separate modelling of the long and short-run dynamics. The properties of the ARFIMA model are exhaustively described by Granger and Joyeux (1980) and Hosking (1981).

In this study the ARFIMA models are estimated by the *GPH* estimator of Geweke and Porter-Hudak (1983), the modified R/S or modified rescaled range (*MRR*, Lo (1991)) and the full information maximum likelihood estimator (*MLE*) of Sowell (1992).

### 2.2.1 The GPH Estimation Technique

Geweke and Porter-Hudak (1983) proposed a two-step procedure for the estimation of fractionally integrated models, based on a non-parametric periodogram regression.

In a *first* step,  $d$  is estimated in the regression equation (where  $I_x(\cdot)$  denotes the sample periodogram)

$$\ln \{I_x(\omega_j)\} = \beta_0 + \beta_1 \ln \{4 \sin^2(\omega_j/2)\} + \eta_j, \quad (2.3)$$

where  $(\omega_j = 2\pi j/T, j = 1, \dots, g(T))$ ,

by ordinary least squares (*OLS*). Prior to estimation of the remaining (ARMA) parameters in a *second* step, the long-memory part of the series is filtered out using the LMF (2.2).  $\hat{\beta}_1$  is under a proper choice of  $g(T)$  a consistent estimator of  $-d$ .  $g(T)$  is set to the integer part of  $T^v$  and the widespread choice of  $v$  is 0.5. The crucial assumption is that the spectrum of an ARFIMA( $p, d, q$ ) process is the same as the spectrum of an ARFIMA(0,  $d$ , 0), for the same value of  $d$ , at low frequencies. However, Agiakloglou *et.al.* (1993) show that large positive AR and large negative MA parameters affect the spectrum at low frequencies and hence cause biased estimates.

### 2.2.2 The MRR Estimation Technique

The modified rescaled range, MRR, approach rests upon the same basic idea as the GPH procedure, that is to estimate  $d$  in a first step and the other parameters in a second one. Instead of using the log-periodogram regression, the fractional differencing parameter is estimated by the R/S statistic, which is also consistent. Lo (1991) robustifies the statistic to short-range dependence in data and this modified R/S statistic is frequently used when testing for fractional integration and for estimation of long-memory models.

The MRR statistic is defined by the ratio

$$Q_T = \frac{R_T}{\hat{\sigma}_T(k)}. \quad (2.4)$$

Following Cheung (1993), the range and standard error are estimated by



$$R_T = \max_{0 < i \leq T} \sum_{t=1}^i (x_t - \bar{x}) - \min_{0 < i \leq T} \sum_{t=1}^i (x_t - \bar{x}) \quad (2.5)$$

$$\hat{\sigma}_T^2(k) = \hat{\sigma}^2 + \frac{2}{T} \sum_{j=1}^k \sum_{i=j+1}^T \left(1 - \frac{j}{k}\right) (x_i - \bar{x})(x_{i-j} - \bar{x}), \quad (2.6)$$

where  $\hat{\sigma}^2$  is the usual maximum likelihood variance estimate and the correction term is similar to that of Newey and West (1987). 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  $(3T/2)^{1/3} \{2\hat{\rho}/(1 - \hat{\rho}^2)\}^{2/3}$ , with  $T$  and  $\hat{\rho}$  denoting the serial length and sample first-order autocorrelation coefficient respectively. Asymptotic results suggest that  $\ln Q_T / \ln T$  approaches  $1/2$  for short-range dependent processes, and thus our estimator of  $d$  is constructed as

$$\hat{d} = \frac{\ln Q_T}{\ln T} - \frac{1}{2}. \quad (2.7)$$

The MRR estimator is biased when the true process is a moving average with large negative parameter values (Cheung (1993)).

### 2.2.3 The ML Estimation Technique

The maximum likelihood estimator<sup>1</sup> differs from the other two; while the GPH and MRR estimators are performed in two steps the ML procedure estimates all parameters in one single step.

Given that the roots of  $\phi_p(B)$  and  $\theta_q(B)$  in (2.1) are all outside the unit circle and the disturbances are normally distributed with mean zero, the likelihood function to be maximized is the well-known

$$L(X_T, \Sigma(\varsigma)) = (2\pi)^{-\frac{T}{2}} |\Sigma(\varsigma)|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} X_T' \Sigma^{-1}(\varsigma) X_T \right\}, \quad (2.8)$$

where  $X_T$  is the  $T \times 1$  data vector and  $\Sigma(\varsigma)$ , a Toeplitz form autocovariance matrix, is a function of the unknown parameter vector

---

<sup>1</sup>A FORTRAN routine for the estimation of the ARFIMA model was generously supplied by Fallaw Sowell.

$\varsigma = [d, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_a^2]'$ . The log of (2.8) is maximized directly with respect to the vector  $\varsigma$ . Explicit expressions for the elements of  $\Sigma(\varsigma)$  and practical issues concerning the ML estimation are discussed by Sowell (1992).

Generally the MLE is efficient; however, it has some drawbacks are. For example, the MLE requires distinct roots, performs less well when  $d$  is close to the non-stationary area (i.e. close to 0.5) and exhibits small-sample bias when  $d$  and AR parameters are estimated jointly (Tschernig (1993)). Furthermore, mis-specification of the likelihood function will in general lead to inconsistent estimates of both the differencing parameter and the short-run parameters.

### 2.3 The Monte Carlo Study

ARFIMA( $p, d, q$ ) processes are generated by the algorithm of Diebold and Rudebusch (1991). The processes are primarily chosen to generate persistent or autoregressive series, but negative  $d$ -values and MA processes are also considered. In each Monte Carlo iteration we estimate ARMA and ARFIMA models, from which predictions are generated. For each prediction horizon  $h$  we calculate the loss functions  $g^m(x_h, \hat{x}_h)$ , where  $x_h$  denotes the actual (simulated) value of the process at time  $T + h$  and  $\hat{x}_h$  the corresponding forecast using model  $m = \{ARMA, ARFIMA\}$ . The accuracies are compared by the mean squared prediction error (*MSPE*) loss function

$$g^m(x_h, \hat{x}_h) = R^{-1} \sum_{r=1}^R (\hat{x}_{h,r} - x_{h,r})^2, \quad (2.9)$$

where  $r = 1, \dots, R$  is the Monte Carlo replicate.  $R$  is selected to 1000. The mean squared prediction error of model  $m$  is expected to increase with the prediction horizon and eventually coincide with, or exceed, the variance of the process. At that particular horizon, that is when the MSPE variance ratio

$$MVR_m = g^m(x_h, \hat{x}_h) / V(y) \quad (2.10)$$

is greater than or equal to one, the model forecasts are not more accurate than when using the process mean. The difference in  $h$ -step performance, expressed by the null-hypothesis of no difference

in MSPE, is evaluated by a usual matched-pair  $t$ -test,

$$t_h = \frac{g_h^{ARMA} - g_h^{ARFIMA}}{\sqrt{V[g_h^{ARMA} - g_h^{ARFIMA}]}} \quad (2.11)$$

along with asymptotically normal critical values, motivated by the large number of replicates and the central limit theorem.

Results for first-order autoregressions and fractionally integrated processes of length 100 and forecasting horizons up to step 20 are reported in the figures of this paper. In addition, larger sample sizes ( $T = 225, 400$  and  $625$ ) and longer prediction horizons (up to  $0.2T$ ) have been investigated. An increase in the serial length implies more accurate estimates (in particular for the fractional procedures), and a relatively better prediction performance for the ARFIMA model. Otherwise, the conclusions drawn from the  $T = 100$  case are not altered by the results obtained for the larger sample sizes.

### 2.3.1 Autoregressions

The forecasting performances for generated first-order autoregressions are examined using the specification

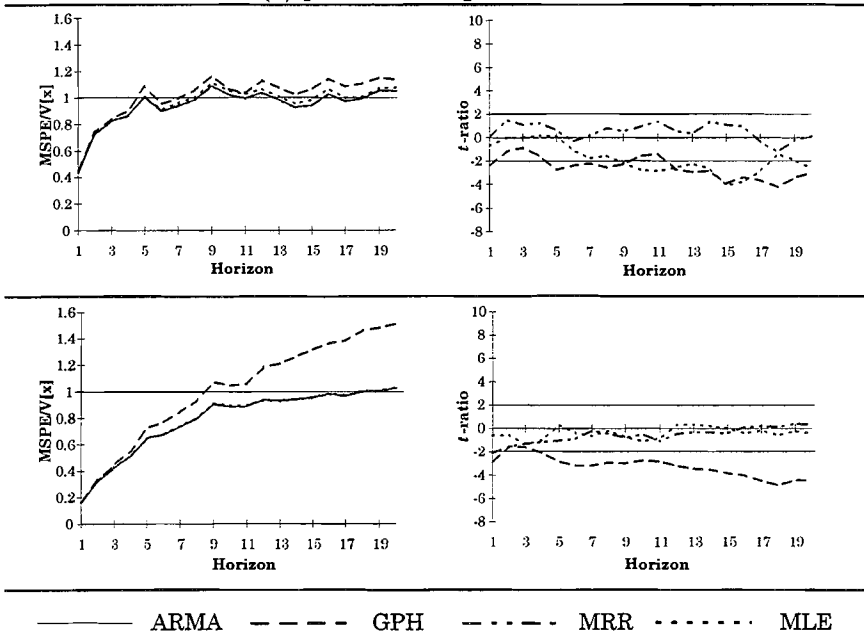
$$x_t = \phi x_{t-1} + a_t, \quad (2.12)$$

where  $\{a_t\}$  is a sequence of  $iidN(0,1)$  random deviates. In order to handle model selection uncertainty, the lag-orders  $p$  and  $q$  of the ARMA and fractional models are selected by the Bayesian information criterion,  $BIC$ , of Schwartz (1978) from the values  $\{0, 1, 2, 3\}$  in each Monte Carlo iteration. The case of no differencing parameter in the ARFIMA model is ruled out.

The study investigates autoregressions with parameter  $\phi = 0.5, 0.7$  and  $0.9$ . Figure 2.1 presents the MSPE variance ratio (2.10) for each of the models together with the associated  $t$ -statistic (2.11) for  $\phi = 0.7$  and  $0.9$  processes.

The maximum possible prediction horizon increases with the size of the AR parameter. In particular, when  $\phi = 0.5$  (not reported in the figures), the MVRs rapidly assume values above one, which implies short prediction horizons. For prediction steps one to three, the model mean squared prediction errors are all quite similar. However, a few significant differences (from the ARMA performance) are found

FIGURE 2.1. Estimated mean squared prediction errors and t-values of the MSPE test for an AR(1) process of length 100.



The figure presents the MSPE variance ratio (2.10) (first column) and the  $t$ -statistic of the MSPE test (2.11) (second column) of the estimation procedures for horizons to 20. The rows correspond to  $\phi = 0.7$  and  $0.9$ . The data are generated according to (2.12).

for the GPH and maximum likelihood estimated models. Using the MRR procedure, the forecasts are as accurate as those produced by the ARMA model. For prediction horizons above three, forecasting is valueless.

As the parameter  $\phi$  becomes larger, the prediction horizon (using the ARMA model) increases to a maximum of five (maybe eight) steps ahead for  $\phi = 0.7$  processes and 18 steps when  $\phi = 0.9$ . For  $\phi = 0.7$ , and in particular  $\phi = 0.9$ , the bias of the GPH procedure is quite large and thus the forecast errors obtained are also quite large. As a consequent, the GPH performance is often significantly worse than that of the non-fractional model, that is the  $t$ -values fall below -2, for  $\phi = 0.7$  processes. For almost unit root processes, the GPH estimator produces prediction errors that are usually significantly larger than those of the ARMA model. The other estimation techniques predict

highly short-term dependent processes (almost) as well as the ARMA models; their  $t$ -ratios usually fall between -2 and 2.

Turning to the mean squared prediction errors, we find no support in favor of the non-fractional specification. In a forecasting context, no substantial loss is experienced when imposing long memory by using ARFIMA models on AR processes. In particular, the MRR procedure performs very well, but the ML estimator is also quite accurate.

### 2.3.2 Persistent Processes

The fractionally integrated series are generated as ARFIMA(1,  $d$ , 0) processes,

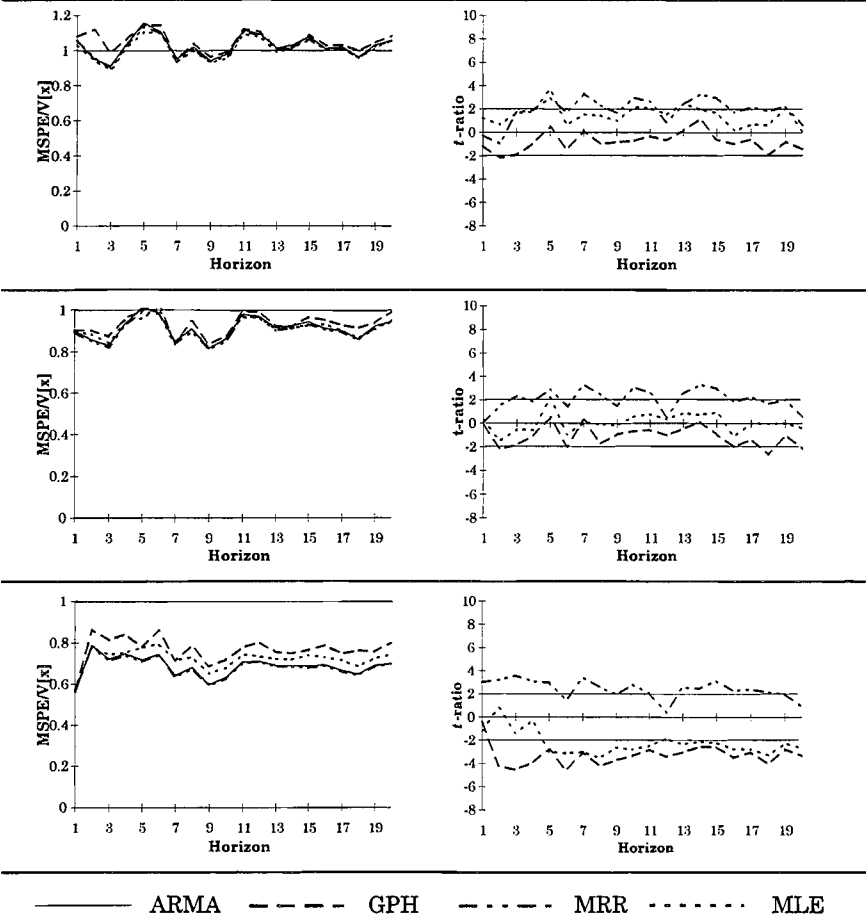
$$(1 - \phi B)(1 - B)^d x_t = a_t, \quad (2.13)$$

where  $a_t$  is again  $iidN(0, 1)$ . The process is denoted fractionally differenced white noise (*FDWN*) when  $\phi$  equals zero. In the case of generated fractionally integrated processes, the models to be estimated are specified by the BIC, where the orders  $p$  and  $q$  are selected from the values  $\{0, 1, \dots, 10\}$ . The MSPE variance ratios (2.10) and  $t$ -values of the MSPE test (2.11) are reported in Figures 2.2 and 2.3.

Weakly persistent FDWN processes with the parameter  $d$  equal to 0.1 or 0.2, being almost white noise, do not contain enough structure to be forecasted. Already at horizon one, the MVRs (see Figure 2.2 for  $d = 0.2$ ) exceed one and subsequently predictions are useless, regardless of estimation technique and model choice.

Results for  $d = 0.4$  in Figure 2.3 suggest that only occasional ARFIMA MVRs of unity are found for a more persistent fractional noise process. In fact, up to step 100 forecasts using the ARFIMA models (with the MLE) are possible. Furthermore, the ARFIMA-MLE is always significantly better than the ARMA model; the  $t$ -values are well above 2. The GPH estimator is better than the ARMA specification at step five and above (significantly at step seven) and the MRR at horizon eight (ten). The relatively weak performance of the MRR procedure is explained by over-adjustment for (non-existing) short-term dependence in data. Therefore, additional AR parameters, and thus mis-specified models, are as a rule required to capture all autocorrelation in the series. Judged by the mean squared errors, the performance of the MRR procedure for ARFIMA(0, 0.4, 0) processes is close to that of the ARMA model, while the GPH pro-

FIGURE 2.2. Estimated mean squared prediction errors and t-values of the MSPE test for an ARFIMA(1,0.2,0) process of length 100.



See note to Figure 2.1. The rows correspond to  $\phi = 0.0, 0.2$  and  $0.5$  respectively. The data are generated according to (2.13).

cedure and in particular the method of maximum likelihood generates notably lower values. The results suggest, when forecasting is worthwhile, that the ARFIMA-MLE should be used for predicting fractional noise processes.

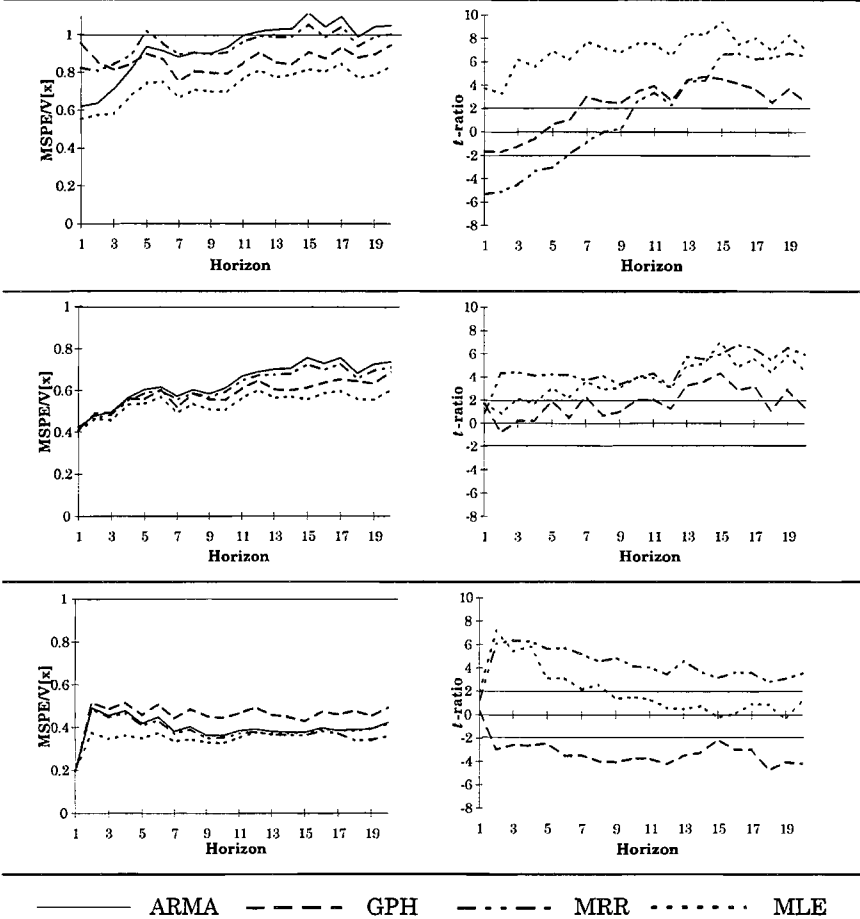
An introduction of AR-type short memory, according to equation (2.13), enhances the predictability compared to FDWN for the same value of  $d$ . Unlike the case of fractionally differenced white noise with  $d = 0.2$ , processes with  $d = 0.2$  and  $\phi = 0.2$  may be predicted several steps ahead. The performances of the ARFIMA model, with all estimation techniques, and the ARMA model are fairly similar, and the  $t$ -ratios suggest that the MRR may be better (by a close margin) than the other estimators. Again for  $d = 0.2$ , when  $\phi$  equals 0.5 we notice that the MRR procedure produces the most accurate forecasts, followed by the non-fractional model. It appears to be the case that the 0.2 differencing power is not large enough to cause problems for the ARMA model. The GPH estimator gives the worst performance in this situation, but even this is not bad.

Proceeding with ARFIMA(1, 0.4, 0) processes, when  $\phi = 0.2$  the GPH predictions are better than those of the ARMA model. However, the variance of  $(g_h^{ARMA} - g_h^{ARFIMA})$  is quite large, leading to small  $t$ -ratios which are usually positive but not always significantly larger than zero. The MRR procedure works well in the presence of AR parameters, and exhibits  $t$ -ratios favoring the ARFIMA model. However significant, the differences in mean squared prediction errors are quite small. The maximum likelihood estimated ARFIMA models display the best performance and the mean squared prediction errors are notably lower than those of the other procedures, especially from horizon seven and onwards.

When  $d = 0.4$  and  $\phi = 0.5$ , the MLE estimated models again have much smaller MSPE figures than the ARMA model. The differences are not that great for the MRR estimated ARFIMA models, but they are strongly significant. As in the case of AR processes with large values of  $\phi$ , the GPH procedure generates poor forecasts for ARFIMA(1, 0.4, 0) processes with intermediate to large positive autoregressive parameters.

The non-fractional ARMA model works fairly well for short-range predictions when the true process is fractionally differenced, whereas the ARFIMA specifications are optimal for long-range forecasting. In general, for the fractionally integrated processes investigated, the

FIGURE 2.3. Estimated mean squared prediction errors and t-values of the MSPE test for an ARFIMA(1,0.4,0) process of length 100.



See note to Figure 2.1. The rows correspond to  $\phi = 0.0, 0.2$  and  $0.5$  respectively. The data are generated according to (2.13).



best predictions are obtained when using the maximum likelihood ARFIMA estimation procedure.

### *2.3.3 Moving Averages and Anti-Persistent Processes*

For MA(1) processes the choice of model hardly matters. The ARMA model is found to have a slightly better performance, and the maximum likelihood estimated ARFIMA models are in turn marginally more accurate than those estimated by the GPH and MRR procedures. The simplicity of the non-fractional specification motivates the use of ARMA models in this case. However, the true process is unknown and the modelling choice hardly affects the forecasting performance.

For anti-persistent ( $d < 0$ ) processes also, the difference in forecasting performance between the ARMA and ARFIMA models is very small. This is in agreement with the results of Smith and Yadav (1994).

## 2.4 Concluding Remarks

This paper investigates and compares the forecasting performance of ARMA and ARFIMA models. The results suggest that the modified rescaled range and maximum likelihood estimators for the fractionally integrated model generate predictions that are almost as accurate as those of the ARMA model, when the true process is a first-order autoregression or moving average. The predictions are also quite similar for negatively fractionally integrated processes. The estimator of Geweke and Porter-Hudak performs badly when the true process is an AR(1) with a large positive parameter.

For persistent processes, the GPH and MRR produce short-term predictions that are worse than the ARMA forecasts. However, the opposite is found for intermediate and long-range forecasts. The MLE is better than the ARMA model for all horizons.

The GPH technique experiences problems when combining fractional integration with an autoregression, especially for intermediate to large AR parameters. In the case of simultaneous short and long memory, the MRR and in particular the MLE procedures generate better predictions than the ARMA model.

In general, it is worse to ignore than to impose long memory when forecasting. Overall, the MLE is the best of the fractional estimators and the ARFIMA model is better than the ARMA.

## Acknowledgements

Suggestions by and discussions with Sune Karlsson and Johan Lyhagen are gratefully acknowledged. I also thank Lars-Erik Öller for comments on previous drafts and participants of the ISF'96, the Sixteenth International Symposium on Forecasting, Istanbul meeting, June 24-27, 1996. 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, 235-246.
- 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, 331-345.
- Diebold, F.X. and G.D. Rudebusch, 1989, Long memory and persistence in aggregate output, *Journal of Monetary Economics*, 24, 189-209.
- Diebold, F.X. and G.D. Rudebusch, 1991a, Is consumption too smooth? Long memory and the Deaton paradox, *The Review of Economics and Statistics*, (Feb), 1-9.
- Diebold, F.X. and G.D. Rudebusch, 1991b, 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, 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, *Biometrika*, 68, 165-176.
- Lo, A. W., 1991, Long-term memory in stock market prices, *Econometrica*, 59, 1279-1313.
- Lyhagen J., 1997, Essays on univariate long-memory models, *Ph. D. Thesis*, Department of Statistics, Uppsala University, Sweden.
- Newey, W. and K. West, 1987, A simple positive definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55, 703-705.
- Ray, B. K., 1993, Modelling long-memory processes for optimal long-range predictions, *Journal of Time Series Analysis*, 14, 511-525.
- Smith, J. and S. Yadav, 1994, Forecasting costs incurred from unit differencing fractionally integrated processes, *International Journal of Forecasting*, 10, 507-514.
- Sowell, F., 1992, Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics*, 53, 165-188.
- Schwartz, G., 1978, Estimating the order of a model, *The Annals of Statistics*, 2, 461-464
- Tschernig, R., 1993, Tackling and understanding the small sample bias in ARFIMA models, *Unpublished manuscript*, Department of Finance, University of Limburg, Maastricht, The Netherlands.



# 3

## 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

### 3.1 Introduction

A time series process  $\mathbf{x}_t$  is said to be cointegrated of order  $CI(\delta, b)$  if the variables of  $\mathbf{x}_t$  individually are integrated of order  $I(\delta)$ , while a linear combination of the variables, denoted  $\beta' \mathbf{x}_t$ , is  $I(\delta - b)$ . The variables are in equilibrium if  $\beta' \mathbf{x}_t$  equals some constant  $\mu$ , 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  $\delta$  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  $\delta$  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  $\delta$  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.

### 3.2 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  $\eta_t$  is interpreted as the equilibrium error. Cheung and Lai (1993) show that the usual least squares estimator  $\hat{\varphi}$  converges in probability to  $\varphi$  also for fractionally integrated  $\eta_t$ . If the series are not cointegrated,  $\eta_t$  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*),

$$\Delta \mathbf{x}_t = \Pi \mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T, \quad (3.1)$$

where  $\Pi$  and  $\Gamma_i$  are  $p \times p$  parameter matrices,  $\boldsymbol{\mu}$  is an intercept and  $\boldsymbol{\varepsilon}_t$  an *iid*  $N(0, \Omega)$  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  $\beta$ . 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  $\alpha$

and  $\beta$  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

$$-2 \log (H(r) | H(p)) = -T \sum_{i=r+1}^p \log (1 - \hat{\lambda}_i), \quad (3.2)$$

where  $1 > \hat{\lambda}_1 > \dots > \hat{\lambda}_p > 0$  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).

### 3.3 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

$$\begin{aligned} x_{1t} + x_{2t} &= u_{1t} \\ x_{1t} + \alpha x_{2t} &= u_{2t}, \end{aligned} \quad (3.3)$$

where  $\alpha$  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 3.1 reports the power of the LR test against autoregressive alternatives at an exact<sup>1</sup> 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.

---

<sup>1</sup> Empirical null distributions of the tests are given in *Appendix*.



TABLE 3.1. Power, in percent, of the cointegration tests against autoregressive alternatives. The size of the tests equals five percent.

Test	$\phi$									
	.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 (3.1) 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  $\phi = 0.85$ .<sup>2</sup>

Results in Table 3.2 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-

<sup>2</sup>Results for  $T = 200$  and the 1% and 10% levels of significance are given in *Appendix*.

TABLE 3.2. Power, in percent, of the cointegration tests against fractional alternatives. The size of the tests equals five percent.

Test	$d$									
	.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 3.1.

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

### 3.4 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 (3.3). The bias and root mean squared error ( $RMSE$ ) are used to evaluate the maximum likelihood estimator of the  $\Pi$  matrix, i.e. the long-run relationship.

A general error correction representation for the system (3.3) is expressed as

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

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  $\Pi$  matrix is given by

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

where  $\phi$  is the AR parameter in the disturbance process  $u_{2t}$  and  $\alpha$  is equal to two.

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

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

Under both specifications the variance-covariance matrix  $\Omega$  is given by

$$\begin{aligned} E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] &= (\alpha - 1)^{-2} \begin{bmatrix} \alpha & -1 \\ -1 & 1 \end{bmatrix} E[\boldsymbol{\xi}_t \boldsymbol{\xi}_t'] \begin{bmatrix} \alpha & -1 \\ -1 & 1 \end{bmatrix}' \\ &= \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}, \end{aligned} \quad (3.4)$$

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 3.3, the bias of the estimated  $\Pi$  matrix is quite small for all specifications and decreases with the sample size. To illuminate further, the relative bias ( $\text{bias}/\pi_{ij}$ ) is almost identical for all estimates of the elements of  $\Pi$ , and it decreases rapidly with the distance between  $\phi$  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 3.3 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  $\phi$  decreases. However, if the RMSE is adjusted for the size of the true parameter, yielding the ratio  $\text{RMSE}/\pi_{ij}$ , we notice that this quantity decreases with  $\phi$ . Moreover, the  $\text{RMSE}/\pi_{ij}$  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 (3.4).

Table 3.4 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 3.3. Bias and RMSE for the maximum likelihood estimation procedure when the equilibrium error follows an AR(1) process. The lag-length  $k=1$ .

$T = 50$					
	$\phi$				
	.85	.65	.55	.45	.25
$\pi_{11}$					
True	.150	.350	.450	.550	.750
Bias	.041	.048	.045	.041	.032
RMSE	.253	.291	.305	.314	.327
$\pi_{12}$					
True	.300	.700	.900	1.10	1.50
Bias	.104	.103	.095	.085	.065
RMSE	.490	.579	.609	.627	.655
$\pi_{21}$					
True	-.150	-.350	-.450	-.550	-.750
Bias	-.051	-.051	-.048	-.043	-.033
RMSE	.167	.187	.193	.199	.206
$\pi_{22}$					
True	-.300	-.700	-.900	-1.10	-1.50
Bias	-.113	-.106	-.098	-.087	-.066
RMSE	.324	.371	.385	.396	.410
$T = 100$					
	$\phi$				
	.85	.65	.55	.45	.25
$\pi_{11}$					
True	.150	.350	.450	.550	.750
Bias	.027	.029	.024	.021	.017
RMSE	.149	.188	.200	.210	.223
$\pi_{12}$					
True	.300	.700	.900	1.10	1.50
Bias	.060	.060	.050	.043	.034
RMSE	.297	.376	.401	.420	.447
$\pi_{21}$					
True	-.150	-.350	-.450	-.550	-.750
Bias	-.030	-.029	-.025	-.022	-.017
RMSE	.098	.120	.127	.133	.141
$\pi_{22}$					
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 3.4. Bias and RMSE for the maximum likelihood estimation procedure when the equilibrium error follows a fractionally integrated process.

$T = 50$					
	$d$				
	.75	.55	.25	.15	.05
$\pi_{11}$					
True	1.00	1.00	1.00	1.00	1.00
Bias	-.832	-.643	-.243	-.126	-.027
RMSE	.879	.723	.417	.361	.335
$\pi_{12}$					
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
$\pi_{21}$					
True	-1.00	-1.00	-1.00	-1.00	-1.00
Bias	.809	.628	.240	.125	.027
RMSE	.832	.669	.326	.249	.211
$\pi_{22}$					
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
$T = 100$					
	$d$				
	.75	.55	.25	.15	.05
$\pi_{11}$					
True	1.00	1.00	1.00	1.00	1.00
Bias	-.881	-.703	-.274	-.145	-.038
RMSE	.897	.735	.359	.272	.233
$\pi_{12}$					
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
$\pi_{21}$					
True	-1.00	-1.00	-1.00	-1.00	-1.00
Bias	.868	.695	.273	.145	.038
RMSE	.876	.713	.313	.208	.150
$\pi_{22}$					
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 3.3.

$d = 0.75$  we notice a bias of 86–88% of the true parameter. Furthermore, the relative bias is constant over true  $\Pi$ , 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*). 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  $\text{RMSE}/\pi_{ij}$  ratios are similar for large  $d$ . For small  $d$ , the  $\text{RMSE}/\pi_{ij}$  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.

### 3.5 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.

## Acknowledgements

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, *Biometrika* 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: Additional Tables

TABLE 3.5. Empirical 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 (3.1) 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 3.6. Empirical 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 (3.5).

TABLE 3.7. Empirical 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 (3.5).

TABLE 3.8. Power, in percent, of the cointegration tests against autoregressive alternatives. Sample size 50.

Test	$\phi$									
	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05
<i>Size 1%</i>										
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
<i>Size 5%</i>										
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
<i>Size 10%</i>										
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 (3.1).

TABLE 3.9. Power, in percent, of the cointegration tests against autoregressive alternatives. Sample size 100.

Test	$\phi$									
	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05
<i>Size 1%</i>										
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
<i>Size 5%</i>										
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
<i>Size 10%</i>										
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 (3.1).

TABLE 3.10. Power, in percent, of the cointegration tests against autoregressive alternatives. Sample size 200.

Test	$\phi$									
	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05
<i>Size 1%</i>										
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
<i>Size 5%</i>										
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
<i>Size 10%</i>										
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 (3.1).

TABLE 3.11. Power, in percent, of the cointegration tests against fractional alternatives. Sample size 50.

Test	$d$									
	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05
<i>Size 1%</i>										
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
<i>Size 5%</i>										
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
<i>Size 10%</i>										
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 (3.1).

TABLE 3.12. Power, in percent, of the cointegration tests against fractional alternatives. Sample size 100.

Test	$d$									
	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05
<i>Size 1%</i>										
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
<i>Size 5%</i>										
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
<i>Size 10%</i>										
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 (3.1).

TABLE 3.13. Power, in percent, of the cointegration tests against fractional alternatives. Sample size 200.

Test	$d$									
	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05
<i>Size 1%</i>										
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
<i>Size 5%</i>										
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
<i>Size 10%</i>										
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 (3.1).

TABLE 3.14. Bias and RMSE for the maximum likelihood estimation procedure when the equilibrium error follows an AR(1) or a fractionally integrated process of length 200. The lag-length  $k=1$ .

		$\phi$				
		.85	.65	.55	.45	.25
$\pi_{11}$						
True		.150	.350	.450	.550	.750
Bias		.015	.014	.014	.011	.008
RMSE		.094	.127	.137	.145	.156
$\pi_{12}$						
True		.300	.700	.900	1.10	1.50
Bias		.032	.029	.027	.023	.015
RMSE		.189	.254	.274	.291	.311
$\pi_{21}$						
True		-.150	-.350	-.450	-.550	-.750
Bias		-.016	-.014	-.013	-.012	-.008
RMSE		.061	.081	.087	.092	.098
$\pi_{22}$						
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
$\pi_{11}$						
True		1.00	1.00	1.00	1.00	1.00
Bias		-.914	-.751	-.295	-.160	-.045
RMSE		.920	.765	.336	.227	.168
$\pi_{12}$						
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
$\pi_{21}$						
True		-1.00	-1.00	-1.00	-1.00	-1.00
Bias		.907	.747	.294	.159	.045
RMSE		.910	.765	.314	.191	.112
$\pi_{22}$						
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 3.3.



# 4

## 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

### 4.1 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.

## 4.2 Testing for Fractional Integration

A time series  $\{x_t\}$  follows an ARFIMA( $p, d, q$ )<sup>1</sup> process if

$$\phi(B)(1-B)^d x_t = \theta(B) a_t, \quad (4.1)$$

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.<sup>2</sup> 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.

### 4.2.1 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:

$$\ln \{I_x(\omega_j)\} = \alpha - d \ln \{4 \sin^2(\omega_j/2)\} + v_j, \quad (4.2)$$

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  $j = 1, \dots, g(T)$ . With a proper choice of  $g(T)$ , the ordinary least squares (*OLS*) estimator of  $d$  is consistent and the distribution of

---

<sup>1</sup> The properties of the fractionally integrated ARMA model are presented by Granger and Joyeux (1980) and Hosking (1981).

<sup>2</sup> Persistence is commonly found in Economic time series, i.e. real exchange rates and unemployment.

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

#### 4.2.2 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

$$\tilde{Q}_T = \frac{R_T}{\hat{\sigma}_T(k)}, \quad (4.3)$$

where the range and standard error are calculated by

$$R_T = \max_{0 < i \leq T} \sum_{t=1}^i (x_t - \bar{x}) - \min_{0 < i \leq T} \sum_{t=1}^i (x_t - \bar{x}) \quad (4.4)$$

$$\hat{\sigma}_T^2(k) = \hat{\sigma}^2 + 2 \sum_{j=1}^k \sum_{i=j+1}^T \left(1 - \frac{j}{k+1}\right) (x_i - \bar{x})(x_{i-j} - \bar{x}) \quad (4.5)$$

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  $(3T/2)^{\frac{1}{3}} \{2\hat{\rho}/(1 - \hat{\rho}^2)\}^{\frac{2}{3}}$ , 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).

#### 4.2.3 A Lagrange Multiplier Test

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

$$\hat{a}_t = \sum_{i=1}^p \beta_i W_{t-i} + \sum_{j=1}^q \gamma_j Z_{t-j} + \delta K_m + u_t, \quad (4.6)$$

where

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

$\hat{a}_t$  and  $\hat{\theta}(B)$  are the estimated residual and MA polynomial from the ARFIMA specification (4.1) 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 (4.6) is fitted by non-linear least squares (the *IMSL* routine DNSLSE) over the time period  $t = m + 1, \dots, T$ . The usual  $t$ -test of the hypothesis  $\delta = 0$  together with asymptotically normal critical values constitutes the LM test.

#### 4.2.4 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.

### 4.3 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,

$$p(\hat{\tau}) = p(\tau \geq \hat{\tau} | \Psi_0, T), \quad (4.7)$$

where  $\Psi_0$  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  $\tau$  based on the original sample  $\mathbf{x} = [x_1, \dots, x_T]'$ . The DGP  $\Psi_0$  is characterized by an unknown ARMA( $p, q$ ) specification. Since the null model, and hence  $\Psi_0$ , 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<sup>3</sup>, 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,

$$(1 - \hat{\phi}_1 B - \dots - \hat{\phi}_{\hat{p}} B^{\hat{p}}) x_t = (1 + \hat{\theta}_1 B + \dots + \hat{\theta}_{\hat{q}} B^{\hat{q}}) \hat{a}_t, \quad (4.8)$$

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

$$(1 - \tilde{\phi} B - \dots - \tilde{\phi}_{\tilde{p}} B^{\tilde{p}}) x_t = \tilde{a}_t, \quad (4.9)$$

---

<sup>3</sup>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 (4.8) and (4.9) 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.<sup>4</sup>

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

$$x_{r,t}^* = \phi^*(B)^{-1} \theta^*(B) a_t^*, \quad (4.10)$$

where  $\phi^*(B)$  and  $\theta^*(B)$  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  $s_a^2$  or  $s_a^2$ .

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

$$p^*(\hat{\tau}) = R^{-1} \sum_{r=1}^R I(|\tau_r^*| \geq |\hat{\tau}|), \quad (4.11)$$

where  $I(\cdot)$  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  $p^*(\hat{\tau})$ .

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.

## 4.4 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

---

<sup>4</sup>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 (4.8) and AR (4.9) 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.

#### 4.4.1 AR and MA Processes

The empirical sizes of the tests are examined for the specifications

$$x_t = \phi x_{t-1} + a_t \quad (4.12)$$

and

$$x_t = a_t + \theta a_{t-1}, \quad (4.13)$$

where the members of  $\{a_t\}$  are *iid*  $N(0,1)$ . The AR and MA parameters  $\phi$  and  $\theta$  are set equal to  $\pm 0.1$ ,  $\pm 0.5$  and  $\pm 0.9$ . Table 4.1<sup>5</sup> 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

---

<sup>5</sup> All results are approximately valid for the 1% and 10% nominal significance level and for  $T=300$  and  $500$ .



TABLE 4.1. Rejection percentage of the nominal 5 percent fractional integration test when the data follow an AR(1) or MA(1) process of length T.

$\phi/\theta$	MRR		GPH		REG	
	Orig.	Boot.	Orig.	Boot.	Orig.	Boot.
<hr/>						
T=50	<i>AR(1) Processes</i>					
-0.9	<b>18.9</b>	<b>2.2</b>	4.7	5.4	<b>6.9</b>	4.7
-0.5	<b>0.9</b>	<b>3.1</b>	4.7	5.3	6.2	4.1
-0.1	<b>6.7</b>	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	<b>2.5</b>	<b>3.1</b>	<b>8.2</b>	5.1	<b>6.8</b>	4.7
0.9	<b>1.2</b>	4.1	<b>63.6</b>	3.6	<b>7.8</b>	4.8
T=100	<hr/>					
-0.9	6.4	<b>2.1</b>	6.4	5.6	5.6	4.8
-0.5	<b>0.8</b>	5.0	5.7	5.0	5.1	4.6
-0.1	<b>6.6</b>	4.8	4.9	4.5	6.2	3.6
0.1	<b>6.8</b>	5.9	4.9	5.2	<b>6.9</b>	5.1
0.5	<b>2.3</b>	4.8	<b>8.3</b>	4.7	6.4	4.7
0.9	<b>0.8</b>	3.9	<b>71.8</b>	3.7	5.1	4.7
T=50	<hr/> <i>MA(1) Processes</i>					
-0.9	4.8	5.4	<b>41.2</b>	<b>25.8</b>	<b>9.3</b>	<b>9.5</b>
-0.5	5.5	5.0	<b>7.7</b>	5.0	<b>3.3</b>	<b>1.6</b>
-0.1	<b>7.3</b>	4.5	5.0	4.5	<b>6.6</b>	5.4
0.1	<b>6.7</b>	4.9	4.3	4.0	<b>7.4</b>	5.3
0.5	3.7	4.3	4.4	4.8	4.0	4.4
0.9	<b>2.0</b>	<b>3.2</b>	5.2	3.7	<b>10.0</b>	6.3
T=100	<hr/>					
-0.9	<b>9.9</b>	<b>10.5</b>	<b>50.1</b>	<b>36.3</b>	<b>8.0</b>	<b>7.2</b>
-0.5	4.2	4.8	<b>7.9</b>	5.6	<b>2.8</b>	<b>2.8</b>
-0.1	<b>6.6</b>	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	<b>3.2</b>	5.7	5.4	5.7	<b>2.7</b>	4.0
0.9	<b>2.3</b>	4.8	6.0	4.4	<b>6.8</b>	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<sup>6</sup>. Agiakloglou *et al.* (1993) show that large positive AR and MA roots bias the periodogram (4.2), 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

---

<sup>6</sup>Positive values of  $\phi$  imply positive roots and positive values of  $\theta$  negative roots.

TABLE 4.2. Rejection percentage of the nominal 5 percent fractional integration test when the data follow an ARFIMA(0,d,0) process of length  $T=100$ .

$d$	MRR		GPH		REG	
	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 4.1. 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.

#### 4.4.2 Fractional Processes

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

$$(1 - B)^d x_t = a_t, \quad (4.14)$$

where the fractional differencing parameter,  $d$ , is set equal to  $\pm 0.05$ ,  $\pm 0.25$  and  $\pm 0.45$ . Table 4.2 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 (4.6) 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 (4.3) increases with  $d$ , i.e. too many autocorrelations are included in the variance correction term (4.5), 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 4.2.

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  $d > 0$ .

## 4.5 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.

## Acknowledgements

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. WOHR (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. *Biometrika* 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

# 5

## 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 seriously 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 that a procedure based on the REG test is nearly exact.

**Key words:** Long-memory; Resampling; Size correction; Skewness and kurtosis; ARCH; Monte Carlo.

**JEL-classification:** C12; C15; C22; C52

### 5.1 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.

## 5.2 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.

### 5.2.1 The Bootstrap Test

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

$$p(\hat{\tau}) = p(|\tau| \geq |\hat{\tau}| | \Psi_0, T) \quad (5.1)$$

where  $\Psi_0$  is the data generating process (*DGP*) under the null hypothesis, and  $\hat{\tau}$  is the realized value of a test statistic  $\tau$  based on a sample of length  $T$ . Since  $\Psi_0$  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  $\Psi_0$  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 (1996a), 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  $\tau_r^*$  are calculated using the same test statistic  $\tau$  as above, the estimated bootstrap  $p$ -value function is defined by the quantity

$$p^*(\hat{\tau}) = R^{-1} \sum_{r=1}^R I(|\tau_r^*| \geq |\hat{\tau}|), \quad (5.2)$$

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).<sup>1</sup>

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

$$\phi(B)(1-B)^d x_t = \theta(B) a_t, \quad t = 1, \dots, T \quad (5.3)$$

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 (5.3) 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.<sup>2</sup>

When testing for fractional integration, the DGP  $\Psi_0$  is characterized by an unknown ARMA( $p, q$ ) specification. Since the null model, and consequently  $\Psi_0$ , is unknown, the estimated (bootstrap) DGP  $\hat{\Psi}_0$  is used to create the bootstrap samples.

### 5.2.2 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

<sup>1</sup> The tests are briefly described in *Appendix A*.

<sup>2</sup> 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 (1996a, 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,

$$(1 - \hat{\phi}_0 - \hat{\phi}_1 B - \dots - \hat{\phi}_{\hat{p}} B^{\hat{p}}) x_t = \hat{a}_t, \quad (5.4)$$

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  $(0, 1, \dots, 5)$  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  $\mathbf{x}_r^*$ ,  $r = 1, \dots, R$ , are created recursively by the equation

$$x_{r,t}^* = \hat{\phi}(B)^{-1} a_t^*, \quad (5.5)$$

where  $\hat{\phi}(B)$  is the polynomial of (5.4) 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 (5.3), 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  $s_a^2$ .

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 cor-

rected 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  $U(\hat{p} + 1, T)$  distributed.

The *third* and *fourth* resampling algorithms are constructed to preserve ARCH(1) dependence in the residuals. ARCH is introduced to the autoregression,  $\phi(B)x_t = a_t$ , 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  $\varepsilon_t$  allows joint estimation of the parameters through maximization of the log-likelihood function

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

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  $N(0, s_{\tilde{\varepsilon}}^2)$  distribution. For the non-parametric ( $b_4$ ) scheme the members of  $\{\tilde{\varepsilon}_t\}$  are drawn from the degrees of freedom adjusted elements of  $\{\hat{\varepsilon}_t\}$ . 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  $\hat{\beta}_0 / (1 - \hat{\beta}_1)$ .

### 5.3 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. 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 1996b). 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,

$$(1 - \phi B) x_t = a_t, \quad (5.6)$$

of length  $T = 100$  are generated and the parameter  $\phi$  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  $a_t$ .

### 5.3.1 Normal Processes

The experiment examining the empirical size of the tests under normality is based on the process (5.6) where the disturbances  $\{a_t\}$  are *iid* normally distributed with mean zero and variance equal to unity. Table 5.1 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 the normality) produces notably large sizes for strongly dependent

TABLE 5.1. Rejection percentage of the nominal 5 percent fractional integration test when the data follow an AR(1), of length 100, with normal errors.

Test		$\phi$				
		0.0	0.1	0.5	0.7	0.9
MRR	<i>o</i>	<b>7.6</b>	<b>6.8</b>	<b>2.3</b>	<b>1.3</b>	<b>0.8</b>
	<i>b</i> <sub>1</sub>	5.3	5.9	4.8	4.7	3.9
	<i>b</i> <sub>2</sub>	4.5	4.2	4.9	5.0	4.4
	<i>b</i> <sub>3</sub>	6.2	6.4	6.0	5.2	3.9
	<i>b</i> <sub>4</sub>	6.0	6.4	5.5	5.3	4.3
GPH	<i>o</i>	4.9	4.9	<b>8.3</b>	<b>17.9</b>	<b>71.8</b>
	<i>b</i> <sub>1</sub>	5.0	5.2	4.7	4.3	3.7
	<i>b</i> <sub>2</sub>	5.0	5.1	4.6	4.2	4.0
	<i>b</i> <sub>3</sub>	5.3	5.8	5.0	4.4	3.6
	<i>b</i> <sub>4</sub>	5.4	5.5	5.0	4.4	3.6
REG	<i>o</i>	5.9	<b>6.9</b>	6.4	5.1	5.1
	<i>b</i> <sub>1</sub>	6.0	5.1	4.7	4.2	4.7
	<i>b</i> <sub>2</sub>	4.5	3.7	5.6	<b>7.9</b>	<b>19.2</b>
	<i>b</i> <sub>3</sub>	5.0	4.6	3.7	<b>3.3</b>	<b>3.4</b>
	<i>b</i> <sub>4</sub>	4.8	4.8	3.6	<b>3.5</b>	<b>3.5</b>

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*<sub>1</sub> – *b*<sub>4</sub> the bootstrap testing procedures.

AR processes. The resamplings that account for the (non-existing) ARCH effects have reasonable estimated sizes, however conservative for  $\phi$  equal to 0.7 and 0.9.

### 5.3.2 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  $\gamma_k$  respectively. The members of  $\{a_t\}$  are generated by the transformation,

$$a_t = c_0 + c_1\alpha_t + c_2\alpha_t^2 + c_3\alpha_t^3 \quad \alpha_t \sim N(0, 1), \quad (5.7)$$

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.  $\gamma_s$  and  $\gamma_k$  are chosen to gen-

TABLE 5.2. 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		$\phi$				
		0.0	0.1	0.5	0.7	0.9
MRR	$o$	<b>7.7</b>	6.0	<b>2.2</b>	<b>2.0</b>	<b>0.5</b>
	$b_1$	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	<b>8.0</b>	<b>16.3</b>	<b>72.2</b>
	$b_1$	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	<b>7.3</b>	<b>7.7</b>	<b>6.6</b>	<b>8.7</b>
	$b_1$	4.9	5.0	3.8	4.7	<b>2.4</b>
	$b_2$	5.1	3.9	4.5	<b>6.9</b>	<b>13.8</b>
	$b_3$	3.9	5.8	4.5	<b>3.1</b>	<b>1.5</b>
	$b_4$	3.8	4.7	4.2	<b>1.8</b>	<b>1.1</b>

See note to Table 5.1. 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.<sup>3</sup> The empirical size of the tests under non-normality is reported in Table 5.2.

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.

<sup>3</sup>The expressions for the determination of  $c_i, i = 0, \dots, 3$  and how the residual skewness and kurtosis depend on those of the time series process are given in *Appendix B*.

### 5.3.3 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  $\beta$  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 5.3 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  $\beta$  and  $\phi$ .

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  $\beta$ . 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  $b_2$ .

### 5.3.4 Size Comparisons and Power

Table 5.4 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 5.3. Rejection percentage of the nominal 5 percent fractional integration test when the data follow an AR(1), of length 100, with ARCH errors.

Test		$\phi$				
		0.0	0.1	0.5	0.7	0.9
$\beta = 0.5$						
MRR	$o$	5.5	4.0	<b>2.4</b>	<b>1.8</b>	<b>0.6</b>
	$b_1$	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	<b>17.6</b>	<b>71.6</b>
	$b_1$	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	$o$	<b>9.8</b>	<b>9.7</b>	<b>9.0</b>	<b>7.4</b>	<b>10.4</b>
	$b_1$	<b>7.6</b>	<b>8.5</b>	5.4	4.6	<b>7.3</b>
	$b_2$	<b>8.7</b>	4.6	6.6	4.6	<b>25.2</b>
	$b_3$	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	$o$	<b>3.4</b>	<b>3.1</b>	<b>1.4</b>	<b>0.8</b>	<b>0.7</b>
	$b_1$	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	<b>6.5</b>	<b>16.8</b>	<b>70.9</b>
	$b_1$	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	$o$	<b>28.6</b>	<b>29.3</b>	<b>29.5</b>	<b>29.2</b>	<b>33.4</b>
	$b_1$	<b>7.9</b>	<b>8.4</b>	<b>7.3</b>	<b>7.1</b>	<b>7.3</b>
	$b_2$	<b>9.2</b>	<b>6.8</b>	<b>8.8</b>	<b>14.4</b>	<b>24.9</b>
	$b_3$	<b>3.1</b>	3.8	<b>3.5</b>	3.8	4.3
	$b_4$	3.6	3.8	<b>3.3</b>	4.1	4.4

See note to Table 5.1. The error processes follow an ARCH process with parameter  $\beta$ .

TABLE 5.4. Best test based on size properties.

	<i>Normal</i>	<i>Non - norm.</i>	<i>ARCH</i>	<i>Overall</i>
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	$b_1, (0, b_3, b_4)$	$b_1, (b_3, b_4)$	$b_4 (b_3)$	$b_3, b_4$

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 (1996b) show that the power of a bootstrap test, based on a pivotal statistic, is generally close to the size-adjusted asymptotic test. Table 5.5 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<sup>4</sup> 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  $b_2$ .

<sup>4</sup>The skewness and kurtosis of the residuals as functions of the time series moments are given in Appendix B.

TABLE 5.5. 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
<i>Normal processes</i>								
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
<i>Non-normal processes</i>								
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
<i>ARCH processes</i>								
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 parametric bootstrap testing procedures.

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.

## 5.4 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.

## Acknowledgements

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. Of course, all errors are of our own making.

## References

- AGIAKLOGLOU, C., P. NEWBOLD and M. WOHR (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 (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*.
- 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

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

where  $I_x(\omega_j)$  is the periodogram at the harmonic frequencies  $\omega_j = 2\pi j/T$ , and  $j = 1, \dots, g(T) = T^{1/2}$ . The ordinary least squares (*OLS*) estimator of  $d$  is then consistent and the distribution of  $(\hat{d}_{OLS} - d)/SE(\hat{d}_{OLS})$  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

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

where the range and the standard error are calculated by

$$R_T = \max_{0 < i \leq T} \sum_{t=1}^i (x_t - \bar{x}) - \min_{0 < i \leq T} \sum_{t=1}^i (x_t - \bar{x})$$

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

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

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

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

$$\hat{a}_t = \sum_{i=1}^p \beta_i W_{t-i} + \sum_{j=1}^q \gamma_j Z_{t-j} + \delta K_m + u_t, \quad (5.8)$$

where

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

$\hat{a}_t$  and  $\hat{\theta}(B)$  are the estimated residuals and MA polynomial under the null and  $m$  is a prespecified truncation lag. The equation (5.8) is fitted by OLS over the time period  $t = m+1, \dots, T$  and the usual  $t$ -statistic for the null hypothesis  $\delta = 0$  follows an asymptotic  $N(0,1)$  distribution.

## B Generation of Non-normal Data

### B.1 The Skewness and Kurtosis Relationships

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  $\Gamma_s$  and  $\Gamma_k$  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  $\delta_i$  is the  $i$ th 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

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

### B.2 The Fleichmann Algorithm

The constants in (5.7) are given as the solutions to the following system of equations,

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

and

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



# 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 företags produkter, relationer och resurser.

**Andersson, H.**, En produkthistoria.

*(separat publicerad bilaga till ovan-stående)*

**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 stor-företag 1993. Delrapport 1. Faktiska tider och påverkande faktorer för månatlig resultat-rapportering. Research Report.

**Nilsson, K.**, Rapportering i svenska stor-företag 1993. Delrapport 2. Rapporteringstider för kvartals/tertiärsrapporter. Research Report.

**Nordenlöv, 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 Distributions-forskning, EFI.

### 1995

**Becker, T.**, Essays on Stochastic Fiscal Policy, Public Debt and Private Consumption.

**Blomberg, J.**, Ordning och kaos i projekt-samarbete - 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 Distributions-forskning, 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 Satisfaction 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 vinstandelsbevis - 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. Männi-skor och informationssystem i sam-verkan.** 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.**, Ledarutveckling för kvinnor - Uppföljning av en satsning 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 hemmamarknadsförvar? De nordiska marknaderna för fruktyghurt 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 organisation.

**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 planering 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 ändrade 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 näringslivet 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 Process of Adopting Lean Production.

**Åkesson, G.**, Företagsledning i strategiskt vakuum. Om aktörer och förändringsprocesser.

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

