STATIONARY AND NONSTATIONARY FARIMA MODELS - MODEL CHOICE, FORECASTING, AGGREGATION AND INTERVENTION

By
Dirk Ocker

DISSERTATION
SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF SOCIAL SCIENCES
AT

UNIVERSITY OF KONSTANZ
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
78457 KONSTANZ, GERMANY

OCTOBER 1999

examining committee:
Prof. Dr. Jan Beran, examiner
Prof. Dr. Siegfried Heiler, co-examiner

To my family

Acknowledgements

I am greatly indebted to my advisor, Prof. Dr. Jan Beran, for having suggested this topic and for his guidance throughout the course of this research, in particular, for his patience and readiness in replying to my endless questions. This dissertation could not have been written without his inspiration as a great teacher and scientist.

Then, I want to express my gratitude to Prof. Dr. Siegfried Heiler for teaching me econometrics and introducing me into the exciting field of time series analysis while I was a graduate student at the University of Konstanz. I also have to thank him and Prof. Dr. Günther Franke for having accepted being part of my jury.

I am very grateful to Dr. Henry-Jean-Paul Hammelbeck for his friendship and computer support, and to Prof. Dr. Miguel Brendl (INSEAD, Fontainebleau) for having admitted me in his lab for one year at the Department of Social Psychology, University of Konstanz.

I would like to thank Dr. Elke M. Hennig (Citibank, Frankfurt) for the stock market series, and Prof. Dr. Werner Antweiler for maintaining an excellent Web-site with exchange rate data at the University of British Columbia, Vancouver.

Finally, I wish to thank my colleagues Dr. Klaus Abberger, Dr. Yuanhua Feng and Dr. Roland Jeske for their technical and moral support during my time at the Department of Economics and Statistics, University of Konstanz.

Dirk Ocker
October, 1999

Contents

Zusammenfassung

Die vorliegende Dissertation befasst sich mit der Modellwahl, Vorhersage, temporalen Aggregation und Interventionsanalyse stationärer und nichtstationärer fraktioneller autoregressiver Prozesse, sowie einer extensiven Anwendung auf weltweite Finanzmarktdaten.

Stationäre fraktionelle autoregressive Modelle wurden zuerst von Granger und Joyeux [48], sowie Hosking [52] eingeführt. Sie dienen vor allem zur stochastischen Modellierung stationärer Zeitreihen mit langfristigen Abhängigkeiten (oder langem Gedächtnis, bzw. Persistenz), die aber nicht so stark sind, dass eine einfache Differenzenbildung im Rahmen traditioneller Box-Jenkins Modelle adäquat wäre.

Unglücklicherweise ist die stochastische Theorie dieser Modelle typischerweise auf den stationären Bereich des Differenzenparameters d beschränkt, d.h. $d \in (-0,5; 0,5)$. In einem aktuellen Artikel zeigte Beran [10] jedoch, dass jedes reellwertige $d > -0,5$ durch einen approximativen Maximum-Likelihood-Schätzer bestimmt werden kann (mit Ausnahme der Werte $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$). Insbesondere kann dadurch die mit der Schätzung des Differenzenparameters $d > -0,5$ verbundene Unsicherheit in den Konfidenzintervallen der autoregressiven Parameter berücksichtigt werden. Beran [10] zeigte dies aber nur für den Fall, dass die autoregressive Ordnung a priori bekannt sei. Eine entsprechende Verallgemeinerung findet sich jedoch in Beran, Bhansali und Ocker [12]. Wir entwickelten eine Version des Akaike-Informationskriteriums (AIC) zur Bestimmung der autoregressiven Ordnung, wenn sowohl d als auch die autoregressiven Parameter simultan geschätzt werden. Die Resultate in Beran und Ocker [17] über die Vorhersage (möglicherweise nichtstationärer) fraktioneller autoregressiver Prozesse rundeten schliesslich diesen vereinheitlichten Ansatz zur simultanen Modellierung und Prädiktion stationärer und nichtstationärer Prozesse mit kurzfristigen und langfristigen Abhängigkeiten ab. Wir zeigten insbesondere, dass die Rate mit der Vorhersageintervalle gegen ihre asymptotische Länge konvergieren (im Falle stationärer Prozesse), bzw. gegen Unendlich divergieren (im Falle nichtstationäre Prozesse), von der Grösse des Differenzenparameters $d > -0,5$ abhängig ist.

Nach wie vor gibt es aber noch eine Reihe ungelöster Probleme und Fragen. Zum Beispiel, welches Modellwahlkriterium (AIC, HIC oder BIC) im Vergleich zu besseren Vorhersageergebnissen führen könnte. Ein anderes ungelöstes Problem ist die Frage, wie temporale Aggregation die Modellstruktur (möglicherweise nichtstationärer) fraktioneller Prozesse asymptotisch beeinflusst und verändert. Ausserdem könnte die Einbeziehung von Interventionen, die typischerweise mit Abweichungen von der Normalverteilung verbunden sind, die Zuverlässigkeit der hier diskutierten Modellklasse in der praktischen Anwendung verbessern. Antworten auf diese Fragen und reale Datenbeispiele finden sich in dieser Doktorarbeit.

Kapitel 1 dient zur formalen Einführung der Themen dieser Dissertation.

Kapitel 2 umfasst einen Überblick der Resultate von Beran [10] sowie Beran, Bhansali und Ocker [12] über Modellanpassung und Modellwahl, und enthält eine neue Simulationsstudie mit kleinen (realistischen) Stichprobengrössen. Es zeigt sich, dass das AIC von der selben allgemeinen Form ist, wie im Falle stationärer autoregressiver Prozesse. Ausserdem können die korrespondierenden Versionen des BICs (Schwarz [74]) und des HICs (Hannan und Quinn [51]) zur (approximativen) konsistenten Schätzung der autoregressiven Ordnung verwendet werden.

Kapitel 3 stellt zusammenfassend die parametrischen Resultate aus Beran und Ocker [17] über die Vorhersage stationärer und nichtstationärer autoregressiver Prozesse mit kurz- und langfristigen Abhängigkeiten dar. Die Ergebnisse einer Simulationsstudie weisen darauf hin, dass Vorhersagen besser die Daten adaptieren, wenn das BIC verwendet wird. Es stellt sich ausserdem heraus, dass Random-Walk Intervalle entweder zu optimistisch (d.h. zu kurz), im Falle langfristiger Abhängigkeiten, oder unnötig lang sind, wenn die Zeitreihe antipersistent ist. Darüberhinaus zeigt sich, dass (gegeben eine hinreichend grosse Stichprobe, beispielsweise $n = 200$) substantielle Verbesserungen in der Punktvorhersagegenauigkeit erreichbar sind für persistente Prozesse.

Das sehr gute Abschneiden und die praktische Relevanz der Resultate über Modellanpassung, Modellwahl und Vorhersage wird in Kapitel 4 anhand verschiedener nominaler Aktienindizes und Wechselkurse demonstriert. Wir finden signifikante langfristige Abhängigkeiten in den Aktienindizes, sowie (temporär) evidente Antipersistenz in den Wechselkursen. Wir beobachten ausserdem überlegene Punkt- (für einige Zeitreihen) und zuverlässigere Intervall-Vorhersagen im Vergleich zur Random-Walk und traditionellen Box-Jenkins ARIMA Prädiktion.

Kapitel 5 befasst sich mit temporaler Aggregation stationärer und nichtstationärer autoregressiver Prozesse mit fraktionellen Abhängigkeiten. Wir zeigen, dass die (zum Teil bekannten) asymptotischen Resultate über temporale Aggregation fraktioneller und nichtfraktioneller, stationärer und nichtstationärer autoregressiver Prozesse unter Verwendung eines einheitlichen Ansatzes hergeleitet werden können. Das Hauptresultat lautet: Kurzfristige Abhängigkeiten sind im Gegensatz zu fraktionellen nicht robust gegenüber temporaler Aggregation. Interessanterweise konvergieren nichtstationäre fraktionelle Prozesse asymptotisch jedoch nicht gegen einen fraktionellen Brown Prozess.

Kapitel 6 präsentiert einen neuen Ansatz zur Modellierung und Schätzung externer Interventionen innerhalb der fraktionellen autoregressiven Modellklasse. Wir leiten notwendige und hinreichende Bedingungen zur konsistenten Schätzung von Interventionsparametern, sowie deren korrespondierende asymptotische Verteilung her. Es zeigt sich insbesondere, dass eine konsistente Schätzung möglich ist, wenn die Intervention langsam hyperbolisch gegen eine Konstante gleich oder ungleich Null konvergiert, oder aber auf einem von Null verschiedenen Niveau verharrt.

Summary

The dissertation deals with model choice, forecasting, temporal aggregation and intervention analysis of stationary and nonstationary fractional autoregressive processes, and provides an extensive application to worldwide financial time series.

The stationary fractional autoregressive model was first introduced by Granger and Joyeux [48], and Hosking [52] for modeling stationary time series with long-range dependence (or long memory, or persistence) and for avoiding the problem of overdifferencing which is often encountered in the usual Box-Jenkins setting. Unfortunately, the stochastic theory known so far has been restricted to the stationary range of the fractional differencing parameter d, i.e. $d \in (-.5, .5)$. Recently, Beran [10] has shown that any real value of $d > -.5$ (except $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, ...$) can be estimated by an approximate maximum likelihood estimator. In particular, the resulting confidence intervals for the autoregressive parameters take into account the additional uncertainty due to the estimation of d. Beran's [10] results are, however, obtained under the assumption that the autoregressive order is known a priori. This problem was solved by Beran, Bhansali and Ocker [12]. We derived a version of the Akaike information criterion, AIC, for determining an appropriate autoregressive order when d and the autoregressive parameters are estimated simultaneously by Beran's [10] maximum likelihood procedure. The findings of Beran and Ocker [17] on forecasting for (possibly nonstationary) fractional autoregressive processes rounded off this unified framework for simultaneous modeling of stationary and nonstationary, fractional and nonfractional autoregressive processes. We have shown that the rate at which forecast intervals converge to the asymptotic length (for stationary processes) or diverge to infinity (for difference stationary processes) depends on the fractional differencing parameter $d > -.5$.

However, a number of open problems remain. For instance, which model choice criterion (AIC, HIC or BIC) may lead, in comparison, to improved forecasts. Another open problem is, how temporal aggregation influences and changes the model structure of (possibly nonstationary) fractional processes asymptotically. Also, incorporating possible interventions, which typically cause deviations from normality, may improve the reliability of the model class considered here in daily applications. Answers to these questions and real data examples are given in this thesis.

Chapter 1 serves as an introduction to the topics of this dissertation.

Chapter 2 summarizes the findings of Beran, Bhansali and Ocker [12] on model selection and provides an additional simulation study for small (realistic) sample sizes. It is shown that the AIC is of the same general form as for stationary autoregressive processes. The corresponding versions of the BIC of Schwarz [74] and the HIC of Hannan and Quinn [51] are shown to yield (approximately) consistent estimators of the autoregressive order.

Chapter 3 reviews the parametric results of Beran and Ocker [17] on forecasting stationary and nonstationary autoregressive processes with short- and long-range dependence . A simulation study indicates that forecasts are better adapted to the data when using the BIC. Also, it turns out that random walk intervals are, in comparison, either too optimistic (e.g. too short) in the presence of long memory or unnecessarily wide if the series is antipersistent. Moreover, substantial improvements of point-forecasts are possible (given a sufficiently large sample size, say $n = 200$) if the degree of persistence is moderate or strong.

The good performance of our methods for estimation, model selection and forecasting is demonstrated in chapter 4 by several nominal stock market indices and exchange rates. We found significant long memory in international stock market indices, and (at least partially) evident antipersistence in foreign exchange rates. Superior point-forecasts (for some series) and more reliable interval-forecasts are achieved in comparison to random walk and traditional Box-Jenkins ARIMA predictions.

Chapter 5 deals with temporal aggregation of stationary and nonstationary fractional autoregressive processes. We show that the (partially known) asymptotic results on temporal aggregation of fractional and nonfractional, stationary and nonstationary autoregressive processes can be derived using one unified treatment. Our main result is that short-memory components vanish under temporal aggregation whereas fractional dependence remains.

Interestingly, nonstationary fractional processes do not converge asymptotically to fractional Brownian motion.

Chapter 6 presents a new approach on modeling and estimating external interventions within the fractional autoregressive framework. We derive necessary and sufficient conditions for estimating the parameters of an intervention consistently, and provide the corresponding asymptotic distribution. It is shown, in particular, that consistency is possible if an intervention decays slowly hyperbolically to zero, a non-zero constant or remains at a non-zero level.

Chapter 1

Introduction

Consider the case that a Gaussian time series X_t $(t = 1, ..., n)$ may be represented by a unified treatment of a nonfractional autoregressive integrated process (ARIMA$(p, m, 0)$) and a stationary fractional ARIMA$(p, \delta, 0)$ model (FARIMA$(p, \delta, 0)$). Beran [10] has shown that the two approaches are special cases of

$$\phi(B)(1-B)^{\delta}\{(1-B)^m X_t - \mu\} = \epsilon_t, \tag{1.1}$$

where ϵ_t are iid zero mean normal with $\sigma_{\epsilon}^2 = Var(\epsilon_t)$, B denotes the backshift operator such that $BX_t = X_{t-1}$. Also, $\phi(x) = \sum_{j=0}^{p} \phi_j x^j$ is a polynomial with $\phi_0 = 1$ and roots outside the unit circle. The integer m is the number of times X_t must be differenced to achieve stationarity. The mth difference $(1-B)^m X_t$ is a stationary FARIMA$(p, \delta, 0)$ process with fractional differencing parameter $\delta \in (-.5, .5)$ and expected value μ. The fractional difference $(1-B)^{\delta}$ is defined by

$$(1-B)^{\delta} = \sum_{k=0}^{\infty} b_k(\delta) B^k, \quad \text{with} \tag{1.2}$$

$$b_k(\delta) = (-1)^k \frac{\Gamma(\delta+1)}{\Gamma(k+1)\Gamma(\delta-k+1)}. \tag{1.3}$$

For $\delta = 0$, (1.1) reduces to an ARIMA$(p, m, 0)$ model (Box and Jenkins [22]). For $m = 0$, (1.1) reduces to a stationary FARIMA$(p, \delta, 0)$ process (Granger and Joyeux [48], Hosking [52]). Moreover, definition (1.1) extends the concept of standard ARIMA$(p,m,0)$ models by allowing the mth difference to be a stationary FARIMA$(p,\delta,0)$ process with arbitrary $\delta \in (-.5,.5)$. Also, (1.1) is an extension of the FARIMA$(p,\delta,0)$ models by allowing the possibility of nonstationarity ($d > .5$). The differencing parameter $d = m+\delta$ determines which, possibly fractional, difference has to be taken in order to obtain a stationary Box-Jenkins ARMA$(p,0)$ process. We refer to (1.1) as FARIMA$(p,d,0)$ model.

For $\delta \in (0,.5)$, the differenced process $Y_t = (1-B)^m X_t$ exhibits long-range dependence. Here, a stationary process Y_t with covariances $\gamma(k) = Cov(Y_t, Y_{t+k})$ is said to have long-range dependence (or long memory) if the spectral density of Y_t, $f(\lambda) = (2\pi)^{-1}\sum_{k=-\infty}^{\infty}\exp(ik\lambda)\gamma(k)$, has a pole at the origin

$$f(\lambda) \sim c_f|\lambda|^{-2\delta} \quad (|\lambda| \to 0) \tag{1.4}$$

for a constant $c_f > 0$ and $\delta \in (0,.5)$, where '$\sim$' means that the ratio of the left and right hand side converges to one (see e.g Mandelbrot [65], Beran [9] and references therein). In particular, this implies that, as $k \to \infty$, the covariances $\gamma(k)$ are proportional to $k^{2\delta-1}$ and hence they are not summable. On the other hand, a stationary process is called antipersistent, if (1.4) holds with $\delta \in (-.5,0)$. This implies that the sum of all covariances is zero. Note that for usual shot-memory processes, such as stationary ARMA processes, (1.4) holds with $\delta = 0$, and the covariances sum up to a nonzero finite value. Thus, (1.1) incorporates stationary and nonstationary, short- and long-range dependent as well as antipersistent autoregressive processes.

Beran [10] has shown that (1.1) allows for maximum likelihood estimation of $d = m+\delta$. Thus, in particular, a confidence interval can be given for d. Moreover, since all parameters are estimated simultaneously from the data, the autoregressive order p can be determined via an automatic model selection criterion such as the Akaike information criterion (Beran, Bhansali and Ocker [12]).

Besides the problems of parameter estimation and model selection, applied data analysts are often confronted with additional nontrivial important questions, such as how to obtain reliable forecasts, how to model aggregates (temporal aggregation) and external interventions within the framework of (1.1). Putting these questions into focus, the outline of the dissertation is as follows.

Chapter 2 reviews the results of Beran [10] and Beran, Bhansali and Ocker [12] on parameter estimation and model selection of (1.1). Finite sample behaviour is studied via an additional simulation study with respect to small sample sizes. The motivation was to assess the consistency properties of model choice criteria (AIC, HIC and BIC), given a realistic sample size, for determining an appropriate autoregressive order when d and the autoregressive parameters are estimated simultaneously by a maximum likelihood procedure according to Beran [10].

Chapter 3 discusses forecasting (point- and interval-forecasts) in a stationary and nonstationary FARIMA$(p, d, 0)$ environment. Parts of this chapter are taken from Beran and Ocker [17]. A simulation study answers the question which model choice criterion (AIC, HIC or BIC) may lead to improved forecasts. Moreover, a relative comparison with the random walk, the benchmark in econometric modeling, is discussed.

Whereas each of the previous chapters follows the same self-contained style of a separate paper, chapter 4 is used as the battlefield for the methods presented in previous sections. In chapter 4 we apply our findings on parameter estimation, model choice and forecasting, within the framework of (1.1), to worldwide nominal stock market indices and foreign exchange rates. We discuss, in particular, the possible gain/loss when using the quite complicated FARIMA$(p, d, 0)$ predictions in comparison to random walk and traditional ARIMA$(p, m, 0)$ forecasts.

Chapter 5 deals with temporal aggregation of (possibly nonstationary) FRIMA$(p, d, 0)$ models. The purpose of this chapter is to show how temporal aggregation influences and changes the model structure of (1.1) asymptotically. We prove that the asymptotic results on temporal aggregation of fractional and nonfractional, stationary and nonstationary autoregressive processes can be derived using one unified treatment. Simulation results and data examples round off our findings.

Chapter 6 focuses on modeling external interventions within the context of FARIMA$(p, d, 0)$ models. A new approach is presented which allows to model and estimate external shocks in (1.1). We derive, in particular, necessary and sufficient conditions for estimating interventions consistently, and derive the corresponding asymptotic distribution.

Finally, chapter 7 is devoted to the final remarks and further research trends from this work.

Chapter 2

Estimation and model selection of FARIMA$(p, d, 0)$ models

2.1 Introduction

The main objective of this chapter is to examine the question of how to select the autoregressive order when a realization of n consecutive observations $X_1, ..., X_n$ following the model (1.1) is available. An additional goal is to obtain reliable estimates of the differencing parameter $d = m + \delta$ and the autoregressive coefficients. Much work has been done on order selection of stationary short-range dependent autoregressive processes (i.e. $d = 0$) using automatic model-selection criteria (see e.g. Bhansali [19] for an overview), such as the AIC (Shibata [75], Hannan [50]), the BIC (Schwarz [74], Akaike [3]) and the HIC (Hannan and Quinn [51]). Similar model selection criteria are also known for nonstationary short-memory autoregressive models (i.e. d=1) (see e.g. Tsay [83]). However, only little can be found on model selection criteria when applied to fractional processes (i.e. $\delta \neq 0$). Recently, Beran, Bhansali and Ocker [12] provided both, the theoretical derivation of the properties of these criteria and a simulation study using Beran's [10] approximate maximum likelihood estimator. We have shown that the consistency properties of the AIC, the HIC and the BIC are analogous to the case of stationary short-range dependent autoregressive processes. Reliable estimates were already obtained for moderately large sample sizes. The results provide a unified treatment of fractional and nonfractional, stationary and nonstationary autoregressive models. This article is the basis of the following chapter.

In section 2, results on maximum likelihood estimation according to Beran [10] are summarized. In section 3, we review the findings of Beran, Bhansali and Ocker [12] on selection criteria and their properties. Finite sample behaviour is illustrated and discussed via an additional simulation study with respect to small sample sizes in section 4. A few general remarks in section 5 conclude the chapter. Tables are provided in the appendix.

2.2 Maximum likelihood estimation

The essential difference between nonfractional Box-Jenkins ARIMA$(p, m, 0)$ models and fractional ARIMA$(p, \delta, 0)$ models is that for the latter, the fractional differencing parameter δ is usually estimated by maximum likelihood and inference about the ARMA-parameters takes into account this estimation (see e.g. Beran [9], Dahlhaus [34], Fox and Taqqu [44], Haslett and Raftery [54], Sowell [76], Yajima [86]). In particular, a confidence interval for the differencing parameter can be given. In contrast, inference for traditional Box-Jenkins models is done in two stages. By applying exploratory tools and some formal tests on unit roots (see e.g. Dickey and Fuller [38]), one decides how many times the time series has to be differenced to obtain stationarity. In a second step, the parameters in $\phi(\cdot)$ are estimated and inference about these parameters is done as if m had been known a priori. No confidence intervals for m are available.

This discrepancy causes serious problems. It is well known that traditional formal tests on unit roots have little power when the alternative is fractionally integrated (see e.g. Hassler and Wolters [55]). In particular, stationary long memory processes often exhibit local spurious trends which are hard to distinguish from purely stochastic nonstationarity (see e.g. Beran [9]). Consequently, using the traditional Box-Jenkins setting, stationary long memory time series may be easily misspecified as nonstationary ARIMA models.

Recently, Beran [10] proposes a unified treatment of nonfractional Box-Jenkins ARIMA$(p, m, 0)$ and fractional ARIMA$(p, \delta, 0)$ models, namely the (possibly nonstationary) FARIMA$(p, d, 0)$ model

$$\phi(B)(1-B)^{\delta}\{(1-B)^{m}X_t - \mu\} = \epsilon_t, \tag{2.1}$$

with fractional difference given by

$$(1-B)^\delta = \sum_{k=0}^{\infty} b_k(\delta) B^k, \quad \text{where} \quad b_k(\delta) = (-1)^k \frac{\Gamma(\delta+1)}{\Gamma(k+1)\Gamma(\delta-k+1)}.$$

Neither the fractional nor the integer part of $d = m + \delta$ need to be known a priori, since both are estimated from the data. Thus, confidence intervals for d can be given and inference about $\phi_1, ..., \phi_p$ takes into account that d is unknown. The results on maximum likelihood estimation are as follows.

Let $X = (X_1, ..., X_n)$ be defined by (2.1). Without loss of generality, we assume μ to be known and equal to zero. The results do not change if μ is replaced by a consistent estimate (see Beran [10]). Denote by $\theta = (\sigma_\epsilon^2, d, \phi_1, ..., \phi_p) = (\sigma_\epsilon^2, \eta)$ the unknown parameter vector, where $d = m + \delta$, $\delta \in (-.5, .5)$ and $m \geq 0$ is an integer. X_t can be written in the infinite autoregressive representation (Beran [10])

$$\sum_{j=0}^{\infty} a_j(\eta) X_{t-j} = \epsilon_t(\eta), \tag{2.2}$$

where the coefficients $a_j(\eta)$ are obtained from (2.1). For example, if $p = 0$, then $a_j(\eta) = a_j(d)$ are obtained from equation (2.1) by

$$a_j(d) = \sum_{k=0}^{min\{m,j\}} (-1)^k \binom{m}{k} b_{j-k}(\delta). \tag{2.3}$$

The innovations $\epsilon_t(\eta)$ may be estimated by the residuals

$$e_t(\eta) = \sum_{j=0}^{t-1} a_j(\eta) X_{t-j}. \tag{2.4}$$

Standardized residuals are defined by $r_t(\theta) = e_t(\eta)/\sqrt{\theta_1}$. The population counterparts are $v_t(\eta) = \epsilon_t(\eta)/\sqrt{\theta_1}$. The autoregressive representation (2.2) suggests that we can estimate θ directly by an approximate maximum likelihood estimator, which minimizes the sum of squared standardized residuals

$$l(X;\theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma_\epsilon^2 - \frac{1}{2}n^{-1}\sum_{j=m+2}^{n} r_t^2(\theta) \tag{2.5}$$

with respect to θ. This results in solving a system of $p+2$ equations

$$\dot{l}(X;\theta) = 0, \tag{2.6}$$

where $\dot{l}$ is the vector of partial derivatives with respect to θ_j $(j = 1, ..., p+2)$. More explicitly, $\hat{\theta}$ is defined by

$$\sum_{t=m+2}^{n} \frac{\partial e_t(\eta)}{\partial \eta_j} e_t(\eta) = 0, \quad j = 1, ..., p+1, \quad \text{and} \tag{2.7}$$

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n} \sum_{t=m+2}^{n} e_t^2(\hat{\eta}) \tag{2.8}$$

The asymptotic distribution of $\hat{\theta}$ is given by the following theorem.

THEOREM 1 *(Beran [10]) Let $\hat{\theta}$ be the solution of (2.5). Then, as $n \to \infty$,*

(i) $\hat{\theta}$ converges almost surely to the true value θ^o;

(ii) $\sqrt{n}(\hat{\theta} - \theta^o)$ converges in distribution to a normal random vector with mean zero and covariance matrix $V = 2D^{-1}$, where

$$D_{ij} = (2\pi)^{-1} \left\{ \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_i} \log f(x) \frac{\partial}{\partial \theta_j} \log f(x) dx \right\} \Bigg|_{\theta = \theta_*^o},$$

$f(x) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma(k) e^{ikx}$ is the spectral density of the differenced process $(1-B)^m X_t$ with $\gamma(k) = Cov[(1-B)^m X_t, (1-B)^m X_{t+k}]$ and $\theta_*^o = (\sigma_{\epsilon,o}^2, \eta_*^o)^T = (\sigma_{\epsilon,o}^2, \delta^o, \phi_1^o, ..., \phi_p^o)^T$.

Remarks:

- Observe that both, the fractional and nonfractional part of $d = m + \delta$, are estimated from the data. As noted in Beran [10], the covariance matrix of $\hat{\eta}$ is especially simple for $p = 0$. From (2.5) it follows that the derivative of $l(\cdot)$ with respect to d is equal to $-\log(2 - 2\cos\lambda)$. Thus, we obtain

$$\lim_{n \to \infty} nVar(\hat{d}) = V_{22} = \frac{4\pi}{\int_{-\pi}^{\pi} \{\log(2 - 2\cos\lambda)\}^2 \mathrm{d}\lambda} = \frac{6}{\pi^2} \approx .6079 \tag{2.9}$$

(see formula 4.224 in Gradshteyn and Rhyzik [49]). This means that the distribution of $\hat{d}$ is asymptotically always the same and does not depend on the unknown m, so that $\theta_2^o = m^o + \delta^o$ is replaced by $\theta_{2,*}^o = \delta^o$ in theorem 1. Generally, the information matrix D does not depend on m and is of the same form as that derived by Dahlhaus [34] for stationary fractional autoregressive processes.

- If μ is unknown, the residuals defined above need to be adjusted. For $m > 0$ it follows directly from (2.1) that $Y_t = (1-B)^m X_t$ is a stationary FARIMA$(p, \delta, 0)$ process with expected value μ. Since Y_t is ergodic, the sample mean $\bar{Y}$ is a consistent estimator of μ and all results hold by replacing the unadjusted residuals by the adjusted residuals (see Beran [10] for a detailed algorithm).

- Note finally that we did not use exact maximum likelihood estimation (MLE) due to its large computational burden. For instance, for a stationary series of length n=6574, Haslett and Raftery [54] reported a CPU time of about 45 hours on a VAX11/780 for the calculation of the exact Gaussian MLE. Also, Cheung and Diebold [29] found evidence in stationary series that Sowell's [76] exact MLE and the approximate frequency-domain MLE of Fox and Taqqu [44] yield very similar results in case of unknown mean.

2.3 Model choice criteria

We now turn to the question how to select the autoregressive order p when a realization of n consecutive observations $X_1, ..., X_n$ following the model (2.1) is available. There has lately been considerable development on the question of order selection for stationary short-range dependent autoregressive processes (see e.g. Bhansali [19] for a review). The typical order determining criterion within the autoregressive Box-Jenkins framework is of the following general form

$$AIC_\alpha(p) = n \log \hat{\sigma}_\epsilon^2(p) + \alpha p, \quad (p = 0, 1, ..., L), \tag{2.10}$$

where $\hat{\sigma}_\epsilon^2$ is the maximum likelihood estimate of the innovation variance $\sigma_{\epsilon,o}^2$, and L is the maximum autoregressive order p. Note that α equals 2 for the AIC (Shibata [75], Hannan [50]), $2c \log\log n$ for the HIC (Schwarz [74], Akaike [3]), with $c > 1$, and $\log n$ for the BIC (Hannan and Quinn [51]). The estimated order $\hat{p}$ is the value of p for which the preferred selection criterion is minimal. However, only little can be found in the literature on these model selection criteria when applied to stationary and nonstationary fractional autoregressive models. Here, we review the recent results of Beran, Bhansali and Ocker [12]. We have shown that the consistency properties of the AIC, the HIC and the BIC are analogous to the case of stationary short-range dependent autoregressive processes. The results on model selection are as follows.

Suppose we observe two independent realizations $X = (X_1, ..., X_n)$ and $Y = (Y_1, ..., Y_n)$ generated by (2.1) with unknown true order p_o. Using (2.5), the approximate maximum likelihood estimate for the series X is given by $\hat{\theta}(X;p)$, where $p \geq p_o$. The loss of using the process with order p and estimated parameter vector $\hat{\theta}(X;p)$ as a model for Y is measured by the loss function $L(p, \hat{\theta}(X;p)) = E_y\{-2l[Y;\hat{\theta}(X;p)]\}$. The corresponding risk function is given by $R(p) = E_x\{L(p, \hat{\theta}(\cdot))\}$. Here, E_y and E_x denote expectation with respect to the joint distribution of Y and X respectively. Both $L(\cdot)$ and $R(\cdot)$ are unknown, because they depend on the unknown true model. Thus, instead of using the unknown risk function $R(\cdot)$ as a criterion for choosing p, a quantity $AIC(p)$ is calculated that is an asymptotically unbiased estimate of $R(p) + C(n;\theta^o)$, where $C(n;\theta^o)$ is a constant that does not depend on p. Essentially, $AIC(p)$ is based on the following considerations: We know from (2.5) that

$$-2l(Y;\theta) = n \log 2\pi + n \log \sigma_\epsilon^2 + \frac{1}{\sigma_\epsilon^2} \sum_{t=m+2}^{n} \sum_{i,j=0}^{t-1} a_i(\eta) a_j(\eta) Y_{t-i} Y_{t-j}. \quad (2.11)$$

Substituting $\hat{\theta}(X;p)$ in (2.11), omitting the arguments $(X;p)$, and taking expectation with respect to the joint distribution Y, yields the loss function

$$\begin{aligned} L(p,\hat{\theta}) &= E_y[-2l(Y;\hat{\theta})] \\ &= n \log 2\pi + n \log \hat{\sigma}_\epsilon^2 + \frac{1}{\hat{\sigma}_\epsilon^2} \sum_{t=m+2}^{n} \sum_{i,j=0}^{t-1} a_i(\hat{\eta}) a_j(\hat{\eta}) \gamma(i-j). \quad (2.12) \end{aligned}$$

Now, since Y is independent of X, we have

$$R(p) = E_x\{L(p,\hat{\theta})\} = L(p,\hat{\theta}). \quad (2.13)$$

In a next step, we approximate $\sum_{i,j=0}^{t-1} a_i(\cdot)a_j(\cdot)\gamma(\cdot)$ by $\sum_{i,j=0}^{\infty} a_i(\cdot)a_j(\cdot)\gamma(\cdot)$. Using Beran's [10] asymptotic central limit theorem 1 for $\hat{\sigma}_\epsilon^2$ and $\hat{\eta}$, a Taylor expansion around the true values $\sigma_{\epsilon,o}^2$ and η^o is applied. As a result we may write (Beran, Bhansali and Ocker [12])

$$R(p) = n \log 2\pi + n \log \sigma_{\epsilon,o}^2 + p + A(n;\theta^o) + o(1), \forall p \geq p_o, \tag{2.14}$$

where $A(\cdot)$ does not depend on p and $o(1)$ denotes a term tending to 0 as $n \to \infty$. Unfortunately, as $\sigma_{\epsilon,o}^2$ is unknown, $R(p)$ cannot be used directly for order selection. At the same time the usual maximum likelihood estimator of $\sigma_{\epsilon,o}^2$ is biased to terms of the same order as those retained in evaluating $R(p)$ (see Akaike [3]). However, an asymptotic expression for the bias in estimating $\log \sigma_{\epsilon,o}^2$ by $\log \hat{\sigma}_\epsilon^2$, for the class of processes satisfying (2.1), is given by (Beran, Bhansali and Ocker [12])

$$E_x(\log \hat{\sigma}_\epsilon^2) = \log \sigma_{\epsilon,o}^2 - \frac{p+1}{n} + o(n^{-1}). \tag{2.15}$$

Thus, the use of AIC for autoregressive model selection may now be justified in the same way as for short-memory processes. We have the following theorem.

THEOREM 2 *(Beran, Bhansali and Ocker [12]) Let* $AIC(p) = n \log \hat{\sigma}_\epsilon^2 + 2p$. *Then* $E_x[AIC(p)] = R(p) + C(n;\theta^o) + o(1)$, *where* $C(\cdot)$ *does not depend on* p.

Theorem 2 states that, up to an additive constant $C(\cdot)$ and asymptotically negligible terms, for all $p \geq p_o$, $AIC(p)$ provides an unbiased estimator of $R(p)$. Moreover, since the constant $C(\cdot)$ is independent of p, the difference $AIC(p_1) - AIC(p_2)$ is an asymptotically unbiased estimate of $R(p_1) - R(p_2)$ for any $p_1, p_2 \geq p_o$. Thus, the AIC may be used in the same way as for classical Box-Jenkins autoregressive models. The estimate $\hat{p}$ is the value of p that minimizes $AIC(p)$.

The AIC defined above has analogous consistency properties as in the context of stationary short-range dependent autoregressive models (see e.g. Shibata [75]). This follows by standard arguments applying the Taylor expansion in theorem 2. The probability of underestimating p_o converges to a zero, that of overfitting to a non-zero constant.

If a consistent estimate of p_o is needed, the penalty term α for the number of parameters must tend to infinity with increasing n. The minimal penalty term for fractional models (2.1) is given by the following theorem

THEOREM 3 *(Beran, Bhansali and Ocker [12]) Let* $c > 1$, $HIC(p) = n \log \hat{\sigma}_\epsilon^2 + 2c \log\log(np)$ *and* $\hat{p}_o = argmin HIC(p)$. *Then* $P(\hat{p} = p_o) \to 1$ *as* $n \to \infty$.

Note that this result is the same as for non-fractional models. Moreover, it follows by theorem 3 that using the BIC, with $\alpha = \log n$, provides a consistent estimator of p_o too.

2.4 Simulations

2.4.1 Simulated models

For d equal to $-.4, -.2, 0, .2, .4, .6, .8, 1.0, 1.2, 1.4$, one hundred series of each of the following six models were simulated:

- Model a: $(1 + .9B)(1 - B)^\delta (1 - B)^m X_t = \epsilon_t$;
- Model b: $(1 - B)^\delta (1 - B)^m X_t = \epsilon_t$;
- Model c: $(1 - .2B)(1 - B)^\delta (1 - B)^m X_t = \epsilon_t$;
- Model d: $(1 - .5B)(1 - B)^\delta (1 - B)^m X_t = \epsilon_t$;
- Model e: $(1 - .9B)(1 - B)^\delta (1 - B)^m X_t = \epsilon_t$;
- Model f: $(1 - 1.42B + .73B^2)(1 - B)^\delta (1 - B)^m X_t = \epsilon_t$.

The choice of the models was based on two considerations.

- To obtain information about the general behaviour of the criteria a broad spectrum of different models is considered. The characteristic features of the models above are (after taking the mth difference): 'Model a' exhibits strong short-range correlations with alternating signs. 'Model b' is the simplest possible model with no autoregressive coefficients. 'Model c' has a very small autoregressive coefficient that model positive short-range dependence which may be difficult to detect. 'Model d' represents a fractional AR(1) process with a medium size autoregressive parameter that models moderate positive short-range correlations. 'Model e' has a very strong positive short-range dependence that may be difficult to distinguish from long-memory behaviour. In particular, it may be difficult to judge whether d is below .5 (stationarity) or not (nonstationarity). The last model represents a case with $p_o = 2$. The spectral density of 'Model f' has a local maximum at a nonzero frequency, implying random short-range periodicities.

- It is interesting to compare our results with existing results in the literature for the corresponding short-memory (see e.g. Pukkila, Koreisha and Kallinen [72], Bierens and Guo [20]) and fractional (see e.g. Crato and Ray [32], Cheung and Diebold [29], Cheung [28], Hauser [56], Taqqu, Teverovsky and Willinger [80]) autoregressive processes. In our simulation study, the autoregressive parameter values correspond to values used in the literature.

The series of the models a-f were generated using the S-Plus function *arima.fracdiff.sim* (and *cumsum* for $d > .5$) with sample sizes

- $n = 50$, a very small one which may be typical for temporarily aggregated (e.g. macroeconomic) time series, and
- $n = 200$, also a small one which corresponds to other simulation studies.

The maximum order L fitted equaled 15.

2.4.2 Results of the simulation study

Estimation of p_o

Tables 2.1, 2.2 and 2.3 show the frequency of the autoregressive order selected for Models a-f for n equal to 50 and 200, based on the AIC, HIC and BIC respectively. For models with $p_o = 1$ and positive short-range correlations (models c, d and e), underestimation often occurs for small sample sizes for the BIC. The reason is that for small sample sizes, it is difficult to distinguish between positive short-range dependence and long-range dependence (which is positive by definition). This is most evident for 'Model c'. Since the BIC penalizes additional parameters very strongly and the autoregressive parameter is very small, the FARIMA$(0, \delta, 0)$ model is often preferred. In contrast, alternating short-range correlations ('Model a') or short-range periodic patterns ('Model f') are much easier to distinguish from long-range dependence. Also, good results are obtained for fractional noise ('Model b'). Our findings are analogous to previous simulation studies for short-memory autoregressive models where, for small sample sizes, the BIC turned out to underestimate p_o quite frequently, if the observed time series was generated by an autoregressive process with weak positive dependence, but performs well if short-range correlations imply short-range periodic patterns (see e.g. Pukkila, Koreisha and Kallinen [72]).

The results for the AIC and HIC illustrate that, with the exception of 'Model c', the two criteria rarely underestimate p_o, whereas overestimation occurs quite frequently. Asymptotically, overestimation is expected to occur in about 30% of the cases using the AIC (Hannan and Quinn [51]). As for the other two criteria, the asymptotic behaviour of the AIC seems to be approached at a slower rate in the presence of strong positive short-range dependence. Also, as for the HIC and BIC, convergence seems to be slower for the nonstationary case ($d > .5$). Overall, as a rule of thumb, the BIC should be used in connection with small sample sizes.

It is also interesting to compare our results with some related models in the paper of Crato and Ray [32]. They conducted a simulation study to investigate the performance of these criteria for different estimation procedures, restricted to the stationary case only. In particular their FARIMA(1,.3,0) model with $\phi_1 = -.65$ is comparable with our 'Model d' and $\delta = .4$. Both models may point towards a near unit root process. In the study of Crato and Ray [32] the true order $p_o = 1$ was detected in only 16% (9.2%) and 21.8% (11.6%) of 490 replications (and $n = 360$) via the AIC (BIC) using the approximate time-domain maximum likelihood procedure of Haslett and Raftery [54] and the frequency-domain approximate maximum likelihood method of Fox and Taqqu [44] respectively. The two approaches are clearly outperformed by Beran's [10] maximum likelihood estimator. Here, the true order was detected in 41% (78%) of the cases via the AIC (BIC) given a sample size of $n = 200$. Also, for the sample size $n = 50$, our method performs much better using the BIC. The results for the FARIMA$(0, .4, 0)$ model are quite close together with, however, different sample sizes. Here, the true order $p_o = 0$ was detected in 80% of the cases using the BIC and $n = 200$, whereas Crato and Ray [32] report success-rates of 92.2% and 78.8% for the Haslett and Raftery [54] method and the Fox and Taqqu [44] procedure respectively. Overall, with respect to estimate the true autoregressive order p_o, the method presented here seems to be highly competitive.

Estimation of d

Putting the estimation of the differencing parameter $d = m + \delta$ into focus tables 2.4, 2.5 and 2.6 give the simulated bias $N^{-1}\sum_{i=1}^{N}(\hat{d}_i - d)$ and the mean squared error (MSE) $N^{-1}\sum_{i=1}^{N}(\hat{d}_i - d)^2$, where $N = 100$ is the number of simulations and $\hat{d}_i$ is the estimate of d for the ith simulated series. Note that sample sizes of 50 and 200 are very small for estimating d. Even if p_o is known, reliable estimation of d is known to require longer series. For instance, for a fractional ARIMA$(0, d, 0)$ process, the (asymptotic) standard deviation of $\hat{d}$ obtained from theorem 1 is equal to $\sqrt{6}/\pi \approx .78$. Thus, the width of an approximate 95%-confidence interval for d is about .43 (.216) for $n = 50$ ($n = 200$). It is therefore difficult to distinguish between $\delta < 0$ and $\delta > 0$, or between stationarity with long-range dependence ($0 < d < .5$) and nonstationarity with a d slightly larger than .5. In view of this, it is not surprising that, for some models, the bias and MSE are relatively large.

However, the results illustrate that the MSE is relatively small if the BIC is used and $n = 200$. It is also interesting to see that for all criteria and $d > 0$, the bias is almost always negative and increases with d. Obviously, the bias is not negligible if an inconsistent criterion, such as the AIC, is used. That is, d tends to be underestimated when p is chosen to large. The intuitive reason is that the autoregressive coefficients tend to model a part of the long-range dependence so that $\hat{d}$ is reduced. This effect is most pronounced in the presence of strong positive short-range dependence. In addition, the results show that, in comparison, small sample estimates of d are most reliable when the BIC is used. Overall, we may thus conclude:

1. If estimation of d is the main purpose, a consistent criterion, such as the BIC, can be recommended;

2. Strong short-range correlations may cause a strong (usually negative) small sample bias of $\hat{d}$, in particular when using an inconsistent criterion.

Estimation of m

There is an extensive literature on theoretical properties of model choice criteria, parameter estimation with unknown order as well as a number of simulation studies for stationary short-memory ARIMA models (see e.g. Pukkila, Koreisha and Kallinen [72], Koreisha and Yoshimoto [59]). Also, several methods have been proposed for the case where d is unknown but restricted to the integers 0 or 1 (see e.g. Bierens and Guo [20], Koreisha and Pukkila [58], Phillips and Ploberger [71], Stock [77]). In particular Koreisha and Pukkila [58] report simulated frequencies of correct identification of m for models a, d, e and f with δ known to equal 0, and the sample size n =100 and 200. The approach presented here differs from the literature in that d is completely unknown and can assume any (integer or fractional) value larger than -.5. This makes estimation and model choice more difficult than a decision between two possibilities $d = 0$ and $d = 1$. Both, the integer part $m = [d + .5]$ and the fractional part $\delta = d - m$ have to be estimated from the data. For instance, $d = .49$ corresponds to $m = 0$ whereas $d = .51$ implies $m = 1$. For the models a, d, e and f with $m = 0$, the simulated frequencies of correctly identifying m, when using the BIC, turned out to be 100, 97, 47, 100 respectively for $n = 50$ and 100, 100, 91, 100 for $n = 200$. Thus, even for $n = 50$, the probability of identifying m correctly appears to be rather high, except for 'Model e'. For 'Model e', the autoregressive polynomial $\phi(B)$ is close to the unit root polynomial $1 - B$. This may explain why, for $n = 50$, the stationary but strong short-range dependence is likely to be confounded with unit root behaviour (i.e. $d = 1$). Overall, the simulated frequencies are comparable to those reported in Koreisha and Pukkila [58]. Thus, for $d = 0$ one may conclude that, in spite of using a wider class of models, the method proposed here yields results that are competitive with those achieved by methods designed specifically for the case where d is an integer. Finally, for $d = 1$, the simulated frequencies of correct identification of m for models a, d, e and f respectively are 84, 64, 49, 81 for $n = 50$ and 89, 87, 48, 93 for $n = 200$. For $n = 200$ these values are very similar to those reported in the literature, except for 'Model e'. In this case the fractional value of δ was apparently highly confounded with the autoregressive order (i.e. $\hat{p} > p_o = 1$).

Estimation of δ

In some situations, the main goal is to estimate the fractional differencing parameter δ. Leaving the question of robustness away, we want to compare our estimation results, using the BIC, with those reported in the recent literature (see e.g. Cheung and Diebold [29], Hauser [56], Taqqu, Teverovsky and Willinger [80]).

In particular Hauser [56] reports simulated mean squared errors for several semiparametric and nonparametric estimation procedures under white noise. Despite of serious losses in efficiency (see Beran [8], [9] and references therein), those methods are still popular in econometric modeling. Observe that we obtain a MSE of .008 for white noise and $n = 200$. In contrast, Hauser [56] reports MSEs that are, overall, twice as large given a sample size of $n = 1000$. A similar simulation study is provided by Taqqu, Teverovsky and Willinger [80]. They also investigated the properties of several widespread semiparametric and nonparametric estimation methods for stationary fractional noise using a sample size of $n = 10000$. We obtained in a former study (Beran, Bhansali and Ocker [12]) the mean squared errors ≈ 0 and .006 for $\delta = .2$ and $\delta = .4$ respectively given a sample size of $n = 1000$. None of the techniques considered in Taqqu, Teverovsky and Willinger [80] performed better for $\delta = .2$. For $\delta = .4$, our procedure performs slightly worse. However, in spite of using a wider class of models and smaller sample sizes, the method proposed here yields results that are at least competitive with those reported in Taqqu, Teverovsky and Willinger [80].

A relative comparison of our results with the simulations reported in Cheungand Diebold [29] may give additional indication of the possible loss/gain in model estimation using our method. The studies are, however, not directly comparable. Again, since we consider a wider class of models that includes autoregressive parameters, the probability of obtaining a wrong estimate of δ when the sample size is small, may be quite high. In view of this, it is interesting to observe that our estimation results for stationary fractional noise ('Model b') are quite competitive in terms of the MSE compared to Sowell's [76] exact time-domain maximum likelihood estimator (MLE) and the approximate MLE of Fox and Taqqu [44]. None of the realistic (e.g. arithmetic mean removed) methods presented there performed better in connection with antipersistence and $n = 50$. For $\delta \in (0, .5)$, our approach yields (as expected) slightly worse results.

2.5 Concluding remarks

In this chapter, we investigated model choice criteria for a class of models that includes nonfractional and fractional, stationary and nonstationary autoregressive processes. Together with Beran's[10] maximum likelihood estimation, the results provide a unified approach to short- and long-range dependent autoregressive modeling. Overall, even for small sample sizes, the method presented here seems to be highly competitive, when using the BIC. Additional simulations with respect to larger sample sizes confirm the investigated findings (see Beran, Bhansali and Ocker [12]).

The theorems remain unchanged if moving average components are included too. Simulations for models with moving average terms pose, however, computational problems. Here, for each d on a grid, the autoregressive parameters were estimated using the S-Plus function *ar.burg*, whereat *arima.mle* turned out to cause inacceptable long CPU times. The development of a more stable and faster algorithm which allows the additional estimation of moving average components is subject to ongoing research.

2.6 Appendix

Table 2.1: Simulated number of estimated order $\widehat{p}$ via the AIC

n=50

Model		d=-.4	-.2	0	.2	.4	.6	.8	1.0	1.2	1.4
a	$\hat{p} < p_o$	0	0	0	0	0	0	0	0	0	0
	$\hat{p} = p_o$	61	66	67	49	32	16	23	32	43	34
	$\hat{p} > p_o$	39	34	33	51	68	84	77	68	57	66
b	$\hat{p} < p_o$	0	0	0	0	0	0	0	0	0	0
	$\hat{p} = p_o$	65	61	59	45	24	22	25	41	33	26
	$\hat{p} > p_o$	35	39	41	55	76	78	75	59	67	74
c	$\hat{p} < p_o$	60	39	26	19	12	20	22	21	15	4
	$\hat{p} = p_o$	11	28	40	45	50	29	22	19	24	32
	$\hat{p} > p_o$	29	33	34	36	38	51	56	60	61	64
d	$\hat{p} < p_o$	25	14	5	6	7	8	8	7	3	0
	$\hat{p} = p_o$	43	48	55	41	28	18	13	20	21	19
	$\hat{p} > p_o$	32	38	40	53	65	74	79	73	76	81
e	$\hat{p} < p_o$	16	14	16	21	18	6	2	1	0	0
	$\hat{p} = p_o$	49	46	21	20	13	15	19	9	4	0
	$\hat{p} > p_o$	35	40	63	59	69	79	79	90	96	100
f	$\hat{p} < p_o$	0	0	0	0	0	0	0	0	0	0
	$\hat{p} = p_o$	65	69	67	48	24	22	25	36	27	19
	$\hat{p} > p_o$	35	31	33	52	76	78	75	64	73	81

n=200

Model		d=-.4	-.2	0	.2	.4	.6	.8	1.0	1.2	1.4
a	$\hat{p} < p_o$	0	0	0	0	0	0	0	0	1	0
	$\hat{p} = p_o$	66	69	55	58	42	8	24	43	53	43
	$\hat{p} > p_o$	34	31	45	42	58	92	76	58	47	57
b	$\hat{p} < p_o$	0	0	0	0	0	0	0	0	0	0
	$\hat{p} = p_o$	67	60	71	60	38	17	36	44	48	23
	$\hat{p} > p_o$	33	40	29	40	62	83	64	56	52	77
c	$\hat{p} < p_o$	27	21	25	18	5	6	9	12	10	2
	$\hat{p} = p_o$	51	55	45	56	68	54	50	42	42	41
	$\hat{p} > p_o$	22	24	30	26	27	40	41	46	48	57
d	$\hat{p} < p_o$	3	1	3	0	0	0	2	1	1	0
	$\hat{p} = p_o$	67	70	58	49	41	28	28	30	27	27
	$\hat{p} > p_o$	30	29	39	51	59	72	70	69	72	73
e	$\hat{p} < p_o$	1	2	0	0	1	2	0	0	0	0
	$\hat{p} = p_o$	77	58	63	44	14	31	22	13	13	8
	$\hat{p} > p_o$	22	40	37	56	85	67	78	87	87	92
f	$\hat{p} < p_o$	0	0	0	0	0	0	0	0	0	0
	$\hat{p} = p_o$	72	61	68	54	19	19	29	37	34	34
	$\hat{p} > p_o$	28	39	32	46	81	81	71	63	66	66

Table 2.2: Simulated number of estimated order $\widehat{p}$ via the HIC

n=50

Model		d=-.4	-.2	0	.2	.4	.6	.8	1.0	1.2	1.4
a	$\hat{p} < p_o$	0	0	0	0	0	0	1	0	0	0
	$\hat{p} = p_o$	83	85	85	79	55	34	38	52	62	54
	$\hat{p} > p_o$	17	15	15	21	45	66	61	48	38	46
b	$\hat{p} < p_o$	0	0	0	0	0	0	0	0	0	0
	$\hat{p} = p_o$	86	79	73	61	49	38	43	52	62	49
	$\hat{p} > p_o$	14	21	27	39	51	62	57	48	38	51
c	$\hat{p} < p_o$	80	60	47	31	26	31	40	43	25	10
	$\hat{p} = p_o$	12	27	45	45	60	41	24	22	35	41
	$\hat{p} > p_o$	8	13	8	24	14	28	36	35	40	49
d	$\hat{p} < p_o$	38	18	19	10	15	15	17	13	5	0
	$\hat{p} = p_o$	50	57	59	56	42	27	24	28	43	28
	$\hat{p} > p_o$	12	25	22	34	43	58	59	59	52	72
e	$\hat{p} < p_o$	31	24	30	29	33	19	6	1	0	0
	$\hat{p} = p_o$	55	57	32	24	13	22	25	24	10	1
	$\hat{p} > p_o$	14	19	38	47	54	59	69	75	90	99
f	$\hat{p} < p_o$	0	0	0	0	0	0	0	0	0	0
	$\hat{p} = p_o$	83	88	80	68	47	40	43	53	58	36
	$\hat{p} > p_o$	17	12	20	32	53	60	57	47	42	64

n=200

Model		d=-.4	-.2	0	.2	.4	.6	.8	1.0	1.2	1.4
a	$\hat{p} < p_o$	0	0	0	0	0	0	0	1	0	0
	$\hat{p} = p_o$	91	85	84	88	65	22	49	69	77	66
	$\hat{p} > p_o$	9	15	16	12	35	78	51	32	23	34
b	$\hat{p} < p_o$	0	0	0	0	0	0	0	0	0	0
	$\hat{p} = p_o$	92	88	92	76	60	38	50	71	70	55
	$\hat{p} > p_o$	8	12	8	24	40	62	50	29	30	45
c	$\hat{p} < p_o$	50	34	45	31	16	17	25	36	36	10
	$\hat{p} = p_o$	45	56	50	67	79	67	61	50	47	57
	$\hat{p} > p_o$	5	10	5	2	5	16	14	14	17	33
d	$\hat{p} < p_o$	8	4	3	0	2	1	5	3	2	0
	$\hat{p} = p_o$	83	87	79	72	64	61	48	55	43	40
	$\hat{p} > p_o$	9	9	18	28	34	38	47	42	55	60
e	$\hat{p} < p_o$	3	5	2	4	11	11	0	0	0	0
	$\hat{p} = p_o$	92	87	87	71	28	49	42	32	28	16
	$\hat{p} > p_o$	5	8	11	25	61	40	58	68	72	84
f	$\hat{p} < p_o$	0	0	0	0	0	0	0	0	0	0
	$\hat{p} = p_o$	91	90	83	73	40	39	57	71	70	55
	$\hat{p} > p_o$	9	10	17	27	60	61	43	29	30	45

Table 2.3: Simulated number of estimated order $\widehat{p}$ via the BIC

n=*50*

Model		d=-.4	-.2	0	.2	.4	.6	.8	1.0	1.2	1.4
a	$\hat{p} < p_o$	0	0	1	0	1	0	1	0	0	0
	$\hat{p} = p_o$	91	98	94	91	76	63	58	76	82	76
	$\hat{p} > p_o$	9	2	5	9	23	37	41	24	18	24
b	$\hat{p} < p_o$	0	0	0	0	0	0	0	0	0	0
	$\hat{p} = p_o$	97	95	91	83	70	65	66	77	83	71
	$\hat{p} > p_o$	3	5	9	17	30	35	34	23	17	29
c	$\hat{p} < p_o$	92	83	68	53	52	52	64	64	53	29
	$\hat{p} = p_o$	4	13	29	34	43	34	21	18	29	40
	$\hat{p} > p_o$	4	4	3	13	5	14	15	18	18	31
d	$\hat{p} < p_o$	64	37	39	27	34	38	38	29	17	2
	$\hat{p} = p_o$	34	54	52	60	43	21	27	28	49	38
	$\hat{p} > p_o$	2	9	9	13	23	41	35	43	34	60
e	$\hat{p} < p_o$	55	51	51	58	55	39	11	1	0	0
	$\hat{p} = p_o$	41	41	32	17	15	26	35	36	21	6
	$\hat{p} > p_o$	4	8	17	25	30	35	54	63	79	94
f	$\hat{p} < p_o$	0	0	0	0	0	0	0	0	0	0
	$\hat{p} = p_o$	92	96	97	87	70	62	67	80	83	59
	$\hat{p} > p_o$	8	4	3	13	30	38	33	20	17	41

n=*200*

Model		d=-.4	-.2	0	.2	.4	.6	.8	1.0	1.2	1.4
a	$\hat{p} < p_o$	0	0	0	0	0	0	0	1	0	0
	$\hat{p} = p_o$	96	97	98	97	84	34	76	86	93	87
	$\hat{p} > p_o$	4	3	2	3	16	66	24	15	7	13
b	$\hat{p} < p_o$	0	0	0	0	0	0	0	0	0	0
	$\hat{p} = p_o$	97	94	98	95	80	65	79	90	88	75
	$\hat{p} > p_o$	3	6	2	5	20	35	21	10	12	25
c	$\hat{p} < p_o$	73	56	62	47	33	51	48	53	56	25
	$\hat{p} = p_o$	26	43	36	53	66	47	48	44	36	52
	$\hat{p} > p_o$	1	1	2	0	1	2	4	3	8	23
d	$\hat{p} < p_o$	16	8	11	4	6	9	12	5	3	0
	$\hat{p} = p_o$	81	90	84	86	78	77	69	79	60	60
	$\hat{p} > p_o$	3	2	5	10	16	14	19	16	37	40
e	$\hat{p} < p_o$	8	16	9	17	34	22	0	0	0	0
	$\hat{p} = p_o$	91	82	89	78	36	61	66	48	44	26
	$\hat{p} > p_o$	1	2	2	5	30	17	34	52	56	74
f	$\hat{p} < p_o$	0	0	0	0	0	0	0	0	0	0
	$\hat{p} = p_o$	98	98	98	87	71	63	83	91	84	73
	$\hat{p} > p_o$	2	2	2	13	29	37	17	9	16	27

Table 2.4: Simulated bias and mean square error of $\widehat{d}$ via the AIC

Bias

Model	n	d=-.4	-.2	0	.2	.4	.6	.8	1.0	1.2	1.4
a	50	.125	-.015	-.146	-.257	-.540	-.694	-.770	-.677	-.456	-.702
	200	.056	-.048	-.153	-.253	-.442	-.809	-.673	-.423	-.358	-.437
b	50	.056	-.068	-.176	-.290	-.523	-.692	-.669	-.693	-.705	-.768
	200	.009	-.085	-.100	-.193	-.457	-.691	-.566	-.466	-.475	-.941
c	50	.086	-.059	-.226	-.430	-.551	-.598	-.688	-.745	-.785	-1.017
	200	.026	-.088	-.165	-.327	-.612	-.617	-.530	-.584	-.641	-.961
d	50	.084	-.123	-.332	-.428	-.535	-.673	-.725	-.789	-.914	-.971
	200	.026	-.152	-.286	-.388	-.484	-.608	-.652	-.722	-.807	-.857
e	50	.228	.080	-.015	-.087	-.198	-.432	-.642	-.762	-.960	-.986
	200	.054	-.055	-.096	-.307	-.543	-.423	-.648	-.823	-.927	-.834
f	50	.107	-.099	-.238	-.320	-.539	-.623	-.716	-.718	-.734	-.870
	200	.017	-.078	-.099	-.266	-.606	-.649	-.600	-.495	-.596	-.703

MSE

Model	n	d=-.4	-.2	0	.2	.4	.6	.8	1.0	1.2	1.4
a	50	.046	.036	.064	.182	.419	.690	.861	.873	.439	.791
	200	.013	.014	.061	.146	.342	.762	.646	.433	.388	.459
b	50	.044	.036	.078	.207	.461	.682	.756	.847	.854	.929
	200	.008	.021	.040	.101	.362	.643	.547	.489	.563	1.297
c	50	.028	.054	.147	.287	.474	.648	.799	.967	.977	1.330
	200	.010	.037	.083	.206	.475	.550	.496	.638	.729	1.190
d	50	.050	.062	.172	.280	.468	.720	.881	.989	1.126	1.140
	200	.016	.042	.120	.212	.381	.578	.693	.821	.959	.977
e	50	.213	.160	.294	.373	.483	.545	.666	.790	1.107	1.122
	200	.027	.045	.087	.199	.468	.455	.650	.839	.453	.887
f	50	.075	.048	.106	.224	.440	.619	.807	.822	.822	1.000
	200	.007	.021	.038	.149	.487	.605	.561	.520	.729	.858

Table 2.5: Simulated bias and mean square error of $\widehat{d}$ via the HIC

Bias

Model	n	d=-.4	-.2	0	.2	.4	.6	.8	1.0	1.2	1.4
a	50	.074	-.030	-.109	-.183	-.423	-.570	-.615	-.540	-.230	-.525
	200	.035	-.041	-.061	-.093	-.310	-.741	-.491	-.259	-.142	-.263
b	50	.023	-.060	-.148	-.272	-.439	-.577	-.564	-.559	-.404	-.589
	200	.002	-.050	-.045	-.135	-.314	-.549	-.461	-.231	-.263	-.628
c	50	.094	-.050	-.195	-.367	-.489	-.504	-.557	-.518	-.637	-.935
	200	.046	-.042	-.096	-.258	-.523	-.494	-.358	-.291	-.354	-.753
d	50	.110	-.121	-.240	-.364	-.430	-.600	-.542	-.695	-.749	-.934
	200	.026	-.145	-.271	-.325	-.376	-.435	-.526	-.553	-.699	-.763
e	50	.308	.207	.155	.032	-.046	-.288	-.609	-.718	-.933	-.990
	200	.091	.020	-.004	-.104	-.309	-.193	-.492	-.687	-.821	-.781
f	50	.053	-.095	-.232	-.256	-.471	-.491	-.607	-.590	-.529	-.760
	200	.007	-.032	-.084	-.186	-.465	-.487	-.399	-.227	-.293	-.479

MSE

Model	n	d=-.4	-.2	0	.2	.4	.6	.8	1.0	1.2	1.4
a	50	.025	.021	.040	.114	.306	.528	.651	.662	.304	.557
	200	.007	.009	.019	.042	.227	.707	.477	.252	.106	.248
b	50	.012	.029	.066	.160	.346	.549	.588	.650	.431	.666
	200	.004	.010	.016	.069	.245	.519	.433	.206	.280	.906
c	50	.025	.038	.109	.248	.402	.518	.640	.685	.774	1.216
	200	.011	.031	.062	.165	.406	.421	.319	.293	.413	.883
d	50	.057	.063	.160	.246	.405	.629	.706	.860	.865	1.073
	200	.017	.042	.112	.166	.269	.383	.553	.608	.812	.840
e	50	.249	.214	.326	.424	.542	.537	.696	.740	1.080	1.132
	200	.046	.051	.054	.132	.445	.404	.516	.726	.410	.844
f	50	.021	.034	.099	.161	.361	.493	.635	.681	.561	.825
	200	.004	.011	.032	.099	.369	.438	.362	.203	.320	.562

Table 2.6: Simulated bias and mean square error of $\widehat{d}$ via the BIC

Bias

Model	n	d=-.4	-.2	0	.2	.4	.6	.8	1.0	1.2	1.4
a	50	.052	-.028	-.095	-.144	-.295	-.394	-.472	-.221	-.054	-.353
	200	.032	1.768	-.036	-.051	-.171	-.626	-.227	-.105	-.057	-.097
b	50	.006	-.059	-.098	-.200	-.314	-.365	-.398	-.288	-.210	-.363
	200	-.001	1.760	-.026	-.065	-.164	-.309	-.205	-.104	-.116	-.372
c	50	.110	1.821	-.085	-.231	-.270	-.308	-.289	-.288	-.356	-.697
	200	.083	1.791	-.040	-.201	-.407	-.203	-.207	-.124	-.205	-.590
d	50	.207	-.006	-.081	-.211	-.234	-.324	-.209	-.464	-.568	-.830
	200	.041	-.120	-.200	-.249	-.281	-.219	-.270	-.305	-.562	-.650
e	50	.496	.434	.400	.355	.236	.015	-.439	-.623	-.865	-.931
	200	.140	.145	.083	.127	.113	.086	-.289	-.538	-.671	-.664
f	50	.040	-.093	-.172	-.223	-.329	-.358	-.433	-.278	-.319	-.578
	200	.007	-.019	-.041	-.118	-.232	-.303	-.182	-.099	-.161	-.316

MSE

Model	n	d=-.4	-.2	0	.2	.4	.6	.8	1.0	1.2	1.4
a	50	.015	.013	.031	.063	.173	.319	.472	.244	.138	.344
	200	.006	.006	.008	.018	.111	.592	.218	.082	.024	.048
b	50	.007	.022	.041	.095	.222	.317	.398	.285	.170	.359
	200	.003	.007	.008	.025	.122	.298	.181	.079	.104	.529
c	50	.026	.032	.074	.169	.257	.344	.342	.428	.430	.870
	200	.016	.028	.053	.148	.333	.203	.210	.127	.280	.690
d	50	.090	.079	.133	.181	.325	.485	.432	.631	.690	.942
	200	.025	.045	.094	.121	.187	.216	.292	.282	.605	.710
e	50	.408	.396	.426	.499	.572	.537	.602	.648	1.005	1.044
	200	.084	.120	.081	.156	.457	.320	.352	.585	.354	.682
f	50	.019	.033	.075	.132	.262	.356	.429	.284	.276	.595
	50	.005	.007	.011	.057	.179	.268	.150	.075	.150	.368

Chapter 3

Forecasting FARIMA$(p, d, 0)$ models

3.1 Introduction

Prediction of future observations and the calculation of valid prediction intervals strongly depends on the type of model that may be assumed. Time series, in particular financial price series (exchange rates, stock market indices, commodities,...) and macroeconomic data, exhibit apparent trends quite often (see e.g. Beran [11], Beran and Ocker [17], Beran and Ocker [18], Beran, Feng and Ocker [15], and Beran, Feng, Franke, Hess and Ocker [16]). A trend may be

(1) deterministic,

(2) stochastic,

(3) spurious, or

a mixture of (1), (2) and/or (3). A deterministic trend (1) is described by a deterministic function $g(t)$. Typical examples of (2) are nonstationary processes, such as random walk or ARIMA$(p, m, 0)$ processes, whose mth difference is stationary. In addition to nonstationary models, stationary long memory processes often exhibit local spurious trends (3) which may be hard to distinguish from deterministic and/or stochastic trends in nonstationary time series.

Forecasts will differ greatly, depending on how these trends are modeled. For instance, for a stationary series, forecasts of a conditional expected value converge to the sample mean with increasing forecasting horizon, and the length of forecast intervals is asymptotically constant. In contrast, for difference stationary series, forecasts converge to the last observation and the length of forecast intervals diverges to infinity. Forecasts for time series with a deterministic trend require a trend extrapolation which can usually not be trusted beyond a short forecast horizon. On a finer scale, the rate at which forecast intervals converge to the asymptotic length (for stationary processes) or diverge to infinity (for difference stationary processes) depends on the fractional differencing parameter (see Beran and Ocker [17]).

In practical applications, it is often very difficult to find the 'right' model and, in particular, to decide whether a series is stationary, has a deterministic or stochastic trend, or whether there may be long-range correlations or antipersistence. Consider, for instance, figure 3.1 (provided in the appendix) where four simulated series are generated respectively by a purely nonstationary process (FARIMA$(0, 1, 0)$), a nonstationary process whose first difference is the sum of a deterministic trend plus a stationary process (FARIMA$(0, 1, 0)$ & $\mu = .1$), a nonstationary long-memory process (FARIMA$(0, 1.2, 0)$), and a nonstationary process whose first difference is the sum of a deterministic trend plus a stationary long-memory process (FARIMA$(0, 1.2, 0)$ & $\mu = .1$). Note that μ denotes the mean of the stationary series $(1 - B)X_t$.

Since a visual assessment of the time series plots appears to be difficult (at least a priori), the large variety of possible models is often confusing to the applied data analyst. Finding an appropriate model and, in particular, defining realistic forecasts, is therefore a challenging task in practice. A possibility to resolve this problem is to set up a unified framework, such as the FARIMA$(p, d, 0)$ model (Beran [10], Beran, Bhansali and Ocker [12])

$$\phi(B)(1 - B)^{\delta}\{(1 - B)^m X_t - \mu\} = \epsilon_t, \tag{3.1}$$

in which flexible modeling of deterministic and stochastic components is possible, and objective data driven inference can be made to decide which of the components (deterministic trend, stochastic trend, spurious trend, stationary components) may be present.

With respect to modeling trends, the model class defined by (3.1) includes stochastic trends ($m > 0$) and local spurious trends generated by long-range dependence ($\delta > 0$). Moreover, if $m = 1$ and $\mu \neq 0$, then a global linear deterministic trend is included. Thus, using (3.1), data are modeled either by

(a) X_t=no deterministic trend + stationary process with short- or long-range dependence, or antipersistence;

(b) X_t=no deterministic trend + difference-stationary process, whose first difference has short- or long-range dependence, or antipersistence;

(c) X_t=deterministic trend + difference-stationary process, whose first difference has short- or long-range dependence, or antipersistence.

Observe that, alternative (a) allows the possibility of local spurious trends ($m = 0, \delta > 0$), while (b) includes stochastic trends ($m = 1$). Alternative (c) allows a mixture of deterministic and stochastic trends ($m = 1$ and $\mu \neq 0$). Also, (b) and (c) allow a combination of stochastic and local spurious trends ($m = 1$ and $\delta > 0$), whereas (c) a mixture of all three kinds of trends ($m = 1$, $\delta > 0$ and $\mu \neq 0$). A compact overview is given by the following table.

Table 3.1: Trends within the FARIMA($p, d, 0$) environment

	$\delta = 0$	$\delta > 0$
$m = 0$	no	spurious
$m = 1$	stochastic ($\mu = 0$) stochastic+deterministic ($\mu \neq 0$)	stochastic+spurious ($\mu = 0$) stochastic+spurious+deterministic ($\mu \neq 0$)

In (3.1), deterministic trends with stationary errors ($m = 0$) and other than polynomial trends are excluded. These possibilities, however, can be incorporated by allowing μ to be a function in time satisfying certain smoothness assumptions (see Beran [11], Beran and Ocker [17], Beran and Ocker [18], Beran, Feng and Ocker [15], and Beran, Feng, Franke, Hess and Ocker [16]).

The presence of a stochastic and/or spurious trend can easily be tested via an appropriate confidence interval for $d = m + \delta$, where $\delta \in (-.5; .5)$ and $m \geq 0$ is an integer. In the simplest case of a (possibly nonstationary) FARIMA$(0, d, 0)$ model, the variance of $\hat{d}$ is $6/(n\pi^2)$ (see Beran [10], Beran, Bhansali and Ocker [12]). Thus, a confidence interval of the form

$\widehat{d} \pm z_{\frac{\alpha}{2}}\sqrt{6}/(\pi\sqrt{n})$ can be used to test the null $d = d_o$ against its alternative $d \neq d_o$. Given such a confidence interval $CI = [CI_{low}; CI_{up}]$, where CI_{low} and CI_{up} denote the lower and upper bound respectively, and imposing the restriction $m \leq 1$, the following conclusions may be drawn (accompanied by a visual inspection of the corresponding time series plot):

- $CI_{low} \in (0, .5) \wedge CI_{up} < .5 \Longrightarrow$ spurious trend,
- $CI_{low} > .5 \wedge CI_{up} < 1 \Longrightarrow$ stochastic trend,
- $CI_{low} > 1 \Longrightarrow$ stochastic + spurious trend

If the observed series X_t is nonstationary, the time path of X_t may be also governed by a deterministic time trend. The presence of a deterministic trend can be tested via $H_o : \mu = 0$ against $H_1 : \mu \neq 0$, where μ denotes the mean of the corresponding stationary time series $Y_t = (1 - B)^m X_t$. If an appropriate test rejects the null, we may conclude that a linear deterministic time trend can be drawn in the observed time series. To illustrate this consider the following. Adding a constant term $\mu \neq 0$ to a random walk yields the so called random walk with drift, i.e.

$$X_t = X_{t-1} + \mu + \epsilon_t.$$

Repeatedly substituting for lagged values of X_t gives

$$X_t = X_0 + \mu t + \sum_{s=1}^{t} \epsilon_s, \quad \text{with} \quad E(X_t|X_0) = X_0 + \mu t,$$

if X_0 is fixed. Thus, besides a stochastic trend, the level of X_t is also governed by a linear deterministic time trend $g(t) = X_0 + \mu t$. That is, testing the null $\mu = 0$ against its alternative $\mu \neq 0$ in the stationary series $Y_t = (1 - B)X_t$ is equivalent to test whether the slope μ of the linear trend function $g(t)$ in the corresponding nonstationary series X_t is significantly different from zero or not.

Generally, imposing the restriction $m \leq 1$ enables us to reconstruct a (possibly zero) trend function $g(t)$ from (3.1) and to express the observed time series X_t explicitly as a sum of the (possibly) present components, i.e.

$$X_t = g(t) + V_t = \begin{cases} \mu + U_t, & \text{if } m = 0 \\ X_0 + \mu t + \sum_{s=1}^{t} U_s, & \text{if } m = 1 \text{ and } X_o \text{ is fixed} \end{cases}, \quad (3.2)$$

where $g(t)$ and V_t denote the deterministic and stochastic part of X_t respectively, and $U_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ is a zero mean stationary fractional infinite moving average process, where ψ_j is obtained by inverting (3.1).

Now what is left to do is to derive an appropriate test and confidence intervals for μ and to show how to obtain reliable point- and interval-forecasts. The chapter is organized as follows. In section 2, we summarize the results of Beran [7],[8],[9] on testing and confidence intervals for μ. Forecasting with FARIMA$(p,d,0)$ models is discussed in section 3. Most of this section is based on results in Beran and Ocker [17]. A new simulation study in section 4 illustrate the results. In particular, the question is put into focus which model choice criterion (AIC, HIC or BIC) may lead to improved (point- and interval-) forecasts. Some concluding remarks are given in section 5. Figures and tables are presented in the appendix.

3.2 Tests and confidence intervals for μ

One standard result of elementary statistics is that, given a sequence of independent, identically distributed (iid) random variables $\{Y_t\}$, with mean $E(Y_t) = \mu < \infty$ and variance $Var(Y_t) = E[(Y_t - \mu)^2] = \sigma_Y^2 < \infty$, the variance of the sample mean $\bar{Y} = n^{-1} \sum_{t=1}^{n} Y_t$ equals

$$Var\left(\bar{Y}\right) = \sigma^2 n^{-1}. \tag{3.3}$$

Hence, the normalized sample mean, the z-statistic, converges asymptotically to a standard normal random variable, i.e.

$$z = \frac{\bar{Y} - \mu}{\sigma_Y} n^{.5} \to_d Z. \tag{3.4}$$

The confidence interval for μ with asymptotic coverage probability $1 - \alpha$ is then given by

$$\bar{Y} \pm z_{\frac{\alpha}{2}} \sigma_Y n^{-.5}. \tag{3.5}$$

In practice, σ_Y has to be estimated and one can use the corresponding confidence interval derived from Student's t-statistic $t = (\bar{Y} - \mu)/s_Y n^{-.5}$, where σ_Y is replaced by the sample standard deviation $s_Y = \sqrt{\frac{1}{n-1} \sum_{t=1}^{n} (Y_t - \bar{Y})^2}$.

Let us now consider what happens to the sample mean's variance if we relax the assumption of independence. Consider a covariance stationary process $\{Y_t\}$. Denoting by $\rho(k) = \gamma(k)/\sigma^2$ the corresponding correlations, where $\gamma(k) = Cov(Y_t, Y_{t+k}), \forall k \geq 0$. Then

$$\begin{aligned} Var(\bar{Y}) &= \sigma_Y^2 \left(1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho(k)\right) n^{-1}, \\ &= \sigma_Y^2 (1 + \xi(\rho)) n^{-1}. \qquad (3.6) \end{aligned}$$

If the correlations decay exponentially (such as for traditional ARMA processes), i.e. $|\rho(k)| \leq ca^{|k|}$, where $0 < c < \infty, 0 < a < 1$, then

$$\begin{aligned} |\xi(\rho)| &\leq 2c\left\{\sum_{k=1}^{n-1} a^k - \frac{1}{n}\sum_{k=1}^{n-1} ka^k\right\}, \qquad (3.7) \\ &\leq 2c\left\{\frac{1-a^n}{1-a} - \frac{1}{n}\left(\frac{na^n - a^n}{1-a} + \frac{a-a^n}{(1-a)^2}\right)\right\}, \qquad (3.8) \\ &\leq \frac{2c}{1-a}, \quad \text{as } n \to \infty. \qquad (3.9) \end{aligned}$$

That is, the variance of $\bar{Y}$ is still proportional to n^{-1} and neither the traditional test statistic (3.4) nor the confidence interval (3.5) do need, apart from a constant, any modification.

In contrast, consider a process with hyperbolically decaying correlations (such as for stationary FARIMA processes), i.e. $\rho(k) \sim_{|k|\to\infty} c_\rho |k|^{2\delta-1}$, where c_ρ is a nonzero finite constant. If $\delta \in (-.5, 0)$, the process is said to be antipersistent and we have

$$\begin{aligned} \xi(\rho) &= 2\left\{\sum_{k=1}^{n-1}\rho(k) - \frac{1}{n}\sum_{k=1}^{n-1}\rho(k)\right\}, \\ &= 2\left\{-\sum_{k=n}^{\infty}\rho(k) - \sum_{k=-\infty}^{0}\rho(k) - \frac{1}{n}\sum_{k=1}^{n-1}\rho(k)\right\}, \\ &= 2\left\{-\sum_{k=n}^{\infty}\rho(k) - \frac{1}{2} - \frac{1}{n}\sum_{k=1}^{n-1}\rho(k)\right\}, \\ &\sim 2\left\{-c_\rho\sum_{k=n}^{\infty}k^{2\delta-1} - \frac{1}{2} - \frac{c_\rho}{n}\sum_{k=1}^{n-1}k^{2\delta}\right\}, \end{aligned}$$

$$\sim \quad 2\left\{-\frac{1}{2} - c_\rho n^{2\delta}\left(\int_1^\infty x^{2\delta-1}dx + \int_0^1 x^{2\delta}dx\right)\right\},$$
$$\sim \quad 2\left\{-\frac{1}{2} + \frac{c_\rho n^{2\delta}}{2\delta(2\delta+1)}\right\} \to -1, \quad \text{as } n \to \infty.$$

For the long memory case $\delta \in (0, .5)$ we have

$$\begin{aligned}
\xi(\rho) &\sim 2c_\rho\left\{\sum_{k=1}^{n-1} k^{2\delta-1} - \frac{1}{n}\sum_{k=1}^{n-1} k^{2\delta}\right\}, \\
&\sim 2c_\rho n^{2\delta}\left\{\int_0^1 x^{2\delta-1}dx - \int_0^1 x^{2\delta}dx\right\}, \\
&\sim \frac{2c_\rho n^{2\delta}}{2\delta(2\delta+1)} \to \infty, \quad \text{as } n \to \infty.
\end{aligned}$$

That is, $\xi(\rho)$ does not converge to a constant but diverge to plus infinity at the rate $n^{2\delta}$.

Therefore, formula (3.3) has to be multiplied not just by a constant but by a constant times $n^{2\delta}$. In view of these results, the statistic

$$z = \frac{\bar{Y} - \mu}{\sigma_\mu} n^{.5-\delta} \to_d Z, \tag{3.10}$$

where $\sigma_\mu = \sqrt{\sigma_Y^2 c_\rho / \delta(2\delta+1)}$, and Z denotes a standard normal random variable, should be used instead of (3.4). The confidence interval for μ with asymptotic coverage probability $1-\alpha$ is then given by

$$\bar{Y} \pm z_{\frac{\alpha}{2}} \sigma_\mu n^{\delta-.5}. \tag{3.11}$$

In practice, both δ and the scale parameter σ_μ have to be estimated. Then, a good approximation of the distribution in (3.10) can be obtained by taking into account the variability of $\hat{\delta}$ in the factor $n^{.5-\hat{\delta}}$ only. Suppose that $\hat{\delta}$ is a $\sqrt{n}$-consistent estimator (such as in Beran [10] and Beran, Bhansali and Ocker [12]). Then, we may approximate the distribution in (3.10) by the distribution of

$$t_n^* = Z_1 \cdot n^{Z_2 s_\delta n^{-.5}}, \tag{3.12}$$

with independent standard normal random variables Z_1 and Z_2. The estimated standard deviation of δ, s_δ, can be obtained from theorem 1. The confidence interval for μ with asymptotic coverage probability $1-\alpha$ is then given by

$$\bar{Y} \pm t^*_{\frac{\alpha}{2}} s_\mu n^{\hat{\delta}-.5}, \tag{3.13}$$

where $s_\mu = \sqrt{s_Y^2 \hat{c}_\rho / \hat{\delta}(2\hat{\delta}+1)}$. Beran [7] has shown that (3.13) provides indeed reliable confidence intervals. If the usually t-test is used in the presence of long-range dependence, the asymptotic coverage probability of the corresponding confidence interval is equal to zero, instead of $1-\alpha$ (Beran [7],[9]). This effect is also present for moderately large sample sizes. For tests of $H_o : \mu = 0$, this implies that the null is rejected with asymptotic probability 1, even if H_o is true.

3.3 FARIMA$(p, d, 0)$ predictions

Let $X_1, ..., X_n$ be observations generated by a FARIMA$(p, d, 0)$ model of order p and parameter vector $\theta = (\sigma_\epsilon^2, d, \phi_1, ..., \phi_p)$, where $d = m + \delta$; also, we impose the restriction $m \leq 1$. The aim is to predict a future observation X_{n+k} for some $k \in (1, 2, 3, ...)$. Recall from equation (3.2) that X_t can be explicitly expressed as sum of a (possibly zero) deterministic and a stochastic component, i.e. $X_t = g(t) + V_t$. Denote by $U_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ a zero mean stationary fractional infinite moving average process, where ψ_j is obtained by inverting (3.1). Then

$$X_{n+k} = g(n+k) + V_{n+k} \tag{3.14}$$

with

$$g(n+k) = \mu, \quad V_{n+k} = U_{n+k} \tag{3.15}$$

if $m = 0$, and

$$g(n+k) = X_1 + \mu \cdot (n+k-1), \quad V_{n+k} = \sum_{s=1}^{k} U_{n+s} \tag{3.16}$$

if $m = 1$. Thus predict X_{n+k} from $X_1, ..., X_n$ two problems need to be solved:

(1) prediction of the stochastic component V_{n+k};

(2) extrapolation of the function $g(t)$ to $n+k$ if $m=1$.

Note again that μ is the expected value of the stationary time series $Y_t = (1-B)X_t$. Since Y_t is ergodic, the sample mean $\bar{Y}$ is a consistent estimator of μ. Then, if $m=1$ and $H_o : \mu = 0$ is rejected, an extrapolation of the deterministic trend function $g(t)$ is given by $X_1 + \bar{Y} \cdot (n+k-1)$. Also, a possibly nonzero mean-level of X_t may be estimated by $\bar{Y} = \bar{X}$ if $m=0$. Prediction of the stochastic component V_{n+k} is slightly more difficult.

Using the mean square criterion, the best linear prediction of V_{n+k} based on past values is defined by $\hat{V}_{n+k} = \beta_{opt}^t U(n)$ where $U(n) = (U_1, ..., U_n)^t$ and the vector $\beta_{opt} = (\beta_1, ..., \beta_n)^t$ minimizes the mean squared error $MSE = E[(V_{n+k} - \hat{V}_{n+k})^2]$. The values of β_{opt} and the corresponding optimal mean squared error prediction error MSE_{opt} are given by the following theorem.

THEOREM 4 *(Beran and Ocker [17]) For all integers $r, s > 0$, define*

$$\gamma_r^{(s)} = [\gamma(r+s-1), \gamma(r+s-2), ..., \gamma(r)]^t, \tag{3.17}$$

$$\tilde{\gamma}_k^{(n)} = \sum_{j=1}^{k} \gamma_j^{(n-1)}, \tag{3.18}$$

and denote by $\Sigma_n = [\gamma(i-j)]_{i,j=1,...,n}$ the covariance matrix of $U(n)$. Then, the following holds.

(i) If $m=0$,

$$\beta_{opt} = \Sigma_n^{-1} \gamma_k^{(n)}, \tag{3.19}$$

$$MSE_{opt} = \gamma(0) - [\gamma_k^{(n)}]^t \Sigma_n^{-1} [\gamma_k^{(n)}]; \tag{3.20}$$

(ii) If $m=1$,

$$\beta_{opt} = \Sigma_n^{-1} \tilde{\gamma}_k^{(n)}, \tag{3.21}$$

$$MSE_{opt} = \sum_{s=1-k}^{k-1} (k-|s|)\gamma(s) - [\tilde{\gamma}_k^{(n)}]^t \Sigma_n^{-1} [\tilde{\gamma}_k^{(n)}]. \tag{3.22}$$

Note in particular that, as $k \to \infty$, the MSE tends to a finite constant in the case of a stationary stochastic component ($m = 0$), whereas it diverges to infinity in the case of a nonstationary stochastic component ($m = 1$). More specifically we have

COROLLARY 1 *(Beran and Ocker [17]) Define* $c_f = lim_{\lambda \to 0} |\lambda|^{2\delta} f(\lambda)$ *where* f *is the spectral density of* U_t*, and let*

$$v(\delta) = \frac{2\Gamma(1-2\delta)\sin \pi\delta}{\delta(2\delta+1)} \tag{3.23}$$

for $|\delta| \in (0, .5)$ *and* $v(0) = \lim_{\delta \to 0} v(\delta) = 2\pi$*. Then, as* $k \to \infty$*, the following holds.*

(i) If $m = 0$*,*

$$MSE_{opt} \to \gamma(0) = Var(U_t); \tag{3.24}$$

(ii) If $m = 1$

$$MSE_{opt} \sim c_f v(\delta) k^{1+2\delta}. \tag{3.25}$$

The intuitive interpretation of this result is that, as k tends to infinity, the past values $U_1, ..., U_n$ do not contribute anything to the prediction so that the MSE approaches the variance of V_{n+k}. For $m = 0$, $Var(V_{n+k})$ is equal to $\gamma(0) = Var(U_t)$, independently of the value of δ. For $m = 1$ (i.e. $d > .5$), the variance of V_{n+k} is proportional to $k^{2d-1} = k^{1+2\delta}$. Note, in particular, that for traditional Box-Jenkins ARIMA models, a discrete choice is made between $d = 0$ (stationarity) and $d = 1$ (unit root). As a result, prediction intervals are either asymptotically constant ($d = 0$) or diverge to infinity at the rate $O(\sqrt{k})$ (for $d = 1$). In contrast, when using fractional models with $d \in (-.5, .5) \cup (.5, 1.5)$, the length of prediction intervals is of the order $O(k^{\tau/2})$ with $\tau = max\{0, 2d - 1\}$. As $d \to .5, \tau \to 0$ and $k^\tau \to k^0 = 1$. Thus, there is a continuous transition between $O(1)$ and $O(k^{d-.5})$ (for $d > .5$). Due to this flexibility, prediction intervals for fractional models are better adapted to the data. In particular, for $.5 < d < 1$, confidence intervals based on classical models will either be asymptotically too narrow, if $d = 0$ is chosen, or unnecessarily wide (by the factor k^{1-d}), if $d = 1$ is chosen. The value of $\tau = max\{0, 2d-1\}$ is estimated from the data by the maximum likelihood method given in theorem 1. Finally, results from theorem 4 and corollary 1 can be used to obtain prediction intervals for X_{n+k} with $k \geq 1$. For known values of $g(\cdot)$ and θ, a $(1-\alpha)100\%$-prediction interval for X_{n+k} is given by $\hat{X}_{n+k} \pm z_{\frac{\alpha}{2}}\sqrt{MSE_{opt}}$. If $g(\cdot)$ and θ are unknown, one uses the corresponding estimated quantities (Beran and Ocker [17]).

3.4 Simulations

3.4.1 Simulated models

For $d = m + \delta$ equal to $-.4, -.2, 0, .2, .4, .6, .8, 1.0, 1.2, 1.4$, one hundred series of each of the following six models (see also chapter 2) were simulated:

- Model a: $(1+.9B)(1-B)^{\delta}(1-B)^{m}X_t = \epsilon_t$;
- Model b: $(1-B)^{\delta}(1-B)^{m}X_t = \epsilon_t$;
- Model c: $(1-.2B)(1-B)^{\delta}(1-B)^{m}X_t = \epsilon_t$;
- Model d: $(1-.5B)(1-B)^{\delta}(1-B)^{m}X_t = \epsilon_t$;
- Model e: $(1-.9B)(1-B)^{\delta}(1-B)^{m}X_t = \epsilon_t$;
- Model f: $(1-1.42B+.73B^2)(1-B)^{\delta}(1-B)^{m}X_t = \epsilon_t$.

The series were generated using the S-Plus function *arima.fracdiff.sim* (and *cumsum* for $d > .5$) with sample sizes

- $n = 50$, a very small one which may be typical for temporarily aggregated (e.g. macroeconomic) time series, and
- $n = 200$, a size which is typical in daily applications such as in banking.

Note that we simulated $k = 1, ..., 120$ additional data points for out-of-sample forecast assessment. For each model and each d, a FARIMA$(p, d, 0)$ model is fitted to each simulation, where p and d are estimated (using the AIC, HIC and BIC) and forecasts (point and interval estimates) are calculated for X_{n+k} (reported for k=1,2,3,4,5,10,15,20,30,40,50) as described above. The maximum autoregressive order equaled 15. The choice of the models was based on the following considerations:

- It is well known that AIC and HIC rarely underestimate the true order p_o, whereas overestimation occurs quite frequently. Also, estimation of d is typically confounded with the selected autoregressive order and estimated parameters. As mentioned in the previous chapter, autoregressive coefficients tend to model a part of the long-range dependence so that $\hat{d}$ is reduced. The simulation results so far indicate that, in comparison, small sample estimates of d are most reliable when the BIC is used (see Beran, Bhansali and Ocker [12] and the previous chapter). It is therefore interesting to see, whether reliable forecasts can be achieved using the 'right' (via the BIC) or the 'wrong' (via the AIC) model.

- The discovery of long-range dependence suggests possibilities for improved forecasting performance within a linear time series framework, especially over longer forecasting horizons (see Granger and Joyeux [48], Geweke and Porter-Hudak [47]). A relative comparison of our results with the random walk, the benchmark in econometric modeling, may give an indication of the possible loss/gain in forecasting when using the more sophisticated FARIMA($p, d, 0$) predictions.

3.4.2 Results of the simulation study

Comparison of the AIC, HIC and BIC

Tables 3.2 through 3.7 give the percentage of 'future' observations contained in the 95%-prediction intervals for k-steps ahead FARIMA forecasts. Note that the percentages are obtained as an average of 100 predictions. If a future value was outside the prediction interval for one of the series, then the observed coverage probability drops from 100 to 99%. For all criteria, the estimated intervals become more reliable as n increases. An explanation of this 'sample size effect' is that the confidence intervals used here do not reflect the uncertainty due to the estimation of the optimal mean squared prediction error. Also, the estimated intervals become less reliable as d increases, in particular, when using an inconsistent criterion such as the AIC. As an explanation note that the AIC typically tends to underestimate the actual value of d. Hence, the confidence intervals of the AIC are quite often too short. In contrast, small sample estimates of d are most reliable when the BIC is used. So that, in comparison, the results are slightly better using the BIC. Thus, as a rule of thumb, given a very small sample size of $n = 50$, the BIC should be used. For n=200 and short forecast horizons ($k \leq 10$), the observed coverage percentages are close to the nominal ones for all model selection criteria. For long-term forecasts, however, using the AIC or HIC appears to be less reliable, whereas BIC predictions yield, in comparison, slightly more realistic prediction intervals. Overall, prediction intervals for fractional models are better adapted to the data, when using the BIC.

The results of the relative empirical mean squared prediction errors (MSE) are given in tables 3.8 through 3.13. Observe that, there are only few cases where AIC and HIC yield substantially better results than the BIC. This is, in particular, most pronounced in the presence of (strong) positive short-range dependence which may point (together with $.2 \leq d \leq .6$) to a near unit-root behaviour. Consider, for instance, 'Model e'. Here, the autoregressive polynomial $\phi(B)$ is close to the unit root polynomial $1 - B$. This may explain why (near) stationary but strong short-range dependence is likely to be confounded with unit root behaviour. That is, the BIC tends, in contrast to the AIC, to overestimate d and yields, as a consequence, worse point-forecasts. A similar explanation holds for 'Model d'. Also, note that d is difficult to detect in 'Model c'. Since the AIC tends to choose large autoregressive orders (i.e. $p = 1$ instead of $p = 0$), the probability of detecting d is higher. In the nonstationary case $d \geq .8$, the most realistic one in modeling financial price series, using the BIC yields (overall) the most reliable forecasts in terms of the point-forecast accuracy for all models. This effect is seen to be particularly pronounced for longer forecasting horizons. Finally, it is found for the remaining cases that AIC-, HIC- and BIC-forecasts yield comparable results in terms of the MSE.

Comparison with the random walk

The theoretical disadvantages of misspecifying long-memory processes as conventional Box-Jenkins ARIMA models have recently been addressed by several authors (see e.g. Beran and Ocker [17], Crato and Taylor [33] and references therein). In practice, however, estimated Box-Jenkins ARIMA models turn out to perform competitively in many cases with the more sophisticated FARIMA forecasts (see Crato and Ray [32] and references therein). It should be noted, however, that the predominant criterion for judging the performance of predictions has been the accuracy of point-forecasts, measured, for instance, by the mean squared prediction error. This is not the only informative criterion. The purpose of a statistical prediction is not only to obtain a point estimate but also to have a confidence interval. Thus, an important criterion for assessing the usefulness of a statistical forecast is whether prediction intervals are realistic, i.e. neither too short nor too long.

We first turn to the question of how the precision of FARIMA point predictions compares to random walk forecasts, the benchmark in econometric modeling. Since we consider for the models a, c, d, e and f confidence intervals for short-range dependence too, these results are not comparable. A relative comparison of the 'Model b', however, may give an indication of the possible loss/gain in forecasting when using FARIMA forecasts. Table 3.14 compares the MSE of FARIMA predictions, when using the BIC, with those of the random walk forecasts. Clearly FARIMA forecasts are quite poor for $n = 50$ if the series is antipersistent or follows a unit root behaviour, but tend to perform better if the amount of persistence is strong (i.e. $d = 1.4$). Thus, in terms of the MSE, not much is gained in connection with $n = 50$ by using the more complicated FARIMA models. In contrast, given a sample size of $n = 200$, FARIMA predictions turn out to perform better or at least competitive for all ds and forecasting horizons $k \leq 10$. In particular, superior results for all lags are achieved for clearly persistent time series (i.e. $d = 1.4$).

In addition, consider the length of FARIMA prediction intervals. Figure 3.2 displays the ratio of the average length of FARIMA prediction intervals for model b with $n = 200$ and $d = .6, .8, 1.0, 1.2, 1.4$ divided by the length of random walk prediction intervals, plotted against k=1,2,...,120. FARIMA intervals turn out to be much shorter if the stationary series is antipersistent, in particular for large lags. The larger the amount of antipersistence, the shorter the relative length of the forecast intervals, while the coverage level of the FARIMA interval appears to be (almost) correct (see table 3.3). At the same time, the length of the confidence intervals of the random walk are unnecessarily wide if the stationary series is antipersistent or too narrow in the long memory case (see table 3.15). In this sense, predictions of FARIMA processes clearly outperform random walk forecasts.

Overall, in comparison to the random walk, we recommend the use of FARIMA forecasting models not only for clearly persistent time series but for antipersistent and moderately long memory behaviour too.

3.5 Concluding remarks

The main advantages of FARIMA models for forecasting may be summarized as follows:

1. In 'stationary versus unit root' approaches, a decision has to be made between $d = 0$ and $d = 1$. A wrong decision has an extreme impact on forecast intervals, since the length of forecast intervals is asymptotically constant for $d = 0$ whereas it diverges to infinity ate the rate $\sqrt{k}$ for $d = 1$. In contrast, for FARIMA models, prediction intervals are of order $O(k^{\tau/2})$ with τ varying in a continuous range, including $\tau = 0$ and $\tau = 1$ as special cases. The value of $\tau = max\{0, 2d-1\}$ is estimated from the data by maximum likelihood. As a result, prediction intervals are better adapted to the observed data, and neither too short (in the presence of long memory) nor too long (in the presence of antipersistence). The extreme decision between $O(1)$ and $O(k^{.5})$ is avoided.

2. The comparison of different model choice criteria suggests, overall, that using the BIC leads to more reliable point- and interval-forecasts. Also, in comparison to the random walk, FARIMA predictions turned out to be highly competitive, in terms of point- and interval-forecasts.

A number of open problems remain. Incorporating possible deviations from normality and in particular long-tailed distributions may improve the reliability of forecasts in practical applications. In addition, better and faster algorithms in theorem 4 may be developed. In our numerical calculations, inversion of Σ_n and calculation of β_{opt} turned out to require very precise evaluation of $\gamma(k)$ for all lags. After trying a number of approaches, Monte Carlo calculation of the covariances turned out to be most reliable. Interpreting $\gamma(k) = \int_{-\pi}^{\pi} f(x) \cos kx dx$ as $2\pi E[f(W) \cos kW]$, where W is uniformly distributed on $[-\pi, \pi]$, $\gamma(k)$ was obtained by 100000 simulations. Finally, more reliable confidence intervals may be obtained by incorporating the uncertainty due to the estimation of the optimal mean squared prediction error.

3.6 Appendix

Table 3.2: Empirical coverage percentages of the 95%-k-step ahead prediction intervals of 'Model a'

	AIC	HIC	BIC	AIC	HIC	BIC	AIC	HIC	BIC	AIC	HIC	BIC
		n=50			n=200			n=50			n=200	
d=		-.4			-.4			-.2			-.2	
k=1	0.89	0.90	0.90	0.93	0.93	0.94	0.90	0.92	0.93	0.93	0.93	0.93
2	0.94	0.96	0.95	0.94	0.95	0.95	0.87	0.95	0.96	0.93	0.92	0.93
3	0.91	0.92	0.92	0.94	0.92	0.93	0.95	0.98	0.98	0.93	0.94	0.93
4	0.85	0.87	0.87	0.94	0.94	0.94	0.93	0.94	0.96	0.94	0.95	0.95
5	0.85	0.89	0.89	0.95	0.95	0.95	0.88	0.89	0.93	0.93	0.93	0.93
10	0.82	0.84	0.85	0.98	0.98	0.98	0.85	0.88	0.91	0.98	0.98	0.97
15	0.84	0.87	0.88	0.96	0.97	0.96	0.90	0.90	0.90	0.94	0.95	0.95
20	0.84	0.84	0.85	0.94	0.93	0.93	0.87	0.88	0.88	0.92	0.91	0.92
30	0.89	0.90	0.91	0.92	0.92	0.93	0.91	0.92	0.92	0.96	0.96	0.96
40	0.91	0.91	0.91	0.92	0.92	0.92	0.92	0.92	0.92	0.89	0.89	0.89
50	0.87	0.86	0.87	0.94	0.94	0.94	0.87	0.87	0.87	0.97	0.97	0.97
d=		0			0			.2			.2	
k=1	0.82	0.84	0.85	0.91	0.91	0.92	0.89	0.93	0.95	0.93	0.91	0.92
2	0.92	0.93	0.92	0.92	0.93	0.92	0.86	0.92	0.92	0.92	0.92	0.92
3	0.92	0.92	0.93	0.93	0.93	0.95	0.87	0.92	0.93	0.90	0.89	0.89
4	0.89	0.90	0.91	0.92	0.94	0.93	0.90	0.95	0.96	0.91	0.91	0.92
5	0.92	0.92	0.93	0.94	0.94	0.94	0.93	0.97	0.97	0.90	0.90	0.92
10	0.88	0.87	0.88	0.92	0.93	0.92	0.85	0.94	0.94	0.92	0.93	0.93
15	0.86	0.87	0.87	0.93	0.93	0.93	0.91	0.93	0.93	0.95	0.96	0.95
20	0.86	0.87	0.87	0.91	0.91	0.92	0.90	0.92	0.93	0.93	0.93	0.93
30	0.86	0.86	0.86	0.89	0.89	0.90	0.89	0.90	0.91	0.90	0.91	0.91
40	0.92	0.91	0.93	0.94	0.94	0.94	0.89	0.90	0.91	0.90	0.90	0.90
50	0.89	0.88	0.88	0.92	0.92	0.92	0.93	0.93	0.93	0.94	0.94	0.93
d=		.4			.4			.6			.6	
k=1	0.88	0.90	0.89	0.92	0.93	0.92	0.94	0.92	0.93	0.97	0.97	0.96
2	0.88	0.89	0.91	0.93	0.92	0.92	0.87	0.88	0.88	0.92	0.91	0.91
3	0.91	0.94	0.94	0.94	0.95	0.95	0.90	0.92	0.93	0.93	0.92	0.93
4	0.91	0.92	0.93	0.93	0.92	0.92	0.91	0.94	0.94	0.90	0.92	0.92
5	0.95	0.95	0.95	0.88	0.88	0.90	0.88	0.89	0.90	0.90	0.91	0.92
10	0.89	0.89	0.91	0.93	0.92	0.92	0.87	0.90	0.90	0.92	0.92	0.92
15	0.83	0.82	0.85	0.96	0.96	0.96	0.81	0.82	0.84	0.94	0.94	0.94
20	0.88	0.88	0.89	0.93	0.94	0.93	0.83	0.90	0.90	0.94	0.95	0.94
30	0.89	0.91	0.92	0.90	0.90	0.92	0.86	0.90	0.89	0.90	0.89	0.91
40	0.88	0.89	0.90	0.94	0.94	0.94	0.85	0.83	0.84	0.92	0.92	0.92
50	0.88	0.88	0.91	0.85	0.85	0.86	0.88	0.86	0.87	0.91	0.91	0.92
d=		.8			.8			1.0			1.0	
k=1	0.89	0.92	0.92	0.92	0.93	0.93	0.92	0.94	0.95	0.92	0.94	0.94
2	0.88	0.89	0.90	0.90	0.90	0.90	0.88	0.90	0.92	0.94	0.93	0.94
3	0.86	0.91	0.92	0.95	0.97	0.97	0.84	0.88	0.90	0.95	0.95	0.96
4	0.83	0.87	0.89	0.95	0.95	0.95	0.84	0.87	0.88	0.93	0.95	0.95
5	0.84	0.86	0.84	0.92	0.93	0.91	0.79	0.85	0.87	0.93	0.92	0.93
10	0.83	0.87	0.89	0.90	0.91	0.93	0.84	0.89	0.95	0.92	0.95	0.94
15	0.77	0.82	0.83	0.88	0.90	0.90	0.75	0.79	0.84	0.86	0.89	0.89
20	0.77	0.83	0.83	0.87	0.86	0.90	0.75	0.80	0.85	0.92	0.92	0.94
30	0.77	0.83	0.83	0.89	0.89	0.94	0.58	0.63	0.74	0.90	0.91	0.90
40	0.70	0.72	0.76	0.86	0.85	0.88	0.64	0.69	0.72	0.86	0.90	0.93
50	0.58	0.63	0.65	0.86	0.87	0.90	0.59	0.65	0.70	0.87	0.88	0.92
d=		1.2			1.2			1.4			1.4	
k=1	0.89	0.90	0.91	0.93	0.92	0.92	0.84	0.88	0.90	0.94	0.96	0.96
2	0.88	0.90	0.92	0.95	0.92	0.94	0.82	0.86	0.88	0.95	0.96	0.95
3	0.82	0.87	0.88	0.91	0.93	0.93	0.86	0.91	0.95	0.92	0.92	0.92
4	0.86	0.88	0.88	0.94	0.93	0.94	0.79	0.85	0.86	0.94	0.94	0.95
5	0.80	0.84	0.86	0.89	0.90	0.90	0.83	0.88	0.92	0.92	0.93	0.94
10	0.70	0.76	0.82	0.84	0.91	0.90	0.65	0.76	0.83	0.89	0.90	0.91
15	0.60	0.66	0.72	0.86	0.90	0.91	0.62	0.72	0.80	0.86	0.88	0.89
20	0.52	0.62	0.70	0.81	0.87	0.88	0.57	0.69	0.73	0.85	0.88	0.89
30	0.52	0.58	0.66	0.77	0.83	0.84	0.55	0.65	0.70	0.76	0.85	0.90
40	0.42	0.50	0.61	0.71	0.78	0.80	0.50	0.66	0.72	0.71	0.79	0.85
50	0.42	0.53	0.61	0.67	0.79	0.84	0.49	0.61	0.67	0.72	0.77	0.86

Table 3.3: Empirical coverage percentages of the 95%-k-step ahead prediction intervals of 'Model b'

	AIC	HIC	BIC	AIC	HIC	BIC	AIC	HIC	BIC	AIC	HIC	BIC
		n=50			n=200			n =50			n=200	
d=		-.4			-.4			-.2			-.2	
k=1	0.91	0.92	0.92	0.95	0.95	0.95	0.95	0.95	0.95	0.94	0.94	0.94
2	0.87	0.89	0.90	0.98	0.99	1.00	0.94	0.94	0.94	0.93	0.93	0.93
3	0.93	0.95	0.95	0.93	0.94	0.94	0.92	0.93	0.93	0.90	0.91	0.92
4	0.94	0.95	0.95	0.95	0.96	0.96	0.91	0.92	0.93	0.95	0.95	0.95
5	0.92	0.93	0.94	0.97	0.97	0.97	0.91	0.93	0.93	0.92	0.92	0.92
10	0.91	0.93	0.92	0.97	0.97	0.97	0.90	0.92	0.91	0.95	0.96	0.96
15	0.95	0.96	0.96	0.96	0.95	0.95	0.97	0.97	0.97	0.98	0.98	0.98
20	0.92	0.92	0.92	0.96	0.96	0.96	0.96	0.97	0.97	0.92	0.92	0.92
30	0.93	0.94	0.94	0.92	0.92	0.92	0.92	0.92	0.91	0.93	0.93	0.93
40	0.93	0.93	0.94	0.94	0.94	0.95	0.96	0.97	0.96	0.94	0.94	0.94
50	0.93	0.93	0.93	0.97	0.96	0.97	0.97	0.97	0.97	0.96	0.96	0.96
d=		0			0			.2			.2	
k=1	0.89	0.91	0.92	0.94	0.95	0.95	0.89	0.90	0.91	0.93	0.95	0.95
2	0.92	0.93	0.93	0.90	0.91	0.92	0.95	0.96	0.97	0.94	0.94	0.94
3	0.96	0.95	0.96	0.97	0.97	0.98	0.89	0.91	0.96	0.98	0.98	0.98
4	0.93	0.94	0.95	0.96	0.96	0.96	0.93	0.93	0.95	0.91	0.92	0.92
5	0.95	0.96	0.97	0.93	0.93	0.94	0.88	0.90	0.91	0.97	0.96	0.96
10	0.92	0.95	0.94	0.97	0.98	0.97	0.89	0.90	0.91	0.93	0.93	0.93
15	0.92	0.92	0.92	0.98	0.98	0.97	0.92	0.91	0.92	0.97	0.96	0.96
20	0.94	0.94	0.95	0.90	0.89	0.89	0.92	0.92	0.96	0.95	0.95	0.95
30	0.95	0.95	0.94	0.94	0.96	0.94	0.94	0.94	0.95	0.95	0.94	0.94
40	0.93	0.93	0.93	0.95	0.95	0.96	0.95	0.96	0.96	0.97	0.96	0.97
50	0.95	0.95	0.95	0.96	0.96	0.96	0.95	0.95	0.97	0.97	0.98	0.97
d=		.4			.4			.6			.6	
k =1	0.85	0.88	0.92	0.96	0.96	0.97	0.92	0.95	0.95	0.94	0.94	0.95
2	0.87	0.89	0.88	0.94	0.93	0.93	0.86	0.88	0.86	0.91	0.93	0.93
3	0.92	0.93	0.95	0.88	0.90	0.89	0.87	0.89	0.90	0.96	0.96	0.95
4	0.93	0.95	0.96	0.95	0.95	0.95	0.80	0.81	0.83	0.91	0.90	0.89
5	0.93	0.94	0.95	0.93	0.93	0.94	0.82	0.85	0.86	0.90	0.90	0.91
10	0.91	0.91	0.92	0.94	0.94	0.94	0.71	0.71	0.73	0.89	0.90	0.91
15	0.84	0.86	0.87	0.97	0.97	0.97	0.74	0.74	0.77	0.95	0.95	0.94
20	0.85	0.87	0.88	0.87	0.88	0.89	0.66	0.66	0.69	0.89	0.89	0.91
30	0.81	0.81	0.82	0.92	0.91	0.92	0.71	0.72	0.74	0.86	0.89	0.87
40	0.91	0.94	0.93	0.93	0.94	0.94	0.65	0.65	0.67	0.83	0.87	0.88
50	0.84	0.87	0.87	0.90	0.92	0.92	0.63	0.64	0.65	0.78	0.80	0.80
d=		.8			.8			1.0			1.0	
k=1	0.87	0.88	0.89	0.96	0.96	0.96	0.86	0.90	0.93	0.93	0.95	0.94
2	0.86	0.92	0.92	0.93	0.92	0.94	0.85	0.88	0.87	0.88	0.91	0.91
3	0.88	0.90	0.90	0.93	0.92	0.94	0.82	0.85	0.86	0.92	0.92	0.93
4	0.82	0.83	0.85	0.93	0.94	0.95	0.84	0.87	0.87	0.93	0.95	0.95
5	0.79	0.85	0.87	0.94	0.95	0.96	0.81	0.85	0.87	0.94	0.95	0.94
10	0.70	0.76	0.78	0.87	0.89	0.91	0.70	0.72	0.78	0.89	0.92	0.95
15	0.67	0.70	0.74	0.83	0.84	0.89	0.62	0.69	0.76	0.82	0.89	0.91
20	0.71	0.73	0.75	0.86	0.88	0.90	0.56	0.67	0.74	0.82	0.89	0.93
30	0.60	0.66	0.66	0.82	0.84	0.91	0.49	0.51	0.61	0.71	0.79	0.83
40	0.56	0.59	0.66	0.78	0.83	0.88	0.46	0.46	0.58	0.68	0.78	0.83
50	0.55	0.56	0.59	0.81	0.79	0.85	0.42	0.44	0.55	0.69	0.75	0.82
d=		1.2			1.2			1.4			1.4	
k=1	0.86	0.91	0.93	0.91	0.93	0.93	0.88	0.91	0.93	0.95	0.94	0.95
2	0.87	0.92	0.92	0.92	0.93	0.94	0.78	0.83	0.88	0.94	0.93	0.93
3	0.87	0.90	0.93	0.91	0.94	0.94	0.76	0.81	0.88	0.91	0.92	0.93
4	0.82	0.87	0.89	0.93	0.95	0.97	0.75	0.81	0.87	0.92	0.91	0.92
5	0.72	0.82	0.86	0.89	0.91	0.94	0.73	0.79	0.83	0.87	0.90	0.90
10	0.64	0.75	0.81	0.90	0.93	0.95	0.63	0.68	0.77	0.86	0.88	0.89
15	0.61	0.74	0.80	0.84	0.86	0.91	0.56	0.63	0.73	0.81	0.84	0.87
20	0.53	0.71	0.79	0.81	0.84	0.90	0.47	0.54	0.66	0.72	0.77	0.83
30	0.49	0.62	0.74	0.84	0.84	0.91	0.42	0.50	0.59	0.71	0.76	0.83
40	0.45	0.58	0.70	0.75	0.78	0.87	0.40	0.45	0.54	0.67	0.76	0.82
50	0.40	0.57	0.67	0.74	0.78	0.86	0.37	0.40	0.52	0.66	0.73	0.81

Table 3.4: Empirical coverage percentages of the 95%-k-step ahead prediction intervals of 'Model c'

	AIC	HIC	BIC	AIC	HIC	BIC	AIC	HIC	BIC	AIC	HIC	BIC
		n=50			n=200			n=50			n=200	
d=		-.4			-.4			-.2			-.2	
k=1	0.91	0.96	0.97	0.94	0.95	0.95	0.88	0.94	0.93	0.93	0.93	0.93
2	0.89	0.94	0.95	0.94	0.95	0.95	0.90	0.92	0.92	0.94	0.94	0.94
3	0.95	0.94	0.96	0.92	0.92	0.92	0.92	0.94	0.93	0.94	0.93	0.94
4	0.90	0.91	0.91	0.97	0.97	0.97	0.92	0.95	0.96	0.91	0.91	0.92
5	0.95	0.97	0.97	0.95	0.95	0.96	0.92	0.94	0.95	0.96	0.97	0.97
10	0.94	0.93	0.94	0.95	0.94	0.94	0.94	0.95	0.97	0.91	0.91	0.92
15	0.98	0.97	0.98	0.96	0.96	0.96	0.96	0.98	0.98	0.91	0.91	0.91
20	0.94	0.94	0.94	0.94	0.94	0.94	0.95	0.93	0.93	0.98	0.98	0.98
30	0.95	0.95	0.95	0.94	0.94	0.95	0.94	0.94	0.95	0.98	0.98	0.98
40	0.95	0.95	0.95	0.94	0.94	0.95	0.88	0.88	0.89	0.95	0.95	0.95
50	0.95	0.95	0.95	0.97	0.96	0.97	0.94	0.94	0.93	0.92	0.91	0.92
d=		0			0			.2			.2	
k=1	0.92	0.93	0.91	0.92	0.92	0.93	0.88	0.87	0.89	0.96	0.96	0.96
2	0.89	0.92	0.93	0.94	0.95	0.95	0.88	0.89	0.90	0.96	0.97	0.97
3	0.89	0.90	0.90	0.94	0.94	0.95	0.86	0.90	0.90	0.97	0.97	0.97
4	0.91	0.91	0.92	0.96	0.97	0.96	0.93	0.94	0.94	0.92	0.92	0.92
5	0.94	0.94	0.95	0.92	0.93	0.93	0.88	0.90	0.92	0.96	0.96	0.96
10	0.95	0.96	0.96	0.93	0.92	0.93	0.83	0.84	0.86	0.90	0.91	0.93
15	0.94	0.94	0.94	0.96	0.96	0.96	0.90	0.90	0.91	0.91	0.91	0.92
20	0.94	0.93	0.93	0.92	0.92	0.91	0.93	0.94	0.94	0.91	0.90	0.90
30	0.94	0.94	0.94	0.95	0.96	0.96	0.90	0.91	0.92	0.94	0.94	0.94
40	0.95	0.95	0.95	0.96	0.96	0.96	0.92	0.93	0.93	0.94	0.94	0.95
50	0.93	0.92	0.92	0.95	0.96	0.95	0.91	0.90	0.92	0.93	0.93	0.95
d=		.4			.4			.6			.6	
k=1	0.90	0.93	0.92	0.95	0.95	0.95	0.88	0.91	0.92	0.91	0.91	0.91
2	0.87	0.90	0.89	0.94	0.93	0.93	0.81	0.83	0.87	0.95	0.94	0.94
3	0.85	0.89	0.90	0.95	0.96	0.95	0.82	0.81	0.81	0.91	0.92	0.94
4	0.82	0.86	0.88	0.94	0.96	0.96	0.82	0.84	0.87	0.91	0.91	0.90
5	0.83	0.86	0.87	0.91	0.93	0.94	0.81	0.84	0.84	0.89	0.87	0.87
10	0.85	0.86	0.86	0.94	0.94	0.95	0.77	0.79	0.81	0.89	0.88	0.88
15	0.86	0.88	0.92	0.93	0.92	0.92	0.75	0.80	0.83	0.88	0.86	0.89
20	0.86	0.87	0.89	0.93	0.91	0.91	0.70	0.74	0.77	0.88	0.88	0.90
30	0.77	0.81	0.81	0.95	0.95	0.97	0.76	0.78	0.78	0.88	0.84	0.90
40	0.80	0.81	0.81	0.91	0.91	0.93	0.69	0.67	0.69	0.80	0.82	0.87
50	0.79	0.83	0.84	0.90	0.91	0.91	0.71	0.74	0.74	0.77	0.79	0.82
d=		.8			.8			1.0			1.0	
k=1	0.86	0.91	0.93	0.89	0.90	0.90	0.89	0.93	0.94	0.96	0.96	0.96
2	0.81	0.88	0.89	0.90	0.90	0.90	0.87	0.93	0.92	0.95	0.96	0.96
3	0.75	0.79	0.82	0.92	0.92	0.93	0.84	0.88	0.89	0.95	0.96	0.95
4	0.77	0.81	0.86	0.96	0.94	0.96	0.79	0.80	0.81	0.91	0.97	0.95
5	0.76	0.79	0.86	0.96	0.96	0.97	0.79	0.83	0.85	0.92	0.92	0.92
10	0.74	0.76	0.84	0.90	0.91	0.90	0.64	0.72	0.78	0.91	0.93	0.94
15	0.62	0.71	0.79	0.86	0.88	0.89	0.63	0.72	0.76	0.83	0.87	0.90
20	0.64	0.67	0.77	0.80	0.85	0.87	0.54	0.65	0.74	0.78	0.84	0.89
30	0.62	0.66	0.75	0.81	0.83	0.85	0.44	0.57	0.69	0.77	0.83	0.88
40	0.57	0.61	0.70	0.76	0.79	0.84	0.50	0.63	0.69	0.69	0.80	0.86
50	0.50	0.58	0.64	0.75	0.79	0.83	0.44	0.54	0.59	0.67	0.76	0.83
d=		1.2			1.2			1.4			1.4	
k=1	0.85	0.86	0.91	0.95	0.95	0.95	0.84	0.87	0.88	0.94	0.94	0.94
2	0.81	0.86	0.88	0.98	0.98	0.98	0.79	0.81	0.85	0.93	0.94	0.94
3	0.77	0.81	0.87	0.95	0.94	0.94	0.80	0.84	0.88	0.90	0.91	0.90
4	0.80	0.84	0.89	0.97	0.98	0.98	0.80	0.85	0.86	0.86	0.87	0.88
5	0.80	0.81	0.85	0.96	0.96	0.96	0.76	0.80	0.84	0.87	0.90	0.91
10	0.66	0.70	0.80	0.90	0.90	0.91	0.56	0.64	0.71	0.87	0.89	0.87
15	0.60	0.70	0.76	0.84	0.83	0.85	0.49	0.54	0.64	0.87	0.89	0.87
20	0.57	0.62	0.71	0.78	0.76	0.80	0.42	0.48	0.58	0.81	0.84	0.81
30	0.52	0.56	0.65	0.79	0.81	0.85	0.32	0.39	0.49	0.69	0.74	0.76
40	0.46	0.50	0.62	0.75	0.77	0.85	0.28	0.35	0.44	0.67	0.70	0.76
50	0.42	0.48	0.59	0.70	0.76	0.83	0.32	0.40	0.47	0.62	0.66	0.71

Table 3.5: Empirical coverage percentages of the 95%-k-step ahead prediction intervals of 'Model d'

	AIC	HIC	BIC	AIC	HIC	BIC	AIC	HIC	BIC	AIC	HIC	BIC
		n=50			n=200			n=50			n=200	
d=		-.4			-.4			-.2			-.2	
k=1	0.90	0.93	0.94	0.94	0.93	0.93	0.86	0.87	0.91	0.96	0.96	0.96
2	0.89	0.91	0.92	0.98	0.99	0.99	0.90	0.93	0.95	0.95	0.95	0.94
3	0.90	0.92	0.94	0.91	0.92	0.93	0.93	0.93	0.96	0.99	0.99	0.99
4	0.91	0.91	0.91	0.96	0.96	0.96	0.90	0.90	0.92	0.95	0.96	0.95
5	0.91	0.91	0.92	0.89	0.90	0.90	0.88	0.89	0.94	0.92	0.94	0.94
10	0.88	0.89	0.89	0.95	0.96	0.95	0.93	0.95	0.95	0.95	0.95	0.95
15	0.96	0.97	0.97	0.96	0.96	0.96	0.91	0.91	0.90	0.96	0.96	0.95
20	0.91	0.92	0.92	0.94	0.95	0.95	0.96	0.96	0.97	0.92	0.92	0.92
30	0.90	0.91	0.92	0.93	0.93	0.92	0.93	0.93	0.93	0.96	0.96	0.96
40	0.95	0.94	0.95	0.94	0.94	0.94	0.90	0.91	0.91	0.94	0.94	0.94
50	0.97	0.97	0.97	0.96	0.96	0.96	0.96	0.96	0.96	0.94	0.94	0.94
d=		0			0			.2			.2	
k=1	0.89	0.92	0.93	0.91	0.92	0.92	0.94	0.95	0.95	0.99	0.99	0.97
2	0.92	0.92	0.93	0.93	0.95	0.94	0.87	0.93	0.94	0.95	0.94	0.95
3	0.90	0.94	0.94	0.93	0.94	0.92	0.90	0.91	0.94	0.92	0.92	0.92
4	0.89	0.90	0.89	0.92	0.92	0.91	0.85	0.88	0.91	0.89	0.90	0.89
5	0.91	0.93	0.92	0.94	0.95	0.94	0.82	0.85	0.86	0.90	0.90	0.91
10	0.91	0.90	0.91	0.91	0.91	0.90	0.87	0.87	0.91	0.95	0.95	0.94
15	0.90	0.92	0.93	0.97	0.97	0.98	0.90	0.91	0.92	0.93	0.94	0.94
20	0.90	0.88	0.89	0.89	0.91	0.90	0.83	0.86	0.87	0.93	0.92	0.93
30	0.90	0.91	0.92	0.91	0.91	0.91	0.86	0.88	0.90	0.94	0.94	0.94
40	0.94	0.93	0.93	0.90	0.91	0.87	0.82	0.83	0.81	0.94	0.94	0.95
50	0.89	0.89	0.88	0.93	0.93	0.92	0.79	0.78	0.80	0.92	0.91	0.92
d=		.4			.4			.6			.6	
k=1	0.86	0.88	0.89	0.95	0.94	0.96	0.83	0.88	0.87	0.94	0.95	0.95
2	0.88	0.86	0.86	0.95	0.95	0.95	0.82	0.82	0.83	0.96	0.97	0.97
3	0.90	0.95	0.94	0.96	0.94	0.95	0.78	0.79	0.81	0.92	0.91	0.92
4	0.79	0.84	0.87	0.94	0.93	0.95	0.78	0.80	0.84	0.89	0.90	0.91
5	0.73	0.79	0.81	0.96	0.95	0.96	0.76	0.81	0.86	0.88	0.88	0.89
10	0.82	0.84	0.85	0.92	0.92	0.90	0.68	0.71	0.78	0.89	0.92	0.91
15	0.85	0.85	0.88	0.97	0.96	0.97	0.63	0.65	0.69	0.83	0.83	0.85
20	0.82	0.83	0.83	0.94	0.95	0.95	0.65	0.68	0.76	0.85	0.89	0.90
30	0.77	0.78	0.78	0.93	0.94	0.94	0.71	0.74	0.77	0.86	0.89	0.91
40	0.78	0.77	0.80	0.97	0.97	0.98	0.63	0.67	0.73	0.79	0.84	0.86
50	0.76	0.77	0.79	0.90	0.90	0.89	0.60	0.61	0.72	0.83	0.87	0.89
d=		.8			.8			1.0			1.0	
k=1	0.89	0.91	0.92	0.92	0.92	0.93	0.92	0.92	0.93	0.91	0.91	0.92
2	0.84	0.87	0.88	0.92	0.92	0.94	0.92	0.92	0.94	0.93	0.94	0.95
3	0.77	0.80	0.82	0.87	0.87	0.88	0.87	0.87	0.88	0.99	0.98	0.99
4	0.77	0.79	0.82	0.89	0.90	0.88	0.89	0.90	0.88	0.97	0.96	0.96
5	0.74	0.76	0.80	0.92	0.95	0.96	0.92	0.95	0.96	0.94	0.95	0.96
10	0.62	0.67	0.74	0.86	0.87	0.88	0.86	0.87	0.88	0.91	0.91	0.92
15	0.59	0.68	0.79	0.83	0.85	0.88	0.83	0.85	0.88	0.87	0.91	0.93
20	0.53	0.62	0.73	0.85	0.84	0.87	0.85	0.84	0.87	0.82	0.87	0.89
30	0.55	0.66	0.74	0.76	0.77	0.83	0.76	0.77	0.83	0.73	0.77	0.84
40	0.60	0.64	0.75	0.75	0.78	0.83	0.75	0.78	0.83	0.75	0.79	0.86
50	0.48	0.57	0.70	0.69	0.74	0.80	0.69	0.74	0.80	0.74	0.79	0.85
d=		1.2			1.2			1.4			1.4	
k=1	0.87	0.91	0.90	0.96	0.97	0.96	0.87	0.89	0.90	0.94	0.94	0.94
2	0.81	0.82	0.81	0.96	0.96	0.95	0.81	0.85	0.87	0.91	0.91	0.91
3	0.76	0.80	0.79	0.94	0.94	0.94	0.80	0.83	0.84	0.91	0.90	0.90
4	0.74	0.79	0.81	0.94	0.94	0.94	0.77	0.80	0.80	0.89	0.89	0.90
5	0.69	0.73	0.75	0.93	0.93	0.93	0.75	0.79	0.80	0.91	0.89	0.89
10	0.56	0.60	0.63	0.83	0.84	0.86	0.64	0.69	0.72	0.88	0.88	0.90
15	0.49	0.59	0.60	0.80	0.82	0.83	0.50	0.59	0.61	0.85	0.85	0.88
20	0.42	0.49	0.56	0.72	0.75	0.77	0.45	0.54	0.57	0.83	0.84	0.87
30	0.41	0.48	0.56	0.66	0.70	0.71	0.40	0.47	0.52	0.74	0.77	0.81
40	0.35	0.45	0.53	0.60	0.67	0.69	0.33	0.41	0.45	0.69	0.70	0.72
50	0.31	0.36	0.48	0.59	0.62	0.66	0.26	0.34	0.39	0.64	0.67	0.67

Table 3.6: Empirical coverage percentages of the 95%-k-step ahead prediction intervals of 'Model e'

	AIC	HIC	BIC	AIC	HIC	BIC	AIC	HIC	BIC	AIC	HIC	BIC
		n=50			n=200			n=50			n=200	
d=		-.4			-.4			-.2			-.2	
k=1	0.88	0.91	0.92	0.89	0.90	0.90	0.91	0.94	0.96	0.88	0.89	0.89
2	0.85	0.86	0.89	0.90	0.90	0.91	0.92	0.95	0.94	0.88	0.89	0.89
3	0.88	0.90	0.91	0.97	0.97	0.97	0.91	0.92	0.93	0.84	0.84	0.85
4	0.85	0.88	0.90	0.93	0.93	0.94	0.89	0.93	0.92	0.88	0.89	0.88
5	0.86	0.89	0.89	0.96	0.97	0.97	0.87	0.91	0.94	0.92	0.93	0.94
10	0.78	0.80	0.81	0.94	0.94	0.95	0.87	0.89	0.88	0.95	0.96	0.95
15	0.89	0.90	0.90	0.93	0.93	0.94	0.79	0.85	0.85	0.92	0.93	0.92
20	0.86	0.89	0.89	0.95	0.95	0.94	0.77	0.75	0.76	0.93	0.93	0.93
30	0.91	0.93	0.91	0.95	0.95	0.95	0.88	0.88	0.88	0.91	0.90	0.90
40	0.85	0.86	0.84	0.92	0.93	0.93	0.83	0.84	0.82	0.94	0.92	0.95
50	0.86	0.88	0.86	0.94	0.93	0.92	0.77	0.78	0.81	0.89	0.88	0.87
d=		0			0			.2			.2	
k=1	0.87	0.90	0.90	0.94	0.95	0.95	0.88	0.87	0.89	0.93	0.93	0.94
2	0.89	0.91	0.91	0.95	0.96	0.96	0.85	0.87	0.90	0.92	0.94	0.93
3	0.84	0.87	0.90	0.92	0.92	0.92	0.80	0.82	0.84	0.93	0.92	0.92
4	0.83	0.87	0.87	0.95	0.95	0.96	0.80	0.82	0.85	0.93	0.91	0.90
5	0.81	0.82	0.88	0.98	0.98	0.98	0.80	0.82	0.85	0.91	0.89	0.89
10	0.68	0.70	0.77	0.91	0.92	0.91	0.71	0.74	0.75	0.83	0.84	0.88
15	0.64	0.70	0.75	0.91	0.92	0.92	0.72	0.73	0.81	0.80	0.82	0.85
20	0.62	0.71	0.77	0.92	0.92	0.93	0.70	0.70	0.76	0.85	0.91	0.92
30	0.77	0.79	0.83	0.93	0.91	0.92	0.63	0.68	0.75	0.86	0.89	0.89
40	0.67	0.69	0.73	0.94	0.95	0.95	0.65	0.70	0.75	0.86	0.85	0.88
50	0.75	0.74	0.75	0.90	0.89	0.90	0.66	0.71	0.76	0.90	0.91	0.92
d=		.4			.4			.6			.6	
k=1	0.91	0.93	0.91	0.97	0.97	0.99	0.84	0.89	0.91	0.97	0.97	0.98
2	0.84	0.85	0.85	0.96	0.96	0.96	0.79	0.86	0.87	0.92	0.93	0.93
3	0.81	0.83	0.84	0.94	0.94	0.94	0.70	0.77	0.78	0.91	0.91	0.91
4	0.82	0.82	0.83	0.92	0.92	0.94	0.72	0.80	0.79	0.90	0.90	0.90
5	0.77	0.78	0.81	0.89	0.89	0.94	0.71	0.78	0.80	0.88	0.90	0.90
10	0.70	0.71	0.76	0.86	0.89	0.94	0.61	0.72	0.76	0.83	0.83	0.88
15	0.61	0.65	0.72	0.83	0.87	0.91	0.50	0.63	0.70	0.85	0.88	0.89
20	0.61	0.64	0.72	0.80	0.84	0.88	0.46	0.58	0.66	0.84	0.86	0.89
30	0.53	0.60	0.68	0.84	0.86	0.89	0.50	0.59	0.64	0.74	0.78	0.80
40	0.51	0.57	0.68	0.84	0.87	0.91	0.50	0.55	0.61	0.70	0.77	0.77
50	0.50	0.55	0.67	0.84	0.85	0.90	0.49	0.56	0.61	0.64	0.73	0.76
d=		.8			.8			1.0			1.0	
k=1	0.88	0.89	0.91	0.92	0.92	0.94	0.91	0.92	0.92	0.93	0.93	0.94
2	0.89	0.89	0.91	0.90	0.90	0.91	0.84	0.85	0.88	0.93	0.95	0.95
3	0.80	0.85	0.84	0.89	0.88	0.90	0.74	0.78	0.81	0.96	0.94	0.95
4	0.74	0.78	0.81	0.91	0.89	0.89	0.70	0.73	0.76	0.94	0.93	0.95
5	0.73	0.76	0.78	0.90	0.88	0.89	0.67	0.70	0.72	0.94	0.92	0.94
10	0.58	0.62	0.66	0.85	0.86	0.88	0.57	0.57	0.60	0.95	0.93	0.94
15	0.54	0.56	0.61	0.86	0.86	0.90	0.50	0.54	0.55	0.87	0.84	0.86
20	0.44	0.46	0.56	0.78	0.78	0.84	0.42	0.45	0.47	0.84	0.85	0.89
30	0.46	0.47	0.54	0.74	0.76	0.82	0.27	0.34	0.43	0.73	0.75	0.84
40	0.39	0.42	0.52	0.70	0.76	0.80	0.22	0.28	0.36	0.71	0.72	0.81
50	0.40	0.43	0.53	0.60	0.66	0.71	0.22	0.29	0.35	0.77	0.75	0.82
d=		1.2			1.2			1.4			1.4	
k=1	0.89	0.89	0.90	0.95	0.95	0.95	0.84	0.88	0.87	0.92	0.93	0.93
2	0.86	0.85	0.86	0.96	0.94	0.93	0.76	0.80	0.80	0.94	0.94	0.96
3	0.78	0.78	0.78	0.95	0.95	0.94	0.71	0.73	0.75	0.94	0.94	0.94
4	0.72	0.72	0.72	0.93	0.94	0.94	0.69	0.71	0.71	0.95	0.94	0.94
5	0.64	0.67	0.69	0.96	0.96	0.95	0.62	0.65	0.66	0.94	0.94	0.94
10	0.53	0.54	0.58	0.93	0.91	0.94	0.55	0.53	0.54	0.85	0.86	0.86
15	0.43	0.46	0.50	0.88	0.86	0.89	0.44	0.47	0.46	0.79	0.79	0.81
20	0.36	0.39	0.42	0.83	0.86	0.88	0.37	0.37	0.40	0.75	0.80	0.82
30	0.30	0.29	0.30	0.78	0.81	0.82	0.26	0.28	0.28	0.74	0.73	0.74
40	0.28	0.29	0.31	0.67	0.68	0.73	0.23	0.24	0.26	0.65	0.65	0.68
50	0.29	0.30	0.31	0.61	0.63	0.68	0.22	0.22	0.25	0.60	0.62	0.64

Table 3.7: Empirical coverage percentages of the 95%-k-step ahead prediction intervals of 'Model f'

	AIC	HIC	BIC	AIC	HIC	BIC	AIC	HIC	BIC	AIC	HIC	BIC
		n=50			n=200			n=50			n=200	
d=		-.4			-.4			-.2			-.2	
k=1	0.89	0.95	0.95	0.97	0.97	0.97	0.86	0.88	0.90	0.90	0.93	0.93
2	0.79	0.86	0.89	0.93	0.92	0.93	0.84	0.85	0.87	0.89	0.90	0.90
3	0.82	0.85	0.86	0.90	0.91	0.91	0.85	0.85	0.87	0.95	0.94	0.94
4	0.84	0.88	0.88	0.96	0.96	0.96	0.85	0.90	0.91	0.93	0.93	0.93
5	0.86	0.90	0.90	0.93	0.93	0.93	0.86	0.87	0.89	0.91	0.91	0.91
10	0.91	0.92	0.93	0.95	0.95	0.95	0.90	0.92	0.94	0.93	0.93	0.94
15	0.86	0.89	0.90	0.98	0.98	0.99	0.87	0.90	0.91	0.91	0.90	0.90
20	0.88	0.90	0.91	0.96	0.95	0.96	0.83	0.86	0.87	0.96	0.95	0.96
30	0.87	0.88	0.88	0.93	0.92	0.92	0.94	0.94	0.94	0.96	0.96	0.96
40	0.89	0.93	0.90	0.95	0.95	0.95	0.88	0.90	0.92	0.90	0.91	0.91
50	0.90	0.91	0.92	0.94	0.93	0.94	0.84	0.83	0.83	0.98	0.97	0.96
d=		0			0			.2			.2	
k=1	0.90	0.91	0.91	0.97	0.97	0.96	0.84	0.87	0.88	0.97	0.97	0.96
2	0.84	0.86	0.88	0.96	0.96	0.96	0.87	0.90	0.90	0.98	0.98	0.97
3	0.78	0.79	0.80	0.94	0.94	0.93	0.85	0.88	0.89	0.98	0.98	1.00
4	0.82	0.83	0.83	0.93	0.93	0.93	0.81	0.81	0.85	0.96	0.96	0.96
5	0.91	0.92	0.93	0.92	0.93	0.94	0.84	0.82	0.85	0.95	0.96	0.94
10	0.90	0.92	0.93	0.91	0.90	0.90	0.81	0.85	0.88	0.86	0.85	0.85
15	0.92	0.92	0.92	0.90	0.90	0.92	0.89	0.89	0.90	0.89	0.90	0.90
20	0.95	0.96	0.96	0.93	0.93	0.93	0.90	0.90	0.92	0.93	0.93	0.92
30	0.94	0.95	0.97	0.95	0.96	0.96	0.88	0.89	0.90	0.94	0.93	0.94
40	0.95	0.95	0.95	0.95	0.95	0.95	0.87	0.87	0.88	0.92	0.91	0.92
50	0.94	0.93	0.95	0.96	0.96	0.96	0.89	0.91	0.92	0.96	0.97	0.97
d=		.4			.4			.6			.6	
k=1	0.90	0.92	0.92	0.94	0.93	0.94	0.90	0.92	0.93	0.93	0.93	0.94
2	0.84	0.84	0.85	0.92	0.92	0.95	0.89	0.88	0.89	0.93	0.93	0.95
3	0.81	0.84	0.86	0.92	0.92	0.91	0.82	0.86	0.85	0.93	0.93	0.95
4	0.80	0.81	0.83	0.94	0.94	0.95	0.76	0.78	0.82	0.92	0.94	0.96
5	0.80	0.83	0.83	0.89	0.90	0.91	0.68	0.72	0.75	0.92	0.93	0.96
10	0.79	0.82	0.85	0.94	0.92	0.93	0.77	0.83	0.85	0.91	0.89	0.91
15	0.84	0.82	0.85	0.90	0.92	0.92	0.79	0.81	0.84	0.84	0.86	0.89
20	0.85	0.85	0.86	0.92	0.93	0.93	0.67	0.74	0.78	0.87	0.89	0.89
30	0.84	0.86	0.87	0.91	0.90	0.90	0.69	0.75	0.77	0.84	0.88	0.89
40	0.88	0.89	0.90	0.95	0.95	0.95	0.65	0.67	0.69	0.83	0.83	0.85
50	0.88	0.90	0.91	0.92	0.91	0.96	0.69	0.70	0.72	0.85	0.85	0.88
d=		.8			.8			1.0			1.0	
k=1	0.86	0.89	0.91	0.91	0.91	0.92	0.92	0.93	0.95	0.92	0.93	0.94
2	0.86	0.86	0.88	0.92	0.95	0.96	0.89	0.91	0.97	0.94	0.94	0.94
3	0.84	0.86	0.88	0.94	0.94	0.96	0.80	0.84	0.87	0.92	0.91	0.93
4	0.81	0.84	0.86	0.92	0.91	0.93	0.77	0.79	0.85	0.90	0.90	0.92
5	0.80	0.83	0.87	0.92	0.92	0.92	0.74	0.77	0.82	0.87	0.92	0.92
10	0.83	0.83	0.85	0.88	0.89	0.87	0.69	0.73	0.85	0.91	0.93	0.93
15	0.69	0.78	0.80	0.86	0.88	0.89	0.56	0.61	0.72	0.82	0.84	0.87
20	0.72	0.80	0.83	0.88	0.89	0.89	0.60	0.61	0.71	0.76	0.82	0.83
30	0.64	0.67	0.73	0.77	0.79	0.80	0.45	0.50	0.63	0.74	0.82	0.86
40	0.51	0.57	0.61	0.84	0.86	0.86	0.45	0.47	0.58	0.71	0.79	0.84
50	0.57	0.61	0.64	0.72	0.75	0.78	0.42	0.44	0.59	0.66	0.76	0.78
d=		1.2			1.2			1.4			1.4	
k=1	0.88	0.90	0.90	0.91	0.93	0.93	0.86	0.86	0.86	0.92	0.92	0.90
2	0.84	0.89	0.90	0.92	0.90	0.90	0.79	0.79	0.80	0.88	0.90	0.89
3	0.82	0.87	0.89	0.93	0.92	0.94	0.78	0.79	0.78	0.90	0.91	0.91
4	0.80	0.85	0.87	0.92	0.94	0.95	0.77	0.77	0.79	0.90	0.91	0.92
5	0.74	0.79	0.83	0.92	0.94	0.93	0.72	0.75	0.76	0.90	0.91	0.92
10	0.60	0.70	0.76	0.89	0.91	0.91	0.58	0.61	0.66	0.85	0.86	0.87
15	0.60	0.71	0.74	0.82	0.87	0.88	0.55	0.54	0.60	0.80	0.83	0.87
20	0.54	0.66	0.72	0.81	0.84	0.89	0.51	0.52	0.59	0.80	0.82	0.83
30	0.50	0.60	0.66	0.75	0.85	0.88	0.41	0.44	0.52	0.73	0.78	0.79
40	0.43	0.55	0.59	0.71	0.81	0.84	0.36	0.38	0.50	0.74	0.80	0.84
50	0.41	0.52	0.58	0.64	0.75	0.82	0.37	0.40	0.52	0.76	0.81	0.89

Table 3.8: Relative empirical mean squared prediction errors of k-step ahead predictions of 'Model a'

	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC
	n=50		n=200		n=50		n=200	
d =	-.4		-.4		-.2		-.2	
k=1	0.983	0.991	1.031	1.000	1.004	1.034	1.062	1.011
2	1.027	1.009	1.011	0.989	1.116	1.045	1.076	1.013
3	1.006	1.006	1.000	0.990	1.166	1.021	1.073	1.014
4	1.051	1.008	1.000	0.990	1.129	1.015	1.029	1.011
5	1.061	1.007	0.979	0.983	1.128	1.037	0.997	0.989
10	1.035	1.018	0.992	0.973	1.145	1.057	1.014	0.999
15	1.041	1.020	1.028	1.004	1.062	1.024	1.005	0.989
20	1.010	1.008	1.012	1.010	1.042	0.999	0.995	1.005
30	0.987	0.993	1.004	0.998	1.009	0.999	1.016	1.005
40	1.004	0.987	0.995	0.999	1.022	1.019	0.999	1.007
50	0.992	0.997	0.991	1.001	0.993	1.009	1.002	1.001
d =	0		0		.2		.2	
k=1	1.020	1.007	1.017	1.000	1.178	1.048	1.000	1.025
2	1.023	1.000	1.029	1.014	1.132	1.007	1.028	1.006
3	0.987	1.017	1.011	1.000	1.175	1.053	1.038	1.023
4	0.946	1.003	1.000	0.991	1.142	1.017	1.032	1.007
5	1.014	1.008	0.988	0.993	1.127	1.033	1.030	1.021
10	1.015	1.021	0.981	1.014	1.156	1.012	1.015	1.002
15	1.034	1.016	1.025	1.007	1.030	1.016	1.002	1.005
20	1.014	1.015	1.007	0.996	1.057	1.027	1.002	1.022
30	1.014	1.011	1.012	0.997	1.071	1.034	0.988	0.994
40	0.993	0.996	1.027	1.007	1.054	1.031	0.986	0.995
50	0.994	0.991	1.014	0.998	1.019	0.996	1.011	1.000
d =	.4		.4		.6		.6	
k=1	1.054	0.955	1.010	1.000	0.990	0.977	1.010	0.990
2	1.050	0.993	1.007	1.007	1.081	1.002	1.036	1.000
3	1.036	0.994	1.005	1.016	0.960	0.990	1.000	1.016
4	1.063	1.038	1.005	1.005	1.152	0.991	1.005	0.986
5	1.000	1.031	0.997	1.010	1.000	1.013	1.041	1.021
10	0.996	1.007	0.984	0.995	0.976	0.971	1.011	1.004
15	1.021	1.013	1.019	0.997	0.959	0.995	1.049	1.021
20	1.046	1.017	0.997	0.997	1.073	1.003	1.075	1.007
30	1.069	0.992	1.011	1.008	1.067	0.996	1.052	1.058
40	1.000	1.003	1.003	1.019	0.905	1.002	1.036	1.036
50	1.009	0.994	1.014	1.014	0.907	0.957	0.995	0.997
d =	.8		.8		1.0		1.0	
k=1	1.073	0.964	1.036	1.018	1.067	0.999	1.007	0.991
2	0.984	1.008	1.016	1.000	1.275	1.167	0.972	0.974
3	1.095	0.990	1.025	0.994	1.119	1.034	1.051	1.020
4	1.066	1.043	1.016	1.000	1.268	1.134	1.074	0.995
5	1.131	1.010	1.015	0.974	1.115	0.995	1.045	1.034
10	1.062	1.031	1.036	1.003	1.379	1.178	1.001	1.014
15	1.045	0.974	1.002	0.964	1.182	1.056	1.029	1.009
20	1.010	1.015	1.027	1.004	1.205	1.062	0.914	0.977
30	0.996	0.968	0.955	1.016	1.085	1.009	0.884	0.974
40	1.001	1.004	0.970	1.020	1.033	0.981	1.002	1.005
50	0.985	0.980	0.901	0.971	1.029	0.971	1.016	1.021
d =	1.2		1.2		1.4		1.4	
k=1	1.087	1.039	1.029	1.010	1.164	1.112	1.010	1.000
2	1.104	1.061	1.009	1.009	1.184	1.078	1.007	1.014
3	1.153	1.009	1.037	1.021	1.394	1.117	1.067	1.049
4	1.167	1.058	1.042	1.008	1.372	1.064	1.022	1.017
5	1.102	1.025	1.092	1.000	1.476	1.146	1.049	1.037
10	1.388	1.085	1.123	1.002	1.621	1.100	1.097	1.094
15	1.407	1.152	1.258	1.002	1.673	1.105	1.203	1.142
20	1.360	1.143	1.304	1.010	1.711	1.067	1.216	1.136
30	1.349	1.156	1.368	1.003	1.654	1.013	1.354	1.172
40	1.202	1.096	1.324	0.972	1.591	1.008	1.375	1.199
50	1.143	1.059	1.303	0.970	1.472	1.015	1.346	1.196

Table 3.9: Relative empirical mean squared prediction errors of k-step ahead predictions of 'Model b'

	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC
	n=50		n=200		n=50		n=200	
d =	-.4		-.4		-.2		-.2	
k=1	0.995	1.013	1.026	1.010	0.983	0.942	0.991	0.999
2	1.206	1.049	1.082	1.018	0.941	0.966	1.006	1.020
3	1.105	0.999	1.047	1.028	0.982	0.987	1.023	0.996
4	1.097	1.007	1.034	1.017	1.023	1.042	1.014	1.005
5	1.174	1.009	0.986	1.021	1.115	1.016	0.996	1.007
10	1.042	1.018	1.022	1.001	0.970	0.988	0.993	1.005
15	1.051	1.008	0.986	1.002	0.983	0.997	0.989	0.988
20	1.005	1.006	1.007	1.002	1.017	1.015	1.008	1.005
30	1.022	1.007	1.006	0.997	1.018	0.995	0.984	0.985
40	1.028	1.005	1.005	1.000	0.961	0.978	1.008	0.999
50	0.992	0.995	1.009	1.008	1.020	1.015	1.005	0.997
d =	0		0		.2		.2	
k=1	1.076	1.056	1.009	0.999	1.023	1.067	0.993	0.990
2	1.008	0.996	1.029	0.998	1.280	1.240	1.001	0.998
3	1.066	1.047	1.025	1.009	1.185	1.150	1.032	0.992
4	1.041	0.996	0.985	1.004	1.238	1.095	1.034	1.031
5	1.042	1.015	0.996	1.014	1.116	1.057	1.000	1.018
10	1.016	1.002	1.005	1.010	1.011	1.030	1.008	0.998
15	1.029	1.007	0.989	0.999	1.036	1.059	0.994	1.010
20	1.059	1.044	1.004	1.007	1.109	1.101	1.000	0.992
30	0.962	0.998	0.998	0.984	1.205	1.134	0.997	1.009
40	1.014	1.039	0.983	0.997	1.120	1.036	0.997	1.003
50	0.988	0.988	0.984	0.991	1.139	1.109	1.028	0.976
d =	.4		.4		.6		.6	
k=1	1.105	1.009	1.013	1.017	1.135	1.030	1.019	1.000
2	1.097	1.016	0.980	0.982	1.056	1.056	1.056	0.979
3	0.942	0.997	1.009	0.998	0.980	1.052	0.937	0.949
4	0.988	1.017	1.007	1.003	1.001	1.028	0.952	0.956
5	1.065	1.023	1.006	0.989	1.010	0.986	0.998	0.995
10	1.099	1.013	1.084	1.030	1.015	1.015	1.034	1.018
15	1.059	0.963	1.050	1.030	0.994	1.007	0.983	0.966
20	1.035	1.007	1.019	1.012	0.980	1.002	1.067	0.965
30	0.997	1.011	1.040	1.016	0.980	0.979	0.942	0.928
40	0.988	0.921	1.055	1.000	0.974	0.971	1.002	0.881
50	1.115	0.937	1.061	1.007	0.971	0.960	0.952	0.872
d =	.8		.8		1.0		1.0	
k=1	0.998	0.977	1.049	1.022	1.106	1.058	1.079	1.032
2	1.032	0.982	1.040	1.000	1.152	1.091	1.065	0.997
3	1.085	1.012	0.971	0.992	1.153	1.025	1.064	0.970
4	1.096	1.020	1.047	1.011	1.116	0.988	1.134	0.989
5	1.178	1.025	1.114	1.060	1.185	1.036	1.154	1.004
10	1.076	0.951	1.053	0.988	1.234	1.167	1.289	1.065
15	1.029	0.971	1.081	1.044	1.253	1.246	1.336	1.057
20	0.954	0.935	1.120	1.022	1.137	1.112	1.418	1.118
30	0.959	0.924	1.112	1.022	1.081	1.079	1.257	1.100
40	0.934	0.933	1.126	1.026	1.024	1.011	1.212	1.088
50	0.902	0.921	1.031	1.008	0.994	0.967	1.167	1.068
d =	1.2		1.2		1.4		1.4	
k=1	1.292	1.049	1.052	1.023	1.176	1.059	0.977	0.994
2	1.221	1.101	1.118	1.013	1.254	1.176	0.966	0.996
3	1.280	1.124	1.126	1.035	1.179	1.145	1.005	0.994
4	1.326	1.097	1.212	1.088	1.208	1.139	1.023	0.990
5	1.412	1.123	1.218	1.105	1.293	1.142	1.047	0.993
10	1.605	1.188	1.265	1.178	1.440	1.199	1.190	1.072
15	1.627	1.176	1.224	1.174	1.486	1.262	1.272	1.148
20	1.579	1.112	1.202	1.166	1.473	1.216	1.298	1.177
30	1.418	1.133	1.220	1.190	1.403	1.138	1.394	1.248
40	1.282	1.093	1.181	1.164	1.284	1.085	1.444	1.252
50	1.186	1.070	1.148	1.103	1.204	1.067	1.428	1.248

Table 3.10: Relative empirical mean squared prediction errors of k-step ahead predictions of 'Model c'

	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC
	n=50		n=200		n=50		n=200	
d =	-.4		-.4		-.2		-.2	
k=1	1.093	1.006	1.034	1.037	1.048	1.006	1.054	1.014
2	1.077	1.006	1.009	1.022	1.173	1.103	0.971	0.990
3	1.015	1.025	1.016	1.004	0.966	0.993	0.987	0.996
4	1.058	1.016	1.020	1.013	1.050	1.027	1.024	1.002
5	1.098	1.065	1.001	1.001	1.071	0.991	0.981	0.981
10	1.045	1.047	1.001	0.998	1.033	1.025	1.016	1.003
15	0.990	1.013	0.985	0.987	1.121	0.993	1.012	1.009
20	1.032	1.008	0.991	1.007	1.001	1.013	1.002	1.002
30	1.014	1.000	1.009	1.026	1.012	1.001	1.017	1.013
40	0.969	0.984	0.996	1.008	1.083	0.995	1.002	1.014
50	0.995	1.004	1.001	0.998	0.969	1.007	1.003	0.994
d =	0		0		.2		.2	
k=1	1.071	1.014	0.965	1.035	0.960	0.973	1.024	1.001
2	1.073	0.999	0.990	0.998	1.011	0.955	1.005	0.982
3	1.032	0.996	0.975	0.973	1.069	0.940	1.022	0.980
4	1.035	0.971	1.003	1.002	1.019	0.972	1.044	1.021
5	0.943	0.988	1.038	1.018	1.098	0.869	0.984	0.976
10	1.015	0.985	1.008	1.013	0.982	0.997	0.997	1.007
15	1.072	1.004	1.025	1.001	0.991	0.970	0.969	0.989
20	0.965	1.006	0.969	0.987	0.944	0.966	1.037	1.003
30	1.025	1.006	1.011	1.002	1.090	1.003	0.977	0.982
40	1.031	0.990	0.997	1.001	1.079	0.979	0.998	1.005
50	1.038	0.994	1.002	1.002	0.957	0.940	1.007	0.992
d =	.4		.4		.6		.6	
k=1	0.969	1.003	0.972	0.998	1.141	1.015	0.968	1.006
2	1.037	0.949	1.014	1.026	1.079	1.010	0.976	1.024
3	1.018	0.972	1.010	1.010	1.022	0.992	0.973	1.011
4	1.082	1.018	1.004	0.992	1.055	1.057	0.962	1.006
5	0.952	0.965	1.046	1.016	0.975	0.994	0.951	1.021
10	0.865	0.878	1.011	1.074	0.974	0.956	1.004	1.100
15	0.939	0.966	0.961	1.029	1.056	0.970	1.045	1.099
20	0.988	1.037	1.002	1.030	1.009	1.004	0.893	0.936
30	0.851	0.854	1.024	1.028	0.795	0.845	1.007	1.047
40	0.871	0.898	0.967	0.967	0.756	0.840	1.022	1.060
50	0.882	0.808	0.834	0.859	0.740	0.810	0.988	1.010
d =	.8		.8		1.0		1.0	
k=1	1.088	0.978	1.016	1.031	1.043	1.006	0.968	0.954
2	1.149	1.027	1.037	1.036	1.209	1.029	0.989	0.988
3	1.184	1.096	1.046	1.025	1.207	1.015	1.019	0.971
4	1.169	1.095	1.080	1.050	1.190	1.039	1.055	0.999
5	1.184	1.102	1.087	1.029	1.309	1.069	1.106	1.025
10	1.243	1.103	0.937	0.982	1.332	1.075	1.287	1.078
15	1.164	1.048	0.987	1.014	1.278	1.058	1.347	1.086
20	1.172	1.082	0.999	1.016	1.201	1.037	1.441	1.097
30	1.046	0.985	1.007	1.003	1.002	1.019	1.387	1.046
40	1.015	1.000	1.005	1.011	0.933	0.950	1.241	1.017
50	1.044	1.041	1.003	0.996	0.896	0.976	1.189	1.010
d =	1.2		1.2		1.4		1.4	
k=1	1.173	1.097	0.968	0.979	1.085	1.096	0.995	0.957
2	1.193	1.094	0.979	1.023	1.097	1.094	0.965	0.946
3	1.248	1.152	0.987	1.044	1.152	1.114	0.957	0.947
4	1.189	1.145	1.019	1.071	1.152	1.103	0.970	0.962
5	1.144	1.161	1.047	1.089	1.158	1.089	0.980	0.969
10	1.252	1.313	1.015	1.051	1.117	1.032	1.055	0.968
15	1.217	1.270	1.026	1.076	1.167	1.072	1.146	0.985
20	1.189	1.257	0.969	1.040	1.207	1.122	1.183	0.995
30	0.958	1.076	0.990	1.056	1.244	1.157	1.243	1.018
40	0.892	0.994	1.057	1.061	1.247	1.159	1.266	1.050
50	0.902	0.992	1.046	1.039	1.250	1.151	1.301	1.077

Table 3.11: Relative empirical mean squared prediction errors of k-step ahead predictions of 'Model d'

	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC
	n=50		n=200		n=50		n=200	
d =	-.4		-.4		-.2		-.2	
k=1	1.175	1.105	1.022	0.995	1.116	1.042	1.019	1.020
2	1.122	1.023	1.031	0.996	1.092	1.000	1.015	0.997
3	1.183	1.051	1.001	0.993	1.075	1.043	0.957	0.993
4	1.106	0.976	0.980	0.964	0.989	1.027	1.005	1.001
5	0.994	0.988	0.995	0.993	1.085	1.075	1.012	0.999
10	1.033	0.993	0.984	1.001	1.068	1.013	0.997	0.993
15	1.045	0.991	1.001	0.986	1.011	1.006	1.002	1.000
20	1.039	0.998	0.994	1.002	1.045	1.031	0.985	0.996
30	1.022	1.002	0.992	0.996	0.981	0.989	0.989	0.990
40	0.972	0.988	1.007	0.992	1.015	1.002	1.001	1.003
50	1.033	1.000	0.996	0.993	1.031	1.018	0.994	1.000
d =	0		0		.2		.2	
k=1	0.937	0.968	1.034	0.998	1.044	0.983	1.023	1.015
2	0.982	1.004	1.041	0.997	1.116	1.040	0.971	0.984
3	0.907	0.902	1.008	0.999	1.112	1.107	0.986	0.986
4	0.982	0.983	0.989	0.979	1.115	1.110	0.997	0.987
5	0.985	0.942	0.996	0.979	1.224	1.120	1.003	1.013
10	0.925	0.941	1.020	1.009	0.998	1.041	0.972	0.998
15	0.880	0.999	0.986	1.003	0.982	0.993	0.966	0.969
20	0.858	0.950	1.007	1.005	0.929	0.958	0.985	0.979
30	0.966	0.953	1.019	1.019	0.901	0.921	0.938	0.922
40	0.950	1.022	0.999	0.997	0.895	0.900	1.053	1.027
50	0.787	0.901	0.996	0.986	0.839	0.868	0.969	0.966
d =	.4		.4		.6		.6	
k=1	1.059	1.024	1.031	1.028	0.991	0.976	0.977	1.004
2	1.008	1.034	1.042	1.043	0.925	0.958	1.072	1.025
3	1.095	1.047	0.984	0.994	0.907	0.978	1.024	1.004
4	1.044	0.994	0.986	0.986	0.925	1.019	0.980	0.984
5	1.111	1.007	1.018	0.994	1.045	1.107	0.991	0.979
10	0.904	0.836	0.970	0.983	1.250	1.214	0.898	0.895
15	0.735	0.755	0.991	0.982	1.044	1.027	0.925	0.893
20	0.817	0.854	0.948	0.977	0.880	0.924	0.905	0.888
30	0.846	0.905	0.993	0.960	0.766	0.830	1.004	0.964
40	0.676	0.780	1.048	0.990	0.686	0.773	1.014	0.957
50	0.741	0.824	1.048	0.986	0.657	0.739	1.007	0.910
d =	.8		.8		1.0		1.0	
k=1	1.238	1.085	1.011	1.022	1.011	1.022	1.047	1.023
2	1.228	1.105	1.014	1.053	1.014	1.053	1.016	1.035
3	1.219	1.159	1.020	1.054	1.020	1.054	1.029	1.041
4	1.213	1.146	1.028	1.048	1.028	1.048	1.062	1.068
5	1.302	1.202	1.032	1.036	1.032	1.036	1.079	1.068
10	1.211	1.188	1.032	1.080	1.032	1.080	1.086	1.100
15	1.144	1.101	1.085	1.074	1.085	1.074	1.085	1.091
20	1.125	1.053	1.087	1.103	1.087	1.103	1.084	1.088
30	1.032	0.955	1.127	1.112	1.127	1.112	1.136	1.085
40	0.997	0.979	1.139	1.045	1.139	1.045	1.285	1.187
50	0.950	0.947	1.077	1.062	1.077	1.062	1.238	1.163
d =	1.2		1.2		1.4		1.4	
k=1	1.143	1.007	1.038	1.031	1.057	1.011	0.943	0.984
2	1.025	0.989	1.059	1.042	1.082	1.036	0.984	1.008
3	1.091	1.021	1.058	1.042	1.135	1.057	0.993	1.018
4	1.171	1.089	1.062	1.043	1.166	1.065	0.991	1.028
5	1.235	1.135	1.060	1.034	1.165	1.064	0.999	1.041
10	1.304	1.206	1.035	1.029	1.126	1.049	1.056	1.078
15	1.258	1.123	1.039	1.043	1.139	1.036	1.042	1.062
20	1.215	1.059	1.082	1.049	1.156	1.020	1.062	1.075
30	1.159	1.032	1.113	1.065	1.155	1.005	1.090	1.080
40	1.065	1.021	1.112	1.044	1.158	1.013	1.117	1.092
50	0.980	0.993	1.108	1.027	1.121	1.005	1.144	1.107

Table 3.12: Relative empirical mean squared prediction errors of k-step ahead predictions of 'Model e'

	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC
	n=50		n=200		n=50		n=200	
d =	-.4		-.4		-.2		-.2	
k=1	1.019	0.965	1.027	1.018	1.121	1.015	1.096	1.024
2	1.044	0.999	1.007	0.994	1.090	0.962	1.015	1.001
3	1.085	0.993	0.984	1.007	1.066	0.953	1.024	1.002
4	1.086	0.997	0.977	0.999	1.031	0.941	1.027	1.004
5	1.026	0.980	0.977	1.007	1.034	0.947	1.084	0.99
10	1.084	1.021	0.984	0.982	0.802	0.852	1.016	1.000
15	1.010	0.940	0.989	1.006	0.868	0.935	0.911	0.936
20	1.019	0.970	0.996	0.995	0.816	0.956	1.007	0.933
30	0.879	0.812	0.988	0.977	0.714	0.825	0.861	0.982
40	0.897	0.891	1.011	1.004	0.705	0.732	0.753	0.778
50	0.894	0.870	0.990	0.987	0.727	0.781	0.819	0.868
d =	0		0		.2		.2	
k=1	1.009	0.992	1.010	1.016	0.943	0.991	1.014	0.974
2	1.033	1.021	0.986	0.992	0.989	0.998	1.034	0.988
3	1.071	1.054	1.016	1.000	0.937	0.949	1.045	1.011
4	1.072	1.030	1.033	0.991	0.956	0.938	1.051	1.021
5	1.017	1.050	1.050	1.025	1.010	0.992	1.059	1.019
10	0.908	0.996	1.017	1.022	1.062	1.061	1.069	1.025
15	1.051	0.958	0.943	0.946	1.027	1.058	1.035	0.997
20	1.005	0.955	0.943	0.954	0.847	0.878	1.006	0.992
30	0.882	0.916	0.915	0.922	0.899	0.884	0.855	0.888
40	0.862	0.964	0.872	0.858	0.867	0.855	0.784	0.820
50	0.800	0.886	0.867	0.829	0.803	0.798	0.745	0.867
d =	.4		.4		.6		.6	
k=1	0.909	0.945	1.050	1.079	1.094	1.021	1.032	1.025
2	0.900	0.968	1.100	1.113	1.137	1.008	1.010	0.999
3	0.923	0.988	1.127	1.110	1.112	0.980	0.991	0.990
4	0.880	0.968	1.169	1.123	1.096	0.970	0.969	0.974
5	0.907	0.990	1.214	1.140	1.113	0.979	0.973	0.975
10	0.894	0.963	1.236	1.126	1.141	0.987	0.981	1.008
15	0.897	0.929	1.076	0.979	1.101	1.025	0.977	0.991
20	0.928	0.964	1.011	0.987	1.050	0.990	0.904	0.936
30	0.910	0.963	0.757	0.817	0.927	0.893	0.784	0.910
40	0.854	0.920	0.639	0.731	0.895	0.890	0.757	0.867
50	0.861	0.911	0.610	0.765	0.872	0.883	0.725	0.844
d =	.8		.8		1.0		1.0	
k=1	1.022	1.005	0.992	1.008	1.122	1.038	1.007	1.033
2	1.048	1.011	0.993	1.022	1.185	1.084	1.036	1.056
3	1.096	1.038	0.994	1.034	1.211	1.096	1.041	1.078
4	1.117	1.066	0.996	1.037	1.203	1.092	1.051	1.100
5	1.124	1.071	1.004	1.036	1.199	1.087	1.067	1.113
10	1.158	1.116	1.101	1.055	1.182	1.056	1.154	1.167
15	1.083	1.077	1.112	1.071	1.162	1.058	1.180	1.171
20	1.062	1.072	1.138	1.086	1.152	1.061	1.156	1.152
30	1.107	1.060	1.092	1.059	1.112	1.050	1.157	1.152
40	1.102	1.048	1.034	0.996	1.096	1.058	1.173	1.172
50	1.049	1.018	1.012	0.980	1.039	1.001	1.160	1.157
d =	1.2		1.2		1.4		1.4	
k=1	1.002	1.029	1.018	0.997	1.084	1.040	1.022	1.014
2	1.033	1.062	1.022	0.995	1.078	1.042	1.029	1.015
3	1.041	1.052	1.011	0.986	1.077	1.042	1.027	1.020
4	1.073	1.062	1.026	0.985	1.070	1.031	1.031	1.031
5	1.089	1.068	1.028	0.985	1.076	1.031	1.038	1.039
10	1.077	1.041	1.063	1.020	1.059	1.015	1.065	1.053
15	1.067	1.025	1.080	1.038	1.031	1.001	1.072	1.062
20	1.043	1.004	1.102	1.051	1.033	1.013	1.078	1.075
30	1.011	0.978	1.114	1.064	1.050	1.040	1.064	1.080
40	0.998	0.974	1.096	1.056	1.069	1.065	1.054	1.071
50	0.994	0.974	1.082	1.048	1.081	1.080	1.043	1.055

Table 3.13: Relative empirical mean squared prediction errors of k-step ahead predictions of 'Model f'

	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC	AIC/BIC	HIC/BIC
	n=50		n=200		n=50		n=200	
d =	-.4		-.4		-.2		-.2	
k=1	1.082	0.987	1.029	1.001	1.010	0.997	1.011	0.979
2	1.100	0.938	1.019	0.998	1.006	0.969	1.015	0.990
3	1.026	0.968	1.015	0.992	1.005	0.948	1.012	0.992
4	1.030	0.975	0.979	0.992	0.988	0.948	1.026	0.992
5	0.992	0.967	0.987	0.994	0.993	0.972	1.013	0.989
10	1.027	0.987	0.982	0.981	1.115	0.948	1.001	1.016
15	1.058	0.988	1.012	0.998	1.068	0.940	1.017	1.013
20	1.075	0.989	0.983	1.012	1.095	1.003	1.015	1.012
30	1.026	1.001	1.005	1.006	1.007	0.964	0.996	0.994
40	1.031	1.007	0.994	0.996	1.023	0.968	0.996	1.000
50	1.024	0.999	0.998	1.010	1.029	0.979	1.010	1.011
d =	0		0		.2		.2	
k=1	0.996	0.968	1.010	0.999	1.048	0.991	1.019	1.019
2	1.047	0.971	1.002	1.002	1.050	0.958	1.035	1.020
3	1.034	0.942	0.991	1.005	1.074	0.942	1.061	1.022
4	1.030	0.951	1.004	1.006	1.027	0.953	1.061	1.011
5	0.999	0.975	1.009	1.005	0.978	0.967	1.045	1.003
10	1.016	0.993	1.017	1.021	1.090	0.959	1.012	0.995
15	1.038	1.032	1.004	1.007	0.965	0.985	1.018	0.997
20	1.039	1.039	1.017	0.991	1.016	0.915	0.978	1.016
30	0.988	0.956	1.009	1.006	1.049	0.981	0.995	0.993
40	1.004	0.994	0.997	0.991	0.933	0.987	1.008	1.009
50	1.012	0.993	0.982	0.997	1.004	0.935	0.993	0.972
d =	.4		.4		.6		.6	
k=1	1.026	1.013	1.061	1.043	0.988	0.983	1.022	1.042
2	1.040	0.990	1.057	1.045	0.980	0.994	1.048	1.055
3	1.059	0.975	1.046	1.030	1.030	0.986	1.046	1.043
4	1.057	0.976	1.033	1.010	1.085	0.987	1.043	1.039
5	1.036	0.988	1.013	0.973	1.114	1.010	1.014	1.021
10	1.049	1.008	1.027	1.007	1.133	0.998	0.950	0.942
15	0.903	1.067	1.009	0.982	0.993	0.995	1.028	1.023
20	0.959	1.056	1.080	1.064	1.130	1.037	1.040	0.927
30	1.108	1.092	0.930	0.957	1.007	0.998	1.066	1.003
40	1.086	1.208	1.035	1.051	0.939	0.961	0.989	0.984
50	0.971	1.091	1.047	1.023	0.931	1.001	0.962	1.021
d =	.8		.8		1.0		1.0	
k=1	1.047	0.948	1.030	1.027	1.018	0.907	1.018	1.019
2	0.999	0.940	1.019	1.025	0.988	0.871	1.035	1.018
3	1.015	0.935	1.012	1.028	0.995	0.891	1.067	1.023
4	1.052	0.932	0.978	1.015	0.994	0.903	1.083	1.026
5	1.086	0.941	0.938	0.985	0.993	0.912	1.092	1.026
10	1.042	0.971	0.925	0.958	1.028	0.766	1.101	0.988
15	1.133	1.009	1.039	1.041	1.043	0.807	1.193	1.036
20	1.188	1.011	0.883	0.921	0.997	0.809	1.182	1.017
30	1.081	1.077	0.960	0.932	0.999	0.880	1.245	1.029
40	1.024	1.139	0.880	0.945	0.958	0.956	1.274	1.056
50	1.026	1.157	0.939	0.967	0.947	1.011	1.228	1.003
d =	1.2		1.2		1.4		1.4	
k=1	1.100	1.010	1.007	0.998	1.036	1.034	0.977	0.970
2	1.145	0.997	1.020	1.018	1.032	1.027	0.999	0.980
3	1.171	0.986	1.036	1.025	1.041	1.006	1.036	0.993
4	1.177	0.968	1.065	1.030	1.049	0.981	1.065	1.005
5	1.169	0.954	1.104	1.033	1.054	0.961	1.088	1.017
10	1.165	0.908	1.220	1.008	1.089	0.896	1.193	1.070
15	1.112	0.931	1.261	1.038	1.119	0.831	1.229	1.106
20	1.223	0.931	1.342	1.071	1.098	0.845	1.239	1.108
30	1.232	0.978	1.358	1.090	1.099	0.874	1.222	1.111
40	1.188	1.002	1.284	1.052	1.080	0.893	1.252	1.153
50	1.102	0.999	1.245	1.068	1.058	0.907	1.299	1.171

Table 3.14: Relative empirical mean squared prediction errors of k-step ahead FARIMA/random walk (RW) predictions of 'Model b'

	n=50					n=200				
	d=.6	.8	1.0	1.2	1.4	d=.6	.8	1.0	1.2	1.4
k=1	1.108	1.072	1.130	1.002	0.488	1.001	0.983	1.017	0.914	0.525
2	0.970	1.029	1.110	1.004	0.528	0.954	0.910	1.032	0.900	0.491
3	1.058	1.088	1.149	1.092	0.561	0.893	0.847	1.038	0.948	0.454
4	1.075	1.036	1.216	1.106	0.566	0.872	0.913	1.045	0.933	0.436
5	0.988	1.091	1.218	1.089	0.592	0.957	0.933	1.052	0.936	0.446
10	1.251	1.131	1.378	1.183	0.682	0.876	1.002	1.081	0.932	0.495
15	1.128	1.215	1.398	1.134	0.652	1.000	0.997	1.136	0.975	0.538
20	1.216	1.400	1.461	1.273	0.689	0.992	0.976	1.202	1.064	0.594
30	1.458	1.593	1.618	1.226	0.718	1.014	1.121	1.174	1.127	0.620
40	1.588	1.971	1.848	1.200	0.722	1.100	1.191	1.287	1.262	0.646
50	2.069	2.268	2.175	1.195	0.756	1.085	1.207	1.392	1.23	0.682

Table 3.15: Empirical coverage percentages of the 95%-k-step ahead random walk prediction intervals of 'Model b'

	n=50					n=200				
	d=.6	.8	1.0	1.2	1.4	d=.6	.8	1.0	1.2	1.4
k=1	0.96	0.93	0.94	0.96	0.85	0.96	0.96	0.95	0.92	0.93
2	0.98	0.96	0.93	0.93	0.73	0.98	0.97	0.93	0.92	0.80
3	0.97	0.98	0.96	0.95	0.71	0.99	1.00	0.93	0.89	0.73
4	0.99	0.96	0.97	0.90	0.71	1.00	0.97	0.97	0.91	0.62
5	0.99	0.99	0.97	0.90	0.68	1.00	0.99	0.99	0.86	0.55
10	0.99	0.98	0.97	0.81	0.56	1.00	0.99	0.97	0.81	0.45
15	1.00	1.00	0.98	0.72	0.46	1.00	1.00	0.96	0.77	0.36
20	1.00	0.99	0.98	0.72	0.35	1.00	1.00	0.94	0.73	0.34
30	1.00	1.00	0.95	0.68	0.28	1.00	1.00	0.92	0.79	0.31
40	1.00	1.00	0.94	0.61	0.24	1.00	1.00	0.94	0.75	0.32
50	1.00	1.00	0.91	0.63	0.23	1.00	1.00	0.93	0.73	0.31

Figure 3.1: Trends within the FARIMA$(p, d, 0)$ environment

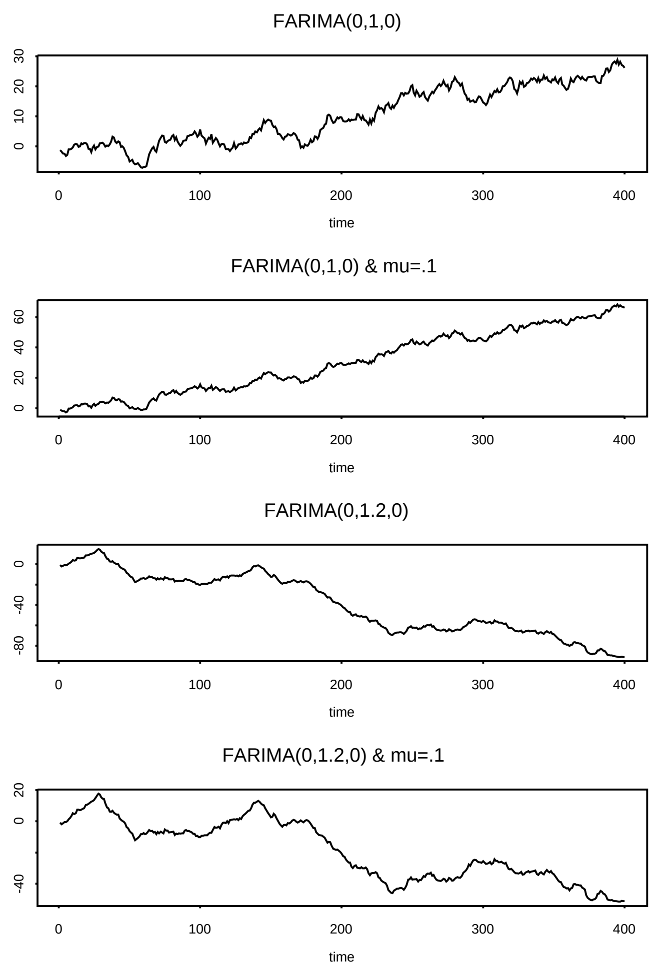
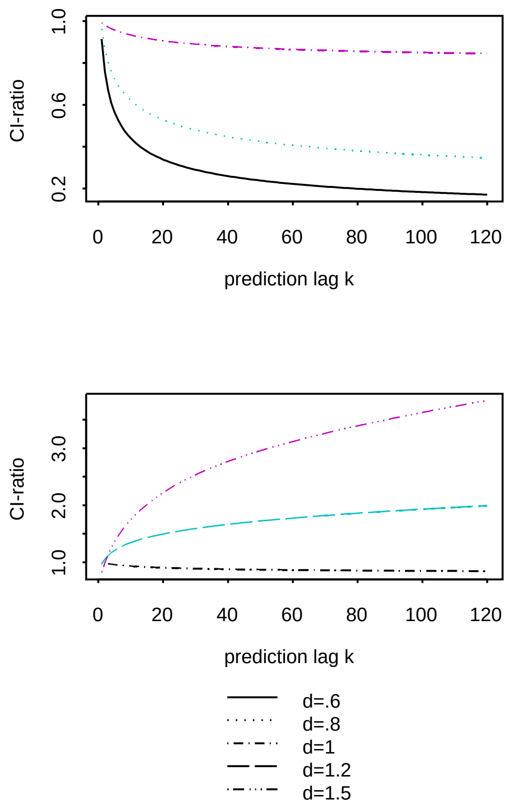


Figure 3.2: Ratio of length of FARIMA and random walk prediction intervals of 'Model b'



Chapter 4

Application of FARIMA$(p, d, 0)$ models to financial data

4.1 Introduction

There has been some discussion in the recent literature about possible unit root behaviour or long memory in financial time series (see e.g. Baillie and Bollerslev [6], Cheung [27], Fong and Ouliaris [42], Liu and He [62], and references therein). Typically, the standard tests for unit roots cannot reject the null hypothesis that there is a unit root in many financial time series (see e.g. Baillie and Bollerslev [6]). On the other hand, Cheung [27] suggested the possibility of long memory in exchange rate data by fitting FARIMA(p, δ, q) models. There, the differencing parameter δ was restricted to the stationary case only (i.e. $\delta \in (-.5, .5)$). In contrast to exchange rates, there is no evidence for long memory in stock market series reported in the literature. For example, Lo [63] finds no evidence to support the presence of long memory in U.S. stock returns using his modified R/S method. Using both Lo's [63] R/S-statistic and the spectral regression of Geweke and Porter-Hudak [47], Cheung and Lai [30] find no evidence of persistence in eighteen international stock return series.[1] As discussed in a previous chapter, these estimation methods are, however, not reliable tools to test for long memory. In view of this, it is interesting to see which hypothesis (long memory or unit root, or antipersistence) may be supported by fitting possibly nonstationary

[1]See Baillie [5] for a survey of fractional integration in finance.

FARIMA$(p,d,0)$ models (Beran [10], Beran, Bhansali and Ocker [12]),

$$\phi(B)(1-B)^{\delta}\{(1-B)^{m}X_t-\mu\}=\epsilon_t, \tag{4.1}$$

to worldwide nominal stock market indices and foreign exchange rates. Specifically, the differencing parameter $d=m+\delta$ (with the integer $m\geq 0$ and $\delta\in(-.5,.5)$) of the model used here is not restricted to the integer domain and can assume any real value. This generalization makes a possibly nonstationary FARIMA$(p,d,0)$ process a parsimonious and flexible model to study stationarity, difference stationarity, polynomial deterministic trends, long memory, short-range correlations and antipersistence simultaneously.

According to the current knowledge, prediction of stock market indices and foreign exchange rates appears to be particularly difficult. There has been a controversial discussion, especially in connection with foreign exchange rates, in how far these financial data can be predicted at all (see e.g. Frankel and Rose [46] and references therein; Meese and Rogoff [69], Baillie and Bollerslev [6], Diebold and Nason [36], Liu and He [62], Meese and Rose [70], Cheung [27], Brooks [25]). Proposed models include for instance random walk, Box-Jenkins ARIMA models, macroeconomic models, nonlinear function models, GARCH models and nonparametric prediction models. In spite of the large variety of models, the success in forecasting the future development of stock market indices and exchange rates seems to have been rather limited. In particular, Frankel and Rose [46] conclude (for exchange rates) that the simple random walk model appears to provide short- to medium term forecasts that are as good as or even better than more sophisticated models. It should be noted, however, that the predominant criterion for judging the performance of predictions has been the accuracy of point predictions, measured for instance by the mean squared prediction error. As mentioned in the previous chapter, this is not the only informative criterion. The purpose of a statistical prediction is not only to obtain a point estimate but also to have a confidence interval. Thus, an important criterion for assessing the usefulness of a statistical forecast is whether prediction intervals are neither too short nor too long. It is therefore interesting to study the forecast performance of the quite sophisticated FARIMA$(p,d,0)$ point- and interval-predictions.

Knowledge of the time series properties of financial data has significant economic implications. In particular, the possibility of long memory has some important impacts.

- If long memory is indeed present in the data, statistical inferences concerning asset pricing models based on traditional testing procedures may no longer be valid (Yajima [86]). Mandelbrot [64] has shown that in the presence of long memory the arrival of new market information cannot be fully arbitraged away and martingale models are not longer valid. Since the most mathematical concepts in finance are based upon the martingale assumption, the pricing of options/futures and the specification of risk in, for instance, asset portfolios, is at least questionable if long memory is indeed present in the data.

- The discovery of long-range dependence (and antipersistence) suggests possibilities for improved price forecasting performance within a linear time series framework. Granger and Joyeux [48] have discussed the forecasting potential of FARIMA models and Geweke and Porter-Hudak [47] have confirmed this by showing that FARIMA models provide more reliable out-of-sample forecasts than do traditional procedures. Also, forecast improvements are demonstrated in Beran and Ocker [17] and in the previous chapter. If the market is weakly efficient, stock prices should behave as a martingale process. The possibility of speculative profits due to superior FARIMA forecasts would cast serious doubt on the basic tenet of market efficiency.

- Also, long-term forecasts and periodical (risk-)reports often require the use of aggregates, e.g. weekly, monthly or quarterly averages. Monthly or quarterly aggregates, however, may not deviate much from a random walk model, apart from a lag-1 correlation of .25, if the original non-aggregated series follows a nonstationary Box-Jenkins model (see e.g. Tiao [82]). In contrast to traditional short-range dependence, long-range dependence and antipersistence are robust with respect to temporal aggregation. Realistic models should therefore explicitly include the possibility of long memory and antipersistence.

The chapter is organized as follows. The next section describes the data. Section 3 presents the empirical estimation and out-of-sample forecast results from the index and exchange rate series using the maximum likelihood and forecasting procedures presented in the second and third chapter respectively. Some remarks in section 4 conclude the chapter. Tables and figures are provided in the appendix.

4.2 The data

The data include eighteen daily nominal stock market closing indices for the period January 4, 1988 to November 10, 1995, and nine daily nominal exchange rates between January 2, 1991 and May 30, 1997. They are, according to definition of the IFC [57], indices for eleven developed markets (DMs: Australia, Belgium, Canada, France, Germany, Hong Kong, Italy, Japan, Switzerland, United Kingdom, and United States), and seven emerging markets (EMs: Brazil, Chile, India, Malaysia, Mexico, South Korea, and Thailand). Table 4.1 presents the names and the exchanges for these indices, together with global ranking by market capitalization in US$ terms as of end-1995 (Euromoney [41]).

Table 4.1: Stock market indices of developed and emerging markets

Country	Exchange	Index	Ranking
DM series			
Australia	Sydney	All Ordinaries	10
Belgium	Brussels	Stock Index	20
Canada	Toronto	TSE 300	7
France	Paris	CAC 40	5
Germany	Frankfurt	DAX	4
Hong Kong	Hong Kong	Hang Seng	9
Italy	Milan	DS General	12
Japan	Tokyo	Nikkei 225	2
Switzerland	Zurich	Swiss Bank Corporation	6
United Kingdom	London	FTSE 100	3
United States	New York	S&P 500	1
EM series			
Brazil	Sao Paulo	BOVESPA	17
Chile	Santiago	IGPA	22
India	Bombay	BSE 30	24
Malaysia	Kuala Lumpur	KLSE	11
Mexico	Mexico City	IPC	21
South Korea	Seoul	KOSPI	18
Thailand	Bangkok	Book Club	23

The indices are expressed in local currency and, overall, are neither adjusted for dividends nor for inflation. The nominal exchange rates chosen here are data in price notation (e.g. US$ per unit of the foreign currency) for the Australian Dollar (aud), Canadian Dollar (cad), Swiss Franc (chf), British Pound (gbp), Hong Kong Dollar (hkd), Japanese Yen (jpy), New Zealand Dollar (nzd), Norwegian Kroner (nok), and Singapore Dollar (sgd).

To study the time series properties, we take a simple pragmatic approach. Missing values in the original index and exchange rate series are replaced by the closest, previous closing value. The analysis of the exchange rate series is then based on the log-transformed daily data, whereas the analysis of the index series uses log-transformed weekly averages. Note, in particular, that we did not consider any corresponding return series. Thus, in contrast to the traditional approach, it is not assumed a priori that the first integer difference has to be taken to make the series stationary. Instead, the possibilities of stationarity, difference stationarity, polynomial deterministic trend, short memory, long memory and antipersistence are left open. It is then decided based on the data which combination of these components may be present.

4.3 Empirical results

4.3.1 Estimation of FARIMA$(p, d, 0)$ models

Tables 4.2 and 4.3 summarize the essential features of the fitted models for the log-transformed weekly averaged stock market indices covering the period January 4, 1988 to November 15, 1990 ($n = 150$), and for the log-transformed daily exchange rates between January 2, 1991 and December 20, 1991 ($n = 250$). The corresponding 95%-confidence intervals are given in brackets. The models were selected using the BIC.

The estimated value of d and the confidence intervals suggest that all series are, overall, nonstationary (i.e. $d > .5$), though a value of d slightly below .5 can not be excluded completely for the stock indices of UK, Brazil and Chile. This is not very surprisingly in view of the visual impressions of the assets under study given in Figures 4.1, 4.2 and 4.3.

More interesting is the fact that $d = 1$ is not contained, or almost not (for the DMs Germany, Hong Kong and US, and for the EMs Malaysia and South Korea), in the confidence intervals of eight out of eleven DMs, and five out of seven EMs. Taking the first, integer, difference may lead to underdifferencing in these cases. Instead of an integer difference, a fractional difference has to be taken to obtain stationary process. The unit-roots hypothesis, $d = 1$, can not be rejected for three DMs (Canada, France and UK) and two EMs (Brazil and Chile) only. For them, however, substantial stochastic short-term dependence is found in form of a strong significant (except France) AR(1) term which is known to model a part of the long-range dependence so that, in effect, $\hat{d}$ is reduced (see Beran, Bhansali and Ocker [12] and the second chapter). Finally, note that a deterministic trend was only found for Brazil. As one may expect (at least a posteriori), for this series, a significant linear trend function is detected due to the relatively stable ascent in the considered time period. For the other stock index series, apparent linear trends do not persist long enough, and can therefore be explained as stochastic. Overall, the estimates indicate that there is strong evidence for long memory in stock market indices, independently of market size and of whether a market is referred to as developed or not. In contrast to the stock indices, the unit-roots hypothesis seems to be an appropriate model for the exchange rates (at least for the period under consideration) as it can be seen from table 4.3. Neither significant long-range dependence nor traditional short memory could be detected by fitting (4.1).

The fits of the models are demonstrated by the correlograms 4.4 and 4.5, and by the normal probability plots 4.6 and 4.7 of the residuals. Recall that one autocorrelation (out of twenty) is expected to lie outside the dotted 95% confidence interval, if the residuals are not correlated. This does (almost) hold for the stock market indices. Some residual series exhibit autocorrelations slightly under the lower confidence bound around lag 15, which is beyond any meaningful interpretation. The fits of the exchange rates are, however, less satisfactoring as it can be seen from figure 4.5. These series exhibit an apparent leptokurtic distribution (see figure 4.7). The fatter tails of the residual's distribution in comparison to the Gaussian case may be due to autoregressive conditional heteroscedasticity (ARCH) (see e.g. Engle [40]). In contrast to the residuals of the stock market series, there is (almost) evidence for ARCH errors in the correlograms of the squared residuals of the exchange rates (compare figure 4.9 with figure 4.8). The presence of ARCH is, though slightly departures from normality (see figure 4.6), negligible for the most stock markets. Thus, our empirical findings are consistent with those reported in the literature. That is, vanishing ARCH dependencies under temporal aggregation (see e.g. Diebold [35]). Note that ARCH is not incorporated in our estimation method. This may cause some losses in efficiency, in particular, for the exchange rates.

4.3.2 Out-of-sample FARIMA$(p, d, 0)$ forecasts

To check the performance of FARIMA predictions, we consider for each stock index/exchange rate, 21/51 partially overlapping subseries $X_1^{(j)}, ..., X_n^{(j)}$ where $n = 150$ and $n = 250$ for the stock indices and exchange rates respectively. $X_i^{(j)}$ are the log-transformed weekly averaged stock indices $X_i^{(j)} = \log X_{i+12j}$ $(j = 0, ..., 20)$ and the log-transformed daily exchange rates $X_i^{(j)} = \log X_{i+25j}$ $(j = 0, ..., 50)$ respectively. Observe that $n = 150$ corresponds to a period of about three years weekly data, whereas $n = 250$ covers a period of about one year daily data. A FARIMA$(p, d, 0)$ model is fitted to each subseries via the BIC (with maximum autoregressive order $p = 5$), where p, μ and the unknown parameters $\theta = (\sigma_\epsilon^2, d, \phi_1, ..., \phi_p)$ are estimated and forecasts (point and interval estimates) are calculated for $X_{n+k}^{(j)}$ as described in the previous chapter. Note that we calculate predictions for the forecast horizons $k = 1, 2, ..., 12$ for the stock indices and $k = 1, 2, ..., 25$ (reported for $k = 1, 2, ..., 10, 15, 20, 25$) for the exchange rates. Also, out-of-sample forecasts were computed in a comparable manner for two benchmark models, namely the random walk and the traditional Box-Jenkins ARIMA$(p, 1, 0)$ model (with $1 \leq p \leq 5$). The ARIMA$(p, 1, 0)$ models were selected via the BIC using the S-Plus function *arima.mle*; forecasts were obtained from *arima.forecast*.

Stock indices

Table 4.4 summarizes the essential features (sample means, standard deviations and ranges of $\hat{d}$) of the fitted models for the stock indices. The only chosen orders were $\hat{p} = 0, 1$ and 2, and in most cases $\hat{p}$ was equal to zero or one. Other values of $\hat{p}$ did not occur. The values below the bold-font entries are the corresponding quantities for the subseries with $\hat{p} = 0, 1$ and 2 respectively. The number of subseries with $\hat{p} = 0, 1, 2$ is given by $n(\hat{p})$. Apparently, the value of $\widehat{d}$ was highly confounded with the autoregressive order $\widehat{p}$. In particular, d tends to be small when $\widehat{p}$ is chosen to be nonzero. As we already know, autoregressive coefficients tend to model a part of the 'long-range dependence' so that $\widehat{d}$ is reduced. Observe, in particular, that $\hat{d} = 1$ is overall not contained in the range of $\hat{d}$ if $\hat{p} = 0$ (except for the US), supporting the investigated evidence for long memory in nominal stock market indices.

Table 4.6 gives the percentages of future observations contained in 95%-prediction intervals for k-step ahead FARIMA, random walk and ARIMA forecasts. As a cautionary remark, it should be noted that the percentages are obtained as an average of 21 (dependent) indicator variables. Thus, in particular, if a future value was outside the prediction interval for one of the subseries, then the observed coverage probability drops from 100 to 95%. Table 4.6 indicates that for the most indices (except Belgium, Japan, Chile, and Mexico) the observed coverage percentages are close to the nominal ones for short-term forecasts (say $k \leq 5$). Medium- to long-term forecasts seem to be more difficult for some of the DMs (Hong Kong, Italy and Japan) and the EMs Chile, India, Malaysia and Mexico. However, sufficient interval forecasts are obtained for the remaining series. On the other hand, there is a tendency that random walk prediction intervals are too narrow in the short- and, in particular, the long-run indicating the presence of long memory. The investigated persistence is, in comparison to the random random walk, better adapted by FARIMA and ARIMA interval predictions. Compare the length of FARIMA prediction intervals with those of random walk and ARIMA forecasts. Figures 4.10 and 4.11 display the ratio of the average length of FARIMA prediction intervals divided by the length of random walk and ARIMA prediction intervals, plotted against the prediction lag $k = 1, 2, ..., 12$. FARIMA intervals turn out to be much wider in comparison to the random walk, in particular, for large lags indicating the presence of long memory. The larger the amount of persistence, the larger the relative length of the forecast intervals, while the coverage level of the random walk intervals appears to be quite often incorrect (see table 4.6). Overall, FARIMA forecasts yield, in comparison to the random walk, (approximately) valid prediction intervals for short-term forecasts and more reliable prediction intervals in the long-run. Also medium and long-term FARIMA prediction intervals are wider, in comparison to ARIMA intervals, if there is a moderately strong and stable degree of long memory over the time (e.g. Australia, Belgium, Germany, Italy, Switzerland and India.) At the same time, the coverage probability of the ARIMA predictions seems to be comparable with those from FARIMA forecasts. Also, FARIMA prediction intervals are rarely worse than ARIMA interval forecasts, but may provide the 'safer choice' if the degree of persistence is strong.

A second question is how the precision of FARIMA point predictions compares to random walk and ARIMA predictions. Table 4.8 compares the mean squared prediction errors (MSE) of FARIMA forecasts with those of random walk and ARIMA forecasts. An entry smaller (larger) than 1 means that the FARIMA prediction yields a smaller (larger) MSE. Random walk and ARIMA predictions turn out to perform better in four DMs (Belgium, Canada, Italy and Japan) and two EMs (India and South Korea). In contrast, FARIMA models yield superior (or at least competitive with respect to ARIMA predictions) forecasts in two DMs (UK and US) and four EMs (Brazil, Chile, Mexico and Thailand). For the other examples, not much is gained in terms of point forecast accuracy by using the more sophisticated FARIMA models. In summary, 12 out of 18 FARIMA point predictions turn out to perform better or at least competitive compared to random walk and ARIMA predictions. Surprisingly, substantial point forecast improvements are partially possible (e.g. US and Brazil). The evidence of better performance in longer horizons, however, does not seem consistent across stock indices.

Exchange rates

Table 4.5 summarizes the essential features (sample means, standard deviations and ranges of $\hat{d}$) of the fitted models for the foreign exchange rates. The only chosen orders were $\hat{p} = 0, 1$ and 2, and in most cases $\hat{p}$ was equal to zero or one. The case of $\hat{p} = 2$ was only observed for nzd and sgd. The values below the bold-font entries are the corresponding quantities for the subseries with $\hat{p} = 0, 1$ and 2 respectively. The number of subseries with $\hat{p} = 0, 1, 2$ is given by $n(\hat{p})$. In contrast to the stock indices, $\hat{d} = 1$ is (just) contained in the range of $\hat{d}$ if $\hat{p} = 0$.

Table 4.7 gives the percentages of future observations contained in 95%-prediction intervals for k-step ahead FARIMA and random walk forecasts. Table 4.7 indicates that FARIMA forecasts yield valid prediction intervals for the foreign exchange rates considered here. Also note that the coverage probabilities of random walk and ARIMA predictions are similar to those of the FARIMA forecasts and close to the nominal ones.

The question of the point forecast accuracy is, in particular, interesting in view of the apparently good performance of random walk predictions for foreign exchange rates reported in the literature (see the references given above). Table 4.9 reports the MSE of the FARIMA models relative to those of the random walk and the ARIMA model. The point forecast accuracy is comparable to random walk and ARIMA predictions. Thus, in terms of the point-forecast accuracy, not much is gained by using the more complicated FARIMA models. This result corresponds to those reported in the literature using different techniques and different data sets.

Consider, however, the length of FARIMA prediction intervals. Figure 4.12 displays the ratio of the average length of FARIMA prediction intervals divided by the length of random walk and ARIMA prediction intervals, plotted against k. FARIMA intervals turn out to be (almost always) considerably shorter, in particular for large lags. The most dramatic improvement is achieved for the sgd. Already for $k = 20$, the average interval is shorter by a factor of about .8, while the coverage level of the interval appears to be correct (see table 4.7). At the same time, a comparison with the random walk and the ARIMA model indicates that, even for long-term forecasts, the coverage probability of random walk and ARIMA prediction intervals is not generally higher, although they are wider. In this sense, FARIMA predictions of exchange rates outperform random walk and ARIMA forecasts. The unit-root hypothesis seems to hold temporarily only. In addition, our findings are consistent with those reported in Beran and Ocker [17]. We have found substantial evidence for antipersistence in the stochastic part of exchange rates resulting in shorter prediction intervals.

4.4 Concluding remarks

Using FARIMA models, we found significant evidence of long memory in nominal stock market indices regardless of the market size and of whether a market is referred to as developed or not. This result is in contrast to those reported in the literature. On the other hand, the unit root hypothesis seems to hold for the nominal exchange rates only. For those series, however, we found evidence of antipersistence too, depending on the period under consideration.

The corresponding out-of-sample forecasts for the stock market indices resulted partially in substantial improvements in forecasting accuracy compared to random walk and ARIMA predictions. Also, it turned out that random walk and ARIMA forecast intervals may be too optimistic (i.e. too short) if the degree of persistence is strong. For the exchange rates, however, shorter prediction intervals were found by using the more sophisticated

FARIMA forecasts, indicating the presence of (at least temporarily) antipersistence. Clearly, as always with forecasting, structural changes in the behaviour of the data that have not occurred in the past cannot be foreseen. Thus, for instance, sudden extreme drops in a particular asset price series are usually due to an artificial intervention that can hardly be predicted from the one observed time series only.

Our findings contradict the martingale model. In particular, the investigated evidence for long memory casts serious doubt on the basic tenet of, for instance, the pricing of options and futures: namely the assumption that asset prices behave like Brownian motion. In particular, the arrival of new market information cannot be fully arbitraged away in the presence of long memory, and martingale models of asset prices cannot be obtained from arbitrage (see Mandelbrot [64]). It would therefore be an interesting task for future research to adapt traditional pricing theory to fractional patterns more rigorously. Moreover, we suggest the use of prediction intervals as a possible measure for future risk. Today's approaches to risk control of a portfolio are typically based upon confidence intervals for the expected return using historical observations. However, forecasting the future development and, in particular, finding an appropriate prediction interval seems to be the more intuitive approach if we want to specify risk in the 'future'. Another application of prediction intervals could be found in the design of derivative products, such as FX warrants. Those products typically consist of bets on the future development of an underlying asset. Clearly, the return paid after a specific holding period depends upon probability considerations. More practical experience will be needed to explore the potential usefulness of FARIMA prediction intervals as a risk measure and product development tool.

4.5 Appendix

Table 4.2: Estimation results: stock market indices

stock index	$\hat{d}$	95%-c.i. d	$\hat{\phi}_1$	95%-c.i. ϕ_1	$\bar{Y}$	95%-c.i. μ
DM series						
Australia	1.200	[1.070, 1.330]	-	-	0.0002	[-0.0089, 0.0093]
Belgium	1.352	[1.230, 1.480]	-	-	0.0019	[-0.0299, 0.0337]
Canada	0.905	[0.562, 1.248]	0.464	[0.075, 0.852]	-0.0002	[-0.0040, 0.0036]
France	0.955	[0.673, 1.237]	0.329	[-0.011, 0.669]	0.0031	[-0.0042, 0.0103]
Germany	1.115	[0.990, 1.240]	-	-	0.0023	[-0.0017, 0.0063]
Hong Kong	1.065	[0.940, 1.190]	-	-	0.0014	[-0.0037, 0.0066]
Italy	1.230	[1.100, 1.350]	-	-	-0.0009	[-0.0145, 0.0128]
Japan	1.210	[1.085, 1.474]	-	-	0.0004	[-0.0109, 0.0116]
Switzerland	1.205	[1.080, 1.330]	-	-	0.0006	[-0.0082, 0.0094]
UK	0.670	[0.230, 1.110]	0.637	[0.204, 1.07]	0.0010	[-0.0033, 0.0052]
US	1.080	[0.955, 1.210]	-	-	0.0015	[-0.0012, 0.0041]
EM series						
Brazil	0.670	[0.225, 1.120]	0.654	[0.224, 1.080]	0.0499	[0.0227, 0.0772]
Chile	0.710	[0.294, 1.130]	0.587	[0.157, 1.020]	0.0061	[-0.0000, 0.0122]
India	1.160	[1.035, 1.285]	-	-	0.0071	[-0.0057, 0.0200]
Malaysia	1.090	[0.965, 1.220]	-	-	0.0035	[-0.0009, 0.0079]
Mexico	1.125	[1.000, 1.250]	-	-	0.0126	[-0.0033, 0.0286]
South Korea	1.105	[0.980, 1.230]	-	-	0.0018	[-0.0029, 0.0065]
Thailand	1.170	[1.040, 1.300]	-	-	0.0049	[-0.0104, 0.0202]

Table 4.3: Estimation results: exchange rates

exchange rate	$\hat{d}$	95%-c.i. d	$\hat{\phi}_1$	95%-c.i. ϕ_1	$\bar{Y}$	95%-c.i. μ
aud	1.000	[0.903, 1.100]	-	-	-0.0000	[-0.0005, 0.0004]
cad	0.985	[0.888, 1.080]	-	-	0.0000	[-0.0002, 0.0002]
chf	1.045	[0.948, 1.140]	-	-	-0.0003	[-0.0013, 0.0007]
gbp	1.090	[0.993, 1.190]	-	-	-0.0002	[-0.0011, 0.0008]
hkd	0.950	[0.853, 1.050]	-	-	0.0000	[-0.0002, 0.0003]
nok	1.010	[0.913, 1.110]	-	-	-0.0002	[-0.0012, 0.0009]
nzd	0.810	[0.713, 0.907]	-	-	-0.0004	[-0.0010, 0.0003]
sgd	1.000	[0.903, 1.100]	-	-	0.0002	[-0.0002, 0.0007]

Table 4.4: Sample means, standard deviations and ranges of $\hat{d}$ of stock market indices

stock index	mean	standard deviation	range
DM series			
Australia	**0.892**	**0.485**	**[-0.450, 1.230]**
$n(\hat{p}=0)=15$	1.175	0.037	[1.115, 1.230]
$n(\hat{p}=1)=5$	0.311	0.020	[0.290, 0.330]
$n(\hat{p}=2)=1$	-0.450	0.000	[-0.450, -0.450]
Belgium	**1.027**	**0.239**	**[0.680, 1.352]**
$n(\hat{p}=0)=11$	1.232	0.042	[1.190, 1.352]
$n(\hat{p}=1)=10$	0.802	0.133	[0.680, 1.025]
Canada	**0.667**	**0.456**	**[-0.450, 1.210]**
$n(\hat{p}=0)=1$	1.210	0.000	[1.210, 1.210]
$n(\hat{p}=1)=17$	0.822	0.073	[0.675, 0.930]
$n(\hat{p}=2)=3$	-0.389	0.107	[-0.450, -0.266]
France	**0.806**	**0.345**	**[0.180, 1.160]**
$n(\hat{p}=0)=9$	1.124	0.034	[1.050, 1.160]
$n(\hat{p}=1)=12$	0.566	0.265	[0.180, 0.955]
Germany	**1.087**	**0.024**	**[1.015, 1.120]**
$n(\hat{p}=0)=21$	1.087	0.024	[1.015, 1.120]
Hong Kong	**0.872**	**0.190**	**[0.670, 1.185]**
$n(\hat{p}=0)=8$	1.098	0.056	[1.010, 1.185]
$n(\hat{p}=1)=11$	0.740	0.065	[0.670, 0.885]
$n(\hat{p}=2)=2$	0.698	0.004	[0.695, 0.700]
Italy	**1.193**	**0.023**	**[1.150, 1.230]**
$n(\hat{p}=0)=21$	1.193	0.023	[1.150, 1.230]
Japan	**0.823**	**0.407**	**[-0.180, 1.210]**
$n(\hat{p}=0)=8$	1.186	0.022	[1.145, 1.210]
$n(\hat{p}=1)=10$	0.732	0.157	[0.310, 0.835]
$n(\hat{p}=2)=3$	0.162	0.562	[-0.180, 0.810]
Switzerland	**1.049**	**0.312**	**[0.300, 1.248]**
$n(\hat{p}=0)=17$	1.182	0.049	[1.105, 1.248]
$n(\hat{p}=1)=4$	0.483	0.329	[0.300,0.975]
UK	**0.816**	**0.293**	**[0.230, 1.180]**
$n(\hat{p}=0)=8$	1.141	0.025	[1.105, 1.180]
$n(\hat{p}=1)=13$	0.615	0.170	[0.230, 0.710]
US	**1.020**	**0.100**	**[0.705, 1.115]**
$n(\hat{p}=0)=20$	1.035	0.072	[0.870, 1.115]
$n(\hat{p}=1)=1$	0.705	0.000	[0.705, 0.705]
EM series			
Brazil	**0.910**	**0.244**	**[0.670, 1.238]**
$n(\hat{p}=0)=9$	1.182	0.038	[1.130, 1.238]
$n(\hat{p}=1)=12$	0.706	0.038	[0.670, 0.785]
Chile	**0.802**	**0.317**	**[0.025, 1.286]**
$n(\hat{p}=0)=2$	1.284	0.003	[1.282, 1.286]
$n(\hat{p}=1)=17$	0.835	0.148	[0.670, 1.065]
$n(\hat{p}=2)=2$	0.035	0.014	[0.025, 0.045]
India	**1.118**	**0.104**	**[0.670, 1.170]**
$n(\hat{p}=0)=20$	1.140	0.017	[1.110, 1.170]
$n(\hat{p}=3)=1$	0.670	0.000	[0.670, 0.670]
Malaysia	**0.908**	**0.312**	**[-0.215, 1.155]**
$n(\hat{p}=0)=10$	1.122	0.022	[1.090, 1.155]
$n(\hat{p}=1)=10$	0.806	0.113	[0.675, 0.970]
$n(\hat{p}=2)=1$	-0.215	0.000	[-0.215, -0.215]
Mexico	**1.084**	**0.218**	**[0.710, 1.261]**
$n(\hat{p}=0)=16$	1.201	0.044	[1.125, 1.261]
$n(\hat{p}=1)=5$	0.710	0.000	[0.710, 0.710]
South Korea	**0.855**	**0.451**	**[-0.450, 1.155]**
$n(\hat{p}=0)=14$	1.093	0.033	[1.035, 1.155]
$n(\hat{p}=1)=5$	0.689	0.033	[0.630, 0.710]
$n(\hat{p}=2)=2$	-0.392	0.082	[-0.450, -0.335]
Thailand	**0.859**	**0.436**	**[-0.331, 1.256]**
$n(\hat{p}=0)=11$	1.172	0.032	[1.140, 1.256]
$n(\hat{p}=1)=9$	0.610	0.305	[0.200, 0.930]
$n(\hat{p}=2)=1$	-0.331	0.000	[-0.331, -0.331]

Table 4.5: Sample means, standard deviations and ranges of $\hat{d}$ of foreign exchange rates

exchange rate	mean	standard deviation	range
aud	**0.947**	**0.103**	**[0.670, 1.040]**
$n(\hat{p}=0)=42$	0.991	0.026	[0.940, 1.040]
$n(\hat{p}=1)=9$	0.715	0.072	[0.670, 0.840]
cad	**0.904**	**0.198**	**[0.310, 1.080]**
$n(\hat{p}=0)=35$	1.020	0.029	[0.975, 1.080]
$n(\hat{p}=1)=16$	0.652	0.174	[0.310, 0.820]
chf	**0.962**	**0.210**	**[-0.040, 1.080]**
$n(\hat{p}=0)=47$	1.010	0.044	[0.935, 1.080]
$n(\hat{p}=3)=4$	0.362	0.428	[-0.040, 0.752]
gbp	**0.951**	**0.177**	**[0.105, 1.160]**
$n(\hat{p}=0)=46$	0.993	0.100	[0.855, 1.160]
$n(\hat{p}=1)=5$	0.558	0.256	[0.105, 0.705]
hkd	**0.894**	**0.191**	**[0.300, 1.040]**
$n(\hat{p}=0)=35$	1.000	0.024	[0.950, 1.040]
$n(\hat{p}=1)=16$	0.661	0.189	[0.300, 0.875]
nok	**0.952**	**0.165**	**[-0.060, 1.080]**
$n(\hat{p}=0)=45$	0.990	0.063	[0.855, 1.080]
$n(\hat{p}=1)=6$	0.668	0.357	[-0.060, 0.860]
nzd	**0.935**	**0.086**	**[0.670, 1.000]**
$n(\hat{p}=0)=41$	0.935	0.052	[0.785, 1.000]
$n(\hat{p}=1)=6$	0.718	0.050	[0.670, 0.805]
$n(\hat{p}=2)=4$	0.928	0.046	[0.880, 0.980]
sgd	**0.894**	**0.199**	**[-0.302, 1.040]**
$n(\hat{p}=0)=45$	0.950	0.068	[0.805, 1.040]
$n(\hat{p}=1)=5$	0.704	0.015	[0.695,0.730]
$n(\hat{p}=2)=3$	0.411	0.622	[-0.302, 0.840]

Table 4.6: Empirical coverage percentages of the 95%-k-step ahead prediction intervals from FARIMA, random walk (RW) and ARIMA predictions of stock market indices

stock index	model	k =1	2	3	4	5	6	7	8	9	10	11	12
DM series													
Australia	FARIMA	0.95	0.90	0.90	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	RW	1.00	0.95	0.81	0.90	0.90	0.95	1.00	0.90	0.90	0.90	0.86	0.86
	ARIMA	0.90	0.90	0.90	0.95	0.90	1.00	1.00	0.95	0.95	0.95	0.95	0.90
Belgium	FARIMA	0.90	0.86	0.86	0.90	0.90	0.90	0.90	0.90	0.95	1.00	1.00	1.00
	RW	0.90	0.81	0.86	0.86	0.90	0.90	0.90	0.86	0.76	0.86	0.81	0.81
	ARIMA	0.95	0.86	0.86	0.90	0.90	0.90	0.90	0.90	0.95	1.00	1.00	1.00
Canada	FARIMA	0.95	0.95	0.95	0.90	0.95	0.95	0.95	0.90	0.90	0.90	0.90	0.90
	RW	0.90	0.95	0.90	0.95	0.95	1.00	1.00	1.00	0.95	0.95	0.90	0.95
	ARIMA	0.95	0.90	0.90	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
France	FARIMA	1.00	0.95	0.95	0.90	1.00	1.00	1.00	0.95	0.95	0.95	1.00	1.00
	RW	1.00	0.95	0.90	0.90	0.95	1.00	1.00	0.95	1.00	1.00	1.00	1.00
	ARIMA	1.00	0.95	0.95	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Germany	FARIMA	0.95	0.90	0.95	1.00	1.00	1.00	1.00	0.95	0.90	1.00	1.00	1.00
	RW	1.00	0.90	0.95	1.00	1.00	1.00	0.95	0.86	0.95	1.00	1.00	0.95
	ARIMA	0.95	0.90	0.95	1.00	1.00	1.00	0.95	0.95	0.95	1.00	1.00	1.00
Hong Kong	FARIMA	0.95	0.90	0.95	0.95	0.95	0.95	0.90	0.81	0.90	0.90	0.86	0.86
	RW	0.95	0.90	0.95	0.95	0.90	0.86	0.86	0.86	0.86	0.86	0.81	0.81
	ARIMA	0.95	0.90	0.95	0.95	0.90	0.90	0.86	0.86	0.90	0.95	0.90	0.90
Italy	FARIMA	0.90	0.95	0.95	0.90	0.81	0.86	0.81	0.81	0.86	0.90	0.90	0.86
	RW	0.90	0.95	0.95	0.90	0.86	0.81	0.81	0.81	0.81	0.81	0.86	0.81
	ARIMA	0.90	0.95	0.95	0.90	0.86	0.86	0.90	0.95	0.90	0.95	0.90	0.90
Japan	FARIMA	0.81	0.86	0.86	0.90	0.95	0.95	0.95	0.90	0.90	0.86	0.86	0.86
	RW	0.81	0.86	0.81	0.86	0.90	0.90	0.95	0.86	0.86	0.86	0.86	0.81
	ARIMA	0.86	0.86	0.90	0.95	1.00	0.95	1.00	0.95	0.95	0.95	1.00	0.95
Switzerland	FARIMA	0.95	0.90	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	RW	0.90	0.86	0.90	0.90	0.90	0.86	0.86	0.81	0.81	0.86	0.86	0.81
	ARIMA	0.95	0.95	0.95	0.90	0.95	0.95	0.95	0.95	1.00	1.00	0.95	0.95
UK	FARIMA	0.95	0.95	0.90	0.90	0.90	0.90	0.90	0.95	0.95	0.95	0.95	0.95
	RW	0.90	0.95	0.90	0.86	0.86	0.86	0.71	0.76	0.86	0.81	0.81	0.86
	ARIMA	0.90	0.95	0.95	0.95	0.95	0.95	0.86	0.95	0.95	1.00	1.00	1.00
US	FARIMA	1.00	1.00	1.00	1.00	0.95	0.95	0.95	0.90	0.95	0.90	0.90	0.86
	RW	0.95	1.00	1.00	1.00	0.95	0.95	0.95	1.00	0.95	0.95	0.90	0.90
	ARIMA	1.00	1.00	1.00	1.00	0.95	0.95	0.95	1.00	0.95	0.95	0.95	0.90
EM series													
Brazil	FARIMA	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	1.00	0.95	0.95
	RW	0.95	0.86	0.81	0.81	0.71	0.62	0.57	0.57	0.52	0.57	0.52	0.48
	ARIMA	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Chile	FARIMA	0.95	0.95	0.90	0.86	0.90	0.86	0.81	0.76	0.81	0.81	0.81	0.86
	RW	0.90	0.86	0.86	0.86	0.86	0.76	0.67	0.57	0.52	0.62	0.71	0.67
	ARIMA	1.00	1.00	0.95	0.95	0.95	0.95	0.95	0.86	0.86	0.90	0.86	0.90
India	FARIMA	0.95	1.00	1.00	0.95	0.90	0.90	0.86	0.81	0.81	0.81	0.86	0.90
	RW	0.95	0.95	0.86	0.86	0.86	0.81	0.86	0.81	0.76	0.81	0.81	0.86
	ARIMA	0.95	1.00	1.00	1.00	0.95	0.86	0.90	0.86	0.90	0.95	0.90	0.95
Malaysia	FARIMA	1.00	1.00	0.95	0.86	0.95	0.90	0.81	0.76	0.86	0.90	0.90	0.90
	RW	1.00	0.90	0.90	0.81	0.86	0.86	0.86	0.76	0.76	0.90	0.95	0.90
	ARIMA	1.00	1.00	0.95	0.86	0.90	0.95	0.90	0.81	0.81	0.95	0.95	0.90
Mexico	FARIMA	0.86	0.90	0.90	0.81	0.90	0.90	0.95	0.90	0.86	0.81	0.86	0.86
	RW	0.86	0.81	0.81	0.76	0.71	0.57	0.62	0.67	0.67	0.67	0.62	0.76
	ARIMA	0.86	0.90	0.90	0.86	0.90	0.86	0.90	0.86	0.90	0.90	0.81	0.90
South Korea	FARIMA	0.95	0.95	1.00	0.95	0.90	0.95	0.90	0.86	0.90	0.90	0.90	0.90
	RW	0.95	0.95	0.95	1.00	1.00	1.00	1.00	0.90	1.00	0.95	1.00	0.95
	ARIMA	0.90	0.95	0.95	1.00	1.00	1.00	1.00	0.95	1.00	1.00	1.00	1.00
Thailand	FARIMA	1.00	0.95	0.95	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.95	0.95
	RW	1.00	0.90	0.95	1.00	0.95	0.95	0.95	0.95	0.90	0.95	0.90	0.95
	ARIMA	1.00	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.95

Table 4.7: Empirical coverage percentages of the 95%-k-step ahead prediction intervals from FARIMA, random walk (RW) and ARIMA predictions of foreign exchange rates

exchange rate	model	k =1	2	3	4	5	6	7	8	9	10	15	20	25
aud	FARIMA	0.98	0.98	0.96	0.98	1.00	0.94	0.94	0.98	0.96	0.90	0.96	0.94	0.92
	RW	0.98	0.98	0.98	1.00	1.00	0.96	0.98	0.98	0.98	1.00	0.96	0.98	0.98
	ARIMA	0.98	0.98	0.98	1.00	1.00	0.96	0.98	0.98	1.00	1.00	0.98	0.98	0.98
cad	FARIMA	0.86	0.88	0.96	0.98	0.96	0.94	0.94	0.94	0.94	0.90	0.92	0.92	0.88
	RW	0.88	0.88	0.96	0.98	0.96	0.94	0.92	0.94	0.94	0.92	0.96	0.98	0.94
	ARIMA	0.88	0.90	0.96	0.98	0.98	0.94	0.94	0.96	0.94	0.92	0.96	0.98	0.94
chf	FARIMA	0.96	0.98	0.98	0.98	0.96	0.96	0.94	0.96	0.98	0.98	0.94	0.92	0.96
	RW	0.94	0.98	0.96	0.98	0.98	0.94	0.94	0.98	0.98	1.00	0.98	0.98	0.96
	ARIMA	0.94	0.98	0.96	0.98	0.98	0.94	0.94	0.98	1.00	1.00	0.98	0.98	0.96
gbp	FARIMA	0.90	0.96	0.96	0.96	0.96	0.98	0.98	0.98	0.96	0.96	0.90	0.94	0.92
	RW	0.92	0.96	0.98	0.96	0.96	0.98	0.98	1.00	0.96	0.96	0.94	0.94	0.98
	ARIMA	0.92	0.98	0.98	0.96	0.98	0.98	0.98	1.00	0.96	0.96	0.94	0.96	0.98
hkd	FARIMA	0.88	0.88	0.96	0.98	0.94	0.94	0.94	0.92	0.92	0.90	0.92	0.92	0.86
	RW	0.88	0.90	0.96	0.98	0.96	0.94	0.92	0.94	0.94	0.94	0.96	0.98	0.94
	ARIMA	0.86	0.88	0.96	0.98	0.98	0.94	0.94	0.94	0.94	0.94	0.96	0.98	0.92
nok	FARIMA	0.98	1.00	1.00	0.98	0.98	0.92	0.94	0.96	1.00	1.00	0.94	0.90	0.96
	RW	0.98	1.00	1.00	0.98	0.98	0.92	0.92	0.98	1.00	1.00	0.94	0.94	0.96
	ARIMA	0.98	1.00	1.00	1.00	0.98	0.92	0.92	0.98	1.00	1.00	0.94	0.92	0.96
nzd	FARIMA	0.94	0.94	0.94	0.98	0.98	1.00	0.96	0.96	0.94	0.94	0.98	0.94	0.92
	RW	0.94	0.96	0.96	1.00	0.98	1.00	1.00	1.00	0.98	0.96	0.98	1.00	0.98
	ARIMA	0.94	0.98	0.96	1.00	0.98	1.00	1.00	1.00	0.96	0.98	0.98	0.98	0.98
sgd	FARIMA	0.90	0.98	0.96	0.98	0.98	0.94	0.94	1.00	0.94	0.92	0.94	0.90	0.90
	RW	0.90	0.96	0.96	0.98	0.96	0.94	0.96	0.98	0.94	0.92	0.98	0.96	0.94
	ARIMA	0.92	0.96	0.96	0.98	0.98	0.94	0.96	0.98	0.94	0.94	0.96	0.96	0.96

Table 4.8: Relative empirical mean squared errors of k-step ahead FARIMA, random walk (RW) and ARIMA predictions of stock market indices

stock index	k =1	2	3	4	5	6	7	8	9	10	11	12
DM series												
Australia												
FARIMA/RW	0.948	0.989	1.043	1.144	1.100	1.026	0.985	1.018	1.005	0.971	1.009	1.012
FARIMA/ARIMA	0.982	1.068	1.065	1.070	1.040	1.010	0.964	0.970	0.968	0.955	1.002	1.016
Belgium												
FARIMA/RW	0.933	0.982	1.022	1.101	1.084	1.094	1.130	1.160	1.159	1.173	1.206	1.262
FARIMA/ARIMA	1.051	1.042	1.057	1.092	1.078	1.090	1.115	1.128	1.132	1.143	1.173	1.221
Canada												
FARIMA/RW	0.875	1.021	1.009	1.016	1.287	1.769	2.033	1.426	1.405	1.441	1.359	1.400
FARIMA/ARIMA	0.989	1.000	1.009	1.183	1.461	1.801	2.158	1.389	1.363	1.447	1.377	1.468
France												
FARIMA/RW	1.146	1.136	1.010	0.930	0.946	0.945	0.981	0.995	1.030	1.064	1.064	0.993
FARIMA/ARIMA	1.078	1.064	0.965	0.910	0.938	0.945	0.976	0.984	1.008	1.039	1.028	0.961
Germany												
FARIMA/RW	0.998	1.011	1.020	1.050	1.021	0.998	1.017	1.041	1.065	1.085	1.108	1.084
FARIMA/ARIMA	0.944	0.964	0.975	1.015	0.992	0.965	0.985	1.017	1.044	1.071	1.095	1.072
Hong Kong												
FARIMA/RW	1.107	1.156	0.937	1.020	1.002	0.996	1.001	1.053	1.116	1.037	1.055	1.001
FARIMA/ARIMA	1.171	1.091	0.910	0.949	0.904	0.938	0.968	1.021	1.083	1.013	1.029	0.979
Italy												
FARIMA/RW	1.028	0.993	1.452	1.512	1.474	1.518	1.503	1.461	1.383	1.270	1.274	1.284
FARIMA/ARIMA	1.061	1.069	1.342	1.368	1.384	1.453	1.445	1.428	1.384	1.271	1.261	1.265
Japan												
FARIMA/RW	1.135	1.130	1.024	0.986	1.003	1.113	1.081	1.109	1.286	1.347	1.347	1.289
FARIMA/ARIMA	1.132	1.100	1.058	0.985	0.975	1.104	1.076	1.103	1.245	1.308	1.317	1.248
Switzerland												
FARIMA/RW	0.965	0.994	1.096	1.100	1.032	1.019	1.023	1.006	0.987	0.983	1.014	1.037
FARIMA/ARIMA	1.084	1.070	1.142	1.162	1.113	1.088	1.093	1.069	1.043	1.044	1.082	1.106
UK												
FARIMA/RW	0.846	0.965	0.970	0.919	0.895	0.898	0.941	1.029	1.052	1.045	1.058	1.099
FARIMA/ARIMA	1.030	0.996	0.991	0.968	0.956	0.947	0.983	1.056	1.085	1.098	1.110	1.150
US												
FARIMA/RW	0.833	0.873	0.897	0.761	0.849	0.936	0.907	1.071	1.033	0.824	0.804	0.736
FARIMA/ARIMA	0.939	0.946	0.967	0.846	0.873	0.945	0.939	1.092	1.070	0.887	0.865	0.779
EM series												
Brazil												
FARIMA/RW	0.636	0.451	0.397	0.329	0.363	0.376	0.336	0.287	0.326	0.286	0.310	0.297
FARIMA/ARIMA	0.850	0.665	0.589	0.492	0.513	0.537	0.475	0.414	0.462	0.402	0.422	0.392
Chile												
FARIMA/RW	0.821	0.923	0.922	0.971	0.945	0.886	0.917	0.989	1.011	0.992	0.968	0.982
FARIMA/ARIMA	1.188	1.191	1.188	1.159	1.118	1.017	1.025	1.083	1.085	1.055	1.023	1.035
India												
FARIMA/RW	1.075	1.080	0.936	1.131	1.277	1.410	1.370	1.237	1.290	1.320	1.251	1.187
FARIMA/ARIMA	1.054	1.102	1.030	1.178	1.291	1.381	1.360	1.225	1.275	1.302	1.237	1.173
Malaysia												
FARIMA/RW	1.238	1.108	1.017	1.039	1.048	1.051	1.031	1.046	1.110	1.090	1.104	1.062
FARIMA/ARIMA	1.158	1.063	1.018	1.015	0.993	0.998	0.981	1.003	1.067	1.063	1.084	1.041
Mexico												
FARIMA/RW	0.822	0.894	0.913	0.832	0.842	0.868	0.863	0.953	1.004	1.013	1.021	0.976
FARIMA/ARIMA	0.942	0.896	0.937	0.919	0.920	0.940	0.929	1.004	1.051	1.060	1.076	1.039
South Korea												
FARIMA/RW	1.087	1.171	1.229	1.417	1.653	1.752	1.706	1.548	1.780	1.721	1.751	1.709
FARIMA/ARIMA	1.083	1.177	1.248	1.403	1.576	1.681	1.633	1.512	1.703	1.652	1.685	1.667
Thailand												
FARIMA/RW	0.781	0.953	1.080	0.885	0.779	0.740	0.775	0.791	0.867	0.991	1.033	0.999
FARIMA/ARIMA	1.014	1.072	1.136	0.942	0.870	0.832	0.850	0.859	0.926	1.032	1.063	1.017

Table 4.9: Relative empirical mean squared errors of k-step ahead FARIMA, random walk (RW) and ARIMA predictions of foreign exchange rates

exchange rate	$k=1$	2	3	4	5	6	7	8	9	10	15	20	25
aud													
FARIMA/RW	1.001	1.056	1.112	1.134	1.181	1.151	1.156	1.154	1.187	1.224	1.118	1.127	1.107
FARIMA/ARIMA	0.972	1.050	1.092	1.145	1.175	1.138	1.142	1.145	1.175	1.225	1.125	1.137	1.117
cad													
FARIMA/RW	1.007	0.997	1.047	1.091	1.070	1.090	1.122	1.102	1.081	1.054	1.048	1.097	1.040
FARIMA/ARIMA	1.006	0.995	1.035	1.076	1.063	1.083	1.116	1.099	1.078	1.054	1.054	1.101	1.041
chf													
FARIMA/RW	1.069	1.048	1.080	1.098	1.089	1.123	1.129	1.157	1.103	1.113	1.198	1.181	1.091
FARIMA/ARIMA	1.066	1.036	1.056	1.082	1.066	1.111	1.126	1.149	1.095	1.106	1.186	1.166	1.089
gbp													
FARIMA/RW	0.999	0.981	0.999	1.041	1.091	1.054	1.089	1.074	1.002	1.009	1.150	1.164	1.214
FARIMA/ARIMA	1.031	1.004	1.013	1.038	1.079	1.050	1.078	1.062	1.000	1.006	1.143	1.162	1.213
hkd													
FARIMA/RW	1.008	1.029	1.063	1.137	1.099	1.118	1.141	1.124	1.089	1.064	1.039	1.071	1.020
FARIMA/ARIMA	1.007	1.020	1.044	1.107	1.091	1.116	1.122	1.112	1.086	1.057	1.044	1.100	1.033
nok													
FARIMA/RW	1.066	1.130	1.189	1.163	1.140	1.120	1.106	1.126	1.098	1.130	1.118	1.151	1.099
FARIMA/ARIMA	1.032	1.091	1.119	1.140	1.109	1.106	1.097	1.112	1.086	1.122	1.111	1.149	1.097
nzd													
FARIMA/RW	0.907	0.963	1.013	1.026	1.035	1.044	1.064	1.088	1.003	0.941	0.949	1.026	0.996
FARIMA/ARIMA	1.006	0.992	1.033	1.049	1.047	1.047	1.057	1.083	1.025	0.964	0.955	1.030	0.997
sgd													
FARIMA/RW	1.048	1.033	1.059	0.995	0.984	1.005	1.061	1.075	0.999	0.964	1.037	1.029	0.888
FARIMA/ARIMA	1.025	1.030	1.072	1.011	1.001	1.015	1.066	1.075	1.006	0.968	1.033	1.031	0.894

Figure 4.1: Stock market indices between January 4, 1988 and November 15, 1990 (log-transformed weekly averages)

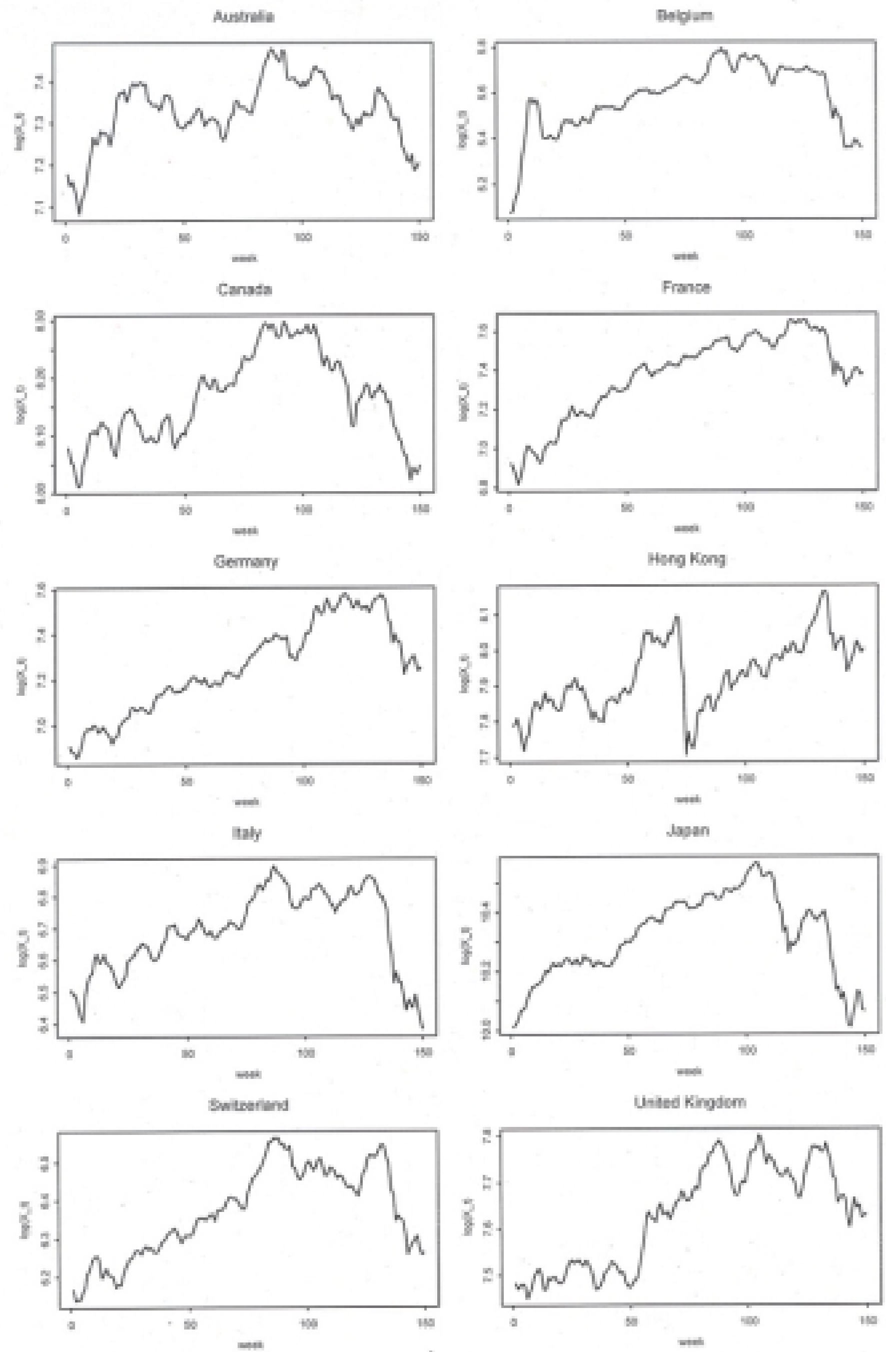


Figure 4.2: Stock market indices between January 4, 1988 and November 15, 1990 (log-transformed weekly averages)

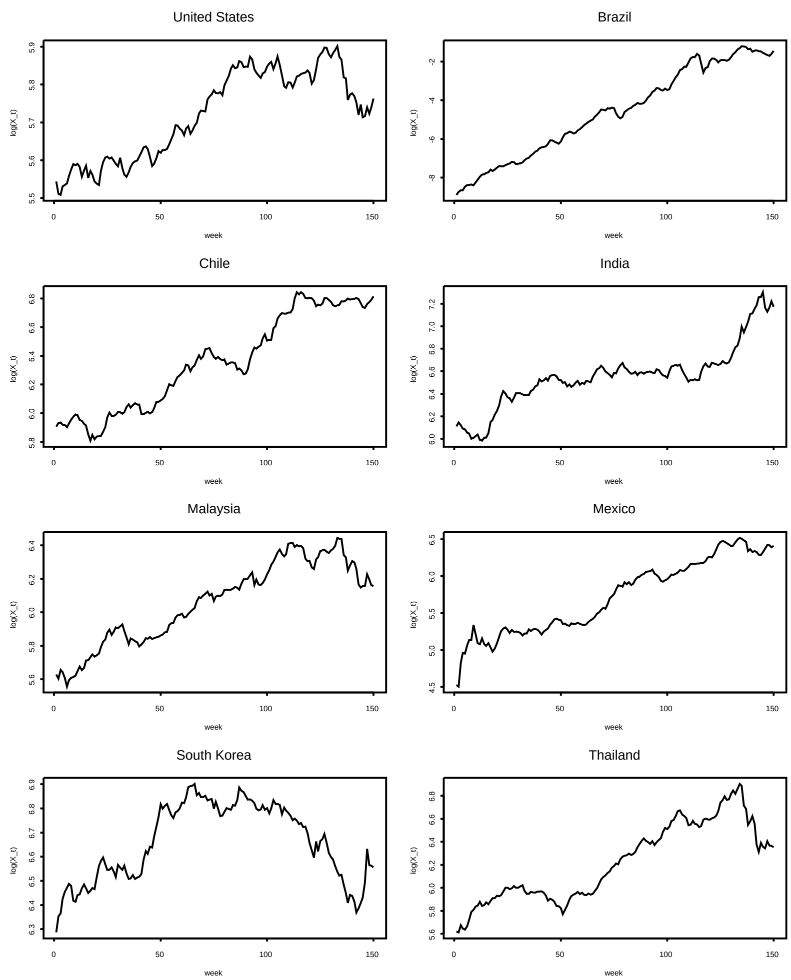


Figure 4.3: Foreign exchange rates between January 2, 1991 and December 2, 1991 (log-transformed daily closing values)

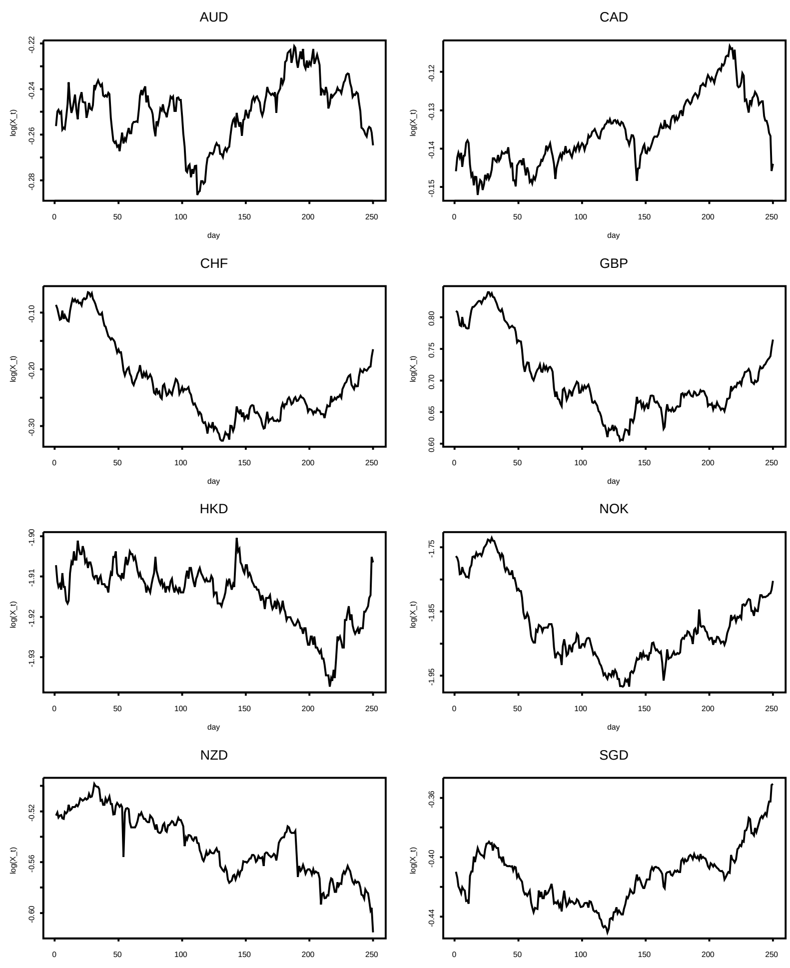


Figure 4.4: Autocorrelations of FARIMA-residuals of the stock market indices between January 4, 1988 and November 15, 1990

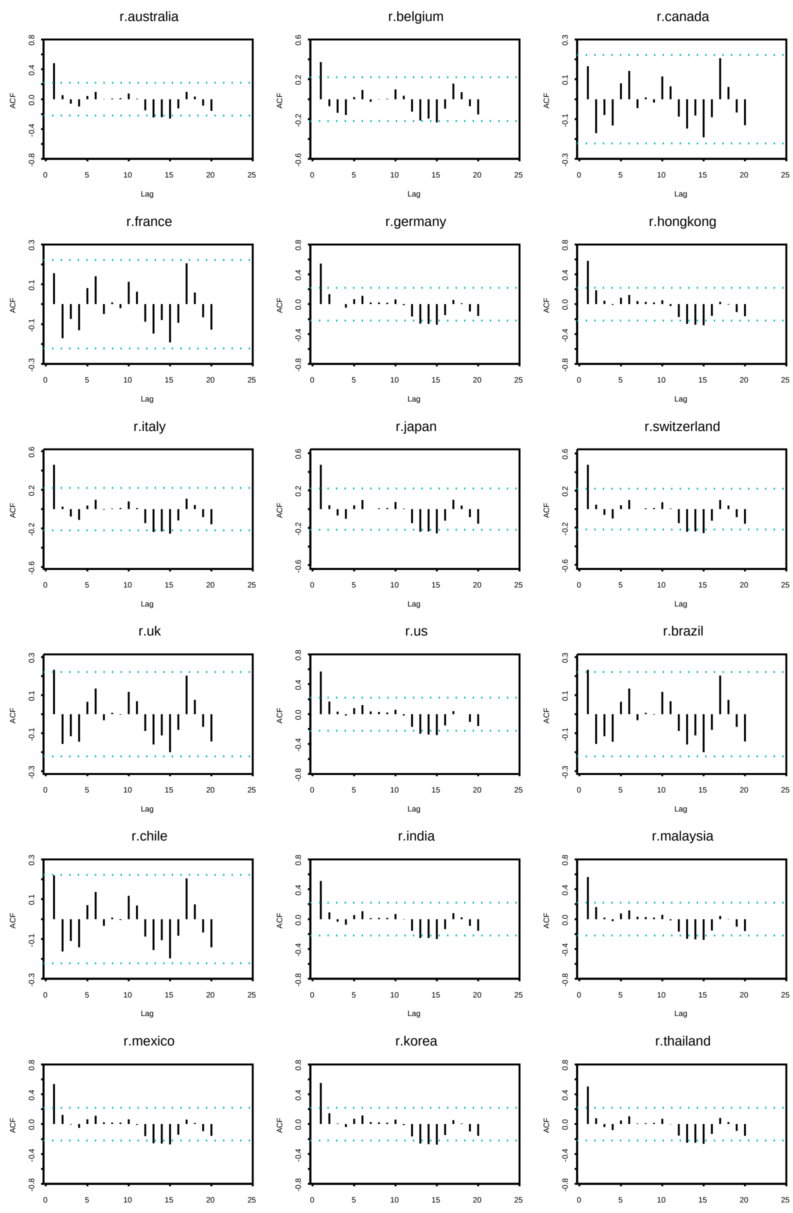


Figure 4.5: Autocorrelations of FARIMA-residuals of the exchange rates between January 2, 1991 and December 2, 1991

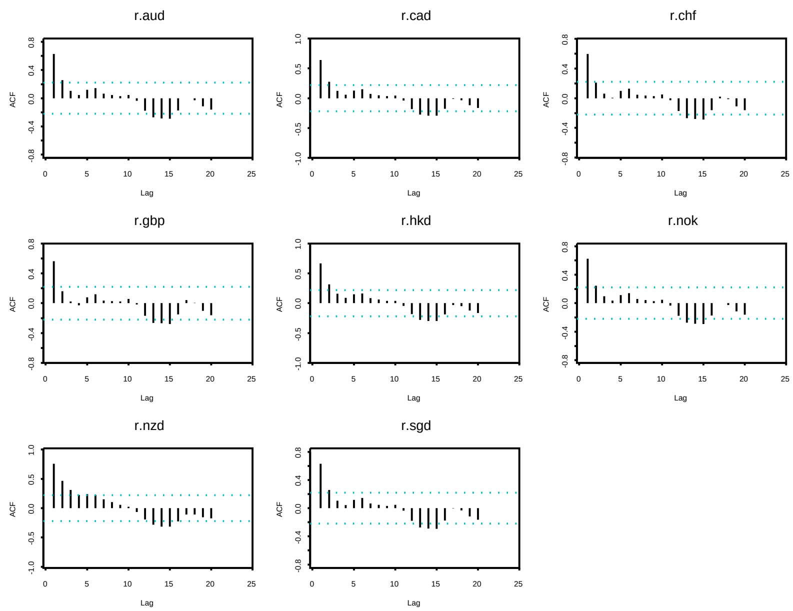


Figure 4.6: Normal probability plots of FARIMA-residuals of the stock market indices between January 4, 1988 and November 15, 1990

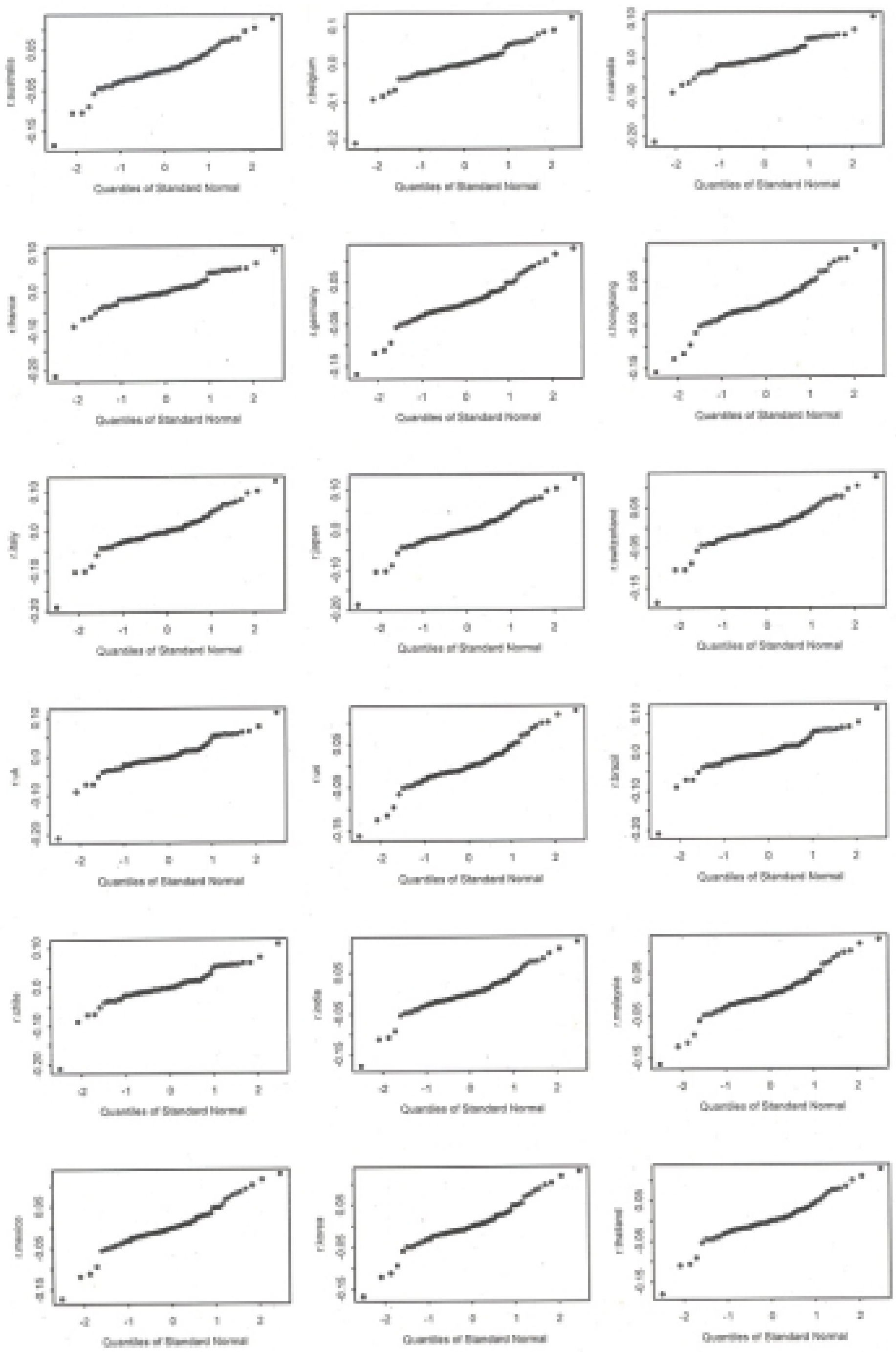


Figure 4.7: Normal probability plots of FARIMA-residual of the exchange rates between January 2, 1991 and December 2, 1991

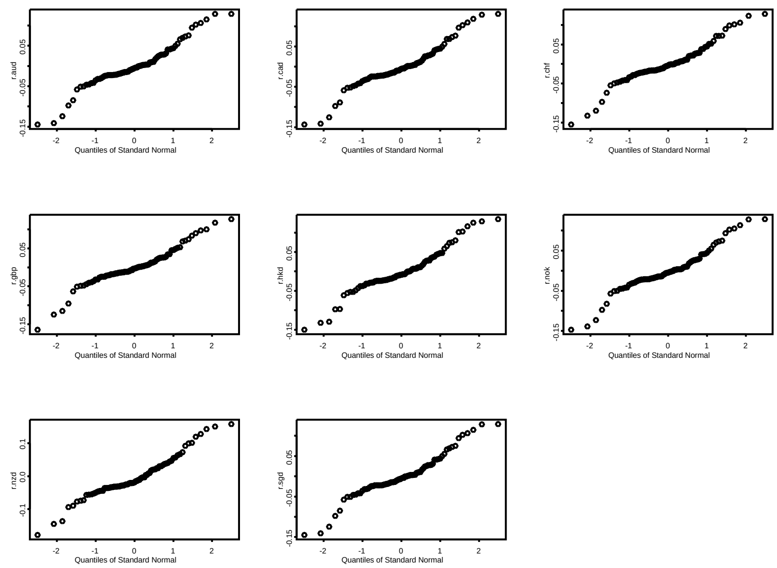


Figure 4.8: Autocorrelations of squared FARIMA-residuals of the stock market indices between January 4, 1988 and November 15, 1990

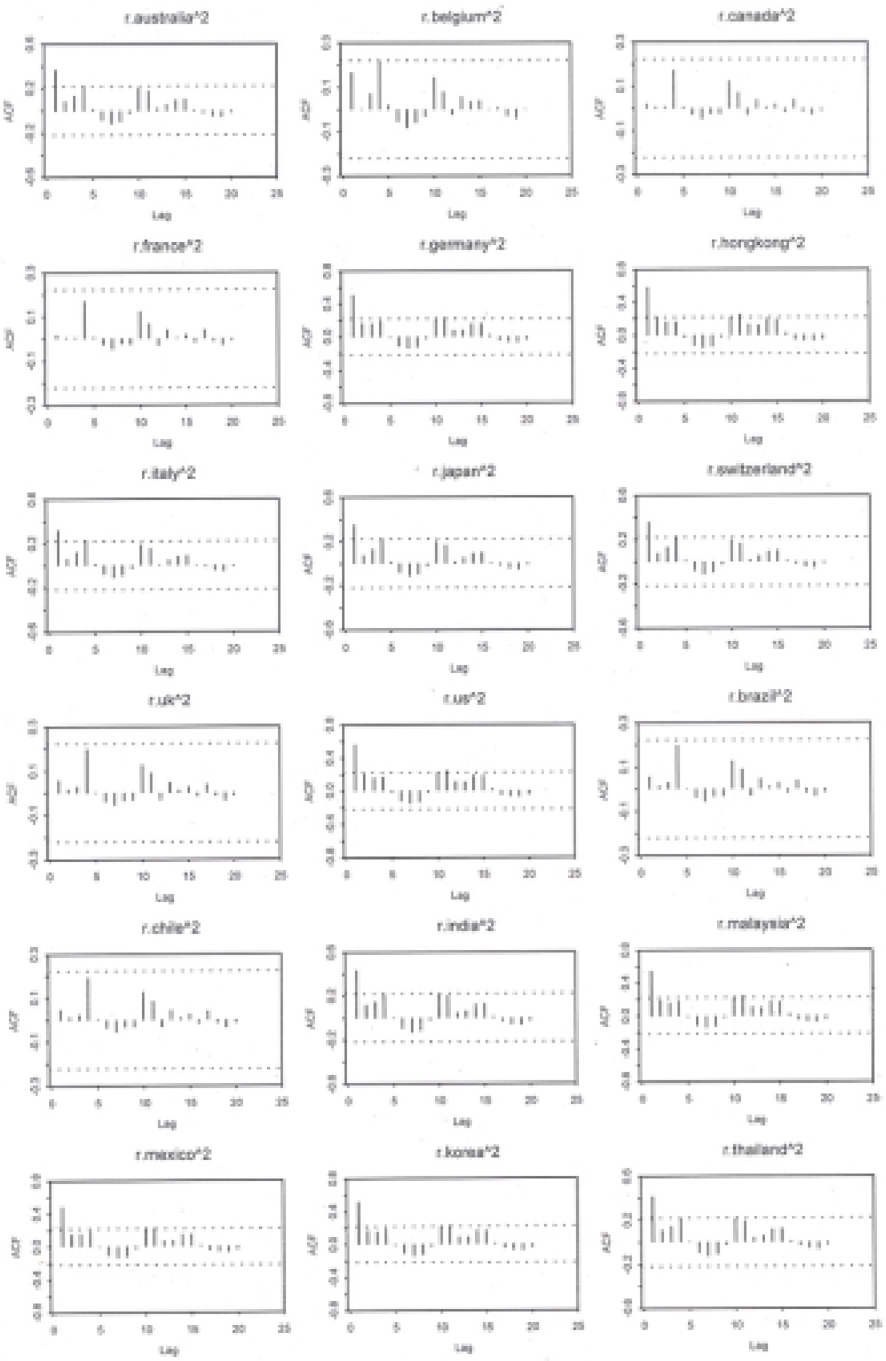


Figure 4.9: Autocorrelations of squared FARIMA-residuals of the exchange rates between January 2, 1991 and December 2, 1991

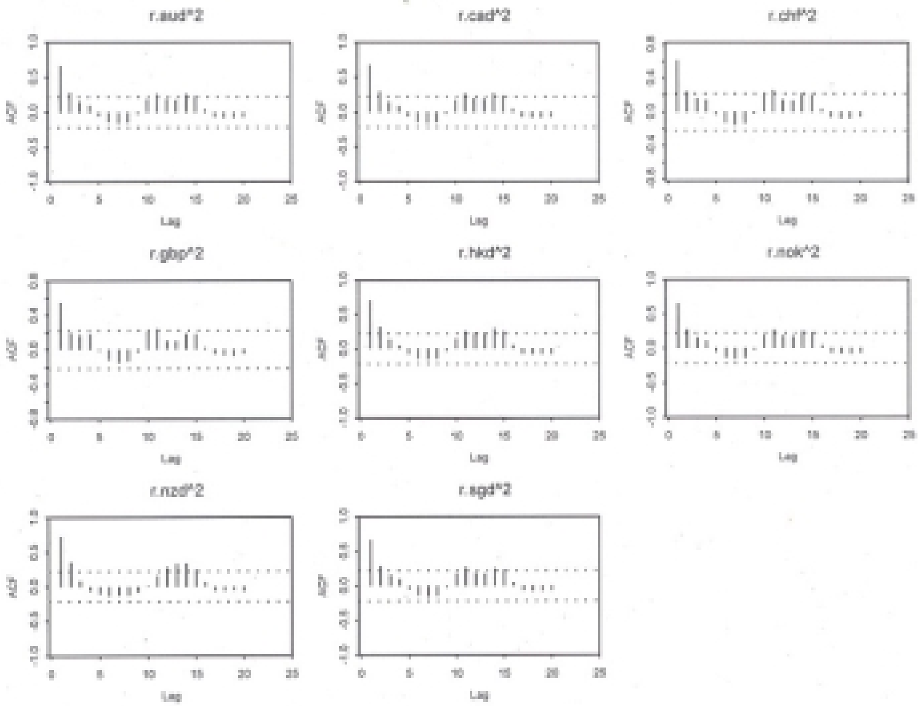


Figure 4.10: Ratio of length of FARIMA, random walk (RW) and ARIMA prediction intervals of stock market indices

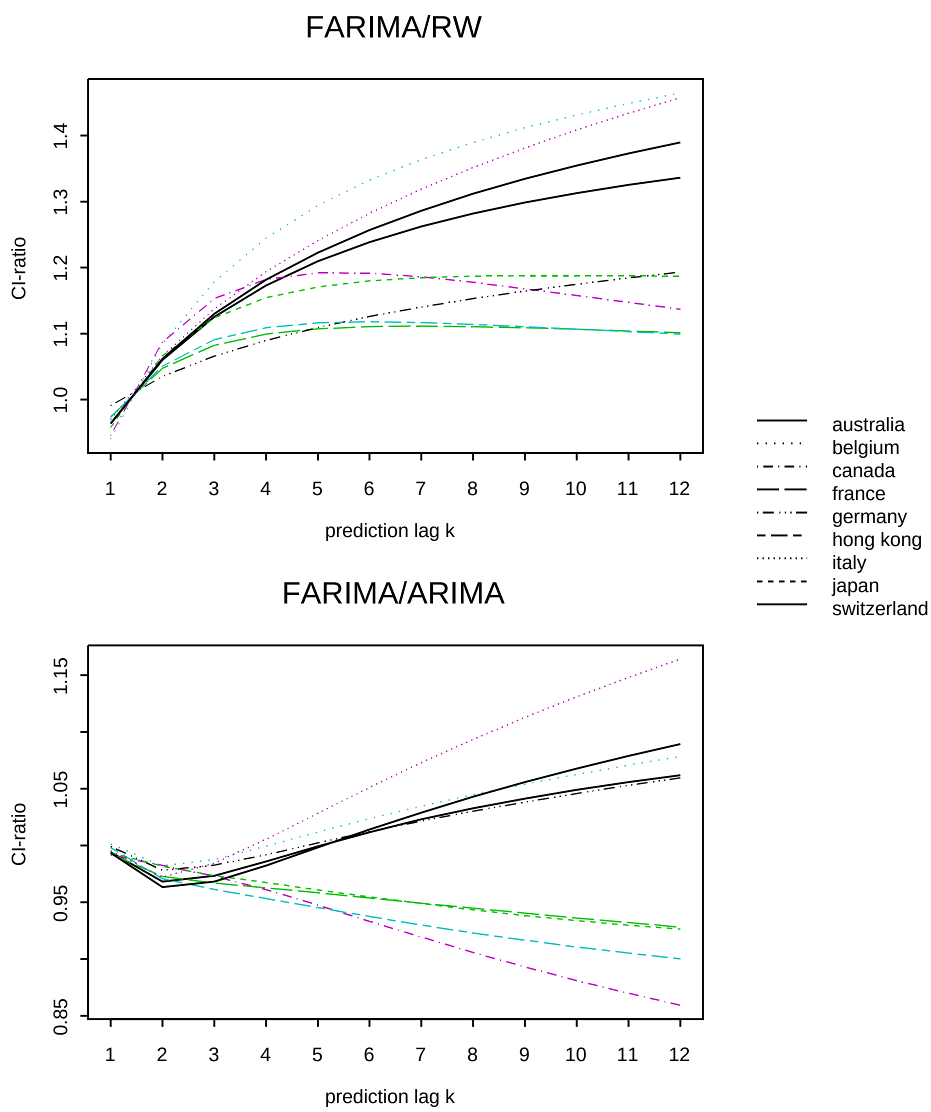


Figure 4.11: Ratio of length of FARIMA, random walk (RW) and ARIMA prediction intervals of stock market indices

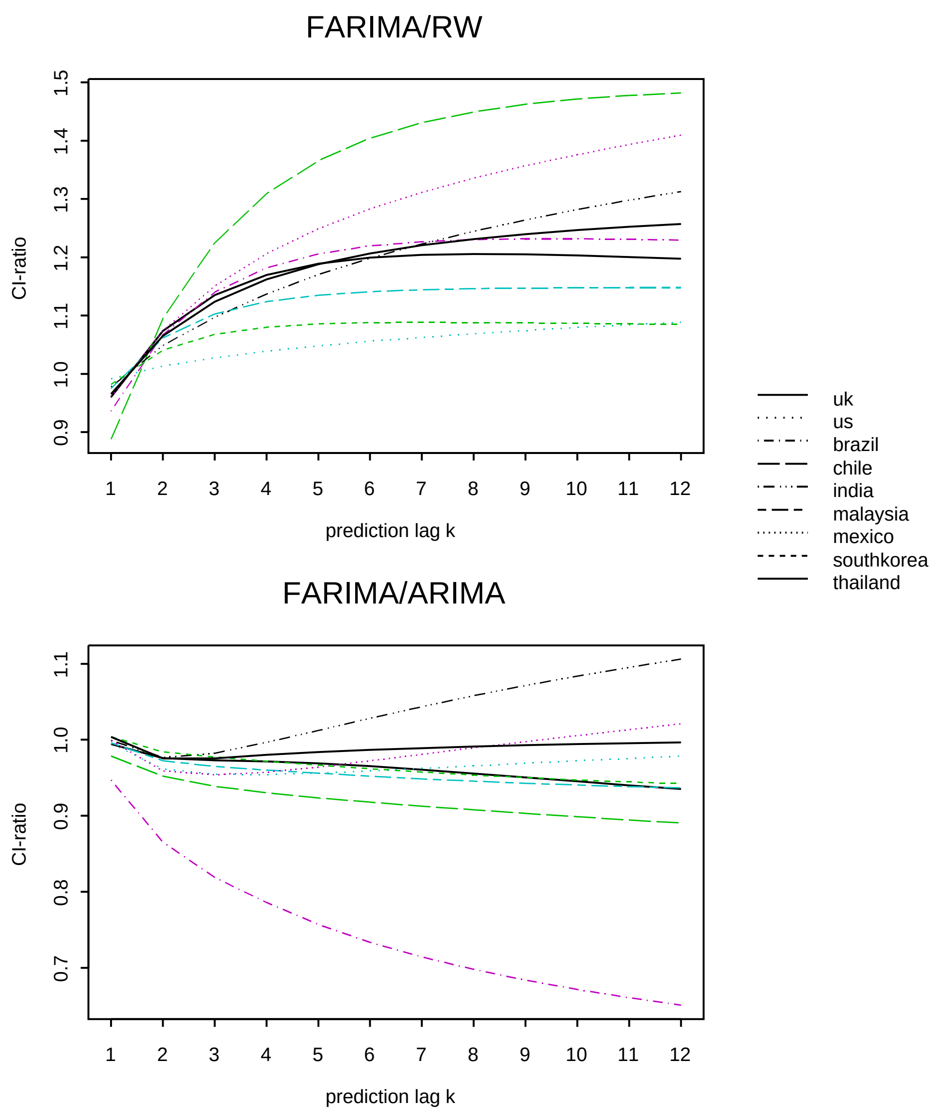
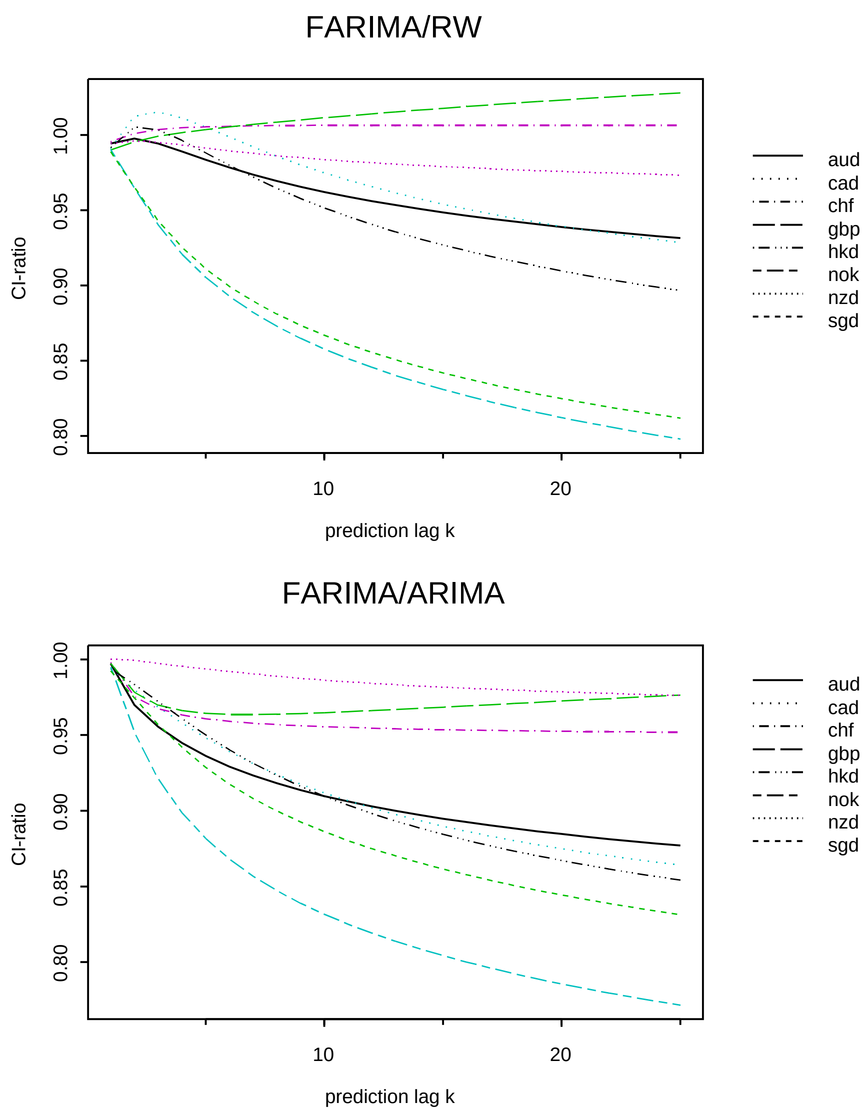


Figure 4.12: Ratio of length of FARIMA, random walk (RW) and ARIMA prediction intervals of exchange rates



Chapter 5

Temporal aggregation of FARIMA($p, d, 0$) models

5.1 Introduction

For many practitioners data are available in aggregated form (temporal aggregation). Typical examples are flow variables, such as industrial production, which exist only through aggregation over a certain time interval. Also, applied data analysts sometimes prefer to analyze aggregated data in order to eliminate seasonal fluctuations and for long-term forecasts. Here, we consider the problem of temporal aggregation for a class of parametric time series models that includes classical nonfractional Box-Jenkins ARIMA($p, m, 0$) models as well as stationary and nonstationary fractional autoregressive processes, namely the stationary and nonstationary FARIMA($p, d, 0$) model (Beran [10], Beran, Bhansali and Ocker [12])

$$\phi(B)(1-B)^{\delta}\{(1-B)^{m}X_t - \mu\} = \epsilon_t. \tag{5.1}$$

The impact of temporal aggregation on nonfractional ARIMA($p, m, 0$) models are well known (see e.g. Amemiya and Wu [4], Stram and Wei [78], Tiao [82], Wei [85]). Also, related results follow directly from functional limit theorems for stationary fractional processes (see Lamperti [60], Taqqu [79] and others). Nothing is, however, known about the effects upon nonstationary fractional models. Many time series, in particular in economics, are nonstationary and appear to have long-range correlations. It is therefore important to know how a process changes when it is aggregated.

The purpose of this chapter is to derive the asymptotic behaviour of temporal aggregates of an observed time series x_t generated by the FARIMA model (5.1). All results on temporal aggregation of nonfractional Box-Jenkins ARIMA$(p, m, 0)$, stationary fractional FARIMA$(p, \delta, 0)$ as well as (possibly) nonstationary fractional FARIMA$(p, d, 0)$ models can be derived using one unified approach. The following definition of temporal aggregation is used.

DEFINITION 1 *Let $t = sT$ and $s \geq 2$, then the series*

$$z_T = \left(\sum_{i=0}^{s-1} B^i\right) x_{sT},$$

represents the s-period non-overlapping aggregates of x_t.

Tiao [82] has shown for Box-Jenkins ARIMA processes that, as the degree of aggregation s tends to infinity, the series z_T approaches an ARIMA$(0, m, m)$ model, irrespective of the ARMA order and the size of the coefficients. Thus the integer differencing parameter m determines the asymptotic process. The question arises whether a similar result holds in connection with fractional differencing.

The chapter is organized as follows. The main result is given in section 2. Simulations and data examples in section 3 illustrate our findings. Some concluding remarks are given in section 4. Figures and proofs are provided in the appendix.

5.2 Asymptotic models

To simplify presentation, the moving average order q is chosen to be equal to zero in the following. The same results hold for $q \neq 0$. Also, without loss of generality, we assume μ to be known and equal to zero. Since a FARIMA$(p, d, 0)$ process has an infinite moving-average representation for $\delta < .5$ (Hosking [52]), we can express (5.1) by

$$(1 - B)^m x_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}, \tag{5.2}$$

where ψ_k is obtained by inverting (5.1). In order to derive the process for the aggregates z_T, we adapt Telser's [81] technique and multiply (5.2) by

$(\sum_{i=0}^{s-1} B^i)^{m+1}$. Using progression we get

$$(1 - B^s)^m \left(\sum_{i=0}^{s-1} B^i \right) x_t = \left(\sum_{i=0}^{s-1} B^i \right)^{m+1} u_t, \tag{5.3}$$

where $u_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}$. Now let

$$y_T = (1 - B^s)^m \left(\sum_{i=0}^{s-1} B^i \right) x_{sT} = (1 - B)^m z_T, \tag{5.4}$$

where B operates now on T by $Bz_T = z_{T-1}$. That is, y_T is the mth difference of the aggregate z_T. The covariances of the process y_T, $\gamma_y(k) = Cov(y_T, y_{T+k})$, are specified in the following lemma for $m = 0$ and 1.

LEMMA 1 *The kth autocovariance of the series $y_T = (1 - B)^m z_T$ equals*

(i) $\gamma_y(k) = \sum_{j,l=0}^{s-1} \gamma_u(j - l - sk)$, *if* $m = 0$,

(ii) $\gamma_y(k) = \sum_{j,l=-(s-1)}^{s-1} (s - |j|)(s - |l|)\gamma_u(j - l - sk)$, *if* $m = 1$,

where $\gamma_u(j - l - sk) = Cov(u_{i(s,T)+j}, u_{i(s,T+k)+l})$, *with* $i(s, T) = s(T - 1) + 1$, *for* $m = 1$, *and* $\gamma_u(j-l-sk) = Cov(u_{sT-l}, u_{s(T+k)-j})$, *for* $m = 0$, *respectively.*

Lemma 1 enables us to derive explicit correlation formulas. In particular in the following theorem it is shown that the correlations of y_T, $\rho_y(k) = \gamma_y(k)/\gamma_y(0)$, depend exclusively on the differencing parameter $d = m + \delta$. This is a generalization of Tiao's [82] result.

THEOREM 5 *The asymptotic formulas, as s tends to infinity, for the covariances and correlations of the series $y_T = (1 - B)^m z_T$ are as follows,*

(i) *if* $d = m = 0$, *then*
$\gamma_y(0) = O(s)$, $\gamma_y(k) = o(s), \forall k \geq 1$, *and*
$\lim_{s\to\infty} \rho_y(k) = 0, \forall k \geq 1$;

(ii) *if* $d = m = 1$, *then*
$\gamma_y(k) = O(s^3), k = 0, 1$, $\gamma_y(k) = o(s), \forall k \geq 2$, *and*
$\lim_{s\to\infty} \rho_y(1) = .25$, $\lim_{s\to\infty} \rho_y(k) = 0, \forall k \geq 2$;

(iii) *if* $d = \delta \in (-.5; .5) \setminus \{0\}$, *then*
$\gamma_y(k) = O(s^{1+2\delta}), \forall k \geq 0$, *and*
$\lim_{s\to\infty} \rho_y(k) = \frac{1}{2}\left\{(k+1)^{2\delta+1} - 2k^{2\delta+1} + (k-1)^{2\delta+1}\right\}, \forall k \geq 1,$
$\rho_y(k) \sim \delta(2\delta+1)k^{2\delta-1}$, *as* $k \to \infty$;

(iv) *if* $d = m + \delta \in (.5; 1.5) \setminus \{1\}$, *then*
$\gamma_y(k) = O(s^{3+2\delta}), \forall k \geq 0$, *and*
$\lim_{s\to\infty} \rho_y(1) = \frac{2^{2\delta+5} - 7 - 3^{2\delta+3}}{8 - 2^{2\delta+4}},$
$\lim_{s\to\infty} \rho_y(k) = \frac{-(k+2)^{2\delta+3} + 4(k+1)^{2\delta+3} - 6k^{2\delta+3} + 4(k-1)^{2\delta+3} - (k-2)^{2\delta+3}}{8 - 2^{2\delta+4}}, \forall k \geq 2,$
$\rho_y(k) \sim \frac{\delta(2\delta+1)(2\delta+2)(2\delta+3)}{(2^{2\delta+3} - 4)} k^{2\delta-1}$, *as* $k \to \infty$.

Remarks:

1. The limiting model of the aggregate y_T (equation (5.4)) of a nonfractional ARIMA process (5.1) (i.e. $\delta = 0$) becomes white noise, for $d = m = 0$, and approaches an MA(1) process, for $d = m = 1$, respectively. This result was first obtained by Tiao [82] using a different approach.

2. When the basic series x_t follows a stationary FARIMA model (5.1), with $d = \delta \in (-.5, .5) \setminus \{0\}$, then the asymptotic process for the aggregate y_T is fractional Gaussian noise (Mandelbrot and van Ness [66], Mandelbrot and Wallis [67]). That is, long memory and antipersistence respectively remain asymptotically whereas short memory model components vanish. This result can also be proved directly by using functional limit theorems for stationary fractional processes (see e.g. Lamperti [60] and Taqqu [79]).

3. When the basic series x_t follows a nonstationary FARIMA$(p, d, 0)$ process, with $d = m + \delta \in (.5, 1.5) \setminus \{1\}$, then the aggregated series y_T becomes a stationary Gaussian process with correlations following the same power law as for the series $x_t - x_{t-1}$. That is, temporal aggregation of a nonstationary fractional autoregressive process (5.1) does not change the property of long-range dependence and antipersistence in the increment process. Interestingly, the process does not converge to fractional Gaussian noise. This is similar to standard integrated ARIMA processes ($\delta = 0$, $m = 1$), where the first difference of the

limiting process has a lag-1 correlation of .25. Note also that the correlation formulas exhibit a continuous transition between $\delta \neq 0$ and $\delta = 0$ (where in the latter case $\rho(1) = .25$ and $\rho(k) = 0, \forall k \geq 2$).

5.3 Simulations and data examples

5.3.1 Results of the simulations

For d equal to $-.3, 0, .3, .7, 1.0, 1.3$, and sample size $n = 40,000$, one hundred series of each of the following four autoregressive models were simulated:

- Model a: $(1 + .9B)(1 - B)^{\delta}(1 - B)^{m} X_t = \epsilon_t$;
- Model b: $(1 - B)^{\delta}(1 - B)^{m} X_t = \epsilon_t$;
- Model c: $(1 - .9B)(1 - B)^{\delta}(1 - B)^{m} X_t = \epsilon_t$;
- Model d: $(1 - 1.42B + .73B^2)(1 - B)^{\delta}(1 - B)^{m} X_t = \epsilon_t$.

'Model c' has a very strong positive short memory and may converge quite slowly to the limiting model with increasing degree of temporal aggregation s. In contrast, 'Model b' is expected to converge quite fast under temporal aggregation. 'Model a' exhibits short-range correlations with alternating signs. 'Model d' has a local maximum at a nonzero frequency implying random short-range periodicities.

Note that simulation by the S-Plus function *arima.fracdiff.sim* poses computational problems for large sample sizes. To generate our simulations we used the function *simARMA0* in Beran [9] to simulate the fractional ARIMA$(0, \delta, 0)$ processes and applied *arima.sim* as a linear filter, together with *cumsum* for $d \geq .5$.

Figures 5.1 and 5.2 (provided in the appendix) display for models a to d, and k fixed equal to 1 and 2 respectively, the simulated average sample correlations $\bar{\rho}_{y,s}(k) = \frac{1}{100}\sum_{i=1}^{100} \hat{\rho}^{i}_{y,s}(k)$, plotted against an increasing degree of aggregation $s = 1, 2, 5, 10, 20, ..., 100$. The straight lines mark the corresponding limiting correlation according to theorem 5. Note that, sample autocorrelations of fractional models are biased (Hosking [53]), and the bias decays very slowly as the sample size increases. Therefore, to avoid an effect of the sample size on the bias, the correlations are estimated using series of a fixed length of 400 for each aggregation level s. Overall, we can observe a slow convergence for 'Model c' to the corresponding limiting values, and a systematic empirical negative bias in the long memory cases $d = .3$ and $d = 1.3$. The observed bias is, however, consistent with the one expected by theory. To illustrate this, consider table 5.1, where the theoretical limiting values of $\rho_y(k)$ (given in theorem 5), the empirical average sample correlation $\bar{\rho}_{y,s}(k)$ for the models a though d (given an aggregation degree of $s = 100$), the corresponding empirical bias ($\hat{b} = \bar{\rho}_{y,s}(k) - \rho_y(k)$) and the theoretical bias (for $n = 400$) according to Hosking [53], i.e.

$$b \sim \begin{cases} (\rho_y(k) - 1)n^{2\delta-1}, & m = 0 \\ (\rho_y(k) - 1)\frac{(2\delta+2)(2\delta+3)}{2^{2\delta+3}-4}n^{2\delta-1}, & m = 1 \end{cases}, \tag{5.5}$$

are listed for the long memory cases $d = .3$ and $d = 1.3$.

Table 5.1: Empirical vs. theoretical bias

	$d = .3$					$d = 1.3$				
	$\rho_y(k)$	$\bar{\rho}_y(k)$	b	$\hat{b}$	$b/\hat{b}$	$\rho_y(k)$	$\bar{\rho}_y(k)$	b	$\hat{b}$	$b/\hat{b}$
Model 1										
$k = 1$	.516	.457	-.044	-.059	.746	.658	.618	-.036	-.040	.900
2	.368	.292	-.058	-.076	.763	.431	.359	-.060	-.072	.833
5	.253	.172	-.068	-.081	.840	.292	.205	-.074	-.087	.851
10	.191	.120	-.074	-.071	1.042	.220	.123	-.082	-.097	.845
15	.163	.080	-.076	-.083	.916	.187	.085	-.085	-.102	.833
20	.145	.062	-.078	-.083	.940	.167	.053	-.087	-.114	.763
Model 2										
$k = 1$	.516	.466	-.044	-.050	.880	.658	.603	-.036	-.055	.655
2	.368	.303	-.058	-.065	.892	.431	.342	-.060	-.089	.674
5	.253	.177	-.068	-.076	.895	.292	.185	-.074	-.107	.692
10	.191	.112	-.074	-.079	.937	.220	.109	-.082	-.111	.739
15	.163	.079	-.076	-.084	.905	.187	.076	-.085	-.111	.766
20	.145	.052	-.078	-.093	.839	.167	.043	-.087	-.124	.702
Model 3										
$k = 1$	.516	.486	-.044	-.030	1.467	.658	.622	-.036	-.036	1.000
2	.368	.308	-.058	-.060	.967	.431	.366	-.060	-.065	.923
5	.253	.176	-.068	-.077	.883	.292	.206	-.074	-.086	.861
10	.191	.105	-.074	-.086	.861	.220	.129	-.082	-.091	.901
15	.163	.075	-.076	-.088	.864	.187	.092	-.085	-.095	.895
20	.145	.057	-.078	-.088	.886	.167	.066	-.087	-.101	.861
Model 4										
$k = 1$	.516	.464	-.044	-.052	.846	.658	.604	-.036	-.054	.667
2	.368	.306	-.058	-.062	.936	.431	.346	-.060	-.085	.706
5	.253	.179	-.068	-.074	.919	.292	.192	-.074	-.100	.740
10	.191	.109	-.074	-.082	.902	.220	.118	-.082	-.102	.804
15	.163	.085	-.076	-.078	.974	.187	.074	-.085	-.113	.752
20	.145	.055	-.078	-.090	.867	.167	.052	-.087	-.115	.756

The empirical bias is always negative, as is expected by equation (5.5), and slightly larger than the theoretical one. The difference is, however, quite small as can be seen from the ratio $b/\hat{b}$. Overall, around 80% of the observed bias are explained by equation (5.5), supporting the practical relevance of theorem 5.

5.3.2 Results for the data examples

As empirical examples we consider two nominal stock market indices, namely the Deutsche Aktienindex (DAX) and the Hang Seng Index. Figure 5.3 displays the logarithm of weekly and monthly averages respectively of the indices under study. The time period considered here is 1 January 1992 to 10 November 1995 ($n = 201$ for the weekly and $n = 50$ for the monthly series respectively). Applying Beran's [10] approximate maximum likelihood procedure, the estimates and 95%-confidence intervals, in brackets, of the FARIMA-fits for the DAX are $\hat{d} = 1.115$ ($[1.007; 1.223]$) and $\hat{d} = 1.150$ ($[0.933; 1.367]$) for the weekly and monthly series respectively. The corresponding results for the Hang Seng are given by $\hat{d} = 1.120$ ($[1.012; 1.228]$) and $\hat{d} = 1.140$ ($[0.923; 1.357]$). The models were selected using the BIC, including autoregressive and moving average terms. The estimated values of d and the confidence intervals suggest that all series are nonstationary. But more interesting is the fact that both indices possess a significant long-memory behaviour with $d \simeq 1.1$ which does, in particular, not vanish under temporal aggregation (though a value of $d = 1$ cannot be excluded completely for the monthly estimates). The good fits of the models are illustrated by satisfactory correlograms and normal probability plots of the residuals in figures 5.4 and 5.5.

5.4 Concluding remarks

In this paper we investigated the effects of temporal aggregation on nonstationary and stationary short- and long-range dependent autoregressive processes. Our main result is that long memory and antipersistence are robust with respect to temporal aggregation, whereas traditional short memory is not. Interestingly, the increments of nonstationary fractional processes do not converge asymptotically to fractional Gaussian noise.

Our findings have some interesting implications for analyzing and forecasting economic time series. Monthly or quarterly aggregates may not deviate much from a random walk model, apart from a lag-1 correlation of .25, if the original nonaggregated series has no or traditional short memory only and $d = m = 1$. The same statement may hold for daily aggregates too, if they are aggregates of tick-data. Thus more interesting is to look for long memory or antipersistence in aggregated series. Also, the fact that $\gamma_y(0) = O(s)$, as $s \to \infty$ and $d = 0$, is used, for example, in option pricing models to annualize volatility and in Value-at-Risk statistics. That is, we are assuming that the standard deviation of a return series increases with the rate $O(s^{.5})$ instead of $O(s^{\delta+.5})$ as it does in the case of fractional Gaussian noise. Thus, there is a danger to over- or underestimate the volatility of temporarily aggregated data systematically in the presence of antipersistence and persistence respectively.

Finally, a formal extension of our results to ARCH models (Engle [40]) would be an interesting task for future research. An ARCH process, and its extensions such as GARCH, EGARCH and IGARCH models, are readily interpreted as ARMA- and ARIMA-type models of the (conditional) variance (Bollerslev and Mikkelsen [21]). That is, they possess exponentially decaying summable correlations or short memory only. Diebold [35] has shown that, under certain moment conditions, the marginal distribution of an ARCH process converges to normality under temporal aggregation. That is (as expected), heteroscedasticity, as a stochastic short-range dependent behaviour, is not robust with respect temporal aggregation. In contrast, recent studies are putting more and more long-range dependence of the (conditional) variance into focus (see e.g. Ding, Granger [39], Bollerslev and Mikkelsen [21], and Beran and Ocker [18]). The new approaches, such as fractionally integrated GARCH (FIGARCH), long memory ARCH (LM(q)-ARCH) and Beran's [11] semiparametric fractionally integrated autoregressive (SEMIFAR) model typically exhibit hyperbolically decaying correlations of an appropriate volatility measure.[1] This is probably an explanation for the robustness of long memory in the variance of stock market return series under temporal aggregation as found by Ding, Granger and Engle [37].

[1] See also Beran and Ocker [18] for a recent application of SEMIFAR models to volatilities.

5.5 Appendix

Proof of Lemma 1:

(i) Noting that $y_T = \sum_{l=0}^{s-1} B^l u_{sT} = \sum_{l=0}^{s-1} u_{sT-l}$ the covariance is given by straightforward calculation as $\gamma_y(k) = \sum_{l,j=0}^{s-1} \gamma_u(j-l-sk)$.

(ii) To prove the case of $m=1$ consider that

$$\begin{aligned} y_T &= \sum_{j=0}^{s-1} B^j \sum_{l=0}^{s-1} B^l u_{sT}, \\ &= \sum_{j=s(T-2)+2}^{sT} (s - |j - s(T-1) - 1|) u_j, \\ &= \sum_{j=i(s,T)-(s-1)}^{i(s,T)+(s-1)} (s - |j - i(s,T)|) u_j, \\ &= \sum_{j=-(s-1)}^{s-1} (s - |j|) u_{i(s,T)+j}, \end{aligned}$$

where $i(s,T) = s(T-1)+1$. Hence, the covariance equals

$$\begin{aligned} \gamma_y(k) &= Cov\left(\sum_{j=-(s-1)}^{s-1} (s-|j|) u_{i(s,T)+j}, \sum_{l=-(s-1)}^{s-1} (s-|l|) u_{i(s,T+k)+l} \right), \\ &= \sum_{j,l=-(s-1)}^{s-1} (s-|j|)(s-|l|)\gamma_u(j-l-sk). \end{aligned}$$

Proof of theorem 5

(i) First note that lemma 1 (i) may be written as

$$\gamma_y(k) = \sum_{j=0}^{s-1} \sum_{v=j-s(1+k)+1}^{j-sk} \gamma_u(v). \tag{5.6}$$

Considering that the covariances $\gamma_u(\cdot)$ of an ARMA process decay exponentially in the sense that there is an upper bound $|\gamma_u(h)| \leq ca^{|h|}$, where $0 < c < \infty, 0 < a < 1$ are constants, we may write (5.6) as $|\gamma_y(k)| \leq c \sum_{j=0}^{s-1} \sum_{v=j-s(1+1)+1}^{j-sk} a^{|v|}$. It follows by standard results on geometric series that

$$
\begin{aligned}
|\gamma_y(0)| &\leq c\sum_{j=0}^{s-1}\sum_{v=j-s+1}^{j} a^{|v|}, \\
&\leq c\sum_{j=0}^{s-1}\left\{\sum_{v=j-s+1}^{-1} a^{-v} + \sum_{v=0}^{j} a^{v}\right\}, \\
&\leq c\sum_{j=0}^{s-1}\left\{\left(\frac{1-a^{s-j}}{1-a}\right) + \left(\frac{1-a^{j}}{1-a}\right)\right\}, \\
&\leq \frac{c}{1-a}\left\{\left(s-\frac{a^{s}-1}{1-a^{-1}}\right) + \left(s-\frac{1-a^{s}}{1-a}\right)\right\}, \\
&\leq s\frac{c}{1-a} = O(s); \quad \text{and}
\end{aligned}
$$

$$
\begin{aligned}
|\gamma_y(k)| &\leq c\sum_{j=0}^{s-1}\sum_{v=j-s(1+k)+1}^{j-sk} a^{|v|}, \\
&\leq c\sum_{j=0}^{s-1}\sum_{v=sk}^{s(1+k)-1} a^{v-j}, \\
&\leq ca^{sk}\left(\frac{1-a^{-s}}{1-a^{-1}}\right)\left(\frac{1-a^{s}}{1-a}\right), \\
&\leq -ca^{s(k-1)}\left(\frac{1}{1-a^{-1}}\right)\left(\frac{1}{1-a}\right) = o(s), \forall k \geq 1.
\end{aligned}
$$

Hence, $\lim_{s\to\infty}\rho_y(k) = 0, \forall k \geq 1$.

(ii) First note that lemma 1 (ii) may be written as

$$
\begin{aligned}
\gamma_y(k) = & \sum_{j=1-s}^{s-1}(s-|j|)\left\{\sum_{v=j-sk}^{j-s(k-1)-1}\left[s(1-k)+j-v\right]\gamma_u(v) + \right. \\
& \left. + \sum_{v=j-s(k+1)+1}^{j-sk-1}\left[s(k+1)-j+v\right]\gamma_u(v)\right\}. \qquad (5.7)
\end{aligned}
$$

Since short memory series are typified by quickly, i.e. exponentially, declining covariances we assume that $\gamma_u(v) = 0, \forall |v| > k_o$. Then, we know from (5.7) that

$$\begin{aligned}
\gamma_y(0) &\approx \sum_{j=1-s}^{s-1}(s-|j|)\left\{(s+j)\sum_{v=j}^{j+s-1}\gamma_u(v)+(s-j)\sum_{v=j-s+1}^{j-1}\gamma_u(v)\right\}, \\
&\approx \sum_{j=-(s-1)}^{s-1}(s-|j|)^2\sum_{v=-k_o}^{k_o}\gamma_u(v), \\
&\approx 2s^3\int_0^1(1-x)^2dx\sum_{v=-k_o}^{k_o}\gamma_u(v), \\
&\approx (2s^3/3)\sum_{v=-k_o}^{k_o}\gamma_u(v)=O(s^3), \quad \text{and}
\end{aligned}$$

$$\begin{aligned}
\gamma_y(1) &\approx \sum_{j=1-s}^{s-1}(s-|j|)\left\{j\sum_{v=j-s}^{j-1}\gamma_u(v)+(2s-j)\sum_{v=j-2s+1}^{j-s-1}\gamma_u(v)\right\}, \\
&\approx \sum_{j=1}^{s}(s-|j|)j\sum_{v=-k_o}^{k_o}\gamma_u(v), \\
&\approx s^3\int_0^1(1-x)xdx\sum_{v=-k_o}^{k_o}\gamma_u(v), \\
&\approx (s^3/6)\sum_{v=-k_o}^{k_o}\gamma_u(v)=O(s^3),
\end{aligned}$$

In addition we can show that

$$\gamma_y(k)=o(s^3), \forall k\geq 2.$$

Putting $\gamma_y(0)$ and $\gamma_y(k)$ together, we obtain $\lim_{s\to\infty}\rho_y(1)=.25$ and $\lim_{s\to\infty}\rho_y(k)=0, \forall k\geq 2$.

(iii) Consider first the long memory case $\delta\in(0,.5)$. Taking into account that the covariances $\gamma_u(\cdot)$ of a FARIMA$(0,\delta,\infty)$ process are asymptotically given by

$$\gamma_u(h)\sim c_\gamma|h|^{2\delta-1}, \text{ as } |h|\to\infty, \tag{5.8}$$

where c_γ is a nonzero finite constant, (5.6) may be written as

$$\begin{aligned}\gamma_y(k) &\sim c_\gamma \sum_{j=0}^{s-1} \sum_{v=j-s(1+k)+1}^{j-sk} |v|^{2\delta-1}, \\ &\sim c_\gamma s^{2\delta+1} \int_0^1 \int_{x-1-k}^{x-k} |v|^{2\delta-1} dx dy = O(s^{2\delta+1}),\end{aligned}$$

as s tends to ininity, which yields

$$\gamma_y(0) \sim \frac{2c_\gamma s^{2\delta+1}}{2\delta(2\delta+1)}, \quad \text{and}$$

$$\gamma_y(k) \sim \frac{c_\gamma s^{2\delta+1}}{2\delta(2\delta+1)} \left\{(k+1)^{2\delta+1} - 2k^{2\delta+1} + (k-1)^{2\delta+1}\right\}, \forall k \geq 1,$$

respectively. The correlations are

$$\rho_y(k) = \frac{1}{2} \left\{(k+1)^{2\delta+1} - 2k^{2\delta+1} + (k-1)^{2\delta+1}\right\}, \forall k \geq 1.$$

The asymptotic behaviour of $\rho_y(k)$ follows by Taylor expansion: First note that $\rho_y(k) = \frac{1}{2}k^{2\delta+1}g(k^{-1})$, where $g(x) = (x+1)^{2\delta+1} - 2 + (1-x)^{2\delta+1}$. The first non-zero term in the Taylor expansion of $g(x)$, expanded at the origin, is equal to $2\delta(2\delta+1)x^2$. Therefore, as k tends to infinity, $\rho_y(k)$ is similar to $\delta(2\delta+1)k^{2\delta-1}$.

To prove the case of $\delta \in (-.5, 0)$, note that, using (5.6) and (5.8), the following holds

$$\begin{aligned}\gamma_y(0) &= \sum_{j=0}^{s-1} \sum_{v=j-s+1}^{j} \gamma_u(v), \\ &= -\sum_{j=0}^{s-1} \left\{ \sum_{v=j+1}^{\infty} \gamma_u(v) + \sum_{v=-\infty}^{j-s} \gamma_u(v) \right\}, \\ &\sim -c_\gamma \sum_{j=0}^{s-1} \left\{ \sum_{v=j+1}^{\infty} v^{2\delta-1} + \sum_{v=-\infty}^{j-s} (-v)^{2\delta-1} \right\}, \\ &\sim -s^{1+2\delta} c_\gamma \int_0^1 \left\{ \int_x^\infty y^{2\delta-1} dy + \int_{-\infty}^{x-1} (-y)^{2\delta-1} dy \right\} dx, \\ &\sim \frac{2s^{1+2\delta}c_\gamma}{2\delta(2\delta+1)} = O(s^{2\delta+1}), \text{ as } s \to \infty, \text{ and}\end{aligned}$$

$$\begin{aligned}
\gamma_y(k) &= \sum_{j=0}^{s-1} \sum_{v=j-s(1+k)+1}^{j-sk} \gamma_u(v), \\
&= \sum_{j=0}^{s-1} \left\{ \sum_{v=-\infty}^{j-sk} \gamma_u(v) - \sum_{v=-\infty}^{j-s(1+k)} \gamma_u(v) \right\}, \\
&\sim c_\gamma \sum_{j=0}^{s-1} \left\{ \sum_{v=-\infty}^{j-sk} (-v)^{2\delta-1} - \sum_{v=-\infty}^{j-s(1+k)} (-v)^{2\delta-1} \right\}, \\
&\sim s^{2\delta+1} c_\gamma \int_0^1 \left\{ \int_{-\infty}^{x-k} (-y)^{2\delta-1} dy - \int_{-\infty}^{x-1-k} (-y)^{2\delta-1} dy \right\} dx, \\
&\sim \frac{s^{1+2\delta} c_\gamma}{2\delta(2\delta+1)} \left\{ (k+1)^{2\delta+1} - 2k^{2\delta+1} + (k-1)^{2\delta+1} \right\} = O(s^{2\delta+1}),
\end{aligned}$$

$\forall k \geq 1$, as s tends to infinity. The proposed result follows by combining $\gamma_y(0)$ and $\gamma_y(k)$.

(iv) Again, we first consider the long memory case $\delta \in (0, .5)$. If (5.8) holds, then lemma 1(ii) may be written as

$$\gamma_y(k) \sim c_\gamma s^{2\delta+3} \int_{-1}^{1} \int_{-1}^{1} (1-|x|)(1-|y|)|x-y-k|^{2\delta-1} dx dy, \quad (5.9)$$

as s tends to infinity, which is of asymptotic order $O(s^{2\delta+3})$. Given (5.9), we know that

$$\gamma_y(0) \sim \frac{c_\gamma s^{2\delta+3}(8 - 2^{2\delta+4})}{2\delta(2\delta+1)(2\delta+2)(2\delta+3)},$$

$$\gamma_y(1) \sim \frac{c_\gamma s^{2\delta+3}(2^{2\delta+5} - 7 - 3^{2\delta+3})}{2\delta(2\delta+1)(2\delta+2)(2\delta+3)},$$

$$\gamma_y(k) \sim \frac{-(k+2)^{2\delta+3} + 4(k+1)^{2\delta+3} - 6k^{2\delta+3} + 4(k-1)^{2\delta+3} - (k-2)^{2\delta+3}}{2\delta(2\delta+1)(2\delta+2)(2\delta+3)(c_\gamma s^{2\delta+3})^{-1}},$$

$\forall k \geq 2$. Thus, the correlations are

$$\rho_y(1) = \frac{2^{2\delta+5} - 7 - 3^{2\delta+3}}{8 - 2^{2\delta+4}}, \quad \text{and}$$

$$\rho_y(k) = \frac{-(k+2)^{2\delta+3} + 4(k+1)^{2\delta+3} - 6k^{2\delta+3} + 4(k-1)^{2\delta+3} - (k-2)^{2\delta+3}}{8 - 2^{2\delta+4}},$$

$\forall k \geq 2$, respectively. Following the proof of theorem 5(iii), the first non-zero term in the Taylor expansion of $g(x)$, expanded at the origin, is equal to $-2\delta(2\delta+1)(2\delta+2)(2\delta+3)x^4$. Therefore, as k tends to infinity, $\rho_y(k)$ is similar to $\delta(2\delta+1)(2\delta+2)(2\delta+3)k^{2\delta-1}/\left(2^{2\delta+3}-4\right)$.

To obtain explicit formulas for the correlations in case of $\delta \in (-.5, 0)$ is slightly more difficult. However, using (5.7), the following property of convergent series, where

$$\sum_{h=1}^{s} \gamma(h) = -\left\{ \sum_{h=s+1}^{\infty} \gamma(h) + \sum_{h=-\infty}^{0} \gamma(h) \right\} = \sum_{h=1}^{\infty} \gamma(h) - \sum_{h=s+1}^{\infty} \gamma(h),$$

and taking integrals, leads to the same correlation formulas as in the long memory case. Hence, the proposed result holds.

Figure 5.1: Lag-1 autocorrelations under temporal aggregation

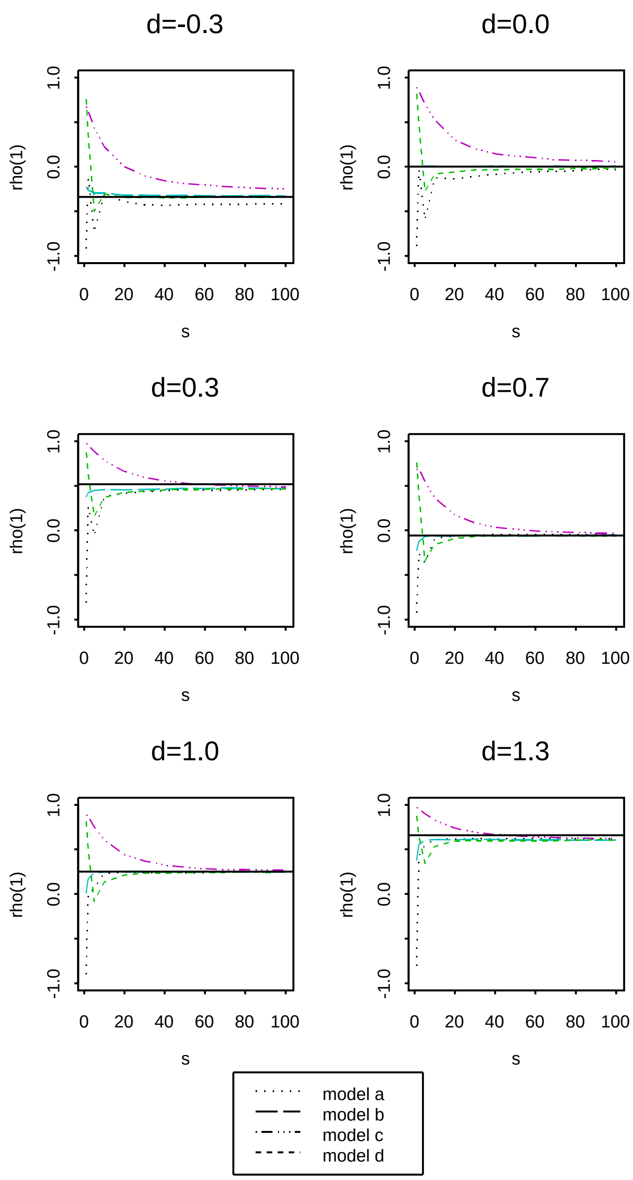


Figure 5.2: Lag-2 autocorrelations under temporal aggregation

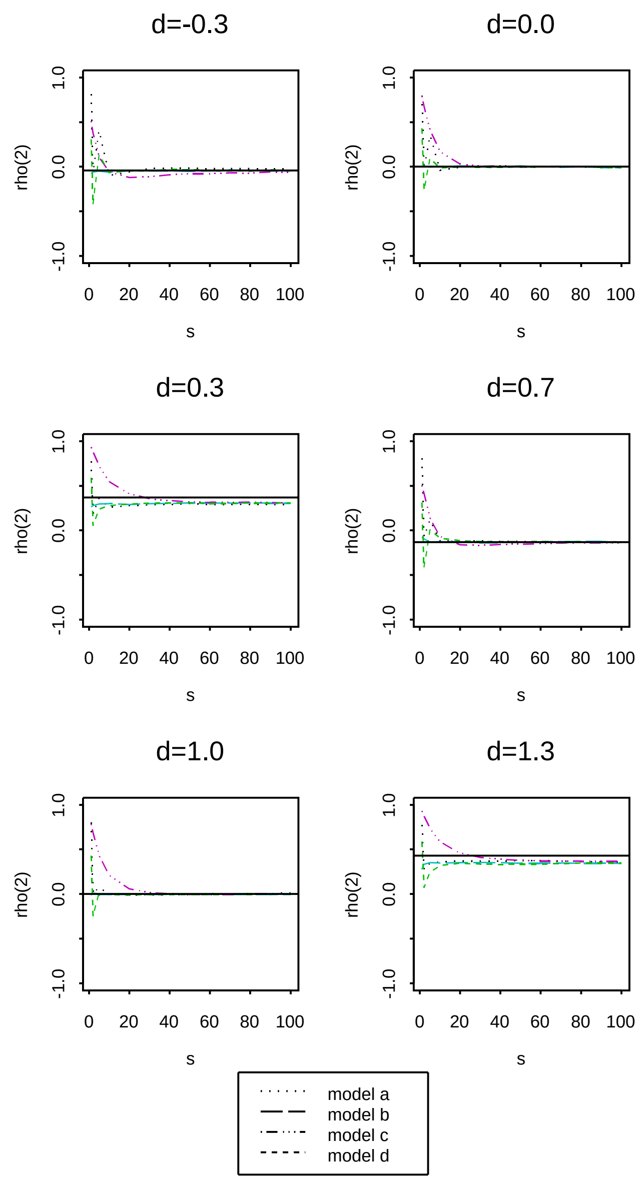


Figure 5.3: Stock market indices between January 1, 1992 and November 10, 1995 (log-transformed weekly and monthly averages)

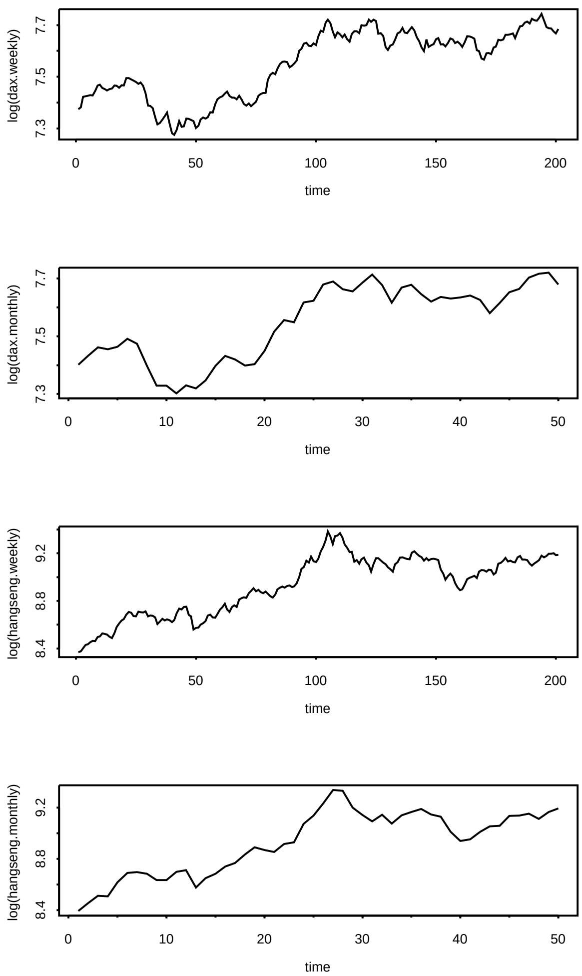


Figure 5.4: Autocorrelations of (squared) FARIMA-residuals of the stock market indices between January 1, 1992 and November 10, 1995

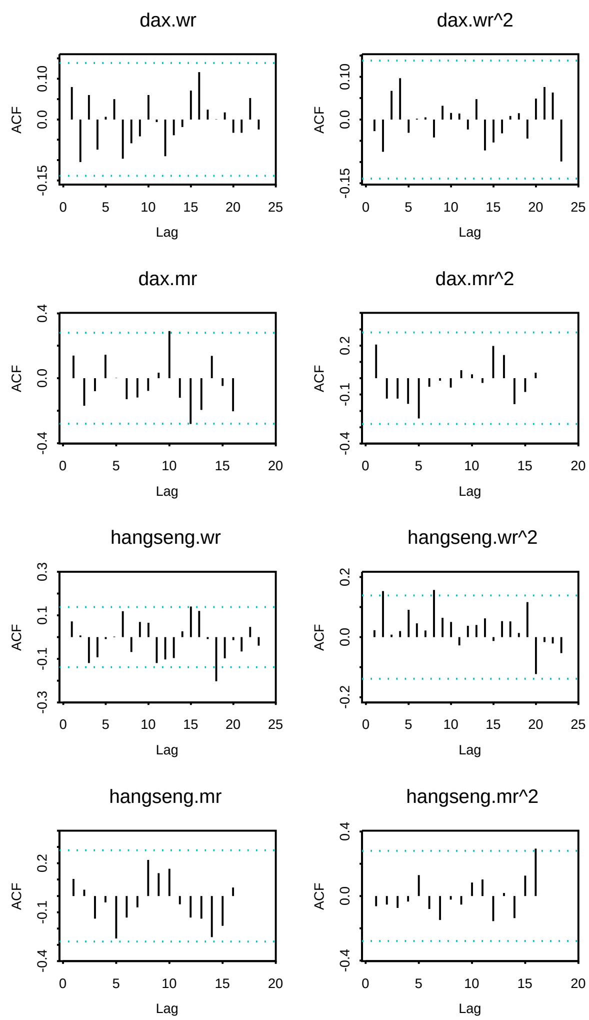
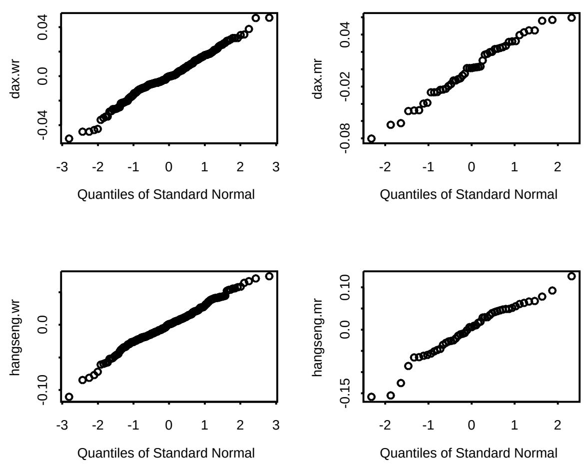


Figure 5.5: Normal probability plots of FARIMA-residuals of the stock market indices between January 1, 1992 and November 10, 1995



Chapter 6

Modeling interventions in FARIMA$(p, d, 0)$ models

6.1 Introduction

Many time series are influenced by eruptive events, such as sudden political and/or economic crises, outbreaks of wars, or unexpected weather disasters. Such disturbances of a time series are usually referred to as interventions. The existence of an intervention in a time series is often indicated by outliers in a corresponding stationary time series. Thus, the traditional Gaussian time series framework is no longer applicable. In order to keep the Gaussian instruments valid, some authors trim back extreme *observable* outliers to, say, three sigmas. There are also some more sophisticated methods dealing with detecting and removing outliers when the timing of a disturbance is unknown (see e.g. Fox [43], Abraham and Box [2], Tsay [84], Chang, Tiao and Chen [26]). However, one assumes that, an intervention can be removed by *pulling back* some few observations. This is, however, only valid as long as the intervention has an effect to a very short time period only. In contrast, an intervention may contaminate a time series over long time periods. It may decay exponentially or hyperbolically to a zero/nonzero constant, or may persist at a nonzero level. An extension of intervention modeling provides a natural means of eruptive disturbances within the Gaussian framework. It provides, in particular, the possibility to model the effect of an intervention explicitly.

Box and Tiao [24] provided a strategy for considering interventions for the nonfractional Box-Jenkins ARIMA-class (see also Wei [85] for an overview). Here, we extend the results to a wider class of models that includes classical nonfractional Box-Jenkins ARIMA models as well as stationary and non-stationary fractional autoregressive processes. The FARIMA$(p, d, 0)$ model (Beran [10], Beran, Bhansali and Ocker [12]) considered here is given by

$$\phi(B)(1-B)^{\delta}\left\{(1-B)^m X_t - \mu\right\} = \epsilon_t, \tag{6.1}$$

A starting point for modeling interventions within the framework of (6.1) is given by

$$\phi(B)(1-B)^{\delta}\left\{(1-B)^m X_t - f(t-T,\beta)\right\} = \epsilon_t, \tag{6.2}$$

where $f(t, T, \beta)$ is an intervention function of an intervention parameter vector $\beta = (\beta_1, ..., \beta_r)^t$ and T is the intervention time. $Y_t = (1-B)^m X_t$ is a FARIMA(p, δ, q) process with expected value $f(\cdot)$. Observe that (6.1) with expected value $f(t, T, \beta) \equiv \mu$ is just a special case of (6.2). $Z_t = Y_t - f(\cdot)$ is a free of intervention zero-mean stationary Gaussian time series, often referred to as noise series.

Given (6.2), the chapter is organized as follows. A general intervention model is introduced and illustrated in section 2. In section 3, results on maximum likelihood estimation are derived. Some remarks in section 4 conclude the chapter. Proofs are given in the appendix.

6.2 A general intervention model

We define the intervention function $f(\cdot)$ in (6.2) by

DEFINITION 2

$$f(t-T,\beta) = \sum_{l=0}^{\infty} b_l(\beta) B^l I_t(T) = \begin{cases} 0, & t < T \\ \sum_{l=0}^{t-T} b_l(\beta) B^l I_t(T), & t \geq T \end{cases}, \tag{6.3}$$

where $I_t(T)$ is a variable indicating the intervention at time $t = T$, B denotes the backshift operator, and $b_l(\beta)$ are coefficients that are characterized by an intervention parameter vector $\beta = (\beta_1, ..., \beta_r)^t$.

There are two common types of intervention indicator variables. If the intervention takes place at one time point only, $I_t(T)$ is usually referred to as pulse input

$$P_t(T) = \begin{cases} 0, & t \neq T \\ 1, & t = T \end{cases} . \tag{6.4}$$

If the intervention takes place at one time point and thereafter, $I_t(T)$ is usually referred to as step input

$$S_t(T) = \begin{cases} 0, & t < T \\ 1, & t \geq T \end{cases} . \tag{6.5}$$

Note that, given the indicator variable $I_t(T)$, it is not difficult to show that, for $l \leq s = t - T \geq 0$, the intervention function (6.3) may be written as

$$f(s, \beta) = \sum_{l=0}^{s} b_l(\beta) I_{T+s-l}(T). \tag{6.6}$$

Thus,

$$f(s, \beta) = \begin{cases} 0, & s < 0 \\ b_s(\beta), & s \geq 0 \end{cases} , \tag{6.7}$$

for $I_t(T) = P_t(T)$, and

$$f(s, \beta) = \begin{cases} 0, & s < 0 \\ \sum_{l=0}^{s} b_l(\beta), & s \geq 0 \end{cases} , \tag{6.8}$$

for $I_t(T) = S_t(T)$ respectively.

Practically meaningful are only intervention effects that do not diverge (as $s \to \infty$ and $T < \infty$). Therefore, we assume

ASSUMPTION 1

$$\lim_{s \to \infty} |f(s, \beta)| < \infty. \tag{6.9}$$

The necessary condition of the intervention function given in (6.7) in order that assumption 1 holds is $|b_\infty(\beta)| < \infty$. Also, a necessary condition of (6.8) so that assumption 1 holds is given by $|\sum_{l=0}^{\infty} b_l(\beta)| < \infty$. Note, in particular, that the condition for $S_t(T)$ is much stronger than for $P_t(T)$, indicating that not all coefficients $b_l(\cdot)$ that are allowed for $P_t(T)$ are allowed for $S_t(T)$. Several types of coefficients that satisfy assumption 1 are given in the following lemma. Note that the trivial case $b_l(\beta) = 0$ ($\forall l \geq 0$) is omitted. Note also that the proof follows by standard results on infinite series.

LEMMA 2 *Assumption 1 holds, if*

(i) $b_l(\beta) \equiv \beta_1 \neq 0$, *and* $I_t(T) = P_t(T)$;

(ii) $b_l(\beta) = \beta_1 \beta_2^l$, *where* $\beta_1 \neq 0$, $|\beta_2| \in (0,1)$, *and a)* $I_t(T) = P_t(T)$ *or b)* $I_t(T) = S_t(T)$;

(iii) $b_l(\beta) \sim L(l,\beta) l^{-\nu(\beta)}$, *where* $L(\cdot)$ *is a slowly varying function of* l, *and either a)* $\nu(\beta) > 0$ *and* $I_t(T) = P_t(T)$, *or b)* $\nu(\beta) > 1$ *and* $I_t(T) = S_t(T)$;

(iv) $b_l(\beta) = \beta_1 + r_l(\beta^*) \to \beta_1 \neq 0$, *as* $l \to \infty$, *where* $\lim_{l\to\infty} r_l(\beta^*) = 0$, $\beta^* = (\beta_2, ..., \beta_r)^t$, *and* $I_t(T) = P_t(T)$.

In addition we can show that, every intervention with step input, such that assumption 1 holds, can be modeled with pulse input. This follows by setting $\tilde{b}_s(\beta) = \sum_{l=0}^{s} b_l(\beta)$. In particular, we have the following lemma.

LEMMA 3 *Let* $I_t(T) = S_t(T)$.

(i) If $b_l(\beta) = \beta_1 \beta_2^l$, *then* $f(s,\beta) = \beta_1 \sum_{l=0}^{s} \beta_2^l \neq 0$ *and* $\lim_{s\to\infty} f(s,\beta) = \beta_1 (1-\beta_2)^{-1}$, $\forall |\beta_2| \in (0,1), \beta_1 \neq 0$.

(ii) If $b_l(\beta) \sim L(l,\beta) l^{-\nu(\beta)}$, *then* $f(s,\beta) \sim \sum_{l=0}^{s} L(l,\beta) l^{-\nu(\beta)} \neq 0$ *and* $\lim_{s\to\infty} f(s,\beta) \neq 0$, $\forall \nu(\beta) > 1$.

Lemma 3 implies that the effects of the step inputs considered in lemma 2 are just special cases of lemma 2 (i). Hence, whether a pulse or a step input is used reduces (asymptotically) just to a question of convenience. Without loss of generality, we consider in the following pulse input interventions only.

For illustration, some simple intervention models (with pulse input) and the corresponding responses are shown in figure 6.1. In all cases, the noise series is white noise. The simulated interventions ($n = 400$ and $T = 100$) in figures 6.1a through d are:

- Figure 6.1a: $b_l(\beta) = -.9$,
- Figure 6.1b: $b_l(\beta) = (.9)^l$,
- Figure 6.1c: $b_l(\beta) = \frac{\Gamma(l+.9)}{\Gamma(l+1)\Gamma(.9)} = (-1)^l \binom{-.9}{l}$,
- Figure 6.1d: $b_l(\beta) = (-.9)^l$.

All interventions force a shift in the mean level of the basic series, with, however, different rates of convergence. As we will see in the next section, it is the rate of convergence that determines whether an intervention can be estimated consistently or not. In particular, the exponentially decaying interventions of lemma 2 (ii) (figures 6.1b and 6.1d) decay too quickly. Also, it should be noted that a simple t-test is typically not adequate to test for a change in the mean-level before and after an intervention, such as in figure 6.1a. This is because the traditional t-test is not robust with respect to the violation of the independence assumption (see, e.g. Box and Tiao [23] and chapter 3). Moreover, the form of the change may not be a constant shift. In contrast, the change of the mean may decay exponentially (figures 6.1b and 6.1d) or hyperbolically (figure 6.1c) to a zero/nonzero constant.

6.3 Maximum likelihood estimation

Let X_t $(t = 1, ..., n)$ be defined by (6.2), and let $Y_t = (1-B)^m X_t$. Also denote by $\theta = (\sigma_\epsilon^2, d, \phi_1, ..., \phi_p, \beta_1, ..., \beta_r) = (\sigma_\epsilon^2, \eta, \beta) = (\sigma_\epsilon^2, \vartheta)$ the parameter vector determining the correlation structure of X_t, where $\delta \in (-.5, .5)$ and $m \geq 0$ is an integer. In this section we assume $(\sigma_\epsilon^2, \eta)$ to be known. Now Y_t admits an infinite autoregressive representation

$$\sum_{j=0}^{\infty} a_j(\eta) \left\{ Y_{t-j} - f(t-T-j, \beta) \right\} = \epsilon_t(\eta, \beta), \tag{6.10}$$

where the coefficients $a_j(\cdot)$ are obtained from (6.2). From (6.2) it follows that $Y_t - f(\cdot)$ is a stationary fractional ARIMA$(p, \delta, 0)$ process with expected value 0. Now let β^o be the true value of β, then (6.10) can be written as

$$\epsilon_t(\eta, \beta) = \epsilon_t(\eta, \beta^o) + u_{t-T}(\eta, \beta^o) - u_{t-T}(\eta, \beta), \tag{6.11}$$

where

$$\epsilon_t(\eta, \beta^o) = \sum_{j=0}^{\infty} a_j(\eta) \{Y_{t-j} - f(t-T-j, \beta^o)\} = \sum_{j=0}^{\infty} a_j(\eta) Z_{t-j},$$

$$u_{t-T}(\eta, \beta^o) = \sum_{j=0}^{\infty} a_j(\eta) f(t-T-j, \beta^o), \quad \text{and}$$

$$u_{t-T}(\eta, \beta) = \sum_{j=0}^{\infty} a_j(\eta) f(t-T-j, \beta).$$

Disregarding constants, the corresponding log-likelihood function is then proportional to

$$l(Y; \eta, \beta) = \sum_{t=2}^{n} \{\epsilon_t(\eta, \beta^o) + u_{t-T}(\eta, \beta^o) - u_{t-T}(\eta, \beta)\}^2 . \tag{6.12}$$

Minimizing (6.12) with respect to β and rearranging the indices, using $s = t - T$ and $\frac{T}{n} \to \zeta \in [0; 1]$, as $n \to \infty$, yields

$$A_n(\eta, \hat{\beta}) + B_n(\eta, \hat{\beta}) \stackrel{!}{=} 0, \tag{6.13}$$

where

$$A_n(\eta, \hat{\beta}) := \sum_{s=0}^{n(1-\zeta)} \epsilon_{s+T}(\eta, \beta^o) \dot{u}_s(\eta, \hat{\beta}) \tag{6.14}$$

is purely stochastic and

$$B_n(\eta, \hat{\beta}) := \sum_{s=0}^{n(1-\zeta)} \left\{ u_s(\eta, \beta^o) - u_s(\eta, \hat{\beta}) \right\} \dot{u}_s(\eta, \hat{\beta}) \tag{6.15}$$

is purely deterministic. For fixed β, $E[A_n(\eta, \beta)] = 0$ and

$$Var[A_n(\eta, \beta)] = \sigma_\epsilon^2 \sum_{s=0}^{n(1-\zeta)} \dot{u}_s^2(\eta, \beta). \tag{6.16}$$

In order to derive a necessary condition for a consistent estimator of β, consider a trial value of $\tilde{\beta}$ and standardize (6.13) by $\sqrt{Var[A_n(\eta,\tilde{\beta})]}$. If $\tilde{\beta}$ is an asymptotic solution of (6.13), then

$$\frac{A_n(\eta,\tilde{\beta})}{\sqrt{Var[A_n(\eta,\tilde{\beta})]}} + \frac{B_n(\eta,\tilde{\beta})}{\sqrt{Var[A_n(\eta,\tilde{\beta})]}} = Z_n + q_n(\eta,\tilde{\beta}) \to 0. \tag{6.17}$$

Thus, if q_n converges to a bounded function, then any $\tilde{\beta} = q_n^{-1}(\eta, -Z_n)$, where $Z_n \xrightarrow{d} N(0,1)$, would be an asymptotic solution of (6.17). Thus, we can only obtain a consistent estimator of β if $q_n(\cdot) \to \pm\infty$, as $n \to \infty$, for all $\tilde{\beta} \neq \beta^o$. We show in the following theorem that consistency just depends on the rate of convergence of the weighted intervention $u_s(\eta,\beta)$.

THEOREM 6 *Suppose that* $u_s(\eta,\beta) = \sum_{j=0}^{\infty} a_j(\eta) f(s-j,\beta) = O[L_1(s,\beta)s^{-\alpha(\delta,\beta)}]$, *as* $s \to \infty$, *where* $L_1(\cdot)$ *is a slowly varying function of* s. *Then the following holds.*

(i) If $\alpha(\delta,\beta) > .5$, *then* $\lim_{n\to\infty} |q_n(\eta,\beta)| < \infty, \forall \zeta \in [0,1]$.

(ii) If $\alpha(\delta,\beta) < .5$, *then, as* $n \to \infty$,

$$|q_n(\eta,\beta)| \sim L_3(n(1-\zeta),\beta,\beta^\star)[n(1-\zeta)]^{.5-\alpha^\star} \to \infty, \forall \zeta \in [0,1),$$

where $L_3(\cdot)$ *is a slowly varying function of* n, $\alpha^\star = min\{\alpha(\delta,\beta^o),\alpha(\delta,\beta)\}$, *and* $\beta^\star$ *is the corresponding value, i.e. either* β^o *or* β.

Thus, the intervention coefficients β can be estimated consistently, if $\alpha(\delta,\beta) < .5$. The question which of the intervention models in lemma 2 correspond to case (i) and (ii) of theorem 6 respectively, is answered by the following corollary.

COROLLARY 2 *Given one of the intervention effects in lemma 2, we have, as* $s \to \infty$,

(i) theorem 6(i) holds, if the intervention effect

(a) decays exponentially to zero, i.e. $f(s,\beta) = \beta_1\beta_2^s$, *where* $\beta_1 \neq 0$ *and* $|\beta_2| \in (0,1)$;

(b) *decays (fast) hyperbolically to zero, i.e.* $f(s,\beta) \sim L(s,\beta)s^{-\nu(\beta)}$, *where* $\nu(\beta) > .5 - \delta$, $\delta \in (-.5, .5)$ *and* $L(\cdot)$ *is a slowly varying function of* s*;*

(ii) *theorem 6(ii) holds, if the intervention effect*

(c) *decays (slowly) hyperbolically to zero, i.e.* $f(s,\beta) \sim L(s,\beta)s^{-\nu(\beta)}$, *where* $\nu(\beta) < .5 - \delta$, $\delta \in (-.5, .5)$ *and* $L(\cdot)$ *is a slowly varying function of* s*;*

(d) *persists at a non-zero level, i.e.* $f(s,\beta) = \beta_1$, *where* $\beta_1 \neq 0$*;*

(e) *decays to a non-zero constant, i.e.* $f(s,\beta) \sim \beta_1 + r_s(\beta^*) \to \beta_1 \neq 0$, *where* $\lim_{s\to\infty} r_s(\beta^*) = 0$ *and* $\beta^* = (\beta_2, ..., \beta_r)^t$.

The asymptotic distribution of a consistently estimated intervention parameter $\hat{\beta} = \hat{\beta}_1$ is given by the following theorem.

THEOREM 7 *Let* $\hat{\beta}$ *be the solution of (6.13), and suppose that case (ii) of theorem 6 holds. Then, as* $n \to \infty$ *and* $\frac{T}{n} \to \zeta \in [0, 1)$,

$$[n(1-\zeta)]^{.5-\alpha(\delta,\beta^o)} L_2(n(1-\zeta),\beta^o)\left(\hat{\beta}-\beta^o\right) \xrightarrow{d} N(0,\sigma_\epsilon^2), \tag{6.18}$$

where $L_2(\cdot)$ *is a slowly varying function of* n.

The concrete expression of the slowly varying function $L_2(\cdot)$ as well as the exact rate of convergence $[n(1-\zeta)]^{.5-\alpha(\delta,\beta^o)}$ given in theorem 7 depend, of course, on the specific behaviour of the intervention under consideration and the corresponding FARIMA$(p, \delta, 0)$ environment.

6.4 Concluding remarks

In this chapter, we introduced a parametric method for modeling interventions in stationary and nonstationary, fractional and nonfractional autoregressive processes, when the timing of an intervention is known. We derive necessary and sufficient conditions for estimating the parameters of an intervention consistently, and provide the corresponding asymptotic distribution. It is shown, in particular, that consistency is possible if an intervention decays slowly hyperbolically to zero, a non-zero constant or remains at a non-zero level. Although we use the assumption that the underlying FARIMA parameters for the free of intervention series are known, our results hold (at least) asymptotically, if all (FARIMA and intervention) parameters are estimated simultaneously.

Also, it is clear that, if necessary, model specification in the presence of outliers can, in principle be done using robust methods (see e.g. Martin and Yohai [68]) or trimmed data. However, an extension of intervention modeling provides a natural means of modeling eruptive disturbances within the Gaussian framework. In particular, the considered method enables data analysts to include interventions due to, for instance, important policy changes explicitly, which is usually more useful in forecasting and control.

Clearly, if the timing of an intervention is unknown, the analysis becomes much more complex. Estimation of intervention and FARIMA parameters simultaneously when the timing of an intervention is unknown will be discussed elsewhere.

6.5 Appendix

Proof of theorem 6

Recall that, the standardized solution of (6.12) is given by

$$\frac{A_n(\eta,\hat{\beta})}{\sqrt{Var[A_n(\eta,\hat{\beta})]}}+\frac{B_n(\eta,\hat{\beta})}{\sqrt{Var[A_n(\eta,\hat{\beta})]}}=Z_n+q_n(\eta,\hat{\beta})\overset{!}{=}0, \tag{6.19}$$

where $Z_n \xrightarrow{d} N(0,1)$ and, for fixed β,

$$Var[A_n(\eta,\beta)]=\sigma_\epsilon^2\sum_{s=0}^{n(1-\zeta)}\dot{u}_s^2(\eta,\beta). \tag{6.20}$$

Since $\hat{\beta}=q_n^{-1}(\eta,-Z_n)$, the condition

$$q_n(\eta,\beta)\to\pm\infty,\ \text{as } n\to\infty, \tag{6.21}$$

for all $\beta\neq\beta^o$, is necessary to estimate $\beta^o=(\beta_1^o,...,\beta_r^o)^t$ consistently. Here β^o denotes the true value of β. Since $u_s(\eta,\beta)=\sum_{j=0}^s a_j(\eta)f(s-j,\beta)=O[L_1(s,\beta)s^{-\alpha(\delta,\beta)}]$ and $\dot{u}_s(\eta,\beta)=O[L_2(s,\beta)s^{-\alpha(\delta,\beta)}]$ we have, as $s\to\infty$,

$$\begin{aligned}B_n(\eta,\beta) &= O\left[\sum_{s=0}^{n(1-\zeta)}\left\{L_1(s,\beta^o)s^{-\alpha(\delta,\beta^o)}-\right.\right.\\ &\quad \left.\left.-L_1(s,\beta)s^{-\alpha(\delta,\beta)}\right\}L_2(s,\beta)s^{-\alpha(\delta,\beta)}\right],\end{aligned} \tag{6.22}$$

and

$$\sqrt{Var[A_n(\eta,\beta)]}=\sqrt{O\left[\sum_{s=0}^{n(1-\zeta)}L_2^2(s,\beta)s^{-2\alpha(\delta,\beta)}\right]}, \tag{6.23}$$

where $L_1(\cdot)$ and $L_2(\cdot)$ are slowly varying functions of s.

(i) For $\alpha(\delta,\beta)>.5$, (6.22) and (6.23) imply $\lim_{n\to\infty}|q_n(\eta,\beta)|<\infty,\forall\zeta\in[0,1]$. This contradicts the necessary condition (6.21).

(ii) By applying Taqqu's [79] lemma 3.1 to (6.22) and (6.23) we have, as $n\to\infty$,

$$\begin{aligned}q_n(\eta,\beta) &\sim \pm\frac{\frac{L_1(n(1-\zeta),\beta^\star)L_2(n(1-\zeta),\beta)}{1-\alpha^\star-\alpha(\delta,\beta)}[n(1-\zeta)]^{1-\alpha^\star-\alpha(\delta,\beta)}}{\sqrt{\frac{L_2^2(n(1-\zeta),\beta)}{1-2\alpha(\delta,\beta)}[n(1-\zeta)]^{1-2\alpha(\delta,\beta)}}},\\ &\sim \pm L_3(n(1-\zeta),\beta^\star,\beta)[n(1-\zeta)]^{.5-\alpha^\star},\end{aligned} \tag{6.24}$$

$\forall \alpha^\star < .5$, where $L_3(\cdot)$ is a slowly varying function of $n(1-\zeta)$, $\alpha^\star = min\{\alpha(\delta, \beta^o), \alpha(\delta, \beta)\}$, and $\beta^\star$ is either β or β^o. Thus, if $\alpha^\star < .5$, we have $\lim_{n\to\infty} |q_n(\beta)| = \infty, \forall \zeta \in [0, 1)$, the necessary condition of estimating β consistently. For $\zeta = 1$, the intervention occurred at time $n = T$ and consistency cannot be achieved. Finally, note that the sign of (6.24) depends on $\alpha^\star$, e.g. it is negative if $\alpha^\star = \alpha(\delta, \beta)$.

Proof of corollary 2

We prove corollary 2 by showing that the listed intervention models (a)-(e) behave asymptotically, as s tends to infinity, like case (i) and case (ii) of theorem 6 respectively. Also note that, without loss of generality, we suppose that $Y_t - f(s, \beta)$ follows a FARIMA$(0, \delta, 0)$ process.

(i) (a) If $f(s, \beta) = \beta_1 \beta_2^s$, where $\beta_1 \neq 0$ and $|\beta_2| \in (0, 1)$, then for any $\alpha \in \Re$ there exists a constant $c > 0$ such that

$$
\begin{aligned}
|u_s(\eta, \beta)| &\leq \beta_1 \sum_{j=0}^{s} |(-1)^j \binom{\delta}{j}||\beta_2^{s-j}|, \\
&\leq c\beta_1 \sum_{j=1}^{s} j^{-\delta-1}(s-j)^{-\alpha}, \\
&\sim c\beta_1 s^{-\delta-\alpha} \int_0^1 x^{-\delta-1}(1-x)^{-\alpha}dx, \text{ and}
\end{aligned}
$$

thus, $\exists\ \alpha < 1$, for which $\int_0^1 x^{-\delta-1}(1-x)^{-\alpha}dx$ exists and $s^{-\delta-\alpha}$ converges to zero too fast. Hence, $\beta = (\beta_1, \beta_2)^t$ cannot be estimated consistently.

Also, for $\delta = 0$, consistency is not possible since $u_s(\eta, \beta) = \beta_1 \beta_2^s$ converges to zero exponentially as $s \to \infty$.

(b) If $f(s, \beta) \sim L(s, \beta)s^{-\nu(\beta)}$, where $\nu(\beta) > .5 - \delta$, $\delta \in (-.5, .5)$ and $L(\cdot)$ is a slowly varying function of s, is given by, for instance, $f(s, \beta) = (-1)^s \binom{\beta_1}{s}$, where $\beta_1 \in]-1, 0[$, then

$$
\begin{aligned}
u_s(\eta, \beta) &= \sum_{j=0}^{s} (-1)^j \binom{\delta}{j} (-1)^{s-j} \binom{\beta_1}{s-j}, \\
&= (-1)^s \binom{\delta + \beta_1}{s},
\end{aligned}
$$

$$\sim \quad \frac{1}{\Gamma(-\delta-\beta_1)} s^{-\delta-\beta_1-1} = O(s^{-\delta-\beta_1-1}), \text{ as } s \to \infty.$$

Thus, if $\delta + \beta_1 + 1 > .5$ (i.e. $\beta_1 > -.5 - \delta$), $\beta = \beta_1$ cannot be estimated consistently.

Also, observe that, for $\delta = 0$, we have $u_s(\eta, \beta) \sim \frac{1}{\Gamma(-\beta_1)} s^{-\beta_1-1}$, as $s \to \infty$. It follows that $\beta = \beta_1$ cannot be estimated consistently if $\beta_1 + 1 > .5$ (i.e. $\beta_1 > -.5$).

(ii) (c) Given the proof of case (b), we know that $\beta = \beta_1$ can be estimated consistently, if $\delta + \beta_1 + 1 < .5$ (i.e. $\beta_1 < -.5 - \delta$).

Also, for $\delta = 0$, consistency is possible for $\beta_1 + 1 < .5$ (i.e. $\beta_1 < -.5$).

(d) If $f(s, \beta) = \beta_1$, where $\beta_1 \neq 0$, then

$$\begin{aligned}
u_s(\eta, \beta) &= \beta_1 \sum_{j=0}^{s} (-1)^j \binom{\delta}{j}, \\
&\sim c\beta_1 \begin{cases} \sum_{j=1}^{s} j^{-\delta-1}, & \delta \in (-.5, 0) \\ -\sum_{j=s+1}^{\infty} j^{-\delta-1}, & \delta \in (0, .5) \end{cases}, \\
&\sim c\beta_1 s^{-\delta} \begin{cases} \int_0^1 x^{-\delta-1} dx, & \delta \in (-.5, 0) \\ -\int_1^{\infty} x^{-\delta-1} dx, & \delta \in (0, .5) \end{cases}, \\
&\sim \frac{c\beta_1}{-\delta} s^{-\delta} = O(s^{-\delta}), \text{ as } s \to \infty,
\end{aligned}$$

where c is a nonzero finite constant. Thus, $\beta = \beta_1$ can always be estimated consistently, since $\delta < .5$.

Also, observe that, for $\delta = 0$, we have $u_s(\eta, \beta) = \beta_1$. That is, $\beta = \beta_1$ can always be estimated consistently.

(e) If $f(s, \beta) = \beta_1 + r_s(\beta^*) \to \beta_1 \neq 0$, as $s \to \infty$, where $\lim_{s\to\infty} r_s(\beta^*) = 0$ and $\beta^* = (\beta_2, ..., \beta_r)^t$, then from the results above we have

$$u_s(\eta, \beta) = O(s^{-\delta}),$$

that is, at least the intervention parameter β_1 can be estimated consistently, since $\delta < .5$.

Proof of theorem 7

Let $\hat{\beta} = \hat{\beta}_1$ be the solution of (6.13) and suppose that case (ii) of theorem 6 holds. Denote again by $A_n(\eta, \beta)$ the stochastic part, and by $B_n(\eta, \beta)$ the deterministic part of (6.13). Applying Taylor's theorem we may express (6.13) by

$$A_n(\eta, \hat{\beta}) + B_n(\eta, \hat{\beta}) = g_n(\eta, \hat{\beta}) \approx g_n(\eta, \beta^o) + \dot{g}_n(\eta, \beta^o)(\hat{\beta} - \beta^o). \quad (6.25)$$

By definition, $g_n(\eta, \hat{\beta})$ is equal to zero, and (6.25) may be rearranged to

$$(\hat{\beta} - \beta^o) \approx -\frac{g_n(\eta, \beta^o)}{\dot{g}_n(\eta, \beta^o)}. \quad (6.26)$$

It follows from (6.13) and (6.15) that

$$g_n(\eta, \beta^o) = A_n(\eta, \beta^o) + B_n(\eta, \beta^o) = A_n(\eta, \beta^o), \quad (6.27)$$

and

$$\dot{g}_n(\eta, \beta^o) = \dot{A}_n(\eta, \beta^o) + \dot{B}_n(\eta, \beta^o) = o_p(1) + \dot{B}_n(\eta, \beta^o). \quad (6.28)$$

Standardizing (6.27) by $\sqrt{Var[A_n(\eta, \beta^o)]}$, we have

$$\frac{g_n(\eta, \beta^o)}{\sqrt{Var[A_n(\eta, \beta^o)]}} = Z_n \xrightarrow{d} N(0, 1). \quad (6.29)$$

Substituting (6.29) into (6.26) and considering that $\dot{g}_n(\eta, \beta^o) = -\sum_{s=0}^{n(1-\zeta)} \dot{u}_s^2(\eta, \beta^o)$ yields

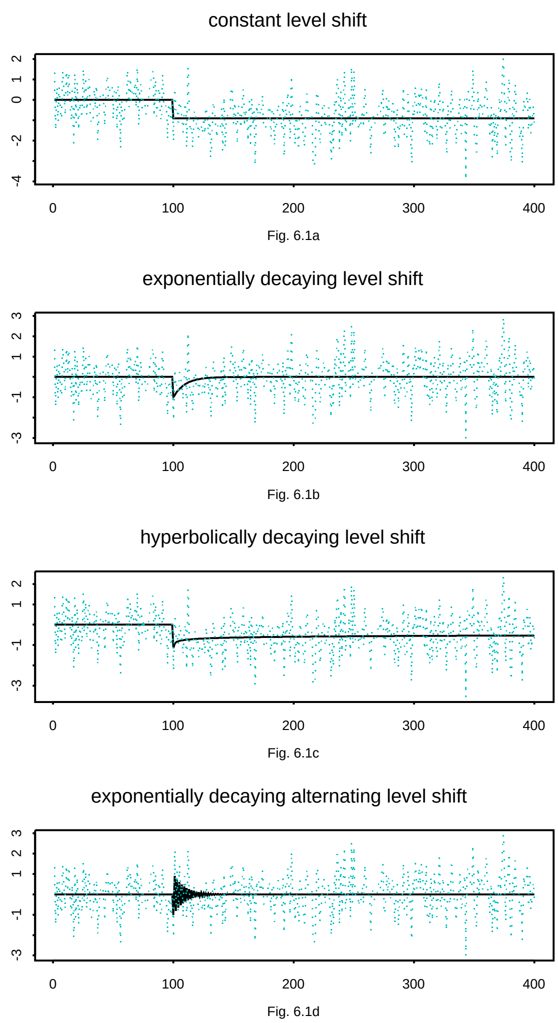
$$\begin{aligned} (\hat{\beta} - \beta^o) &\approx \frac{\sigma_\epsilon}{\sqrt{\sum_{s=0}^{n(1-\zeta)} \dot{u}_s^2(\eta, \beta^o)}} Z_n, \\ &\approx \frac{\sigma_\epsilon}{\sqrt{Var[A_n(\eta, \beta^o)]}} Z_n. \end{aligned} \quad (6.30)$$

Applying Taqqu's [79] lemma 3.1, we obtain the asymptotic variance

$$\begin{aligned} Var[A_n(\eta, \beta^o)] &= O\left[\sum_{s=0}^{n(1-\zeta)} L_2^2(s, \beta^o) s^{-2\alpha(\delta, \beta^o)}\right], \\ &\sim L_2^2(n(1-\zeta), \beta^o)[n(1-\zeta)]^{1-2\alpha(\delta, \beta^o)}, \end{aligned} \quad (6.31)$$

which completes the proof.

Figure 6.1: Simulated examples: intervention (full line) and contaminated (dotted line) series



Chapter 7

Final remarks and suggestions for further research

The dissertation has addressed model choice, forecasting, temporal aggregation and intervention analysis of stationary and nonstationary fractional autoregressive processes. The proposed methods provide a unified framework for modeling (polynomial deterministic) trends, long-memory, short-memory, antipersistence and nonstationarity simultaneously. The techniques help the data analyst to answer the question which components are present in the observed time series and to calculate the corresponding point- and interval-forecasts. In order that the considered method is effective in general, the observed series must not be too short, say $n = 200$. The main goals of this dissertation, in particular, can be summarized as follows.

Chapters 2 and 3 have put parameter estimation, model choice and forecasting into focus. We found that:

1. The AIC is of the same general form as for stationary autoregressive processes. The corresponding versions of the BIC of Schwarz [74] and the HIC of Hannan and Quinn [51] are shown to yield (approximately) consistent estimators of the autoregressive order.

2. In 'stationary versus unit root' approaches, a decision has to be made between stationarity and difference-stationarity. A wrong decision has an extreme impact on forecast intervals, since the length of forecast intervals is asymptotically constant for $d = 0$ whereas it diverges to infinity at the rate $\sqrt{k}$ for $d = 1$. In contrast, for FARIMA models,

prediction intervals are of order $O(k^{\tau/2})$ with τ varying in a continuous range, including $\tau = 0$ and $\tau = 1$ as special cases. The value of $\tau = max\{0, 2d-1\}$ is estimated from the data by maximum likelihood. As a result, prediction intervals are better adapted to the observed data, and often shorter. The extreme decision between $O(1)$ and $O(k^{.5})$ is avoided.

3. The comparison of different model choice criteria suggests that, the usage of the BIC leads to more reliable point- and interval-forecasts. Also, in comparison to the random walk, using FARIMA predictions turned out to be highly competitive, in terms of point- and in terms of interval-forecasts. Random walk intervals are, in comparison, either too optimistic (e.g. too short) in the presence of long memory or unnecessarily wide if the series is antipersistent. Moreover, substantial improvements of point-forecasts are possible if the degree of persistence is moderate or strong.

Chapter 4 was entirely devoted to applications of the methods presented above. We fit FARIMA models to several nominal stock market indices and exchange rates and calculated the corresponding forecasts. In particular:

4. We found significant long memory in international stock market indices, regardless of the market size and of whether a market is referred to as developed or not, and (at least temporarily) evidence of antipersistence in foreign exchange rates. These findings are in sharp contrast to those reported in the literature.

5. The corresponding out-of-sample forecasts yielded superior point-forecasts (for some series) and more reliable interval-forecasts in comparison to random walk and traditional Box-Jenkins ARIMA predictions, indicating the failure of the martingale hypothesis.

In chapters 5 and 6, we presented new theorems on temporal aggregation and modeling external interventions of/in (non-)stationary fractional autoregressive processes respectively. The main goals were as follows:

6. We demonstrated that the (partially known) asymptotic results on temporal aggregation of fractional and nonfractional, stationary and

nonstationary autoregressive processes can be derived using one unified treatment. Our main result was that short-memory components vanish under temporal aggregation whereas long-range dependence remains. Interestingly, we showed that the increments of nonstationary fractional processes do not converge asymptotically to fractional Gaussian noise. This is in contrast to the stationary case and has important implications for analyzing (and forecasting) economic time series.

7. We derived necessary and sufficient conditions for estimating the parameters of an intervention consistently, and provide the corresponding asymptotic distribution when the timing of an intervention is known. We showed, in particular, that consistency is possible if an intervention decays slowly hyperbolically to zero, a non-zero constant or remains at a non-zero level.

Further refinements and extensions, however, will be worth pursuing in the future. We think that the following problems deserve further study. Some of them are the subject of our current research.

- The theorems presented in chapter 1 on estimation and model choice of (non-)stationary fractional autoregressive processes respectively remain unchanged if moving average components are included too. Simulations for models with moving average terms pose, however, computational problems. Here, for each d on a grid, the autoregressive parameters were estimated using the S-Plus function *ar.burg*, whereat *arima.mle* turned out to cause inacceptable long CPU times. The development of a more stable and faster algorithm which allows the additional estimation of moving average components is subject to ongoing research.

- Better and faster algorithms for forecasting FARIMA models, as discussed in the chapters 2 and 3, may be developed. In our numerical calculations, inversion of Σ_n and calculation of β_{opt} turned out to require very precise evaluation of $\gamma(k)$ for all lags. After trying a number of approaches, Monte Carlo calculation of the covariances turned out to be most reliable. Interpreting $\gamma(k) = \int_{-\pi}^{\pi} f(x)\cos kx dx$ as $2\pi E[f(W)\cos kW]$, where W is uniformly distributed on $[-\pi,\pi]$, $\gamma(k)$ was obtained by 100000 simulations. Also, more reliable confidence intervals may be obtained by incorporating the uncertainty due to the estimation of the optimal mean squared prediction error.

- So far, we did not consider processes incorporating deterministic trends with stationary errors and other then polynomial trends within the (non-)stationary FARIMA environment. These possibilities, however, can be incorporated by allowing μ to be a function in time satisfying certain smoothness assumptions. The corresponding theory on so called semiparametric fractional autoregressive (SEMIFAR) models, together with real and simulated data examples, can be found in Beran [11], Beran and Ocker [17], Beran and Ocker [18], Beran and Feng [13], Beran, Feng and Ocker [15], and Beran, Feng, Franke, Hess and Ocker [16].

- Also, a more sophisticated analysis may be obtained by applying ARCH extensions of FARIMA models to high-frequency (e.g. daily) (log-)price series. The mathematical theory for such extensions is subject of current research. In a recent paper Beran and Feng [14] extended the related (but more flexible) SEMIFAR approach to GARCH error processes.

- Additional extensions of FARIMA (and SEMIFAR) models are obvious. Incorporating seasonal components as well as parametric and nonparametric explanatory variables are the subject of current research.

- Moreover, a formal extension of our results on temporal aggregation to ARCH models deserve further study.

- Also, the inclusion of more complex intervention models may improve the reliability of the model class considered here in daily applications. Clearly, if the timing of an intervention is unknown, the analysis becomes much more complex. Estimation of intervention and FARIMA parameters simultaneously when the timing of an intervention is unknown should be discussed elsewhere.

Bibliography

[1] Abraham, B., Application of intervention analysis to a road fatality series in Ontario, *Journal of Forecasting*, 6, 211-220, 1987.

[2] Abraham, A., and Box, G.E.P., Bayesian analysis of some outlier problems in time series, *Biometrika*, 66, 229-236, 1979.

[3] Akaike, H., A Bayesian extension of the minimum AIC procedure of autoregressive model fitting, *Biometrika*, 66, 237-242, 1979.

[4] Amemiya, T., and Wu, R.Y., The effect of aggregation on prediction in the autoregressive model, *Journal of the American Statistical Association*, 67, 628-632, 1972.

[5] Baillie, R.T., Long memory processes and fractional integration in econometrics, *Journal of Econometrics*, 73, 5-59, 1996.

[6] Baillie, R.T., and Bollerslev, T., The mesage in daily exchange rates: a conditional-variance tale, *Journal of Business & Economic Statistics*, 7, 297-305, 1989.

[7] Beran, J., Estimation, testing and prediction for self-similar and related processes, *PhD Thesis*, ETH Zurich, 1986.

[8] Beran, J., A test of location for data with slowly decaying serial correlations, *Biometrika*, 76, 261-269, 1989.

[9] Beran, J., *Statistics for long-memory processes*, Chapman & Hall, New York, 1994.

[10] Beran, J., Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models, *Journal of the Royal Statistical Society*, 57, 659-672, 1995.

[11] Beran, J., Estimating trends, long-range dependence and nonstationarity, *Preprint*, 1998.

[12] Beran, J., Bhansali, R.J., and Ocker, D., On unified model selection for stationary and non- stationary short- and long-memory autoregressive processes, *Biometrika*, 85, 921-934, 1998.

[13] Beran, J., and Feng, Y., Local polynomial fitting with long-memory, short-memory and antipersistent errors, *Preprint*, 1999.

[14] Beran, J., and Feng, Y., Local polynomial estimation with a FARIMA-GARCH error processs, *Preprint*, 1999.

[15] Beran, J., Feng, Y., and Ocker, D., SEMIFAR models, Working Paper presented at the Workshop on 'Fractional integration and cointegration with application to empirical finance', University of Dortmund, 1998.

[16] Beran, J., Feng, Y., Franke, G., Hess, D., and Ocker, D., SEMIFAR models, with applications to commodities, exchange rates and the volatility of stock market indices, *Preprint*, 1999.

[17] Beran, J., and Ocker, D., SEMIFAR forecasts, with applications to foreign exchange rates, *Journal of Statistical Planning and Inference*, 80, 137-153, 1999.

[18] Beran, J., and Ocker, D., Volatility of stock market indices - an analysis based on SEMIFAR models, *Preprint*, 1999.

[19] Bhansali, R.J., Recursive order selection for an ARMA process, *Proceedings of the first US/Japan conference on the frontiers of statistical modeling: an informational approach*, H. Bozdogan (ed.), Kluwer Academic Publishers, Netherlands, 1994.

[20] Bierens, H.J., and Guo, S., Testing stationarity and trend stationarity against the unit root hypothesis, *Econometric Reviews*, 12, 1-32, 1993.

[21] Bollerslev, T, and Mikkelsen, H.O., Modeling and pricing long memory in stock market volatility, *Journal of Econometrics*, 73, 151-184, 1996.

[22] Box, G.E.P., and Jenkins, G.M., *Time series analysis: forecasting and control*, Holden Day, San Francisco, 1976.

[23] Box, G.E.P., and Tiao, G.C., A change in level of a non-stationary time series, *Biometrika*, 52, 181-192, 1965.

[24] Box, G.E.P., and Tiao, G.C., Intervention analysis with applications to economic and environmental problems, *Journal of the American Statistical Association*, 70, 70-79, 1975.

[25] Brooks, C., Linear and non-linear (non-)forecastability of high-frequency exchange rates, *Journal of Forecasting*, 16, 125-145, 1997.

[26] Chang, I., Tiao, G.C., and Chen, C., Estimation of time series parameters in the presence of outliers, *Technometric*, 30, 193-204, 1988.

[27] Cheung, Y.-W., Long memory in foreign-exchange rates, *Journal of Business & Economic Statistics*, 11, 93-101, 1993.

[28] Cheung, Y.-W., Test for fractional integration: a Monte-Carlo investigation, *Journal of Time Series Analysis*, 14, 331-345, 1993.

[29] Cheung, Y.-W., and Diebold, F.X., On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean, *Journal of Econometrics*, 62, 301-316, 1994.

[30] Cheung, Y.-W., and Lai, K.S., A search for long memory in international stock market returns, *Journal of International Money and Finance*, 14, 597-615, 1995.

[31] Crato, N., Some international evidence regarding the stochastic behaviour of stock returns, *Applied Financial Economics*, 4, 33-39, 1994.

[32] Crato, N., Ray, B.K., Model selection and forecasting for long-range dependent processes, *Journal of Forecasting*, 15, 107-125, 1996.

[33] Crato, N., Taylor, H.M., Stationary persistent time series misspecified as nonstationary ARIMA, *Statistical Papers*, 37, 215-223, 1996.

[34] Dahlhaus, R., Efficient parameter estimation for self-similar processes, *Annals of Statistics*, 17, 1749-1766, 1989.

[35] Diebold, F.X., *Empirical modeling of exchange rate dynamics*, Springer, New York, 1988.

[36] Diebold, F.X., and Nason, J.A., Nonparametric exchange rate predicition, *Journal of International Economics*, 315-332, 1990.

[37] Ding, Z., Granger, C.W.J., and Engle, R.F., A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, 1, 83-106, 1993.

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

[39] Ding, Z., and Granger, C.W.J., Modeling volatility persitence of speculative returns: a new approach, *Journal of Econometrics*, 73, 185-215, 1996.

[40] Engle, R.F., Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica*, 50, 987-1008, 1982.

[41] Euromoney, *The LGT guide to world equity markets 1996*, Euromoney, London, 1996.

[42] Fong, W.M., and Ouliaris, S., Spectral test of the martingale hypothesis for exchange rates, *Journal of Applied Econometrics*, 10, 255-271, 1995.

[43] Fox, A.J., Outliers in time series, *Journal of the Royal Statistical Society*, Ser. B, 43, 350-363, 1972.

[44] Fox, R., and Taqqu, M., Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Annals of Statistics*, 14, 517-532, 1986.

[45] Franke, G., Stapleton, R., and Subrahmanyan, M.G., When are options overpriced? The black-shooles model and alternative characterisations of the pricing kernel, to appear in *European Finance Review*.

[46] Frankel, J.A., and Rose, A.K., A survey of empirical research on nominal exchange rates, CIDER, University of California, Berkeley, Working Paper No. C95-051, 1995.

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

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

[49] Gradshteyn, I.S, and Ryzhik, I.M., *Tables of integrals, series and products*, Academic Press, London, 1965.

[50] Hannan, E.J., The estimation of the order of an ARMA process, *Annals of Statistics*, 8, 1071- 1081, 1980.

[51] Hannan, E.J., and Quinn, B.G., The determination of the order of an autoregression, *Journal of the Royal Statistical Society*, 41, 190-195, 1979.

[52] Hosking, J.R.M., Fractional differencing, *Biometrika*, 68, 165-176, 1981.

[53] Hosking, J.R.M., Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series, *Journal of Econometrics*, 73, 261-284, 1996.

[54] Haslett, J., and Raftery, A.E., Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource, *Applied Statistics*, 38, 1-50, 1989.

[55] Hassler, U., and Wolters, J., On the power of unit root tests against fractional alternatives, *Economics Letters*, 45, 1-5, 1994.

[56] Hauser, M.A., Semiparametric and nonparametric testing for long memory: A Monte Carlo study, *Empirical Economics*, 22, 247-271, 1997.

[57] International Finance Corporation (IFC), *Emerging stock markets factbook 1996*, IFC, Washington, 1997.

[58] Koreisha, S.G., and Pukkila, T.M., New approaches for determining the degree of differencing necessary to induce stationarity in ARIMA models, *Journal of Statistical Planning and Inference*, 36, 399-412, 1993.

[59] Koreisha, S.G., and Yoshimoto, G, A comparison among identification procedures for autoregressive moving average models, *Int. Statist. Rev.*, 59, 37-57, 1991.

[60] Lamperti, J., Semi-stable stochastic processes, *Z. Wahrsch. verw. Geb.*, 22, 205-225, 1972.

[61] Ling, S.-Q., and Li, W.K., Fractional ARIMA-GARCH time series models, *Journal of the American Statistical Association*, 92, 1184-1194, 1997.

[62] Liu, C.Y., and He, J., A variance-ratio test of random walks in foreign exchange rates, *Journal of Finance*, 36, 773-785, 1991.

[63] Lo, A.W., Long-term memory in stock market prices, *Econometrica*, 59, 1279-1313, 1991.

[64] Mandelbrot, B.B., When a price be arbitraged efficiently? a limit to the validity of the random walk and Martingale models, *Review of Economics and Statistics*, 53, 225-236, 1971.

[65] Mandelbrot, B.B., *The fractal geometry of nature*, Freeman, New York, 1983.

[66] Mandelbrot, B.B., and van Ness, J.W., Fractional brownian motions, fractional noises and applications, *SIAM Review*, 10, 422-437, 1968.

[67] Mandelbrot, B.B., and Wallis, J.R., Computer experiments with fractional gaussian noises, *Water Resources Research*, 5, 228-267, 1969.

[68] Martin, R.D., and Yohai, V.J., *Robustness in time series and estimating ARMA models*, Handbook of Statistics, Volume 5: Time Series in the Time Domain, E.J. Hannan et.al. (ed.), 119-156, Amsterdam: North-Holland, 1985.

[69] Meese, R.A., and Rogoff, K., Empirical evidence rate models of the seventies - do they fit out of sample?, *Journal of International Economics*, 14, 3-24, 1983.

[70] Meese, R.A., and Rose, A.K., An empirical assessment of nonlinearities in models of exchange rate determination, *Review of Economic Studies*, 58, 603-619, 1991.

[71] Phillips, P.C.B., and Ploberger, W., Posterior odds testing for a unit root with data-based model selection, *Econometric Theory*, 10, 774-808, 1994.

[72] Pukkila, T, Koreisha, S., and Kallinen, A., The identification of ARMA models, *Biometrika*, 77, 537-548, 1990.

[73] Rogers, L.C.G., Arbitrage with fractional brownian motion, *Mathematical Finance*, 7, 95-105, 1997.

[74] Schwarz, G., Estimating the dimension of a model, *Annals of Statistics*, 6, 461-464, 1978.

[75] Shibata, R., Selection of the order of an autoregressive model by Akaike's information criterion, *Biometrika*, 63, 117-126, 1976.

[76] Sowell, F., Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics*, 53, 165-188, 1992.

[77] Stock, J.H., Deciding between $I(1)$ and $I(0)$, *Journal of Econometrics*, 63, 105-131, 1994.

[78] Stram, D.O., and Wei, W.W.S., Temporal aggregation in the ARIMA process, *Journal of Time Series Analysis*, 7, 279-292, 1986.

[79] Taqqu, M.S., Weak convergence to fractional brownian motion and to the Rosenblatt process, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 31, 287-302, 1975.

[80] Taqqu, M.S., Teverovsky, V., and Willinger, W., Estimators for long-range dependence: an empirical study, appeared in *Fractals*, 3, 785-788, 1995.

[81] Telser, L.G., Discrete samples and moving sums in stationary stochastic processes, *Journal of the American Statistical Association*, 67, 484-499, 1967.

[82] Tiao, G.C., Temporal aggregation of time series, *Biometrika*, 59, 525-531, 1972.

[83] Tsay, R.S., Order selection in nonstationary autoregressive models, *Annals of Statistics*, 12, 1425-1433, 1984.

[84] Tsay, R.S., Time series model specification in the presence of outliers, *Journal of the American Statistical Association*, 81, 132-141, 1986.

[85] Wei, W.W.S., *Time series analysis: univariate and multivariate methods*, Addison-Wesley, New York et.al., 1990.

[86] Yajima, Y., On estimation of long-memory time series models, *Australian Journal of Statistics*, 27, 303-320, 1985.

List of Tables

List of Figures