ERASMUS UNIVERSITY ROTTERDAM
ECONOMETRIC INSTITUTE

UNIVARIATE FRACTIONALLY INTEGRATED TIME SERIES MODELING

Pieter J.A. Stam

Rotterdam, August 1993

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my parents Rina and John, brother José and sister Karin. Each in their own way they encouraged me in good and bad times. Special thanks I owe to my girlfriend Ria. Since we met there was always this thesis to be finished, which made it difficult for me to concentrate on other things in life. She suffered from my moods when finishing this thesis and cared for me nonetheless.

I am very grateful to dr. Marius Ooms who took the time for reading and correcting the thesis. I appreciate his flexibility in finding time even for discussion at evenings. Furthermore, I would like to thank prof. dr. Bernard van Praag who gave me the opportunity to acquire plenty of working skills as a student-assistant.

Also I would like to thank Rob Flik, Marcel Warnaar and Arménio Bispo who created a great atmosphere to work in. Thanks as well to John Morssink, Wilbert Balk and Steven Kensenhuis with whom I lived together in Rotterdam. Finally I would like to thank all other persons who have cared for me and made it possible to reach this memorable point in my life.

Thank you all very much!

Piet

CONTENTS

CONTENTS (continued) *page*

1. INTRODUCTION

Limits of ARMA models of long-run dynamics

The emphasis in this thesis is on the long-run dynamics of univariate data generating processes. Stationary univariate stochastic processes are traditionally described by autoregressive moving-average (*ARMA*) models (see e.g. Box and Jenkins 1976; Priestley 1981; Judge et al 1982; Harvey 1981). However, these models are only applicable for describing long-run behavior if the *ARMA* parameter values are near the boundary of the parameter space. In this case the asymptotic distributions of estimated parameters and test statistics turn out to be inadequate as approximations to the finite sample distributions thereof. Hypothesis testing based on these asymptotic distibutions is invalid then. This is for example the case if the autoregressive polynomial function contains a so-called unit-root. Models of this type are called integer integrated *ARMA* models (*ARIMA*).

Furthermore, as is indicated by Sowell (1992a), an *ARMA* model that is designed to describe the long-run behavior is necessarily less appropriate to describe the shorter-run behavior of a process. This can be observed in the frequency domain from the spectrum which not only has power at the low frequencies associated with the long-run behavior but also at somewhat higher frequencies. Consequently, restrictions are placed on the possibilities to describe the short-run behavior of the process with a model designed for description of long-run behavior. Ideally, no restrictions are placed at other frequencies than the specific long-run frequencies.

A third drawback of using *ARMA* models for describing long-run dynamics is the impossibility to direct the fit of the parameters to the long-run characteristics of a series. As pointed out by Cochrane (1988), maximum likelihood (asymptotically) chooses parameter values to minimize the difference of the periodogram of the realization and the spectral density of the parametric model weighted at different frequencies. Consequently, the maximum likelihood parameter estimates may have been sacrificed to obtain a better description of the short-run dynamics although our interest centers on long-run dynamics.

Fractional models

When investigating the long-run behavior of a time series, a model should be considered that allows the long-run behavior to be captured. Clearly, the

ARMA models are not typically suited for modeling long-run dynamics. We therefore introduce univariate fractionally integrated *ARMA* models (*ARFIMA*), which turn out to have some desirable properties for this purpose (Granger and Joyeux 1980; Hosking 1981). *ARFIMA* models are a generalization of *ARIMA* models in that the integration parameter is not restricted to take integer values only. It turns out that processes that are integrated of an order less than a half are stationary nonetheless: we can define an autocorrelation function and a pseudo spectral density in this case. The autocorrelation function decays hyperbolically to zero and thus more slowly than autocorrelation functions of stationary *ARMA* processes which converge to zero exponentially. In the frequency domain we observe a generalization of the spectral density behavior near zero frequency as compared to behavior typical for *ARFIMA* processes. The integration parameter dictates how long shocks are felt in the system, i.e. the persistence of shocks. Furthermore, the long-run dependence given a fractional order of integration is achieved with less restrictions on the higher frequency behavior of the series than with integer order of integration. A third comparative advantage of *ARFIMA* models is the possibility to direct the parameter estimates for purposes one wants the model to be used for. Hosking (1981) shows that in the *ARFIMA* model the short-run dynamics of a process are captured by the autoregressive and moving-average parameters, whereas the fractional integration parameter captures its long-run behavior.

As a measure of long-run dependence we apply the cumulative impulse response measure. Given a Wold decomposition (e.g. Box and Jenkins 1976) the value of this measure is equal to the moving-average polynomial function evaluated at frequency zero, i.e. at $exp\{0\}=1$. Cochrane (1988) states that measures based on this quantity are the only measures of the presence of a zero frequency unit root in a finite sample.

ARFIMA models appear useful for other purposes than long-run dynamics description as well. Traditionally, (augmented) Dickey-Fuller tests are applied to a realization in order to test the null hypothesis of an autoregressive unit-root against a whole class of stationary alternatives: rejection of the null hypothesis then implies stationarity. Fitting an *ARFIMA* model without any autoregressive and moving-average parameters to the sample allows one to determine the fractional order of integration, i.e. any real value between zero and one is allowed (Hassler 1993). Given the Dickey-Fuller test of a point-hypothesis against a whole class of points, this amounts to a more symmetrical treatment of the null and alternative hypothesis.

Furthermore, it should be noted that rejection of the Dickey-Fuller null hypothesis does not rule out non-stationary behavior. Only a parameter value between zero and a half implies stationarity. If the Geweke and Porter-Hudak (1983) two-stage semi-parametric estimation procedure is applied an additional advantage of this alternative unit root testing strategy is that no model structure has to be specified preceding the estimation stage.

As an example of an application of the generalization described above we note that deterministic long-run behavior of processes is modeled as a polynomial of time. Linear trends are commonly observed in econometric and economic literature. Discrimination between stochastic and deterministic trend behavior amounts to specifying a model that nests both types of long-run dynamics and subsequently testing the order of integration. Allowing for fractional integration parameter values enriches the class of admissible possibilities such that mixtures of both types are allowed as well. Compare for instance Nelson and Plosser (1982) to Sowell (1992a).

Outline of the study

In Section 2 we describe the theory underlying *ARMA*, *ARIMA* and *ARFIMA* modeling of univariate stochastic processes. We highlight the unit root testing strategies and the prediction theory associated with the three model types. The consequences of deterministic trend behavior for modeling the processes are treated as well.

In Section 3 we present the estimation methods. In this thesis the nonfractional *ARMA* parameters are estimated using three estimation methods, i.e. Yule-Walker, ordinary least-squares and maximum likelihood. The fractional parameter in the *ARFIMA* model is estimated using two methods. The first method is the two-stage estimation procedure as proposed by Geweke and Porter-Hudak (1983) the integration parameter is estimated first by using a frequency domain property of *ARFIMA* processes; the *ARMA* parameters of the transformed process are modeled afterwards. The other method is maximum likelihood estimation as proposed by Sowell (1992b), such that the integration parameter and *ARMA* parameters are estimated simultaneously.

In Section 4 we illustrate the theory with economic and non-economic data sequences. The economic series is quarterly real US gross national product (seasonally adjusted, 172 data points); the non-economic series is annual Trier oak tree ring widths (1143 data points). Section 5 concludes.

2. TIME SERIES MODELS

In Section 2.1 we discuss the traditional so-called *ARMA* models. The concept of asymptotic stationarity plays a crucial role here. In Section 2.2 we present the natural extension to integrated *ARMA* models, where the so-called order of integration is traditionally restricted to integer values. In Section 2.3 we come to the main body of this thesis, fractionally integrated *ARMA* processes which allow the integration order to take any real value. These kinds of models are particularly useful to model long-term dynamics. Simulation experiments are presented in Section 2.4 in order to get a grasp of the empirical implications of the presented fractional integration theory.

2.1 *ARMA* MODELS

Suppose that a univariate stochastic process y with zero mean has the following *AutoRegressive Moving Average* model, i.e. $y \sim ARMA(p+d,q)$:

$$(1) \qquad \alpha(L)\, y_t \;=\; \theta(L)\, \varepsilon_t$$

where

$$\alpha(L) \;=\; 1-\alpha_1 L-\alpha_2 L^2-\ldots-\alpha_{p+d}L^{p+d}$$

and

$$\theta(L) \;=\; 1-\theta_1 L-\theta_2 L^2-\ldots-\theta_q L^q$$

where y_t is a realization of the process y $(t=1,2,\ldots,T)$, L is a lag-operator such that $Ly_t \equiv y_{t-1}$ and ε_t is the t-th observation of a covariance stationary stochastic error process.

A process ε will be called <u>covariance stationary</u> if its first two moments exist, are finite, and are independent of time; covariances only depend on the time span between two observations of the process:

$$\begin{aligned}
&\mathcal{E}(\varepsilon_t) && = \mu_\varepsilon && |\mu_\varepsilon| < \infty \\
&\mathcal{E}(\varepsilon_t-\mu_\varepsilon)^2 && = \sigma_\varepsilon^2 && \sigma_\varepsilon^2 < \infty \\
&\mathcal{E}(\varepsilon_t-\mu_\varepsilon)(\varepsilon_{t-k}-\mu_\varepsilon) && = \gamma_\varepsilon(k) && |\gamma_\varepsilon(k)| < \infty,\ k=1,2,\ldots
\end{aligned}$$

A stationary process ε is said to be <u>ergodic</u> if the sample mean of every function of finite observations tends to its expected value in mean square, at least when the expected value of the square of the function exists (Griliches and Intriligator 1983, p. 243, footnote 7). The following

quantities are consistent estimators of the mean μ_ε, variance σ_ε^2 and covariances $\gamma_\varepsilon(k)$ respectively (Harvey 1981, except for the correction factors for the variance and covariances):

$$\begin{aligned}\hat{\mu}_\varepsilon &= T^{-1}\sum_{t=1}^{\infty}\varepsilon_t \\ \hat{\sigma}_\varepsilon^2 &= (T-1)^{-1}\sum_{t=1}^{\infty}(\varepsilon_t-\hat{\mu}_\varepsilon)^2 \\ \hat{\gamma}_\varepsilon(k) &= (T-k-1)^{-1}\sum_{t=1}^{\infty}(\varepsilon_t-\hat{\mu}_\varepsilon)(\varepsilon_{t-k}-\hat{\mu}_\varepsilon)\end{aligned}$$

In this case, a single very long realization of the stationary process allows us to infer everything about the probability law generating that process, i.e. the finite sample moments converge to the infinite sample moments, which are equal to the population moments with probability one (Nerlove, Grether and Carvalho 1979). If we say that ε is stationary, we hereafter assume that it is ergodic as well and that it has finite moments. Furhermore, Unless stated otherwise in this thesis, ε_t will be assumed to be a normally distributed series of independent random shocks with zero mean and variance σ_ε^2, i.e. ε_t-$NID(0,\sigma_\varepsilon^2)$. We speak of "white noise".

Given this covariance stationary process ε, what can we say about the stochastic properties of the process y? To study these properties we have to investigate the properties of the lag-polynomials $\alpha(L)$ and $\theta(L)$. Because the algebra of these polynomial operators is isomorphic to the algebra of the polynomial functions $\alpha(z)$ and $\theta(z)$ (see Franses 1991 and the references therein), we can study the properties of the polynomial operator $\alpha(L)$ by looking at the polynomial functions $\alpha(z)$ and $\theta(z)$, z being a complex variable. The behavior of the discrete time series y_t is different as the roots of the equations $|\alpha(z)|=0$ and $|\theta(z)|=0$ — the zeros of the determinants of the polynomials $\alpha(z)$ and $\theta(z)$ — fall in different regions of the complex plane.

The polynomial function $\alpha(z)$

Let us first concentrate on the polynomial function $\alpha(z)$. Denoting the smallest zero by ξ, we can classify *ARMA* processes as follows.[1]

- If $|\xi|>1$, the process y is called <u>asymptotically stationary</u>, i.e. the time series y_t converges to a covariance stationary time series if t tends to

[1] In this thesis we abstain from processes with an autoregressive polynomial function that has at least one zero inside the complex unit circle, i.e. $|\xi|<1$. These processes are called explosive.

infinity.

However, it is not guaranteed that y is covariance stationary if ξ is outside the complex unit circle. In the $AR(1)$ case[2] for example, the latter condition boils down to the condition $|\alpha_1| < 1$. Following Dickey, Bell and Miller (1986), we demonstrate in Appendix A that mean and variance depend on the properties of the initial observation. Therefore, to ensure covariance stationarity, the following additional conditions are necessary: the starting point of the series is independently distributed from the disturbances ε_t, the initial shock ε_0 is zero, the initial observation y_0 has zero mean, and y_0 has zero variance or the same variance as the realizations $\{y_t\}_{t=1}^{\infty}$ of the process y. The five conditions together are sufficient to ensure covariance stationarity. In appendix A it is noted that similar arguments hold for higher order models.

- If ξ is a zero of α with $|\xi| = 1$, then ξ is called a unit root. A distinction is usually made between the real root -1, the real root +1 and the complex roots $exp(-i\lambda)$ with $\lambda \in (0,2\pi)\backslash\{\pi\}$. The existence of unit roots in the autoregressive structure equal to +1 has led to a distinct class of models, so-called *ARIMA* models (see Sections 2.2 and 2.3).

In general, we can classify a stochastic process y by putting it in moving-average form, i.e.[3]

$$y_t = \sum_{k=0}^{\infty} h_k \varepsilon_{t-k}.$$

The stochastic process y is called asymptotically stationary if the variance of the process y is finite and (asymptotically) non-stationary in case of infinite variance. (Gourieroux and Monfort 1990).[4]

The variance of the process y is finite if the coefficients h_k are squared summable, i.e. $\sum_{k=0}^{\infty} h_k^2 < +\infty$.

For the $AR(1)$ model the moving-average approximation is

---

[2] *ARMA(p,0)* and *ARMA(0,q)* processes are usually called *AR(p)* and *MA(q)* processes respectively.

[3] Putting the model of y as given in (1) in moving-average form is equivalent to finding a particular solution to this stochastic difference equation.

[4] The condition $\sum_{k=0}^{\infty} h_k^2 < \infty$ implies $lim_{k\to\infty} h_k = 0$.

$$y_t = \sum_{k=0}^{\infty} h_k . \varepsilon_{t-k},$$

where $h_k = \alpha_1^k$ if $|\alpha_1| < 1$ and so the process y is asymptotically stationary if $\sum_{k=0}^{\infty} \alpha_1^{2k} < +\infty$. From equation (A3) in Appendix A it is seen that in this case any new observation at t causes the variance of the process y to increase with only $\alpha_1^{2(t-1)}\sigma_\varepsilon^2$.

If $\alpha_1 = 1$ the $AR(1)$ process has a unit root $+1$; we call such a process a random walk. From $(A3)$ it can be seen that in this case any new observation adds variance σ_ε^2 to the process. Notice that a random walk process can not be put in a finite moving-average form.

The polynomial function $\theta(z)$

We now turn to the polynomial function $\theta(z)$. If the smallest zero of the moving-average polynomial lies outside the complex unit circle, the process y is called invertible. In this thesis we only deal with invertible *ARMA* processes.

2.1.1 The autocovariance function

The autocovariance generating function of a stationary process y (a.c.g.f.), say $g_y(z)$, is defined as (Harvey 1989)

$$(2) \qquad g_y(z) = \sum_{k=-\infty}^{\infty} \gamma_y(k) . z^k,$$

where $\gamma_y(0) = \sigma_y^2$ and the autocovariance $\gamma_y(k)$ corresponding to z^k is as defined before. In case the observations y_t are real, as is the case with most economic time-series, there holds $\gamma_y(-k) = \gamma_y(k)$. For the stationary *ARIMA* process, a so-called *ARMA* process, the a.c.g.f. is

$$g_y(z) = \sigma_\varepsilon^2 . \frac{\theta(z)\theta(1/z)}{\phi(z)\phi(1/z)}$$

The k-th order autocorrelation function is defined as $\rho_y(k) = \gamma_y(k)/\gamma_y(0)$. For *ARMA* processes the autocorrelation function can be approximated by

$$\rho_y(k) \sim a^k$$

where $0 < a < 1$ and $k \to \infty$. The autocorrelation function shows an exponential decay.

2.1.2 The power spectrum

In the frequency domain a powerful tool is the power spectrum of a stationary process, usually called the spectrum. The spectrum is defined as the Fourier transform of the autocovariances:

$$(3) \qquad f_y(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma_y(k).exp\{-i\lambda k\}, \qquad -\pi \leq \lambda \leq \pi$$

provided the right-hand side converges, which can be written as

$$f_y(\lambda) = (2\pi)^{-1} \left\{ \gamma_y(0) + 2 \sum_{k=1}^{\infty} \gamma_y(k).cos(\lambda k) \right\}$$

if y is a process that generates real observations. In this case the spectrum is symmetric around zero frequency. Note that λ is a continuous variable.

The autocovariance function and the spectrum contain the same information about the process; the difference is in the representation of this information. This can be seen by the inversion of (3) which gives

$$\gamma_y(k) = \int_{-\pi}^{\pi} f_y(\lambda).exp\{-i\lambda k\}\, d\lambda$$

from which we observe that $\sigma_y^2 = \gamma_y(0)$ can be decomposed into the spectrum values at the frequencies $\lambda \in [-\pi, \pi]$. We can therefore interpret the power spectrum as a measure of the contribution of a cycle with a particular angular frequency to the variance of the process.[5]

Moreover, the spectrum $f(\lambda)$ is proportional to the a.c.g.f. $g_y(.)$ evaluated along the complex unit circle in the following way (Harvey 1989):

$$f_y(\lambda) = (2\pi)^{-1} g_y(exp\{-i\lambda\}), \qquad -\pi \leq \lambda \leq \pi$$

where 2π is a normalization constant. This expression for the spectrum seems to be more convenient for the evaluation of spectra than (1). Harvey calls the right-hand side the 'spectral generating function' (s.g.f.).

Hence, for $ARMA(p,q)$ processes the s.g.f. is given by

$$(4) \qquad f_y(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\theta(exp\{i\lambda\})|^2}{|\phi(exp\{i\lambda\})|^2}$$

For example, the spectrum of an $AR(1)$ process is

---

[5] If we divide (3) by the variance σ_y^2 of the process we speak of spectral density because then there holds

$$\int_{-\pi}^{\pi} f_y(\lambda)\, d\lambda = 1$$

$$f_y(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{1-2\phi cos(\lambda)+\phi^2}$$

whereas the spectrum of a white noise process is constant:

$$f_y(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi}.$$

The latter spectrum indicates that the contribution of each cycle with a particular frequency to the variance of the process is just as much as that of any other. In the time domain this is equivalent with zero autocovariances at positive lags.

In the figures 2.1, 2.2, 2.3 and 2.4 we illustrate the relationship between the s.g.f. and the a.c.g.f.. In each figure we observe five panels: three panels with $T = 300$ realizations of a particular process for three different parameter values, a panel with the a.c.g.f. and a panel with the s.g.f., both for the three different parameter values. The dotted line in the latter panel is the spectrum of a white noise process. The variance of the error process is assumed to be one.

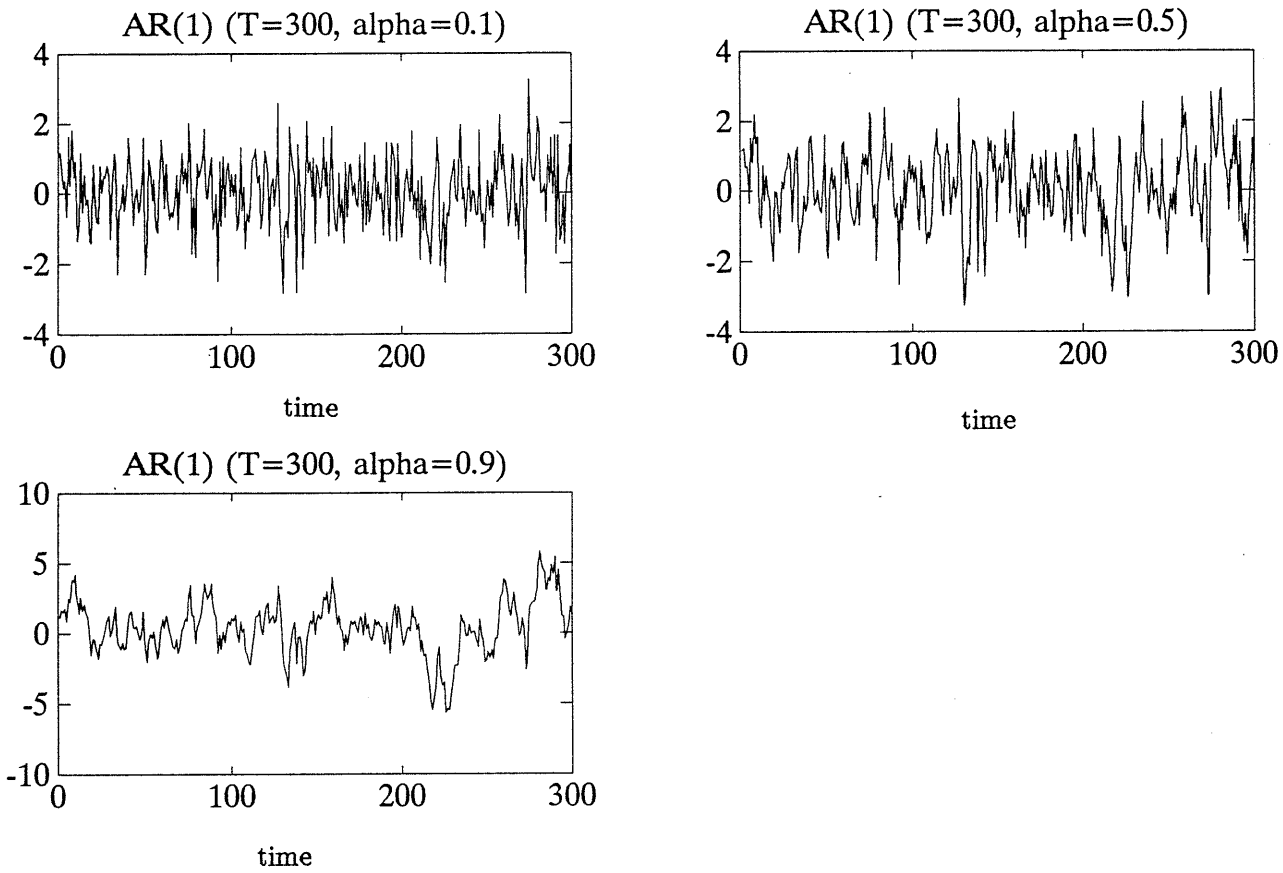
From figure 2.1 we observe realizations of an $AR(1)$ process for the parameter values $\alpha_1 = 0.1, 0.5, 0.9$. To obtain a realization of the AR(1) process we used zero as a starting-up value. It can be seen that the higher the parameter value, the more the observations depend on observations in the past. This longer memory property for a higher parameter value is also observed from the a.c.g.f. and the s.g.f., where the spectral density near zero tends to infinity for parameter values close to one.

From figure 2.2 we observe realizations of an $AR(1)$ process for neagtive parameter values. Apart from the negative correlation between adjacent observations as can be seen from the a.c.g.f., we observe that short-run cycles contribute more to the variance of the process the more negative the parameter value.

Figure 2.3 reveals the consequence of the a.c.g.f. of an $MA(1)$ process for the s.g.f.: the positive association between adjacant observations is equivalent to the predominance of long-run cycles in the series. The realizations of the $MA(1)$ process are obtained by an indirect Cholesky decomposition of the variance-autocovariance procedure (see Section 2.4.1).

Figure 2.4 deals with negative $MA(1)$ coefficients. The predominance of the

Figure 2.1


AR(1) (T=300, alpha=0.1)
AR(1) (T=300, alpha=0.5)
AR(1) (T=300, alpha=0.9)
time
time
time

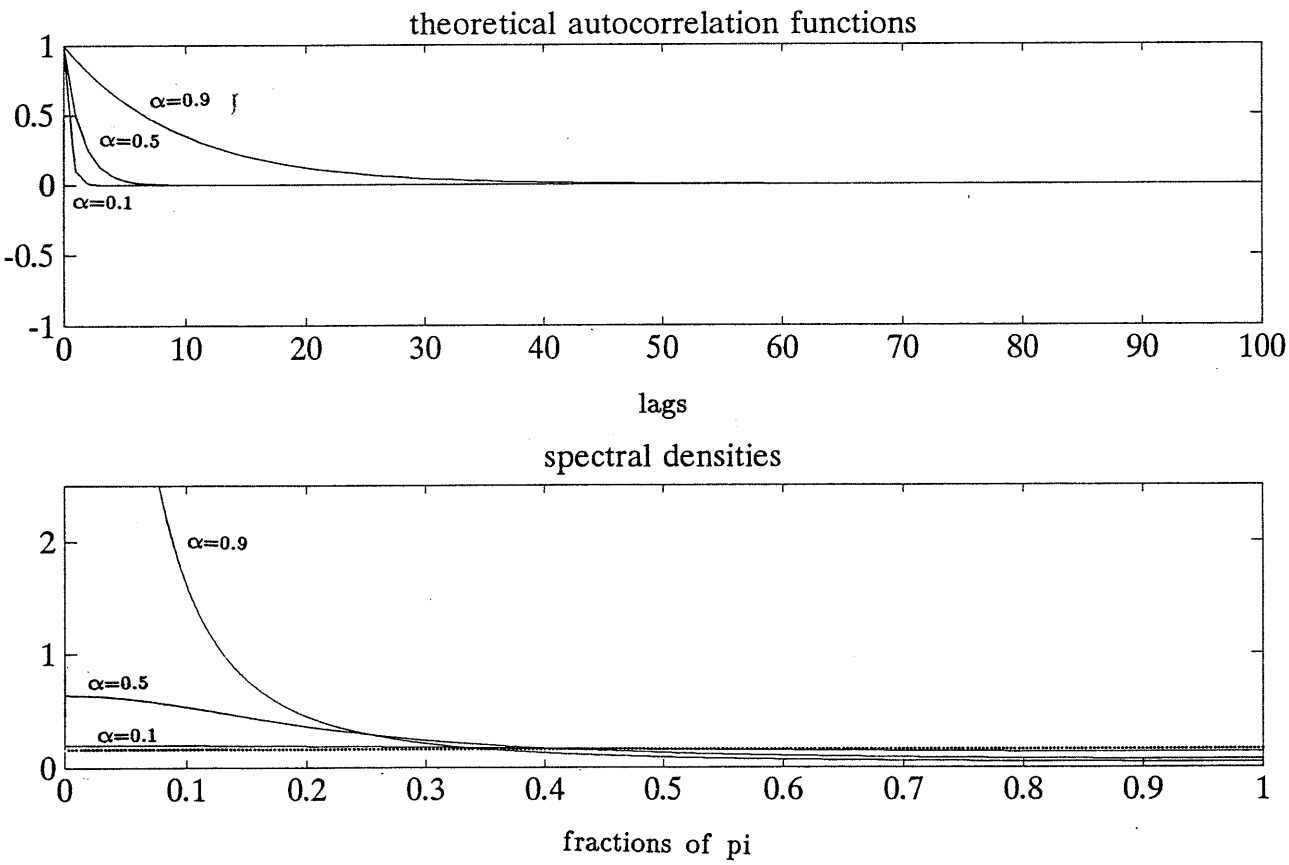
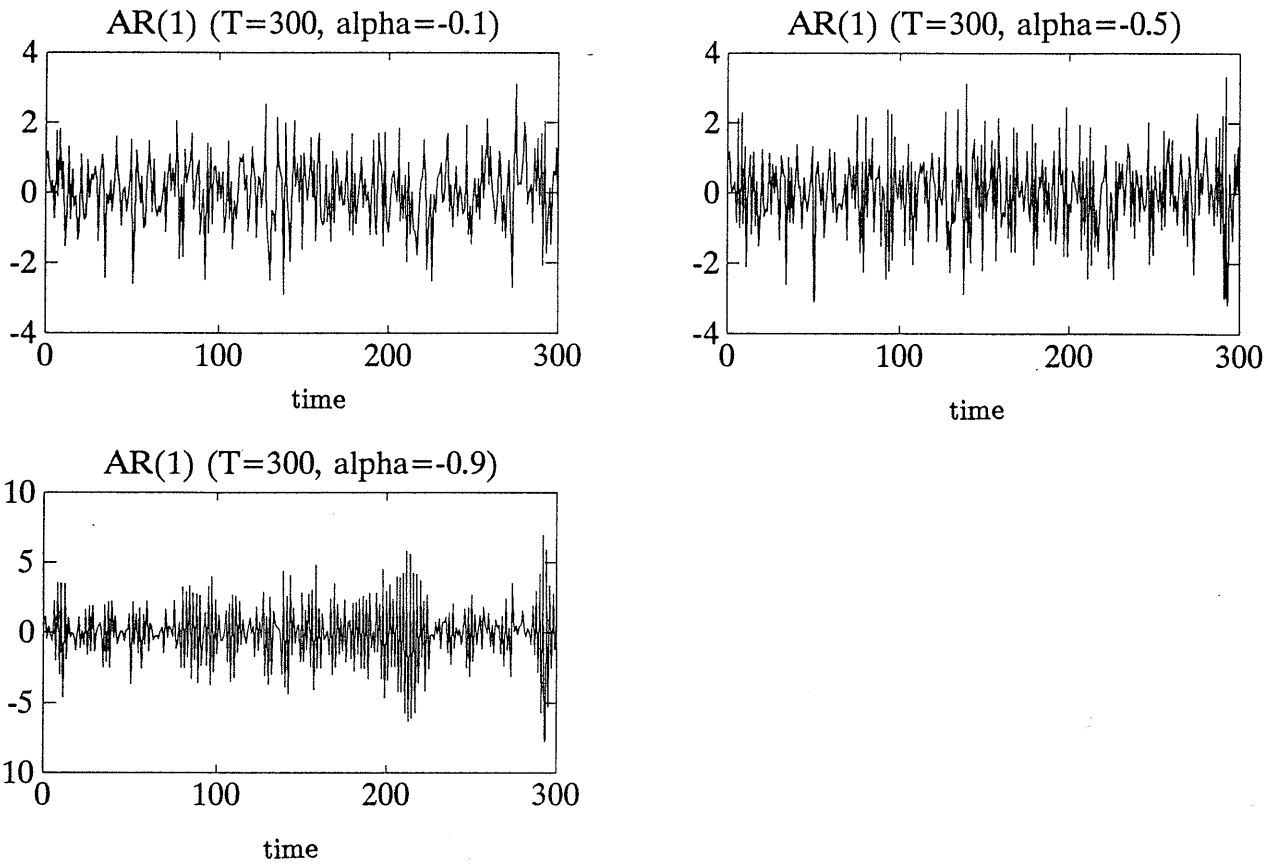

theoretical autocorrelation functions
α=0.9
α=0.5
α=0.1
lags
spectral densities
α=0.9
α=0.5
α=0.1
fractions of pi

Figure 2.2


AR(1) (T=300, alpha=-0.1)
AR(1) (T=300, alpha=-0.5)
AR(1) (T=300, alpha=-0.9)
time
time
time

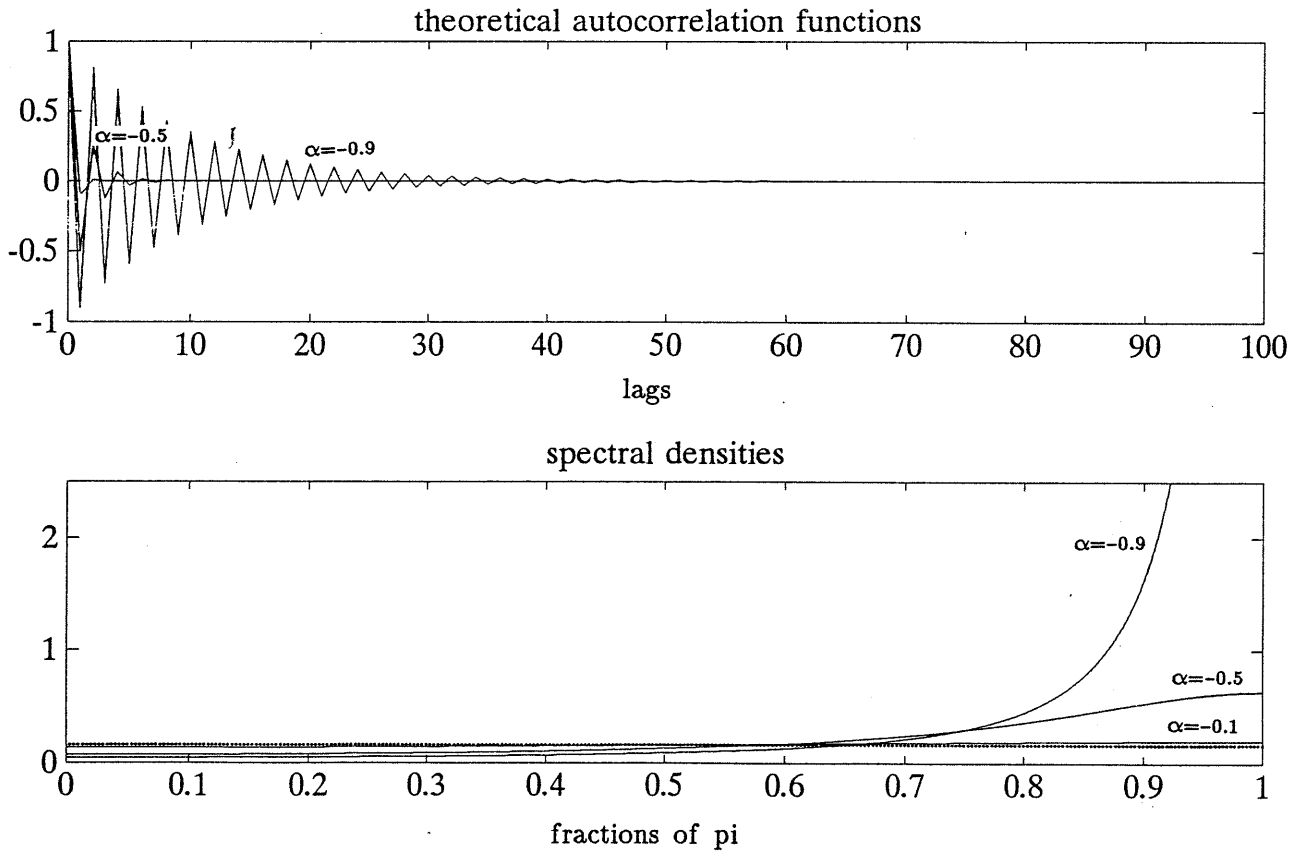
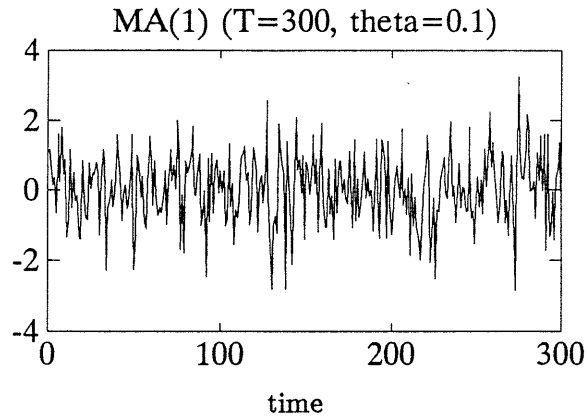
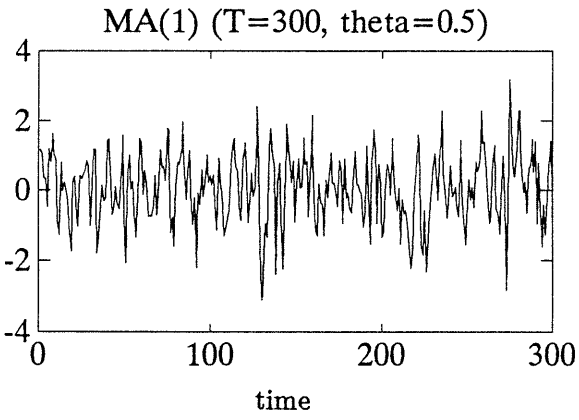
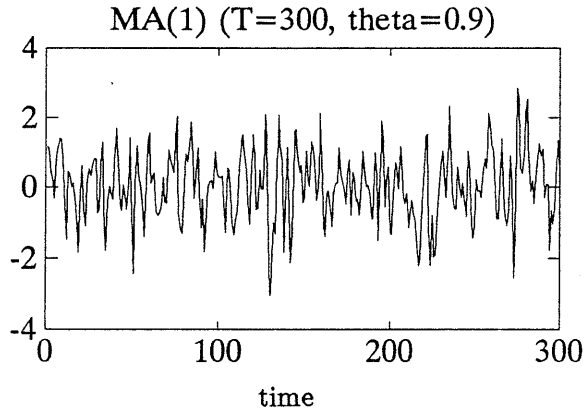
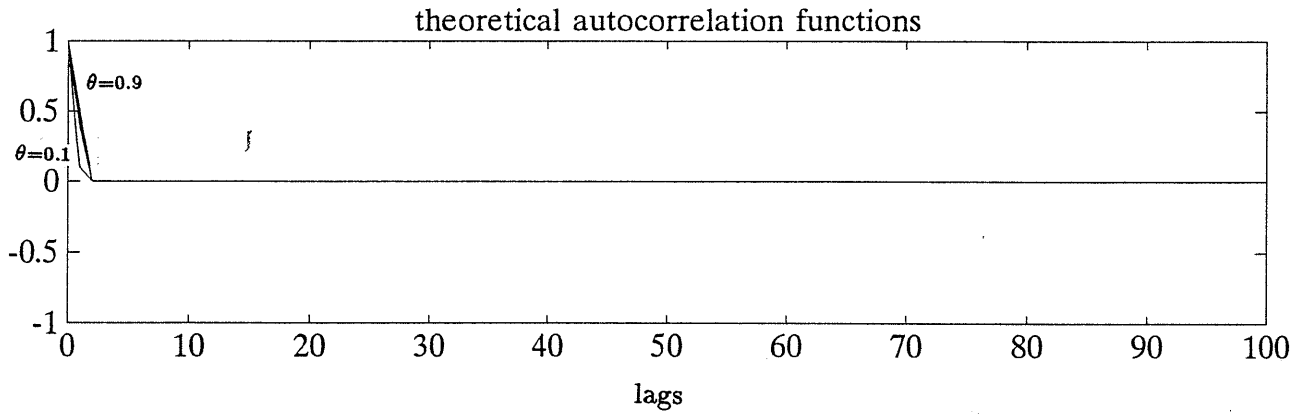

theoretical autocorrelation functions
α=-0.5
α=-0.9
lags
spectral densities
α=-0.9
α=-0.5
α=-0.1
fractions of pi

Figure 2.3


MA(1) (T=300, theta=0.1)
4
2
0
-2
-4
0
100
200
300
time


MA(1) (T=300, theta=0.5)
4
2
0
-2
-4
0
100
200
300
time


MA(1) (T=300, theta=0.9)
4
2
0
-2
-4
0
100
200
300
time


theoretical autocorrelation functions
θ=0.9
θ=0.1
1
0.5
0
-0.5
-1
0
10
20
30
40
50
60
70
80
90
100
lags

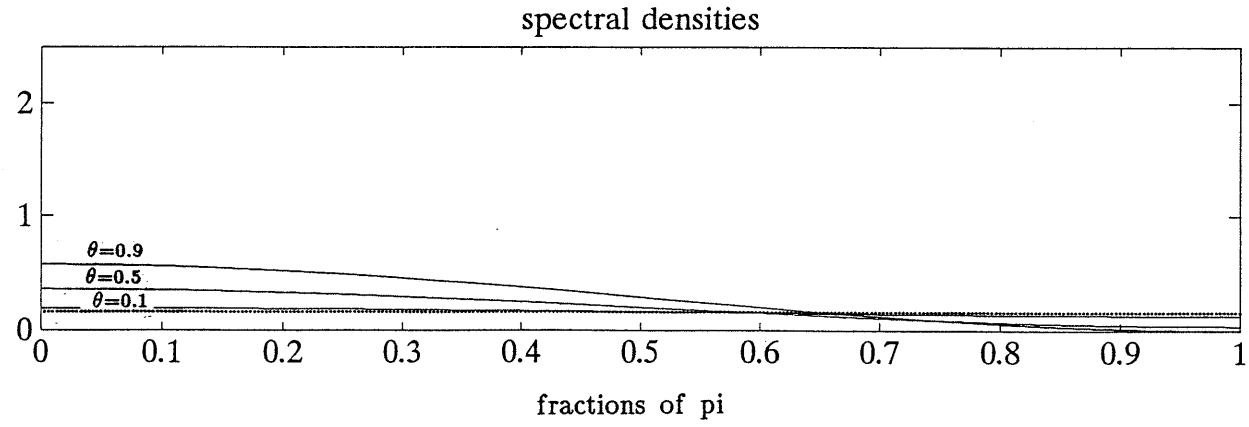
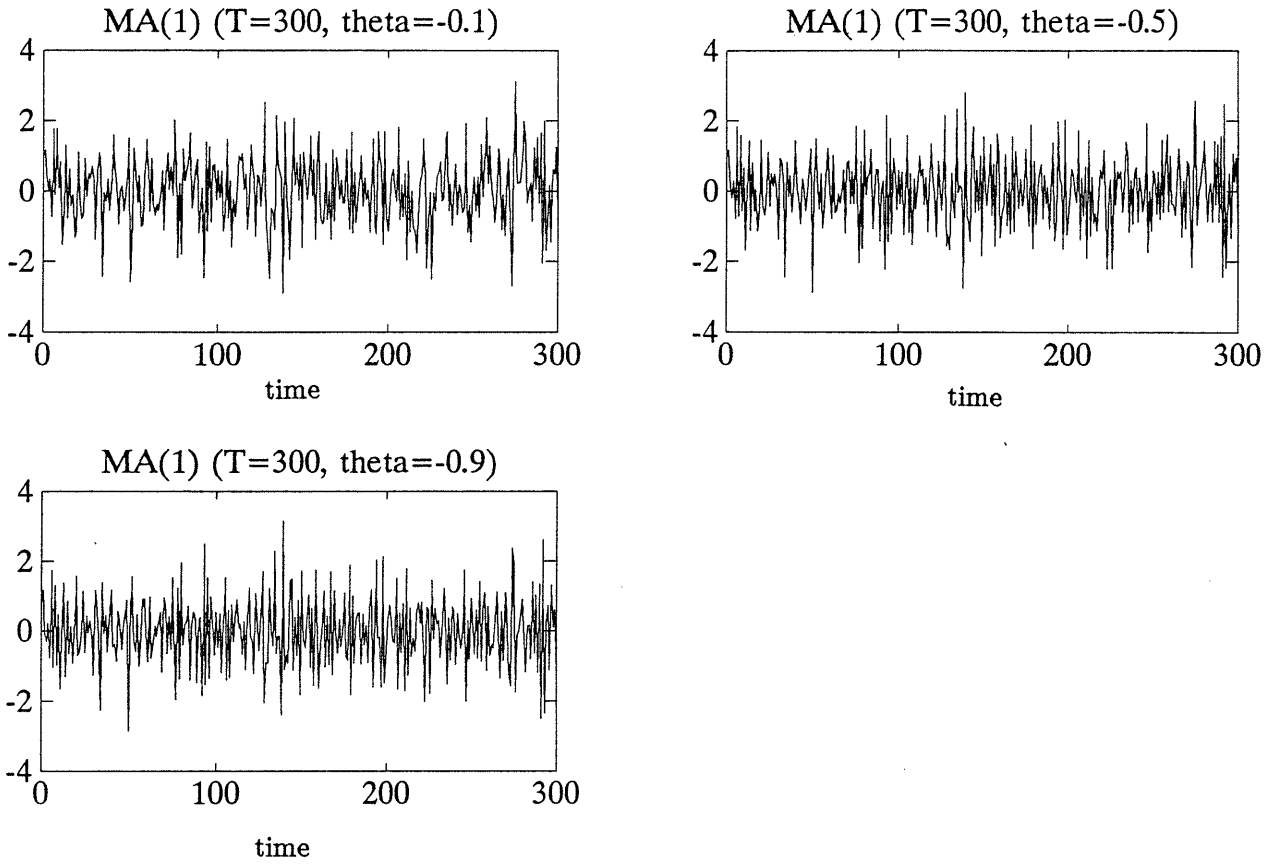
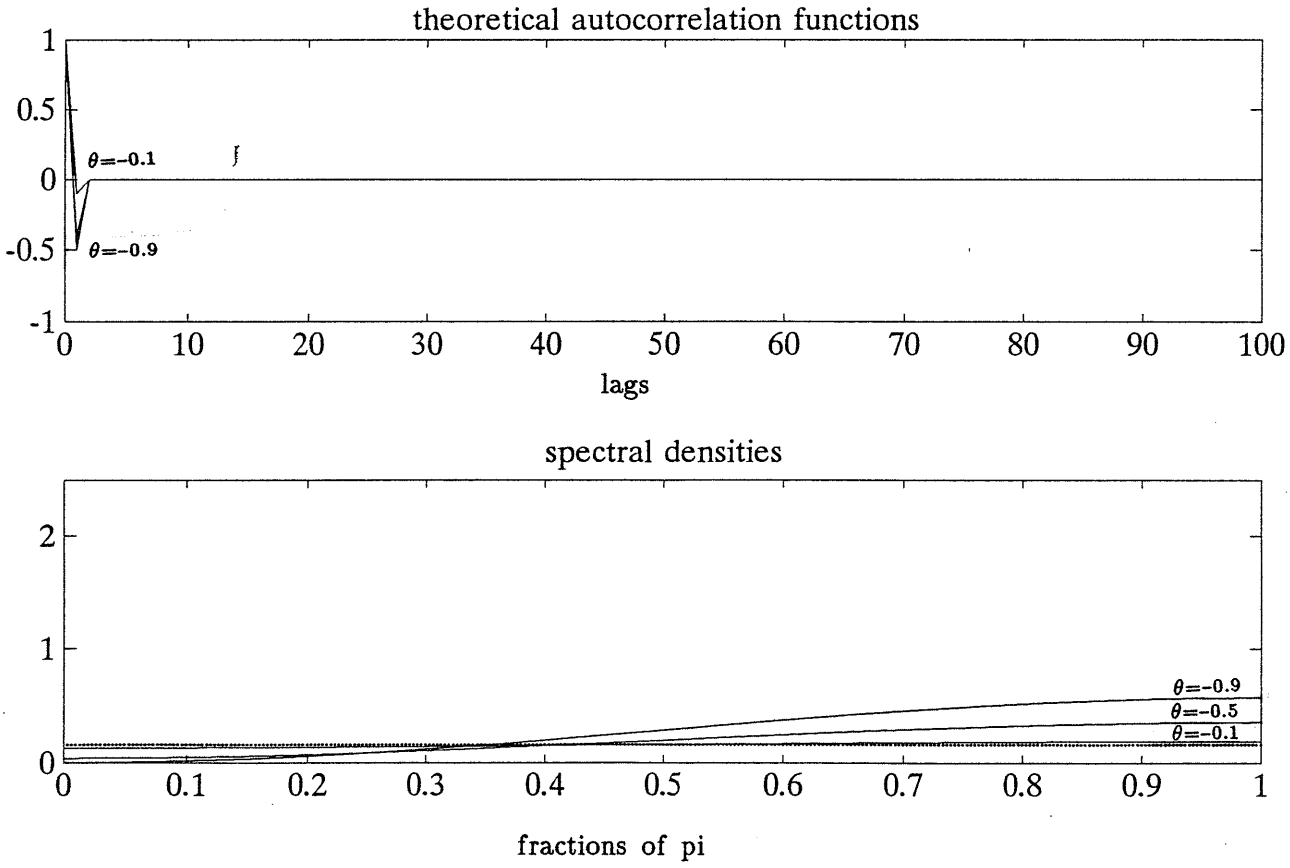

spectral densities
θ=0.9
θ=0.5
θ=0.1
2
1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
fractions of pi

Figure 2.4


MA(1) (T=300, theta=-0.1)
MA(1) (T=300, theta=-0.5)
MA(1) (T=300, theta=-0.9)
time
time
time


theoretical autocorrelation functions
θ=-0.1
θ=-0.9
lags
spectral densities
θ=-0.9
θ=-0.5
θ=-0.1
fractions of pi

shorter-run cycles is equivalent to the negative association between adjacent observations.

Sample equivalent of spectral density function

In practice we only observe realizations of a process and therefore need an estimator for the power spectrum. The sample equivalent of the power spectrum is the sample spectrum, say $I_y(\lambda)$, where the autocovariances are replaced by their sample estimates $\hat{\gamma}_y(k)$, with a denominator equal to T instead of $(T-k-1)$. The summation is restricted to $T-1$ in this case and the sample spectrum appears to be proportional to the squared absolute value of the Fourier transform of the realizations $\{y_1,\ldots,y_T\}$:

$$(5) \qquad I_y(\lambda) = (2\pi)^{-1} \sum_{k=-T+1}^{T-1} \hat{\gamma}_y(k).exp\{-i\lambda k\}, \qquad -\pi \leq \lambda \leq \pi$$

$$= (2\pi T)^{-1} \left| \sum_{k=1}^{T} y_t exp(-i\lambda k) \right|^2$$

The power spectrum can be seen as the average value of the sample spectrum in repeated realizations, i.e. there holds

$$lim_{T\to\infty} \mathcal{E}(I_y(\lambda)) = f_y(\lambda)$$

However, the sample spectrum is not a consistent estimator of the power spectrum: the variance does not decrease as the sample size increases.

Usually the frequency λ in (5) is restricted to be discrete, in particular $\lambda_j = 2\pi j/T$ is the jth harmonic of the fundamental frequency $2\pi/T$. Although the periodogram of a process is proportional and not identical to the sample spectrum where λ is discrete instead of continuous, usually one speaks of (5) as the periodogram.[6] We follow this convention.

It is interesting to note that for the discrete λ case Box and Jenkins (1976) showed that equation (5) is equal to

$$I_y(\lambda_j) = \frac{T}{2} (a_j^2 + b_j^2), \qquad j = 1,2,\ldots,q$$

If T is even $I(\lambda_q) = Ta_q^2$. The interesting point is that the coefficients (a_j, b_j) seem to arise from an equation where y_t is described as a linear combination of cosine and sine terms (Box and Jenkins 1976), the so-called real Fourier transform:

$$(6) \qquad y_t = a_0 + \sum_{j=1}^{q} \left\{ a_j cos(\lambda_j t) + b_j sin(\lambda_j t) \right\} + \varepsilon_t$$

---

[6] The periodogram is defined as $4\pi I_y(\lambda)$.

where $\lambda_j = 2\pi j/T$ and $q = [T/2]$.[7] As such the periodogram evaluated at the frequency λ_j indicates the contibution of the corresponding cycle to the variance of the process, as measured by the squared amplitude. Finally, it can be derived that

$$\sum_{t=1}^{T}(y_t - \bar{y}) \;=\; \sum_{j=1}^{q} I_y(\lambda_j).$$

2.1.3 Prediction Theory

The General Prediction Problem

Suppose y and x are two random variables with joint probability density $f(y,x)$. If we want to predict y from some function of x, say $u(x)$, Priestley (1981, Theorem 2.12.2) shows that $u(x) = \mathcal{E}(y \mid x)$ is the optimal predictor in a mean square sense, i.e. in this case the objective function

$$\mathcal{S}(u(x)) \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[y-u(x)]^2 f(y,x)\; dy\; dx$$

is minimized over all functions u of x.

Suppose for the moment that at time $(t-1)$ we have p observations $(y_{t-1}, y_{t-2}, \ldots, y_{t-p})$ and want to predict y_t. Following Priestley's (1981) argument the minimum mean square error predictor of y_t given this information, denoted by $y_{t-1,1}$, is

$$y_{t-1,1} \;=\; \mathcal{E}(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_{t-p}).$$

The first index implies that he realizations before and inclusive the period $(t-1)$ are known, the second index reflects the prediction horizon.

In practice the density function f is hardly ever known in advance and therefore we cannot say anything about this conditional mean. We therefore have to make an assumption. We assume $y_t, y_{t-1}, \ldots, y_{t-p}$ to be jointly normally distributed for which the following linear relationship holds

$$(7) \qquad y_t \;=\; a_{1,p}\cdot y_{t-1} \;+\; \ldots \;+\; a_{p,p}\cdot y_{t-p}$$

where the subscript p in $a_{k,p}$ denotes that the coefficient corresponds to an autoregressive polynomial of finite order and is introduced for future

---

[7] If T is odd, the least squares estimates of the coefficients a_0 and (a_i, b_i) where $i=1,\ldots,q$ are: $a_0 = \bar{y}$, $a_i = 2T^{-1}\sum_{t=1}^{T} y_t cos(\lambda_j t)$ and $b_i = 2T^{-1}\sum_{t=1}^{T} y_t sin(\lambda_j t)$. If T is even, the least squares estimates of the coefficients a_0 and (a_i, b_i) where $i=1,\ldots,q-1$ are as mentioned above, whereas $(a_q, b_q) = (T^{-1}\sum_{t=1}^{T}(-1)^t y_t, 0)$.

purposes. The optimal prediction in a mean squared error sense is obtained by minimizing $\mathcal{E}[(y_{t-1,1}-y_t)^2]$.[8]

Differentiating the latter expectation with respect to $a_{k,p}$, $k=1,\ldots,p$, gives the following set of equations for which the expectation is minimized:

$$(8) \qquad \gamma_y(k) = a_{1,p}\cdot\gamma_y(k-1) + \ldots + a_{p,p}\cdot\gamma_y(k-p)$$

where $k=1,2,\ldots,p$ (see e.g. Box and Jenkins 1976).[9] These equations are usually called the Yule-Walker equations and can be obtained directly by multiplying (7) throughout by y_{t-k} and taking expectations. In matrix notation the set of linear equations for $a_{1,p},a_{2,p},\ldots,a_{p,p}$ in terms of the autocovariances $\gamma_y(0),\gamma_y(1),\ldots,\gamma_y(p)$ reads:

$$(9) \qquad \begin{bmatrix}\gamma_y(1)\\ \vdots \\ \gamma_y(p)\end{bmatrix} = \begin{bmatrix}\gamma_y(0) & \ldots & \gamma_y(-p+1)\\ \vdots & & \vdots \\ \gamma_y(p-1) & \ldots & \gamma_y(0)\end{bmatrix}\begin{bmatrix}a_{1,p}\\ \vdots \\ a_{p,p}\end{bmatrix}$$

In practice the population values of the autocovariances are usually unknown. Estimates of the finite order autoregressive parameters are then obtained by substitution of estimates of the theoretical autocovariances (see also chapter 3 on estimation of time series models). A formula to obtain the standard errors for $\hat{a}_{1,p},\hat{a}_{2,p},\ldots,\hat{a}_{p,p}$ is described in Box and Jenkins (1976, eq. 7.3.8).

In general we make predictions of $y_{n,h}$, i.e. at time n we make a prediction of y_{n+h} which lies h periods ahead. The finite order coefficients of this "shifted" autoregressive model are then obtained from (9), where $(\gamma_y(1),\ldots,\gamma_y(p))'$ is then replaced by $(\gamma_y(h),\ldots,\gamma_y(p+h))'$.

Predictions made from a specified model

If we assume that the the realization results from a certain type of model we must procede in a different way to obtain predictions. Contrary to the previous discussion we now highlight the prediction problem when the predictions are made from a certain type of model. If we assume that y is an *MA(q)* process then there holds

---

[8] The predictor $y_{t-1,1}$ is optimal in the class of linear predictors. If the normality assumption holds as well, it is an optimal non-linear predictor as well.

[9] The equations can also be expressed in terms of the autocorrelations $\rho_y(1),\rho_y(2),\ldots,\rho_y(p)$.

$$y_{t,h} = \sum_{k=0}^{q-h} \theta_{k+h}\varepsilon_{t-k}, \qquad h \leq q,$$
$$= 0, \qquad h > q.$$

with corresponding prediction variance

$$V(y_{t,h}) = \sum_{k=0}^{q-h} \theta_{k+h}\sigma_\varepsilon^2$$

Note that $\varepsilon_t = y_t - y_{t-1,1}$ can be interpreted as a one-step ahead prediction error, also called innovation error. Making use of this fact and given $h \leq q$ we can write

$$y_{t,h} = \sum_{k=0}^{q-h} \theta_{k+h}(y_{t-k} - y_{t-k-1,1}),$$

which is a useful expression for updating purposes.

Granger and Newbold (1986) show that if y is an *AR(p)* process the optimal predictor $y_{t,h}$ is as follows:

$$y_{t,h} = \sum_{k=1}^{p} \alpha_k y_{t,h-k}$$

where $y_{t,k} = y_{t+k}$ for $k \leq 0$. In particular, if $p = 1$ then

$$y_{t,h} = \alpha_1 y_{t,h-1} = \alpha_1^h y_t$$
$$var(y_{t,h}) = (1-\alpha_1^{2h})\sigma_\varepsilon^2/(1-\alpha_1^2)$$

where the variance of the prediction $var(y_{t,h})$ tends to the variance of the process if h tends to infinity.

If the process under consideration is an *ARMA(p,q)* process Granger and Newbold (1986) show that

$$y_{t,h} = \sum_{k=1}^{p} \alpha_k y_{t,h-k} + \sum_{k=0}^{q-h} \theta_{k+h}(y_{t-k} - y_{t-k-1,1}), \qquad q-h \leq 0,$$
$$= \sum_{k=1}^{p} \alpha_k y_{t,h-k}, \qquad q-h > 0,$$

where $y_{t,h-k} = y_{t+h-k}$ for $h-k \leq 0$. In general, the variance of the prediction can be obtained by writing the *ARMA* model in moving-average form. Judge (1982, p. 699) presents a recursive formula to obtain the coefficients of the corresponding *MA*-representation from the autoregressive parameters ϕ and the moving-average parameters θ.

2.1.4 Modeling non-zero processes

In equation (1) we have assumed that the process y has zero mean. This assumption boils down to exactly the same model specification as that of a

non-zero mean process for which its mean has been subtracted from its realization:

$$(1)' \qquad \alpha(L)\ (y_t - \mu_y) = \theta(L)\ \varepsilon_t$$

It should be noted that the realization has to be corrected for its mean when calculating the a.c.g.f. of the non-zero mean process y as well, but that the periodogram is independent from this mean. This means that standard periodograms do not differ at non-zero frequencies for processes that only differ with respect to their mean.

Instead of modeling a mean-corrected non-zero mean process as in (1)' we can model the mean explicitly as follows

$$(1)'' \qquad \alpha(L)\ y_t = c + \theta(L)\ \varepsilon_t$$

The mean μ_y of the process y is then $c/\alpha(1)$; an estimate $\hat{\mu}_y$ can be obtained after c and the autoregressive parameters have been estimated. The possibility of extension to higher order polynomial deterministic behavior is obvious.

2.2 INTEGER INTEGRATED *ARMA* MODELS

To turn to the subject of this thesis, suppose that $\alpha(z)$ contains a zero-frequency unit root of multiplicity d, where $d \in \mathbb{N} \cup \{0\}$. Then there holds

$$\alpha(z) = \phi(z)(1-z)^d$$

so that equation (2.1.1) can be written as

$$(1) \qquad \phi(L)\ (1-L)^d y_t = \theta(L)\varepsilon_t$$

where

$$\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)$$

and $\theta(L)$ is as defined before; $t = 1, \ldots, T$. If all zeros of the polynomial $\phi(z)$ are outside the unit circle, equation (1) is called an *AutoRegressive Integrated Moving Average* model of the process y with parameters p, d and q, i.e. $y \sim ARIMA(p,d,q)$.[10] This type of presentation of discrete time series was introduced by Box and Jenkins (1970).[11]

In the special case where $\underline{d=0}$, y_t is said to be a realization of an asymptotic stationary process. We get a model analogous to equation (2.1.1), under the assumption that $|\xi| > 1$. Asymptotic stationary processes can be distinguished as short-memory processes and long-memory processes:

An asymptotic stationary process is of $\underline{short-memory}$ *if* $\sum_{k=0}^{\infty} |h_k| < +\infty$ *and of* $\underline{long-memory}$ *if* $\sum_{k=0}^{\infty} |h_k| = +\infty$, *where* h_k *is the* $k-th$ *moving-average coefficient if* (1) *is put in moving-average form.*

If the roots of the polynomials $\phi(z)$ and $\theta(z)$ are outside the unit circle and $d = 0$, the coefficients h_k of the infinite order moving-average model decay exponentially. In particular, $\sum_{k=0}^{\infty} h_k^2 < +\infty$ *and* $\sum_{k=0}^{\infty} |h_k| < +\infty$, i.e. *ARMA* processes have short memory.

On the other hand, if the integer $\underline{d > 0}$ the process is non-stationary, except under trivial assumptions like $\sigma = 0$. Engle and Granger (1987) call a process y with no deterministic component $\underline{\text{integrated}}$ of order d, denoted $y \sim I(d)$, if taking first differences d times leaves a process that has a stationary and

---

[10] In this thesis we assume that the smallest zero of the determinant of the polynomial function $\theta(z)$ is outside the unit circle in the complex plane.

[11] The *ARMA*(p,q) model is an *ARIMA*$(p,0,q)$ model.

invertible *ARMA* model. Notice that this concept reflects zero-frequency integration only.

The level of an integer integrated process is non-stationary. Random walk realizations for example, where $d=1$, can be thought of as summations of covariance stationary observations like ε_t. However, a positive value of d does not necessarily imply that the level of the integrated process is non-stationary: in this thesis an argument will be given for d to be any real number and it can be shown that for a subset of positive values we deal with a stationary process nonetheless.

It should be clear that the autocovariance and spectral generating function are not defined for integer integrated *ARMA* processes, because of non-stationarity.

2.2.1 Prediction Theory

For an $ARIMA(p,d,q)$ process y where $d>0$, prediction of a realization at time $t-1$ one step ahead necessarily has to be preceded by prediction of the realization of the stationary dth differenced process. Along the lines of Section 2.1.3 we obtain the prediction for the latter realization, denoted $\Delta^d y_{t-1,1}$. The prediction $y_{t-1,1}$ can be derived easily hereafter.[12]

2.2.2 Testing Unit Roots Against Stationary Alternatives

AR*(1) *processes

To determine the order of integration of a process y Fuller (1976) proposes the so-called Dickey-Fuller (DF) test on unit roots. The unit root null hypothesis H_0 is tested against the alternative H_1 of no unit root, i.e. it is a test of $I(1)$ integration against $I(0)$:

$$
\begin{aligned}
& H_0\colon\ y_t = y_{t-1} + \varepsilon_t, \\
(2)\qquad & H_1\colon\ y_t = \phi . y_{t-1} + \varepsilon_t, \qquad\qquad \phi \neq 1,
\end{aligned}
$$

where in both cases $y_0 = c$ is a constant term and ε_t is independently and identically distributed (Fuller 1976; Dickey and Fuller 1979).

---

[12] Suppose $d=1$. Then $y_{t-1,1}=\Delta y_{t-1,1}+y_{t-2,1}$, where $y_{t-2,1}$ can be replaced by y_{t-1} which is known at time $t-1$.

As we exclude explosive processes from our analysis, a realistic one-sided alternative hypothesis would be (2), where $|\phi| < 1$. It is common use to estimate equation (2) in first differences, i.e.

$$(3) \qquad \Delta y_t = a_0 . y_{t-1} + \varepsilon_t$$

where there is a unit root if $a_0 = 0$. The advantage of this expression is that $\hat{\tau}$ is the t-ratio corresponding to a_0 and is therefore readily available.

If we can not reject the hypothesis H_0 it is possible that zero-frequency unit root has multiplicity two or more. A DF test of $I(2)$ against $I(1)$ should then be applied to the first differences of the process y. Rejection of the $I(2)$ hypothesis then usually implies that we continue modeling first differences instead of the levels of the process. Notice that this testing strategy starts with the most general (unrestricted) model and continues testing more restricted models subsequently.

Size and power of the Dickey-Fuller test

From asymptotic theory we know that $(\hat{\phi} - \phi) = \mathcal{O}_p(T^{-1/2})$, or

$$T^{1/2}(\hat{\phi} - \phi) \overset{asy}{\sim} N(0, 1 - \phi^2),$$

provided $|\phi| < 1$. Under the null hypothesis the asymptotic distribution is degenerate as $plim_{T \to \infty} \; T^{1/2}(\hat{\phi} - 1) = 0$. Fuller (1976) showed $\hat{\phi}$ to be super-consistent: $(\hat{\phi} - 1) = \mathcal{O}_p(T^{-1})$. As a consequence of this difference in convergence the probability that $\hat{\phi} < 1$ given $\phi = 1$ approaches 0.6826 as sample size gets large, given testing on the basis of standard normal limit distributions: empirical size 0.6826 is much larger than nominal size (0.50). This size is not often used in testing however. Ooms (1993, p. 23) gives another example of the magnitude of error one can make by incorrectly ignoring unit root characteristics. example. In table 8.5.1 of Fuller (1976) finite sample and asymptotic critical values of $\hat{\phi}$ based on $T(\hat{\phi} - 1)$ are given, derived from Monte Carlo studies.

Another test-statistic is defined as $\hat{\tau} = (\hat{\phi} - \phi)/\hat{\sigma}_\rho$, where $\hat{\sigma}_\rho$ is the least-squares estimator of the standarddeviation of $\hat{\phi}$. As the standard t-distribution is not a good finite sample distribution of $\hat{\tau}$ under the nullhypothesis, for this statistic as well Fuller (1976) presented percentage points for the empirical distribution of $\hat{\tau}$ (table 8.5.2). In Dickey and Fuller (1979) analytical representations for the limit distributions of the estimator of ϕ and of the regression t test are derived.

De Jong et al (1992) show that powers of integration tests against plausible trend-stationary alternatives[13] can be quite low (see also Dickey and Fuller 1981, table IX), just as the powers of trend-stationarity tests against integrated alternatives can be. Furthermore, there are many cases in which neither test will reject. Evans and Savin (1981) derive power functions from exact distributions of $\hat{\phi}_\tau$ and show that a large value of T is required to achieve reasonable power at alternatives near the null hypothesis, particularly for values less than unity. They prove that for ϕ near but below unity the cumulative distribution function is very poorly approximated by the limiting normal, even for large values of T. Cochrane (1991) warns that application of unit root tests without consideration for their low power and for the restrictions that they inevitably impose in finite samples can be misleading. Furthermore, properties of unit root processes can be arbitrarily close to those of any given stationary process and vice versa.

Blough (1992) explains that the minimum power against any stationary alternative can be no greater than the maximum nominal size of the DF unit root test for all elements of the composite null. Hassler (1992) concludes from Monte Carlo experiments that in general unit root tests are not very powerful except when the probability of a type I error is high. Therefore, such tests can have power against any stationary alternative only if they also have excessive probability of false rejection for some unit root processes.

AR(p) processes

Equation (3) applies if we deal with an (integrated) $AR(1)$ process. Said and Dickey (1984) assume the residuals in equation (2) to be a stationary and invertible $ARMA(p,q)$ process, i.e.

$$H_0\colon\ y_t = y_{t-1} + u_t$$
$$H_1\colon\ y_t = \phi y_{t-1} + u_t, \qquad \phi \neq 1$$

where $\phi(L)u_t = \theta(L)\varepsilon_t$ and ε_t is i.i.d. again. If the data generating process of u_t is $\underline{AR(p)}$ we simply augment (3) with lagged first-differences of y. The stochastic process y can then be modeled as follows:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

Subtracting y_{t-1} from both sides and reordering terms results in the

---

[13] See appendix D.

following equation:

$$(3)' \qquad \Delta y_t = a_0 y_{t-1} + \sum_{k=1}^{p-1} a_k \Delta y_{t-k} + \varepsilon_t$$

where $a_0 = \sum_{k=1}^{p} \phi_k - 1$ and $a_k = -\sum_{i=k+1}^{p} \phi_k$ for $k = 1, \ldots, p-1$, $t = 1, \ldots, T$ and $p \geq 2$.

ARMA(p,q) processes

If the residual process u_t is (integrated) ARMA(p,q) the DF test can still be applied if we accept a finite *AR*-approximation of the *ARMA*-process (Said and Dickey 1984).[14] The autoregressive approximation assumption and the accompanying fact that in practice the approximation is finite reduce the reliability of the outcomes of the ADF-test. Firstly, use of the finite order autoregressive approximation can influence the outcome. For example, given the stationary and invertible *ARMA*(1,1) process we have

$$(1-\alpha L)\, y_t = (1+\gamma L)\, \varepsilon_t,$$

Under the assumption that the generating process has an autoregressive structure we have:

$$\frac{(1 - \alpha L)}{(1 + \gamma L)}\, y_t = \varepsilon_t$$

$$(1-\alpha L)\,(1-\gamma L-\gamma^2 L^2-\ldots)\, y_t = \varepsilon_t$$

$$(1-(\gamma+\alpha)L-\gamma(\gamma-\alpha)L^2-\ldots)\, y_t = \varepsilon_t$$

If $\alpha = 0.8$ and $\gamma = 0.2$ the autoregressive order p will probably be taken equal to one. The autoregressive root then equals $1/(\gamma+\alpha)$, i.e. one. Use of a finite order autoregressive approximation thus can lead to wrongly detecting a unit-root.

Secondly, several criteria can be used to choose the autoregressive order thereby possibly leading to alternative choices of p. The outcome of the test therefore depends on the choice of the criterion. One criterion is based on the t-value of the coefficient a_{p-1}. Sequentially, a lower order is chosen if equality to zero can not be rejected. Another criterion is the variance of the estimated residuals. In that case the minimal variance order p is chosen. Additionally, you can take account of the order p when seaking minimal variance. Such criteria are the Akaike- and Schwartz-criteria.

Thirdly, if we try to describe an *ARMA* process by a finite order *AR* model,

---

[14] A necessary condition to validate an AR-approximation of an ARMA process is that of invertibility: the smallest zero of the moving average polynomial function $\phi(z)$ should lie outside the unit circle in the complex plane.

the residuals of this augmented DF test will necessarily be correlated with the regressors: inconsistent parameter estimates are obtained if ordinary least squares is applied.[15] One of the solutions proposed is adding lagged first-differences to the DF regression (Said and Dickey 1984).

In this thesis the following test equation is estimated when applying the ADF test for zero frequency unit roots (compare (3') and (*D2b*)):

$$(3)'' \qquad \Delta y_t = c + \delta t + a_0 y_{t-1} + \sum_{k=1}^{p-1} a_k \Delta y_{t-k} + \varepsilon_t$$

It should be noted that the outcome of the test depends on the choice of the estimation period. If equation (3)'' is estimated recursively[16] it is possible that the null hypothesis is not repeatedly rejected and not repeatedly not rejected. An application of this test procedure can be found in e.g. Ooms (1993).

2.2.3 Measuring shock persistence

Once it is known that there exists a unit root, a measure of the permanent component is needed. One measure is the sum of the coefficients of the moving-average model of the first-differenced series, also known as the cumulative impulse response. Specifically, consider

$$(5) \qquad \Delta y_t = B(L)\ \varepsilon_t = (1 + b_1 L + b_2 L^2 + \ldots)\ \varepsilon_t,$$

where $B(L) = \phi(L)^{-1}\theta(L)$.

The impact of a unit shock in period t on the growth rate of Y at time $t+k$ is b_k, while the impact on the level of y at time $t+k$ is $c_k \equiv 1 + b_1 + b_2 + \ldots + b_k$.[17] In the limit we obtain c_∞, which is the effect of a unit shock today on the level of Y infinitely far in the future. For stationary series $c_\infty = 0$, because the effect of any shock is transitory as reversion to mean or trend

---

[15] If we approximate an *ARMA*(1,1) process by an *AR*(p) model, we get the following model:

$$(1 - \gamma L - \gamma^2 L^2 - \ldots - \gamma^p L^p) y_t = u_t$$

where the residual $u_t = (\gamma^{p+1} L^{p+1} + \gamma^{p+2} L^{p+2} + \ldots) y_t + \varepsilon_t$, ε_t being the error term of the *ARMA*(1,1) model.

[16] Recursive estimation means that an observation is added to the observation period each time the model is estimated.

[17] The measure c_k equals $B(1) = B(exp\{-i\lambda\})$, where $\lambda = 0$. Therefore c_k has to do with the zero frequency long-run characteristics of the process.

eventually dominates. For a random walk c_∞=1, i.e. the effect of a shock is permanent. In general, unit roots lead to a nonzero long-run response; however, the particular value of c_∞ depends on the specific parameterization of the process.

Cochrane (1988, 1991) states that measures based on the quantity c_∞ are the only measures of the presence of a unit root in finite samples. Thus, we can construct a trend-stationary series that is 'just like' a given difference-stationary series in every respect except c_∞. Or, by changing the periodogram ordinate of the first differences at frequency zero without changing the other ordinates we can construct a non-stationary series with a zero frequency unit root from a stationary series and vice versa.

As Diebold and Nerlove (1988) note, it is possible that the *infinite* horizon corresponding to $B(1)$ is not likely to correspond to the *economic* horizon of interest. Diebold and Rudebusch (1989) not only use c_∞ to measure persistence, but the sequence $C=\{1, c_1, c_2, \ldots, c_\infty\}$. They use C to study persistence because then it is possible to answer the question: 'How does a shock today affect the level of output in the short, medium, long and very long run?'. In the long-memory models that Diebold and Rudebusch consider, the cumulative impulse response can differ substantially from c_∞, even at quite long horizons. Note that c_∞ is not measurable for finite series.

2.3. FRACTIONALLY INTEGRATED *ARMA* MODELS

The issue of modeling persistence in economic series is addressed in various ways in the literature. Campbell and Mankiw (1987) used an $ARIMA(2,1,2)$ representation to model persistence in real output, whereas Watson (1986) and Clark (1987) used unobserved components (UC) models, which are (nonlinearly) restricted *ARIMA* models (see e.g. Harvey 1989). Although it is true that *ARIMA* models are designed to represent the short- rather than the long-run dynamics of an economic series, one can have doubts about a restricted *ARIMA* model as being favorable to describe persistence as well. In this thesis we investigate the merits of a generalized instead of a restricted *ARIMA* representation of long-run dynamics, i.e. we will allow for fractional integratedness of economic series. This generalization has desirable properties, to which we will return after having formalized matters.

Recall equation (2.1.1), the standard *ARIMA* representation of an economic process y that is integrated of the order d:

$$(1) \qquad \phi(L)\ (1-L)^d y_t \ = \ \theta(L)\varepsilon_t$$

where $\phi(z)$ and $\theta(z)$ are as defined before, both polynomials having zeros that are all outside the unit circle. If we assume that the parameter d can take any real value instead of restricting its value to be an integer, we have the case of <u>fractional integration</u>: the zero frequency unit root has fractional multiplicity. Geweke and Porter-Hudak (1983) called these generalized models *AutoRegressive Fractionally Integrated Moving Average*, denoted $ARFIMA(p,d,q)$.

2.3.1 *ARFIMA*(0,*d*,0) PROCESSES

Suppose for the moment that the *ARMA* part of the process can be neglected, i.e. that we deal with an $ARFIMA(0,d,0)$ process or fractional noise. Geweke and Porter-Hudak (1983) called these processes simple integrated series. To describe the properties of such a process we first have to get the binomial expansion of the operator $(1-z)^d$:

$$\begin{aligned}(2) \qquad (1-z)^d \ &= \ \sum_{k=0}^{\infty} \binom{d}{k} (-z)^k \\ &= \ \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)} \frac{z^k}{k!} \\ &= \ 1 \ - \ dz \ - \ \frac{d(1-d)}{2!}z^2 \ - \ \frac{d(1-d)(2-d)}{3!}z^3 \ + \ \ldots\end{aligned}$$

where $\Gamma(d)$ denotes the gamma, or generalized factorial, function and is defined as follows

$$\Gamma(d) = \begin{cases} \int_0^\infty s^{d-1} e^{-s} ds & \text{if } d>0 \\ \infty & \text{if } d=0 \end{cases}$$

For $d<0$, $\Gamma(d)$ is defined in terms of the above expressions and the recurrence formula $d\Gamma(d)=\Gamma(d+1)$ which holds for all values of d. Note that the recurrence formula and $\Gamma(0)=\infty$ imply that $\Gamma(d)$ has poles at the nonpositive integers. Thus, $(1-z)^d$ provides an infinite-order polynomial function with slowly and monotonically declining weights.

When $d>-\frac{1}{2}$ the process y is invertible and has an infinite autoregressive representation:

$$\sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)} \frac{L^k}{k!} y_t = \varepsilon_t$$

Hosking (1981) obtains this representation from the binomial expansion of $(1-z)^d$. From Stirling's formula it follows that $\Gamma(k-d)/(\Gamma(-d)\,k!)$ can be approximated by $k^{-d-1}/\Gamma(-d)$ for $k\to\infty$.

When $d<\frac{1}{2}$, the process y is stationary and has an infinite moving-average representation. Hosking (1981) obtains the moving-average representation by inversion of (3.1) and by making use of the binomial expansion of $(1-z)^{-d}$:

$$\begin{aligned} y_t &= (1-L)^{-d}\,\varepsilon_t \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(k+d)}{\Gamma(d)} \frac{L^k}{k!} \varepsilon_t \end{aligned}$$

From Stirling's formula it follows that $\Gamma(k+d)/(\Gamma(d)\,k!)$ can be approximated by $k^{d-1}/\Gamma(d)$ for $k\to\infty$. The factor k^{d-1} is summable or not depending on the value taken by d. The fractional process is asymptotically stationary if and only if $d<\frac{1}{2}$; the (asymptotically) stationary fractional process is of short-memory if $d\leq 0$ and of long-memory if $0<d<\frac{1}{2}$.

In short, an ARFIMA$(0,d,0)$ process is stationary and invertible if and only if d is less than $\frac{1}{2}$ in absolute value.

It is obvious that the autoregressive polynomial of infinite order has to be approximated by a polynomial of finite order, given that a realization of a process y consists of a finite amount of observations. We can choose to simply truncate the infinite autoregressive polynomial and ignore the

remaining part. Alternatively, following Levinson (1947) and Hosking (1981), we can derive the autoregressive coefficients of a finite order autoregressive polynomial approximation to the polynomial of infinite order.

Let us denote the coefficient of the infinite order autoregressive polynomial function (2) corresponding to z^k by $a_{k,\infty}$. Thus

$$a_{k,\infty} = -\frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}$$

Notice the minus sign that follows from the definition of an autoregressive polynomial function in this thesis. Approximating the infinite order polynomial function by a finite order p polynomial function gives equation (2.1.7) which is represented here:

$$(3) \qquad y_t = a_{1,p}y_{t-1} + \ldots + a_{p,p}y_{t-p}$$

where the second index p indicates that it is an autoregressive coefficient of a finite order p polynomial function. Hosking (1981) derived the analytical representation of the finite order linear regression coefficients by solving the Yule-Walker equations for the autoregressive model of finite order p with the Levinson-Durbin-Whittle recursion formulas (see appendix B) for the autoregressive coefficients of (3) for $p=1,\ldots,T-1$:

$$(4) \qquad a_{k,p} = a_{k,\infty}\cdot\frac{p(p-1)\ldots(p-k+1)}{(p-d)(p-d-1)\ldots(p-d-k+1)} \qquad k=1,2,\ldots,p-1$$

$$(5) \qquad a_{p,p} = \frac{d}{p-d}$$

where

$$(6) \qquad a_{k,\infty} = -\left[\frac{-d(1-d)\ldots(k-1-d)}{k!}\right] \qquad k=1,2,\ldots$$

The discrepancy between the infinite and finite order autoregressive coefficients which are necessary for applying the infinite polynomial operator to a finite realization of the process y is clearly seen from figure 2.5. Given that we approximate the infinite polynomial by a polynomial of the order $p=100$ we see that the coefficients $a_{k,100}$ seem to account for the truncation of the polynomial at high k. This reveals the short-coming of simply truncating a polynomial of infinite order.[18]

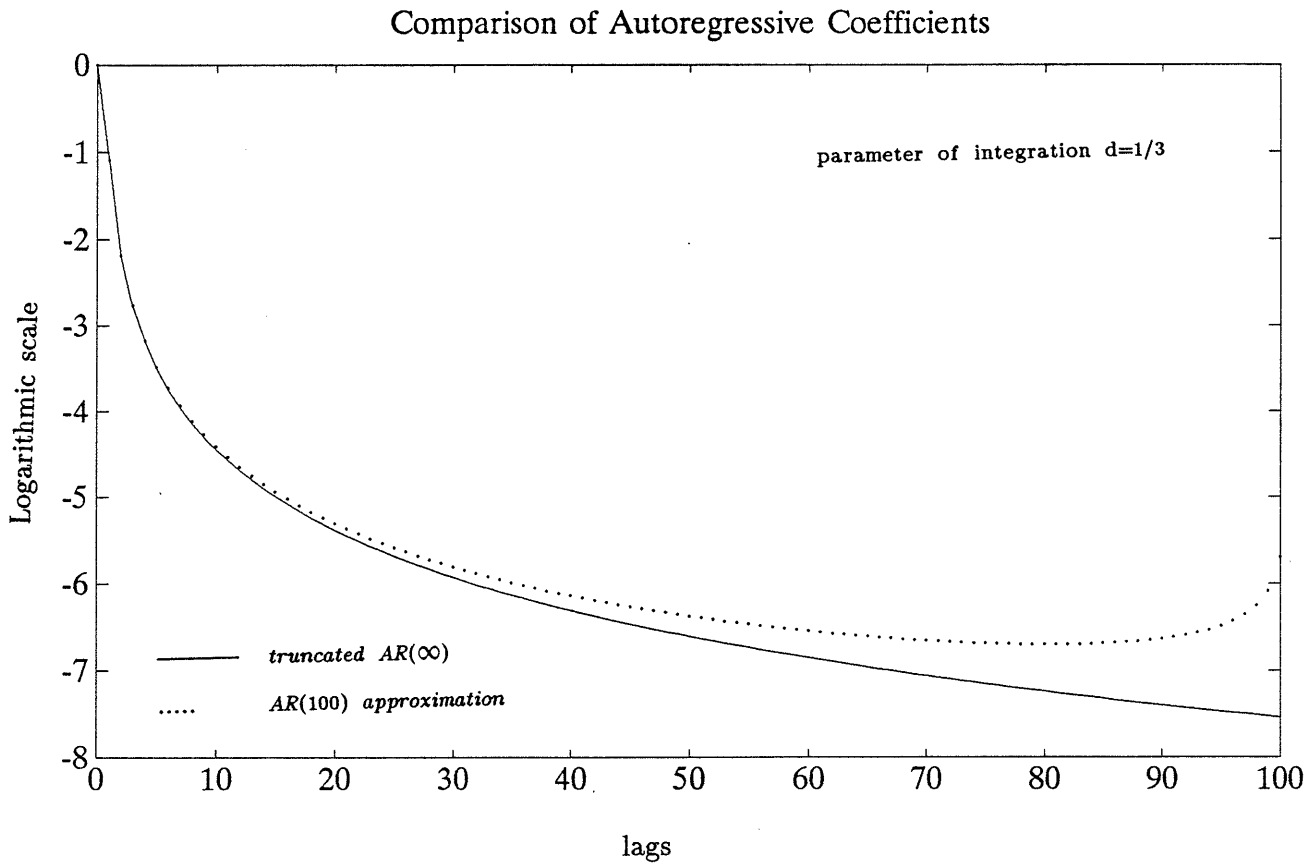
---

[18] From (4) and (6) the following recursions can be derived:

(6′) $a_{k,\infty}=a_{k-1,\infty}\cdot(k-1-d)/k$; $a_{1,\infty}=d$; $k\geq 2$; $a_{0,\infty}=1$;

(4′) $a_{k,p}=a_{k-1,p}\cdot(p-k+1).(k-1-d)/((p-d-k+1).k)$; $a_{1,p}=d.p/(p-d)$; $k\geq 2$; $a_{0,p}=1$.

Figure 2.5


Comparison of Autoregressive Coefficients
parameter of integration d=1/3
truncated AR(∞)
AR(100) approximation
Logarithmic scale
lags
0
-1
-2
-3
-4
-5
-6
-7
-8
0
10
20
30
40
50
60
70
80
90
100

Instead of applying the finite version of the infinite lag operator or simply truncating the infinite lag operator in the time domain, Geweke and Porter-Hudak (1983) propose to premultiply y_t by the infinite operator in the frequency domain after which the result is transformed back to the time domain. However, Sowell (1987, 1992b) proves that this procedure suffers from the time domain truncation problem as well, as was not expected by Geweke and Porter-Hudak.

2.3.1.1 The autocovariance function

To derive the autocovariances we make use of the moving-average representation of the $ARFIMA(0,d,0)$ process. It can be shown (Gourieroux and Monfort 1990) that the k-th order autocovariance is

$$\gamma_y(k) = \sigma_\varepsilon^2 \cdot \frac{\Gamma(k+d)}{\Gamma(k+1-d)} \cdot \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)},$$

having the same sign as the order of integration d for $k \geq 1$.[19] Making use of Stirling's formula we can write for large k

$$\gamma_y(k) \sim \sigma_\varepsilon^2 . k^{2d-1} \cdot \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}$$

It shows that the convergence to zero is hyperbolic and thus slower than the exponential decay of the autocovariance function of $ARMA$ processes. Indeed, the k-th order autocorrelation for $ARFIMA$ processes is

$$\rho_y(k) = \frac{\gamma_y(k)}{\gamma_y(0)} \sim k^{2d-1}$$

where $d < \frac{1}{2}$ and $k \to \infty$.

The long-memory property of fractionally integrated processes can be illustrated by comparing the autocorrelations of two stationary series where we allow one to be fractionally integrated[20]. The two are an $ARIMA(1,0,0)$ process with coefficient $\alpha_1 = 0.5$ and an $ARFIMA(0,\frac{1}{3},0)$ process. The example is taken from Diebold and Rudebusch (1989).[21] The autocorrelations $\rho_y(k) = \alpha_1^k$ for

---

[19] A recursive formula to compute the autocovariances is:

$$\gamma_y(k) = \gamma_y(k-1).(k-1+d)/(k-d), \quad k \geq 1;$$

given $\gamma_y(0) = \sigma_\varepsilon^2 . \Gamma(1-2d)/\{\Gamma(1-d)\}^2$.

[20] Notice that stationarity does not exclude series to be zero frequency integrated of some order d, where d is a fraction. However, if d is restricted to be an integer then integrated processes are non-stationary and vice versa.

[21] In their Table 1 Diebold and Rudebusch (1989) mistakenly associate the given

the *ARFIMA*(1,0,0) process can be derived from the a.c.g.f.; for the *ARFIMA*$(0,\frac{1}{3},0)$ process it can be derived that the autocorrelation function is given by $\rho_y(k)=\Gamma(1-d)\Gamma(k+d)/\{\Gamma(d)\Gamma(k+1-d)\}$.

Table 3.1

Theoretical Autocorrelations of an ARIMA and an ARFIMA Process

lag k	1	2	3	4	5	10	25	50	100
$(1-0.5L)y_t=\varepsilon_t$	0.50	0.25	0.13	0.06	0.03	0.00	0.00	0.00	0.00
$(1-L)^{1/3}y_t=\varepsilon_t$	0.50	0.40	0.35	0.32	0.30	0.24	0.18	0.14	0.11

From Table 3.1 we see that the decay in correlation between observations k periods apart is much slower for the fractionally integrated process. It is therefore that those processes are usually called 'long-memory' processes, whereas the usual *ARIMA*$(p,0,q)$ processes are labeled 'short-memory'. Notice that the first order autocorrelations for the two processes are the same.

2.3.1.2 The pseudo-spectrum

In Table 3.1 a comparison between two processes has been made in the time domain. In the frequency domain we can make a comparison between the two processes from Table 3.1 as well. However, a problem arises with the *ARFIMA*$(0,\frac{1}{3},0)$ process. If $y\sim I(d)$, d nonzero, the process y does not strictly possess a spectrum but we can define a pseudo-spectrum as follows:

$$(7) \qquad f_y(\lambda) = (2\pi)^{-1}.g_y(exp\{-i\lambda\}).\,|1-exp\{-i\lambda\}|^{-2d}$$

This follows from the fact that differencing a series once in the time domain is equivalent to multiplying its spectrum by $|1-exp\{-i\lambda\}|^2$ in the frequency domain.

As $g_y(exp\{-i\lambda\})=\sigma_\varepsilon^2$ for *ARFIMA*$(0,d,0)$ processeses (see (2.1.4)) the pseudo-spectrum reads

---

correlation function with an ARFIMA(0,0.3,0) process.

$$f_y(\lambda) = (\sigma_\varepsilon^2/2\pi).\,|1-exp\{-i\lambda\}|^{-2d}$$

which equals

$$f_y(\lambda) = (\sigma_\varepsilon^2/2\pi).\,(2sin(\lambda/2))^{-2d}.$$

If λ tends to zero we have

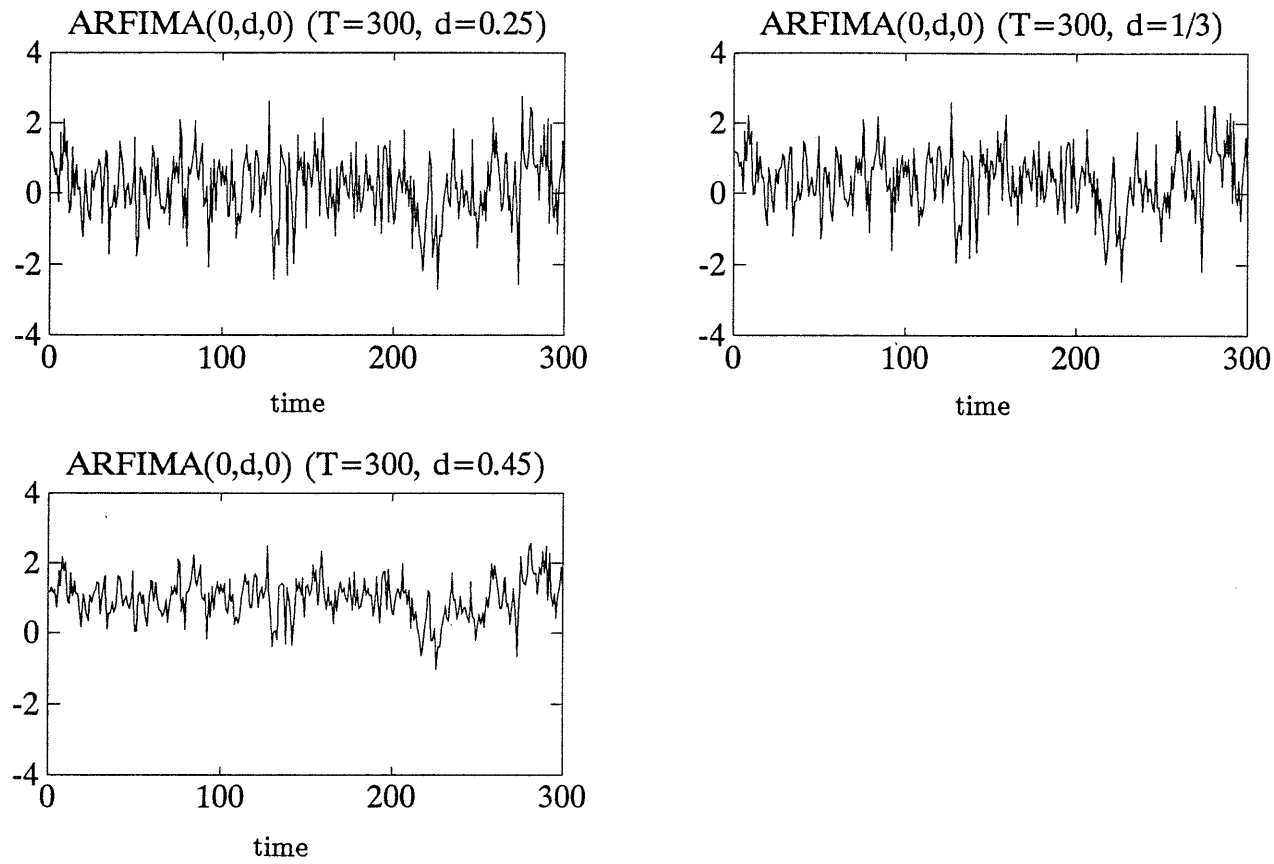
$$(8) \qquad f_y(\lambda) = (\sigma_\varepsilon^2/2\pi).\,|1-exp\{-i\lambda\})|^{-2d} \sim (\sigma_\varepsilon^2/2\pi).\lambda^{-2d}$$

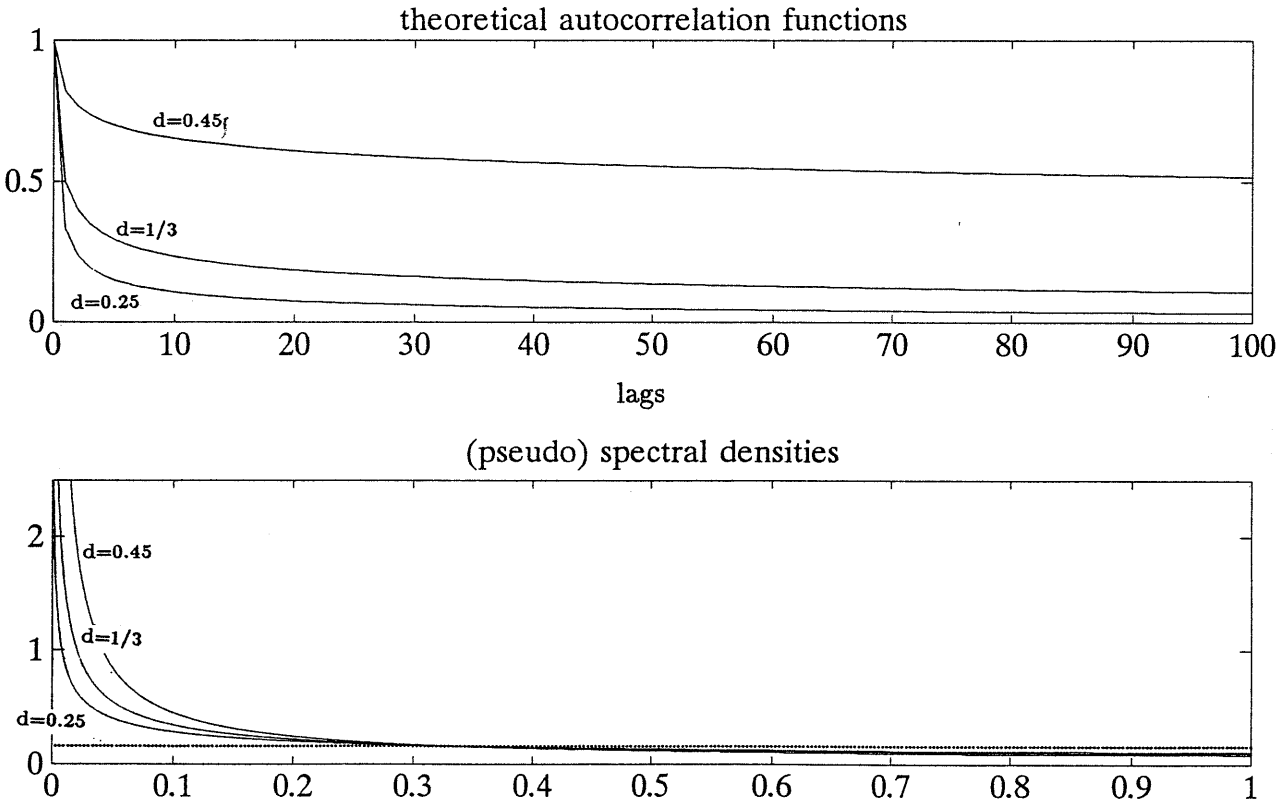
Thus, the behavior of $ARFIMA(0,d,0)$ processes is proportional to λ^{-2d} for λ near zero frequency, whereas for nonfractional *ARIMA* processes that are multiplicity one integrated it is proportional to λ^{-2} (substitute $d=1$ in (8)). As such, *ARFIMA* processes can be seen as a generalization of the zero frequency unit-root case (i.e. the unit root $+1$).

The pseudo-spectrum is also useful to understand the difference between short memory and long memory (or non-stationary) processes. For the latter processes the pseudo-spectrum tends to infinity if λ tends to zero. The velocity depends on the exponent d: the closer d to $\frac{1}{2}$ the higher the velocity. This can be seen from the figures 2.6 and 2.7. The long-memory property of *ARFIMA* processes is very clearly revealed by the autocorrelation function. Although in Table 3.1 the autocorrelation function revealed this typical behavior already for $d=\frac{1}{3}$, this property is even more clear from figure 2.6 for $d=0.45$. Furthermore, it can be seen from the s.g.f. that the restriction of the long-memory property is on fewer frequencies than the restriction of other types of behavior on frequency points. Application of a fractional filter to a realization of a process therefore has a less disturbing effect on the behavior of that process — in terms of number of affected frequencies — than e.g. the (stationary) autoregressive filter.

The short-memory property of *ARFIMA* processes, i.e. $d\leq 0$, turns up from the time domain as well as from the frequency domain. From figure 2.7 we see that a relatively small part of the variance of such processes is explained by the low-frequency behavior and a relatively large part by the high-frequency behavior. The autocorrelation function of short-memory fractional processes turns out to compare well with that of an $MA(1)$ process with negative parameter (compare e.g. $d=-0.25$ with $\theta=-0.1$). The realizations of the figures 2.6 and 2.7 are obtained by an indirect Cholesky decomposition of the varianc-covariance matrix procedure (see Section 2.4.1).

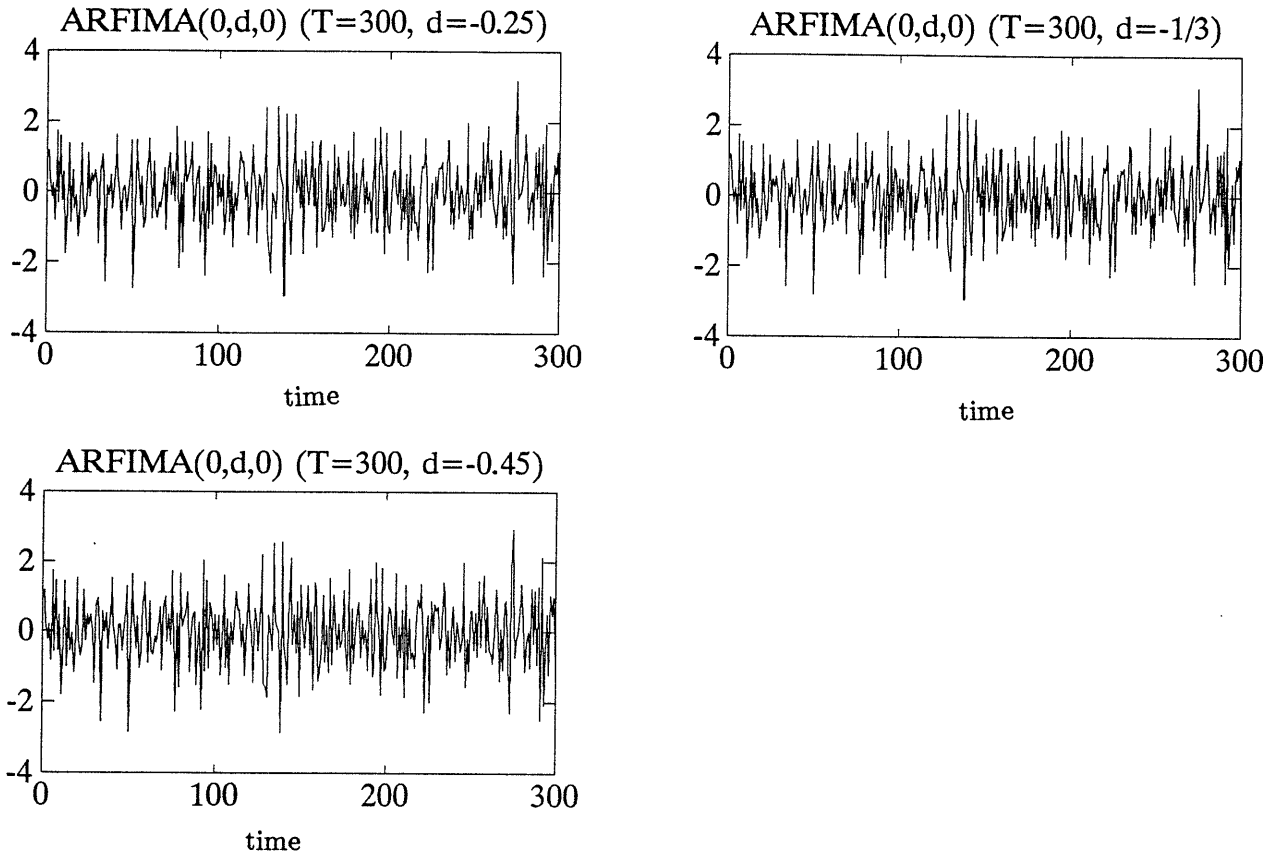
Figure 2.6

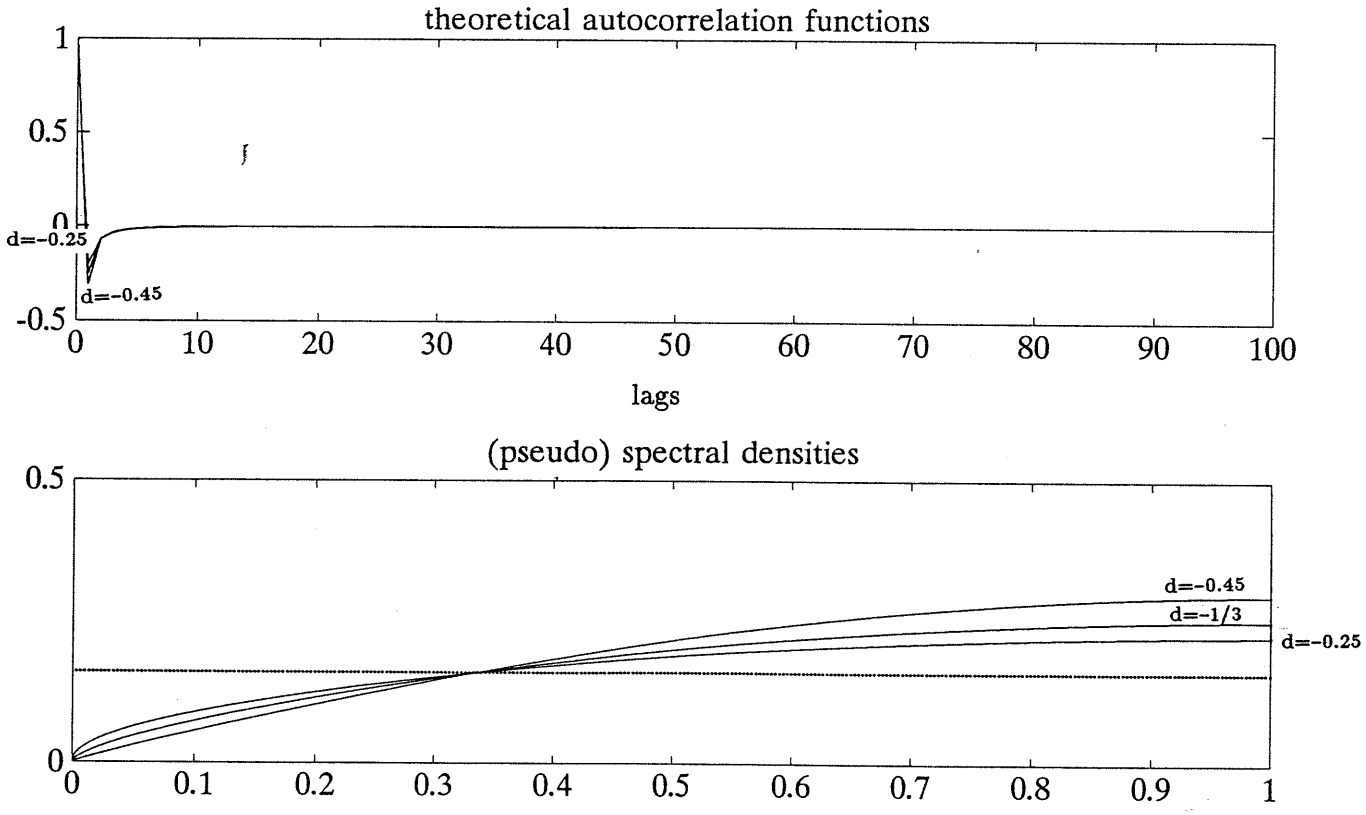

ARFIMA(0,d,0) (T=300, d=0.25)
ARFIMA(0,d,0) (T=300, d=1/3)
ARFIMA(0,d,0) (T=300, d=0.45)
time
time
time


theoretical autocorrelation functions
d=0.45
d=1/3
d=0.25
lags

(pseudo) spectral densities
d=0.45
d=1/3
d=0.25
fractions of pi

Figure 2.7


ARFIMA(0,d,0) (T=300, d=-0.25)
ARFIMA(0,d,0) (T=300, d=-1/3)
ARFIMA(0,d,0) (T=300, d=-0.45)
time
time
time


theoretical autocorrelation functions
d=-0.25
d=-0.45
lags
(pseudo) spectral densities
d=-0.45
d=-1/3
d=-0.25
fractions of pi

2.3.1.3 Prediction theory

Making a prediction given a realization of a fractionally integrated process without any autoregressive and/or moving-average structure requires that we take account of the infiniteness of the autoregressive polynomial expression of the model. We choose to approximate the infinite autoregressive polynomial by the finite order polynomial expression (3) where the autoregressive coefficients are derived by Hosking (1981). Along the lines of section 2.1.3 we obtain predictions from the autoregressive model thus obtained.

2.3.2 *ARFIMA(p,d,q)* PROCESSES

$ARFIMA(0,d,0)$ processes as described in the previous section have a very distinctive auto-correlation structure and therefore can only be of limited value to model long memory data. However, if we extend these models to account for short-run dynamics as well, a very broad class of possible autocorrelation functions results.

The low-lag correlation structure of fractionally integrated processes is modeled by an *ARMA* part which we neglected in Section 2.3.1. The parameters ϕ and θ can be chosen to describe this low-lag correlation structure, whereas the parameter d accounts for the correlation between observations that are very distant from each other (see equation (1)).

To ensure stationarity the order of integration d should be less than $\frac{1}{2}$ and the roots of the equation $|\phi(z)| = 0$ should all lie outside the complex unit circle. If $d > -\frac{1}{2}$ and the root of the equation $|\theta(z)| = 0$ lies outside the unit circle the $ARFIMA(p,d,q)$ process is invertible.

The long-term behavior of an $ARFIMA(p,d,q)$ process may be expected to be similar to that of an $ARFIMA(0,d,0)$ process with the same value d, if for very distant observations the effects of the ϕ and θ parameters will be negligible. Hosking (1981) shows that this is indeed the case:

Given that y is stationary and invertible, the spectral density $f_y(\lambda)$ *and the autocorrelation function* $\rho_y(k)$ *have the respective properties*:

$$lim_{\lambda \to 0} f_y(\lambda) = \lambda^{-2d}$$
$$lim_{k \to \infty} \rho_y(k) = k^{2d-1}$$

This theorem expresses that the effect of the parameter d decays hyperbolically as the lag increases, whereas the effect of the *ARMA* parameters decays exponentially.

2.3.2.1 The Autocovariance Function

Sowell (1987, 1992b) shows that for *ARFIMA*$(0,d,q)$ processes the autocovariances have the form

$$\gamma_y(k) = \sigma_\varepsilon^2 . \sum_{l=-q}^{q} \Psi(l) . \frac{\Gamma(1-2d).\Gamma(d+k-l)}{\Gamma(d).\Gamma(1-d).\Gamma(1-d-k+l)}$$

where

$$\Psi(l) = \sum_{i=max\{0,l\}}^{min\{q,q-l\}} \theta_i \theta_{i-l}$$

For *ARFIMA*(p,d,q) processes the autocovariances look like

$$\gamma_y(k) = \sigma_\varepsilon^2 . \sum_{l=-q}^{q} \sum_{j=1}^{p} \Psi(l) . \zeta_j . C(d, p+l-k, \rho_j)$$

where $\Psi(l)$ is as defined before,

$$\zeta_j = \left[\rho_j . \prod_{i=1}^{p} (1-\rho_i \rho_j) . \prod_{m \neq j} (\rho_j - \rho_m) \right]^{-1}$$

and the factor $C(d, p+l-k, \rho_j)$ can be computed recursively from

$$C(d,h,\rho) = \frac{\Gamma(1-2d).\Gamma(d+h)}{\Gamma(d).\Gamma(1-d).\Gamma(1-d+h)} * \left[\rho^{2p} F(d+h,1;1-d+h;\rho) + F(d-h,1;1-d-h;\rho) - 1 \right]$$

where the hypergeometric function $F(a,b;c;x)$ is defined by

$$F(a,b;c;x) = \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c+n)\Gamma(n+1)} . x^n$$

for which the following recursion holds

$$F(a,1;c;\rho) = \frac{c-1}{\rho(a-1)} [F(a-1,1;c-1;\rho) - 1].$$

For each evaluation of the likelihood function only p different hypergeometric functions need to be evaluated. The ρ_j's result from the factorization of the autoregressive polynomial, i.e. $\phi(z) = \prod_{j=1}^{p} (1-\rho_j z)$.[22]

---

[22] For this factorisation to hold Sowell (1992b) assumes that the autoregressive polynomial roots are simple. By Monte Carlo simulation he shows that this is not a binding restriction at an empirical level. In addition Sowell assumes that $d<\frac{1}{2}$, the autoregressive and moving-average

2.3.2.2 The Pseudo-Spectrum

From equation (7) we derive the pseudo-spectrum for the general $ARFIMA(p,d,q)$ process

$$(9) \qquad f_y(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \cdot \left| \frac{\theta(exp\{i\lambda\})}{\phi(exp\{i\lambda\})} \right|^2 \cdot \left| 1 - exp\{-i\lambda\} \right|^{-2d}$$

If λ tends to zero we have

$$f_y(\lambda) \sim g_y(1) \cdot \left| 1 - exp\{-i\lambda\} \right|^{-2d} \sim g_y(1) \cdot \lambda^{-2d}$$

which means that near zero frequency behavior of $ARFIMA(p,d,q,)$ processes is proportional to λ^{-2d}.

From (9) it follows that if we look at the lower angular frequencies of $f_y(\lambda)$ we are able to say something about the order of integration d. Because the shape of the pseudo-spectrum uniquely determines the relationship $f_y(\lambda)/\left| 1 - exp\{-i\lambda\} \right|^{-2d}$ we can determine the second order characteristics of the process $(1-L)^d Y_t$ (for large t) if we know the pseudo-spectrum (e.g. Geweke and Porter-Hudak 1983).

From the frequency domain it is clarifying to see that neither not differencing, nor differencing a fractionally integrated series once —— i.e. applying integer differencing —— are good ways to proceed. Furthermore it reveals that fractional differentiation has attractive features that seem to be tailor-made for the 'typical spectral shape' problem, mentioned by Granger (1966).

2.3.2.3 Prediction Theory

If we deal with a fractionally integrated process (1) with autoregressive parameters we have to make use of the concept of convolution in order to derive predictions from this model, i.e. convolution of the finite order autoregressive coefficients (4) and (5) corresponding to the fractional filter (2) and those of the autoregressive polynomial $\phi(z)$. Specifically, given the finite $p1$ order autoregressive polynomial $a(z)$ and the finite $p2$ order polynomial $\phi(z)$ defined as

---

polynomials are of an order less than or equal to p and q respectively, all polynomial roots are outside the unit circle and ε_t is normally distributed with zero mean and variance σ_ε^2.

$$a(z) = 1 - a_{1,p1}z - a_{2,p1}z^2 - \ldots - a_{p1,p1}z^{p1}$$

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_{p2}z^{p2}$$

the k-th coefficient of the convolution $\xi(z) = \phi(z)a(z)$ is given by

$$\xi_k = \sum_{j=0}^{k} \phi_j a_{k-j,p1}$$

where $k = 0,1,\ldots,p1+p2$ and $a_{0,p1} = \phi_0 = 1$.

Geweke and Porter-Hudak (1983) use a truncated version of the infinite order autoregressive coefficients $a_{k,\infty}$ as given in (6) to approximate the fractional filter (2). Given the finite order autoregressive polynomial function $\xi(z)$ we can make the predictions as usual. After convolution, predictions from an $ARFIMA(p,d,q)$ process can be obtained by application of the prediction theory with respect to $ARMA(p,q)$ processes as described in Section 2.1.3.

2.3.2.4 Testing Unit Roots Against Stationary Alternatives

In section 2.2.2 we described the (augmented) Dickey-Fuller test for zero frequency unit roots. This test amounts to the comparison of one point, i.e. $\rho = 1$, against a whole range of stationary alternatives $|\rho| < 1$. We started by assuming i.i.d. residuals ε_t and allowed for stationary $ARMA(p,q)$ residuals u_t later on. In both cases the residuals are zero frequency integrated of the order zero, i.e. $\varepsilon_t \sim I(0)$ and $u_t \sim I(0)$.

In this section we assume u_t to be general fractional noise (Geweke and Porter-Hudak 1983), i.e. $u_t \sim I(\delta)$ where the parameter of zero frequency integration δ can be any real number less than one half in absolute value. The null hypothesis is:

$$H_0: y_t = y_{t-1} + u_t$$

where $\phi(L)(1-L)^{\delta}u_t = \theta(L)\varepsilon_t$ which implies that under this hypothesis

$$(10) \qquad \phi(L)\,(1-L)^{d}y_t = \theta(L)\varepsilon_t$$

where $d = 1-\delta$, $d \in (-1/2, 3/2)$ and ε_t i.i.d. again.

Dickey-Fuller unit-root testing now amounts to regressing y_t on y_{t-1}, the estimator $\phi(1)$ in (1) converging to its asymptotic distribution at a rate

crucially depending on $\tilde{d}$ (Sowell 1990). For $\tilde{d}\in[0,\frac{1}{2})$ the convergence result is

$$(\hat{\phi}_1-1)=\mathcal{O}_p(T^{-1})$$

whereas if $\tilde{d}\in(-\frac{1}{2},0)$ the convergence is faster or slower than $\mathcal{O}(T^{-1/2})$ as $\tilde{d}$ is greater or smaller than $-\frac{1}{4}$, respectively. In this case

$$(\hat{\phi}_1-1)=\mathcal{O}_p(T^{-2\tilde{d}-1})$$

Sowell (1990) proofs that the (augmented) Dickey-Fuller tests are invalid if one allows for residual (general) fractional noise, except when $\delta=0$ or $d=1$. Diebold and Rudebusch (1991) conclude that the power of the Dickey-Fuller tests (Fuller 1976) against fractionally integrated alternatives is quite low. Diebold and Rudebusch (1991) illustrate this with a Monte Carlo experiment. Sowell (1990) as well conjectures that commonly applied Dickey-Fuller unit root tests under fractionally integrated alternatives have quite low power. Although asymptotically the integer unit root distribution differs severely from its fractional equivalent, finite sample distributions are rather similar.[23]

As an alternative test for zero frequency unit roots it is possible to estimate the parameter of integration d in equation (10). As the parameter values under the null hypothesis reflect non-stationarity, the model for the first differences Δy_t has to be estimated instead and an estimate of d can be obtained as $1-\hat{\delta}$, where $\hat{\delta}$ is the estimated δ. Notice that these tests are valid in case of fractionally integrated processes as well as in case of *ARMA* processes, as opposed to the (augmented) Dickey-Fuller tests.

2.3.2.5 Measuring shock persistence

Analogue to equation (2.2.5) we define the cumulative impulse response function in case of fractional integration as follows:

$$\Delta y_t \;=\; B(L)\;\varepsilon_t \;=\; (1+b_1L+b_2L^2+\dots)\;\varepsilon_t$$

where now $B(L)=\phi(L)^{-1}(1-L)^{-\delta}\theta(L)$, where $\delta=d-1$. Under the assumption $\frac{1}{2}<d<1\frac{1}{2}$ there holds $-\frac{1}{2}<\delta<\frac{1}{2}$, i.e. the first differences follow a fractionally

---

[23] Sowell (1990) shows that the asymptotic fractional unit root distribution may be severely misleading in all but very large samples. This is because the distribution of ϕ_1 in the $AR(1)$ case depends on two underlying random variables, the convergence of one of which to its asymptotic distribution is very slow for plausible d values of fractional integration.

integrated stationary and invertible *ARMA* process, and therefore the coefficients b_j are square summable.

2.4 SIMULATION EXPERIMENTS

In this section we first describe two procedures to generate a realization from a theoretical stochastic process. Secondly, using one of these procedures we perform simulation experiments in order to be able to identify the quality of variance- and autocovariances-estimates for different kinds of models. Variances and covariances play a crucial role in model identification.

2.4.1 *Generating a realization from some process*

If we want to generate a realization $y_1, y_2, \ldots, y_T$ of a stochastic process y the general idea is to start from a normally distributed white noise sequence which under some transformation gives us a realization of y. This transformation can be carried out in two ways.

Given that $y \sim ARMA(p,q)$ an obvious and easy way to procede is to generate a white noise sequence $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T$ and just combine these values according to the formulated model. Note that for this procedure we need $p+q$ start-up values in order to generate a realization of length T. For T not too small the influence of these start-up values is negligible because of the short-memory property of *ARMA* processes.

Another procedure is presented in McLeod and Hipel (1978, p. 497). Suppose we want to generate a T dimensional realization from a zero mean stochastic process y that is characterized by the theoretical variance-autocovariance matrix Q, i.e.

$$Q = \begin{bmatrix} \gamma_y(0) & \cdots & \gamma_y(-p+1) \\ \vdots & & \vdots \\ \gamma_y(p-1) & \cdots & \gamma_y(0) \end{bmatrix}$$

Premultiplication of the vector of white noise observations $\underline{\varepsilon}$ by a transformation matrix, say M, gives us a realization of y if an only if

$$Q = MM'$$

which is called a direct Cholesky decomposition of the variance-autocovariance matrix Q of the process y. The matrix M is lower triangular. In general, if $y \sim ARMA(p,q)$ it is possible to generate a vector of realizations $\underline{y}$ of a process y with mean vector $\underline{\mu}$ and autocovariance-matrix Q as follows

$$\underline{y} = \underline{\mu} + M\underline{\varepsilon}$$

If we procede in this way we do not need start-up values to generate a realization of y.

For *ARMA* processes both procedures can be pursued. However, if we want to generate a realization of $y \sim ARFIMA(p,d,q)$ only the second procedure is relevant: the first procedure depends on the influence of start-up values which only dies out very slowly because of the long-memory property of fractionally integrated processes.

The transformation matrix M is found by a direct Cholesky decomposition of the variance-autocovariance matrix Q. Alternatively, the matrix M can be computed from an indirect Cholesky decomposition of Q as suggested by Jonas (1981). Whittle (1983, §3.4 and §7.1) shows the correspondence between the autocovariance matrix Q and the finite order autoregressive coefficients $a_{k,p}$

$$(1) \qquad Q^{-1} = A\,\Sigma^{-2}A'$$

where

$$(2) \qquad A = \Sigma\,Q^{-1/2} = \begin{bmatrix} 1 & -a_{1,1} & -a_{2,2} & \cdots & -a_{p,p} \\ 0 & 1 & -a_{1,2} & \cdots & -a_{p-1,p} \\ 0 & 0 & 1 & \cdots & -a_{p-2,p} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & -a_{1,p} \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$(3) \qquad \Sigma^2 = diag(\sigma_0^2, \sigma_1^2, \ldots, \sigma_p^2)$$

such that if $Q = MM'$ then $M^{-1} = \Sigma^{-1}A'$. The autoregressive coefficients of the matrix A and the elements of the diagonal matrix Σ are computed recursively from the Levinson-Durbin-Whittle algorithm (see appendix B).[24] Storage requirements are proportional to T^2 for bothdirect and indirect Choleky decompositions. The advantage of performing the Cholesky decomposition

---

[24] In Jonas (1981) and Jonas (1983) there are some notational inconsistencies in the formulas concerning the indirect Cholesky decomposition. The correct formulas are stated in appendix B. Corresponding to these formulas they give the formulas to obtain a realization of the process y recursively. These formulas should read

$$y = (I - M^{-1})y + \varepsilon$$

where $M^{-1} = \Sigma^{-1}A'$ and therefore there holds

$$\begin{cases} y_1 = \sigma_0 \varepsilon_0 \\ y_t = \sum_{k=1}^{t-1} a_{k,t-1} y_{t-k} + \sigma_{t-1}\varepsilon_{t-1} & t \geq 2 \end{cases}$$

indirect is in terms of computation time which is proportional to T^2, contrary to T^3 in case of a direct Cholesky decomposition. Notice that $p=T-1$ for a series of length T.

2.4.2 *A simulation study on variance and autocovariances*

The variance and autocovariance structure of a process have to be estimated in practice. The importance of the autocovariances in model identification on the one hand and their relationship with the spectral density function on the other hand makes it interesting to investigate the quality of the empirical estimates of the population variance and autocovariances. We use simulations to investigate these empirical properties given the theoretical mean and covariance-structure of the relevant process (e.g. McLeod and Hipel 1978; Granger and Joyeux 1980; Geweke and Porter-Hudak 1983). In general, we want to investigate the empirical properties of a certain statistic $\mathcal{S}$ using simulations.

Suppose we generate N independent simulations of a covariance-stationary time-series $y_1, y_2, \ldots, y_T$ and the statistic $\mathcal{S}=\mathcal{S}(y_1, y_2, \ldots, y_T)$ is calculated in each simulated series. The empirical mean $\overline{\mathcal{S}}$ is then the average over N simulations. If each realization $y_1, y_2, \ldots, y_T$ is independent of the other realizations so that the empirical averages of the statistic $\mathcal{S}$ are mutually independent, we can calculate the variance V_S of $\mathcal{S}$. By the central limit theorem, $\overline{\mathcal{S}}$ will be distributed very nearly normally with mean equal to $\mathcal{E}[\mathcal{S}]$ and with variance approximately equal to V_S/N. The standard deviation and confidence intervals for the expected value being estimated are readily obtained from V_S/N (McLeod and Hipel 1978, p. 501).

In the context of simulation of fractionally integrated processes the following point is of great importance. Suppose we generate series of length NT and then subdivide it into N series of T data-points. This procedure has bad consequences if adjacent data-points are heavily correlated and subsequently the N resulting estimations of $\mathcal{S}$. The resulting estimate for $\mathcal{E}[S]$ will be less precise (i.e. has larger variance) and the estimate of the variance will be underestimated, so that the correct standard deviations and confidence intervals for $\mathcal{E}[\mathcal{S}]$ will not be available. For a simulation study on fractional processes we therefore explicitly generate N realizations of dimension T.

The Geweke and Porter-Hudak* (1983) *experiment replicated

Let us start with a replication of the Geweke and Porter-Hudak (1983, Table 1) simulation study of a fractionally integrated process. In their experiment they generate $N=4000$ realizations of dimension $T=265$ from zero mean $ARFIMA(0,d,0)$ processes for $d=0.25$ and $d=0.45$. They apply the indirect Cholesky decomposition of the variance-autocovariance matrix procedure in order to generate a single realization, i.e. they use the Levinson-Durbin-Whittle algorithm. As the variance of the generated processes equals one we can read autocorrelations when we discuss the autocovariances. The computer code we used is written in Convex/Unix Fortran.

In figure 2.8 we present graphically the results of the replication of the Geweke and Porter-Hudak (1983) experiment in the upper two panels.[25] The corresponding tables can be found in appendix C. The mean estimated autocovariance function seems to coincide almost exactly with its population equivalent in both cases. The mean variance and the first 25 mean autocovariances are slight underestimates if $d=0.25$, whereas they are slight overestimates of the population variance and autocovariances if $d=0.45$ as can be observed as well from Table 1 (Geweke and Porter-Hudak 1983). The confidence interval is larger if $d=0.45$ than in the other case.

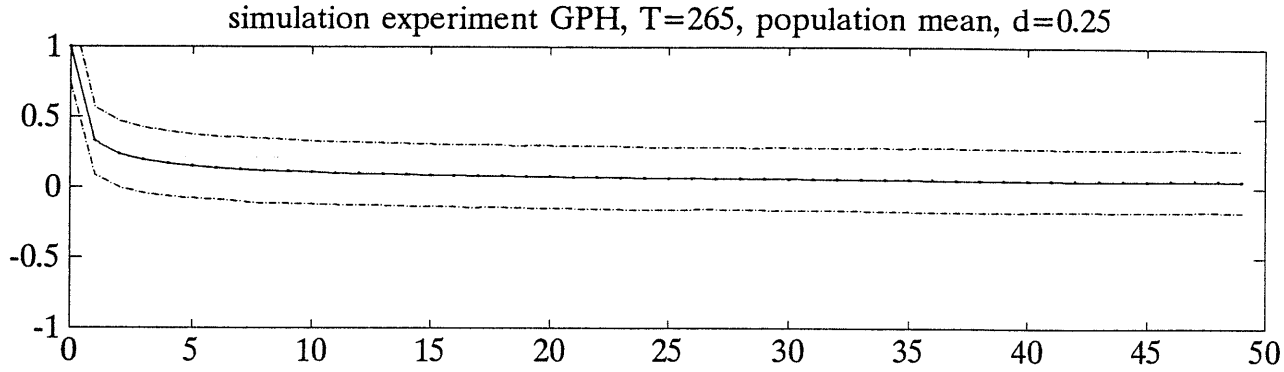
From personal communication with Geweke (1992) we discovered that the autocovariances in Geweke and Porter-Hudak (1983) are computed given the population mean of the simulated process. Although not mentioned in Geweke and Porter-Hudak (1983) this information appears to be crucial for the identification of the quality of the estimate of the autocovariance function. Of course, in practice such information is hardly ever known.

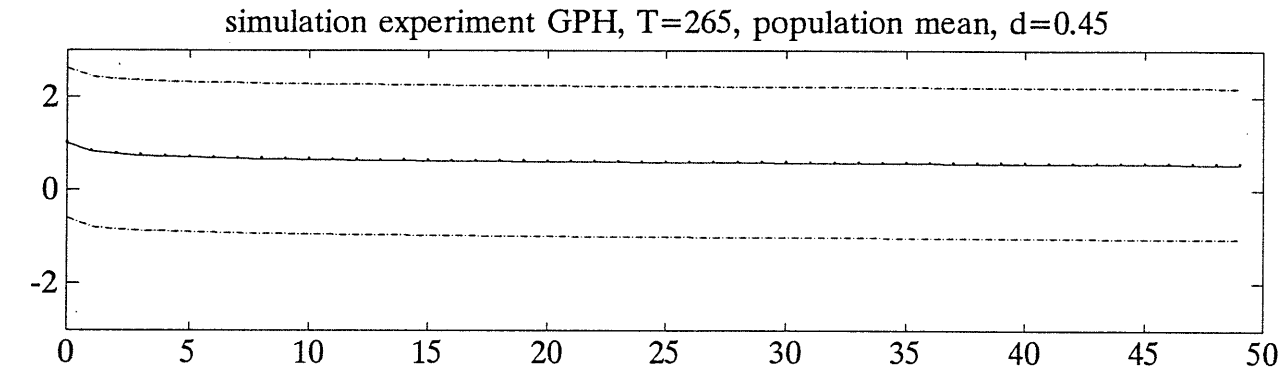
The consequence for the estimated autocovariance functions of <u>not</u> making use of the population mean of the process but estimating it in each replication can be observed from the lower two panels of figure 2.8. The table can again be found in appendix C. For $d=0.25$ processes the mean estimated autocovariance function appears to lie below the theoretical autocovariance function, although they do not differ significantly. The mean estimated

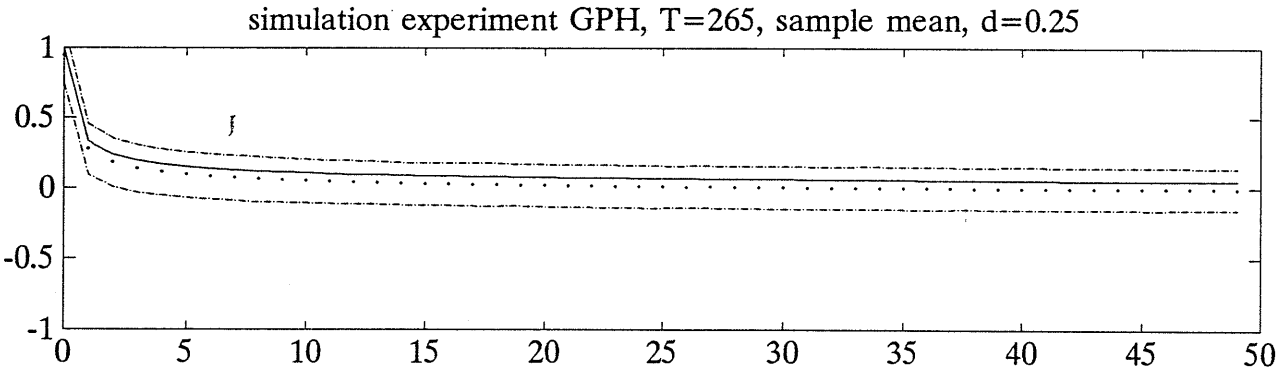
---

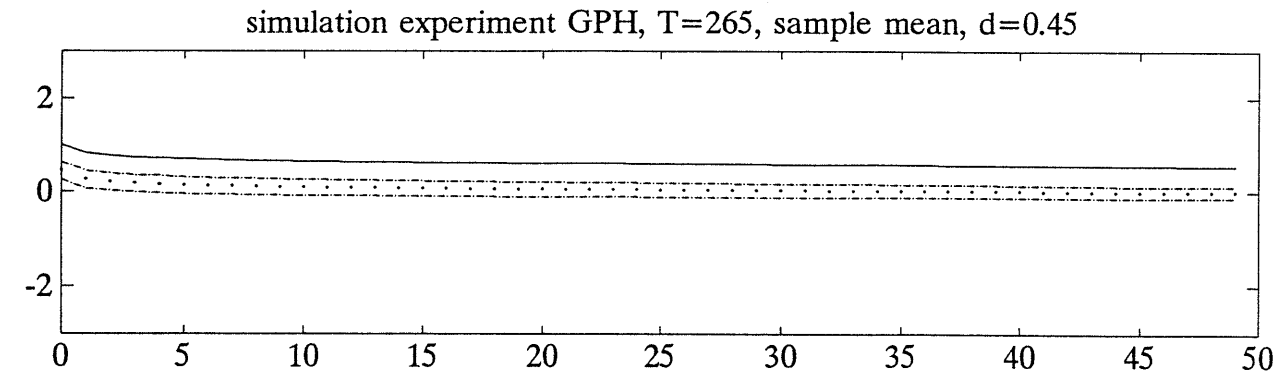
[25] The line represents the theoretical autocovariance function, the dots represent the mean estimated autocovariances from 0 up to the order 49 over N=4000 replications and the dot-dash lines represent the mean minus two times sigma and mean plus two times sigma confidence intervals.

Figure 2.8


simulation experiment GPH, T=265, population mean, d=0.25
1
0.5
0
-0.5
-1
0
5
10
15
20
25
30
35
40
45
50


simulation experiment GPH, T=265, population mean, d=0.45
2
0
-2
0
5
10
15
20
25
30
35
40
45
50


simulation experiment GPH, T=265, sample mean, d=0.25
1
0.5
0
-0.5
-1
0
5
10
15
20
25
30
35
40
45
50


simulation experiment GPH, T=265, sample mean, d=0.45
2
0
-2
0
5
10
15
20
25
30
35
40
45
50

autocovariance function for $d=0.45$ processes differs significantly from its theoretical equivalent. What has happened?

Underestimation of the sample autocovariance function

It is possible that the phenomenon described above only takes place if we deal with fractionally integrated processes. Therefore it seems useful to investigate the phenomenon and make a comparison with an alternative process at the same time. We have already noted that the autocovariance functions for the $ARFIMA(1,\frac{1}{3},0)$ process and the $ARIMA(1,0,0)$ process where $\alpha=0.5$ are comparable.

The set-up is as follows. We generate $N=1000$ replications of realizations of length $T=300$ for both processes and calculate the mean of the estimated variance and autocovariances over 1000 replications, first given that the population mean of the sample is known and second given that we have to estimate the mean of the sample each time the autocovariance function has to be computed. For both processes we employ the indirect Cholesky decomposition procedure.

The consequence of using the sample instead of the population mean of the process for the calculation of the autocovariance function is the same for both processes. The underestimation problem in case of the $ARIMA(1,0,0)$ process turns out to be less severe than in case of the $ARFIMA(1,\frac{1}{3},0)$ process. The reason of the underestimation is evident if we observe the histograms that are depicted in the figures 2.9, 2.10 and 2.11 on estimated means, variances and covariances.[26]

From figure 2.9 we see that the $N=1000$ estimated means of the generated samples from the $ARFIMA(0,1/3,0)$ process have a much larger spread around zero than the estimated means of the $ARIMA(1,0,0)$ samples, though for both processes about symmetrically distributed with respect to zero. The consequences are observed from the next two figures.

In figure 2.10 we have depicted the the $N=1000$ estimated variances of the $ARFIMA(0,1/3,0)$ process and the $ARIMA(1,0,0)$ process given the population mean in the first and second panel respectively, whereas in the lower two

---

[26] The text "d=1/3" in the title of a depicted histogram indicates that the process under consideration is $ARFIMA(0,1/3,0)$; if the text reads "alpha=1/2" the process is $ARIMA(1,0,0)$.

Figure 2.9

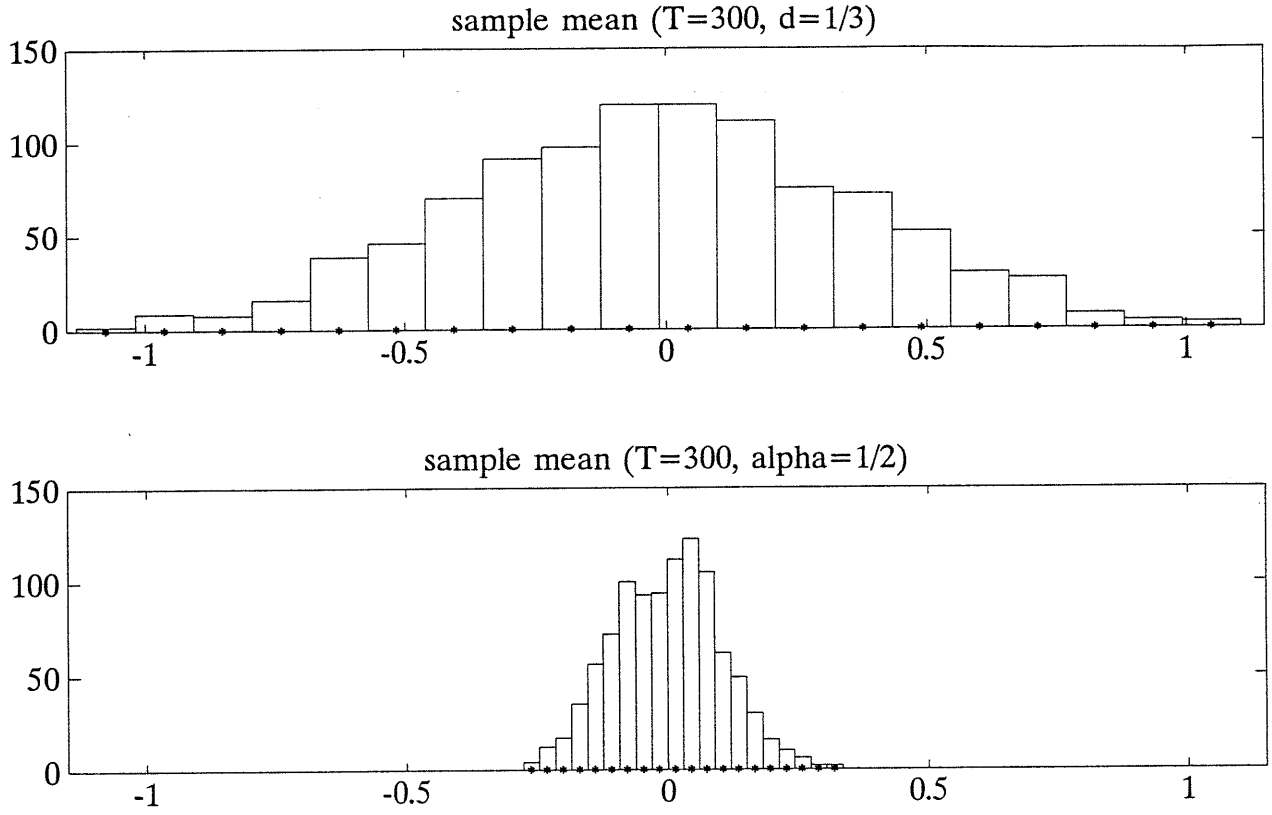
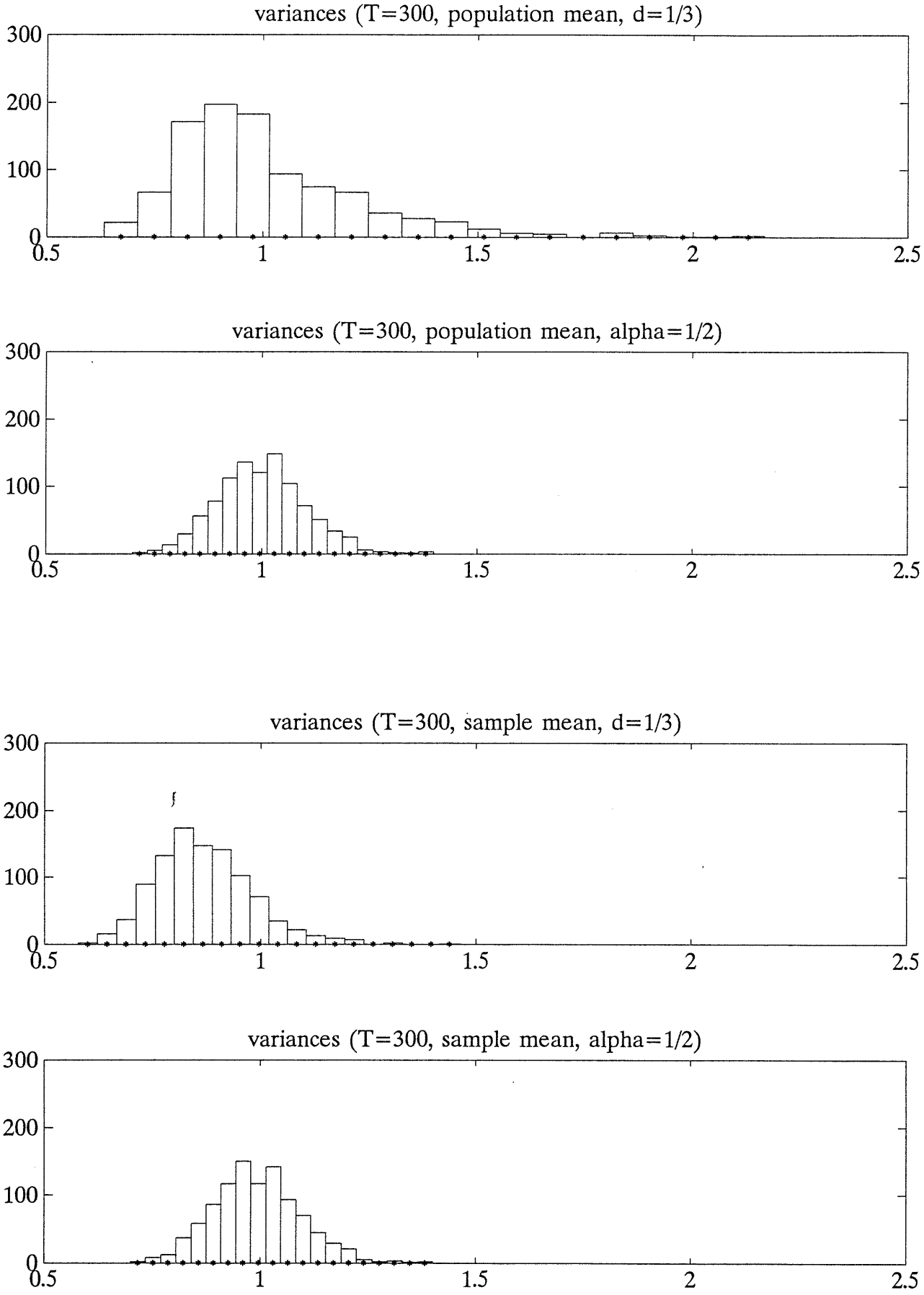

sample mean (T=300, d=1/3)
150
100
50
0
-1
-0.5
0
0.5
1
sample mean (T=300, alpha=1/2)
150
100
50
0
-1
-0.5
0
0.5
1

Figure 2.10


variances (T=300, population mean, d=1/3)
300
200
100
0
0.5
1
1.5
2
2.5
variances (T=300, population mean, alpha=1/2)
300
200
100
0
0.5
1
1.5
2
2.5
variances (T=300, sample mean, d=1/3)
300
200
100
0
0.5
1
1.5
2
2.5
variances (T=300, sample mean, alpha=1/2)
300
200
100
0
0.5
1
1.5
2
2.5

panels the histograms are given if we estimate the mean of each generated sample. We conclude first that histograms of the estimated variances, whether the population or sample means are used, are skewed to the left for the *ARFIMA* process and symmetrical around their population value for the *ARIMA* process. Secondly, with respect to the *ARFIMA* process, the estimated variances reveal a much larger spread if the population mean is used than when the mean is estimated. Consequently, the mean of the variances over 1000 replications coincides more with its theoretical value if the population mean information is available. Thirdly, the histograms for the *ARIMA* process have a comparable spread of the estimated variances. The histograms for the first-order autocovariances in figure 2.11 show equivalent results.

In conclusion: Suppose that in a simulation experiment where N samples are generated from a fractionally integrated process the mean of the N estimated autocovariance functions is taken as an estimator of the population autocovariance function. Then, if we do not have information about the population mean of the process, this mean function falls severely short of the population function not because the estimated variances and autocovariances are skewed to the left, but because the spread is much smaller and therefore the skewness is more important.[27] In Brockwell and Davies (1991) the slow convergence of the estimator of the mean to the population mean for long-memory processes is indicated.

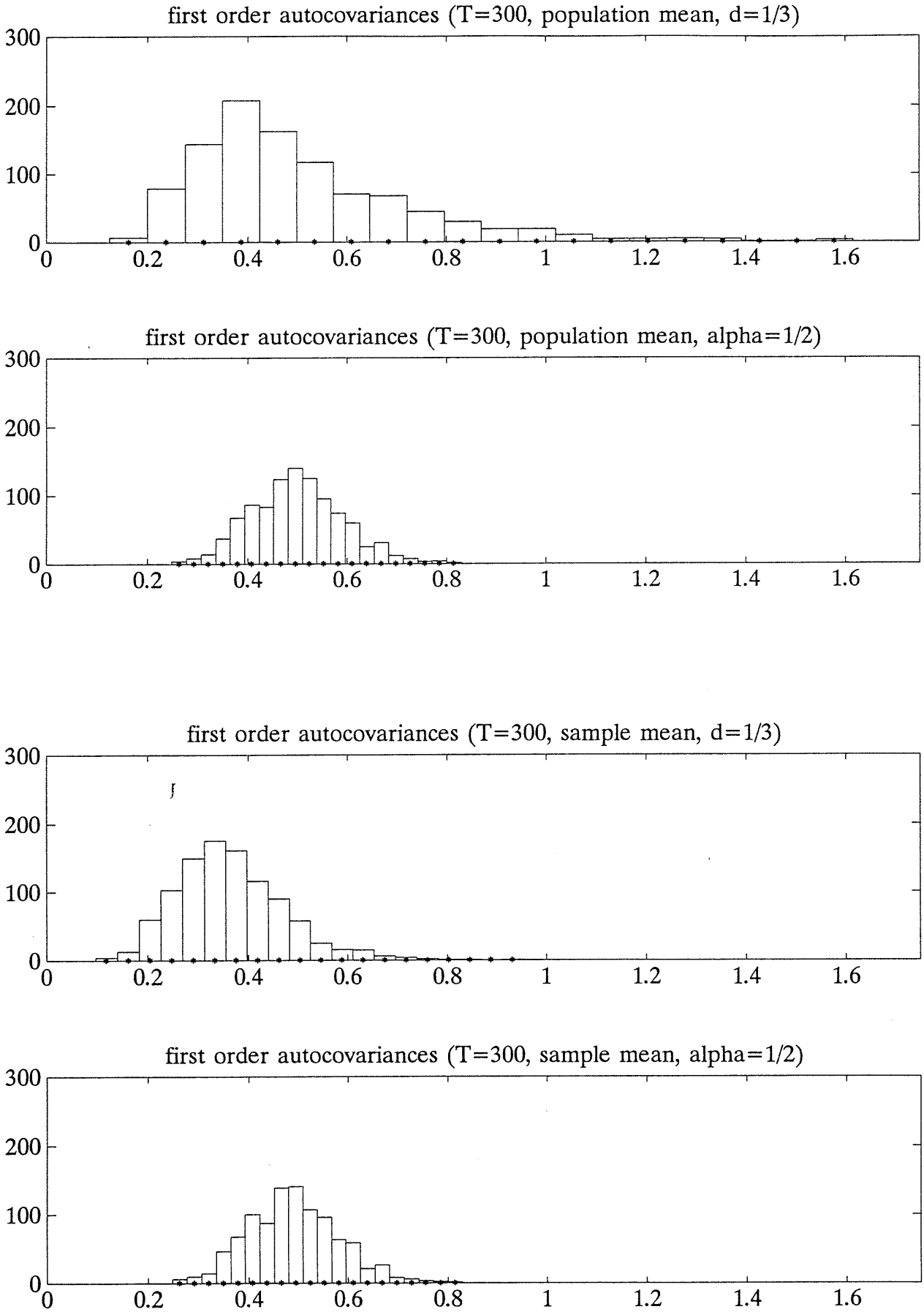
The Granger and Joyeux* (1980) *experiment replicated

A second type of simulation experiments has been performed by Granger and Joyeux (1980). They recognized the fact that the McLeod and Hipel (1978) method to generate a sample of observations could be time-consuming for T about 200 or more because it includes the direct Cholesky decomposition of the variance-autocovariance matrix Q. Instead of using the alternative

---

[27] Granger and Joyeux (1980) also encountered the problem of serious underestimated variance and autocovariances. Geweke (1992) believes that their problem is caused by taking a denominator of 4000 instead of 4000-k when calculating a k-th order autocovariance. However, it is possible that their problems can be attributed to the use of the sample mean when calculating the autocovariance functions.
Granger and Joyeux (1980) used the McLeod and Hipel (1978) procedure to generate 100 start-up values for an *AR*(100) approximation to an *ARFIMA*(0,0.25,0) and an *ARFIMA*(0,0.45,0) process. Series of length 400 were generated. The estimated and theoretical autocorrelations matched more closely for d=0.25 than for d=0.45.

Figure 2.11


first order autocovariances (T=300, population mean, d=1/3)
300
200
100
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
first order autocovariances (T=300, population mean, alpha=1/2)
300
200
100
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
first order autocovariances (T=300, sample mean, d=1/3)
300
200
100
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
first order autocovariances (T=300, sample mean, alpha=1/2)
300
200
100
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6

Jonas's (1981, 1983) indirect Cholesky decomposition method they proposed to approximate the infinite order autoregressive polynomial by a finite order autoregressive polynomial. Notice that the finite order polynomial coefficients are obtained from solving the Levinson-Durbin-Whittle recursions.

Given the autoregressive coefficients it is easy to calculate a new observation from the past observations. However, to procede in this way start-up values are needed. Granger and Joyeux (1980) used the McLeod and Hipel (1978) procedure to generate the start-up values necessary for the finite autoregressive approximation.

Specifically, Granger and Joyeux (1980) used the McLeod and Hipel (1978) procedure to generate 100 start-up values for an AR(100) approximation to an ARFIMA(0,0.25,0) and an ARFIMA(0,0.45,0) process. Series of length 400 were generated. The estimated and theoretical autocorrelations matched more closely for d=0.25 than for d=0.45.[28] From their tables it is observed that for $d = 0.25$ the mean estimated autocorrelations of each order are overestimated, whereas these are underestimated for $d = 0.45$. The estimated and theoretical autocorrelations matched more closely for $d = 0.25$ than for $d = 0.45$.

Geweke (1992) believes that their underestimation problem is caused by some kind of tapering or windowing, e.g. taking a denominator of N instead of N-k when calculating a k-th order autocovariance, where N denotes the number of simulations. However, it is possible that their problems can be attributed to the use of the sample mean when calculating the autocovariance functions, as appeared to be the case when dealing with our underestimation problems. In order to check this we replicated the Granger and Joyeux (1980) experiments. Series of length $T = 400$ given 100 start-up values were generated 1,000 times. However, we had difficulty even getting a satisfatory estimate of the unit variance of the process, as can be verified in appendix C. The reason for this is yet unknown.

The influence of sample size on the results

In this experiment we try to determine the influence of sample size on the results of the simulation studies reported above, i.e. for $T = 50$, 100, 200 and

---

[28] Granger and Joyeux (1980) did not mention how many series of length T were generated for each process. It could be that they generated one for each process.

300. Thereby we make a comparison between the results of the experiment when the autocovariance function is computed given the theoretical mean of the process and those given the sample mean of the process. There will be two processes under consideration: the $ARIMA(1,0,0)$ and the $ARFIMA(1,\frac{1}{3},0)$ processes from Table 3.1, respectively.

The figures 2.12 and 2.13 deal with the $ARFIMA(1,\frac{1}{3},0)$ process. The mean of the autocorrelation functions that are estimated given the population mean show a pattern as expected: for all four sample sizes they approximate the theoretical autocovariance function satisfactory. The influence of sample size is on the width of the confidence interval. However, if the sample mean is used for computation of the autocovariance functions we are confronted with the familiar problem of underestimation of the theoretical autocovariance function for all four sample sizes. Although the confidence intervals are smaller the higher sample size the theoretical autocovariance function is almost on their upper border and therefore almost significantly different from the mean of its empirical equivalents.

The figures 2.14 and 2.15 deal with the $ARIMA(1,0,0)$ process. Comparing these figures reveals that the confidence intervals of the mean of the estimated autocovariance functions for the $ARIMA(1,0,0)$ process are smaller than for the $ARFIMA(1,\frac{1}{3},0)$ process for each sample size. For the $ARIMA(1,0,0)$ process the underestimation becomes smaller as sample size grows if the sample mean is used, as is observed from figure 2.15. The width of the confidence intervals are comparable with that of the $ARFIMA(1,\frac{1}{3},0)$ process.

Figure 2.12

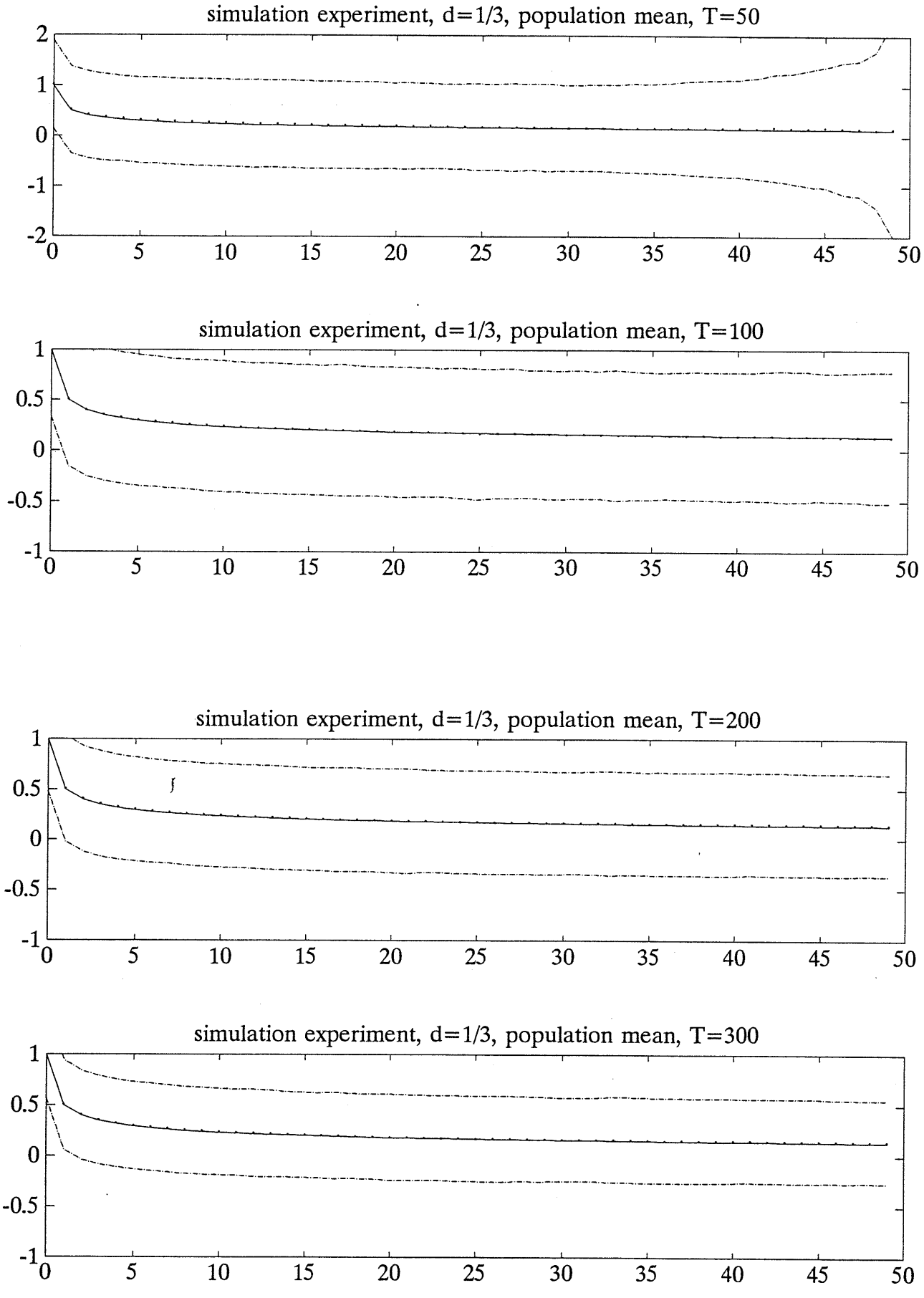

simulation experiment, d=1/3, population mean, T=50
2
1
0
-1
-2
0 5 10 15 20 25 30 35 40 45 50
simulation experiment, d=1/3, population mean, T=100
1
0.5
0
-0.5
-1
0 5 10 15 20 25 30 35 40 45 50
simulation experiment, d=1/3, population mean, T=200
1
0.5
0
-0.5
-1
0 5 10 15 20 25 30 35 40 45 50
simulation experiment, d=1/3, population mean, T=300
1
0.5
0
-0.5
-1
0 5 10 15 20 25 30 35 40 45 50

Figure 2.13

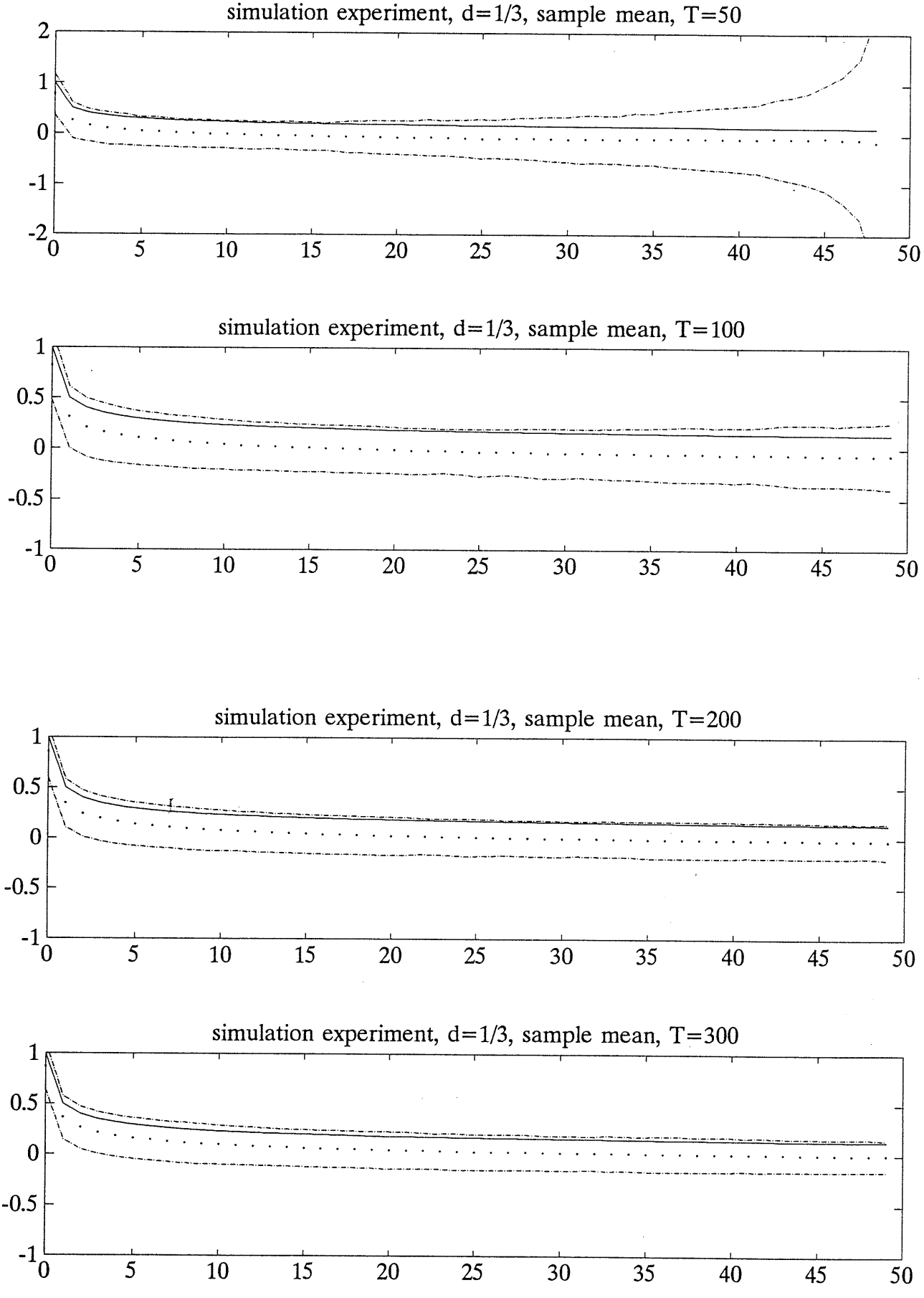
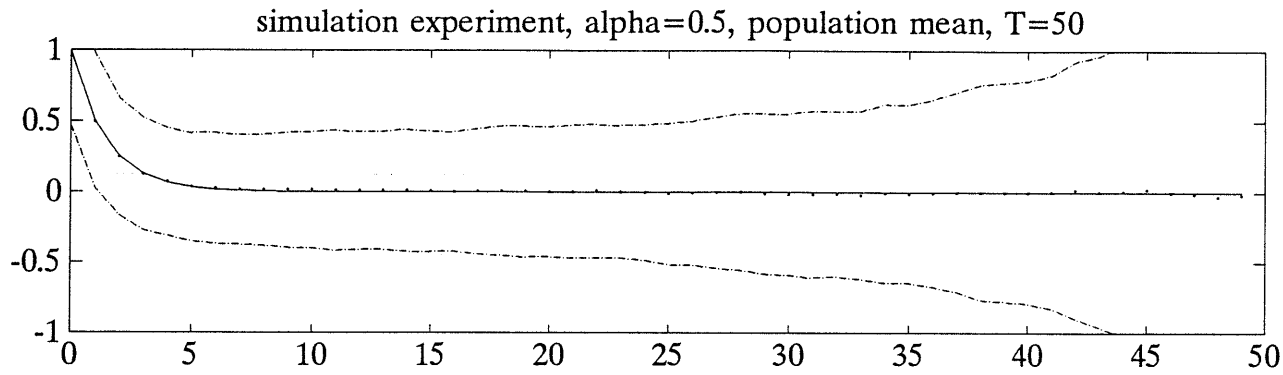
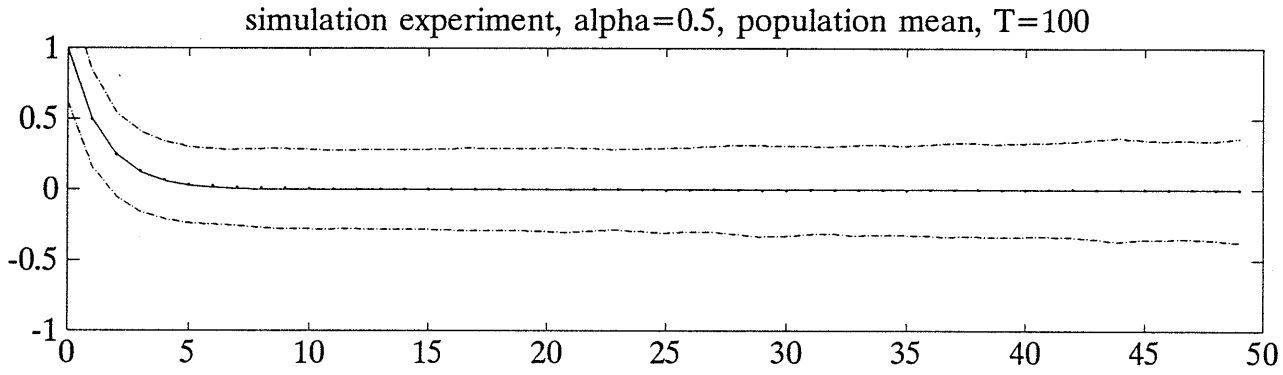
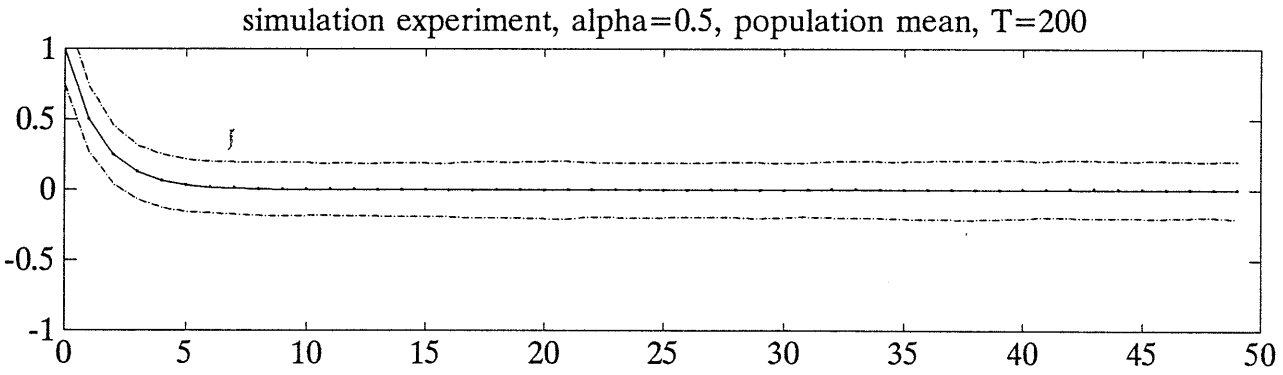

simulation experiment, d=1/3, sample mean, T=50
2
1
0
-1
-2
0 5 10 15 20 25 30 35 40 45 50
simulation experiment, d=1/3, sample mean, T=100
1
0.5
0
-0.5
-1
0 5 10 15 20 25 30 35 40 45 50
simulation experiment, d=1/3, sample mean, T=200
1
0.5
0
-0.5
-1
0 5 10 15 20 25 30 35 40 45 50
simulation experiment, d=1/3, sample mean, T=300
1
0.5
0
-0.5
-1
0 5 10 15 20 25 30 35 40 45 50

Figure 2.14


simulation experiment, alpha=0.5, population mean, T=50
1
0.5
0
-0.5
-1
0
5
10
15
20
25
30
35
40
45
50


simulation experiment, alpha=0.5, population mean, T=100
1
0.5
0
-0.5
-1
0
5
10
15
20
25
30
35
40
45
50


simulation experiment, alpha=0.5, population mean, T=200
1
0.5
0
-0.5
-1
0
5
10
15
20
25
30
35
40
45
50

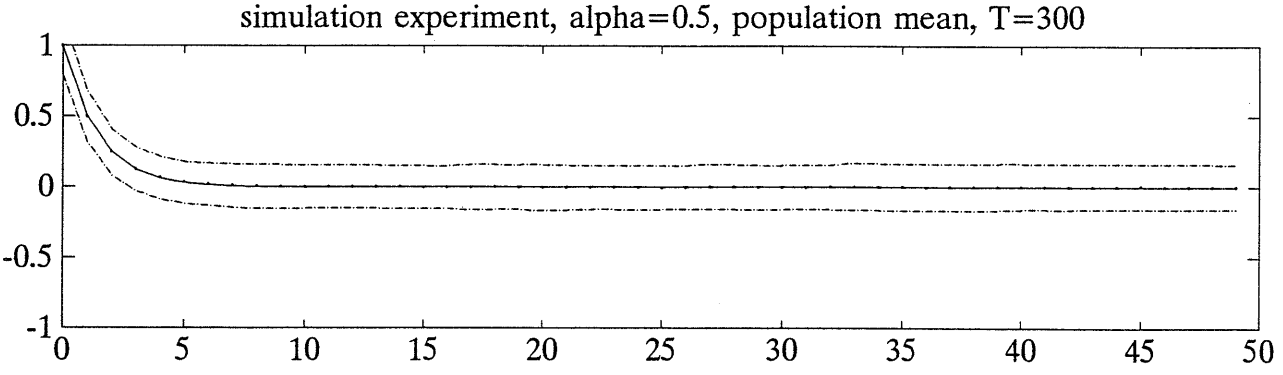
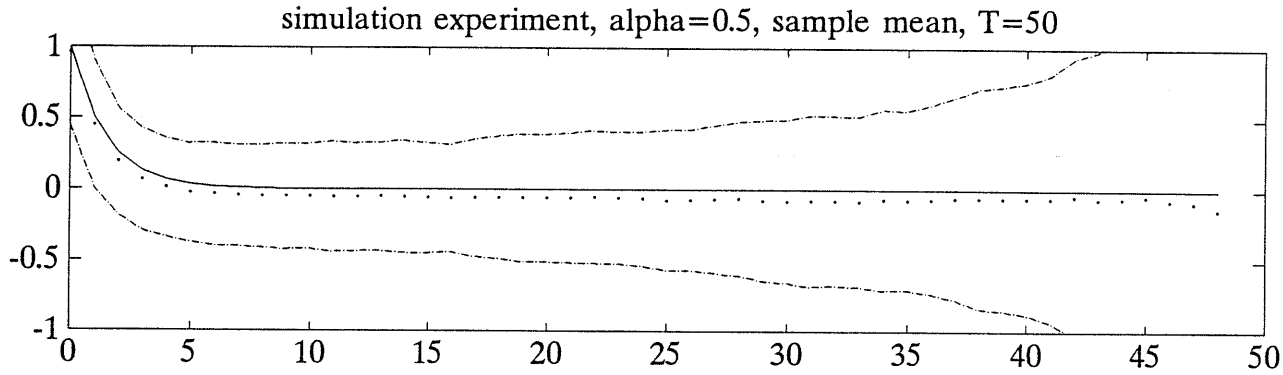
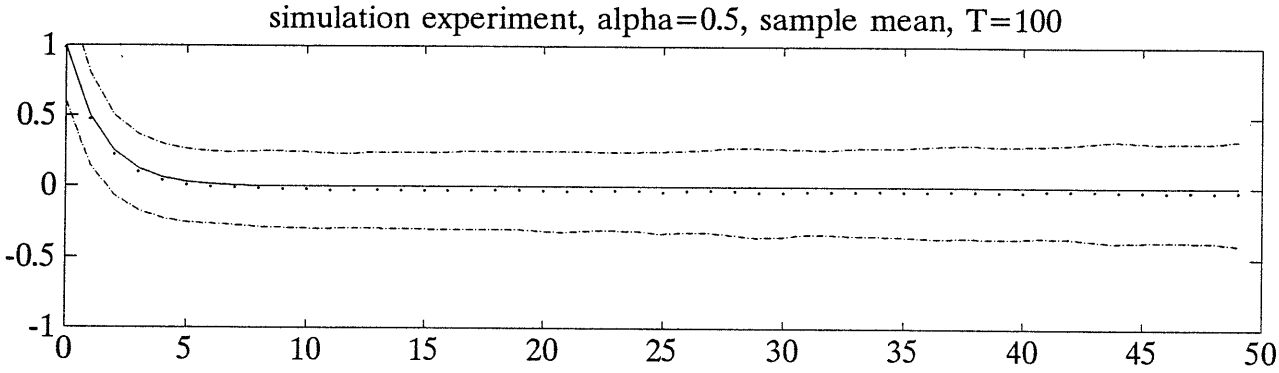
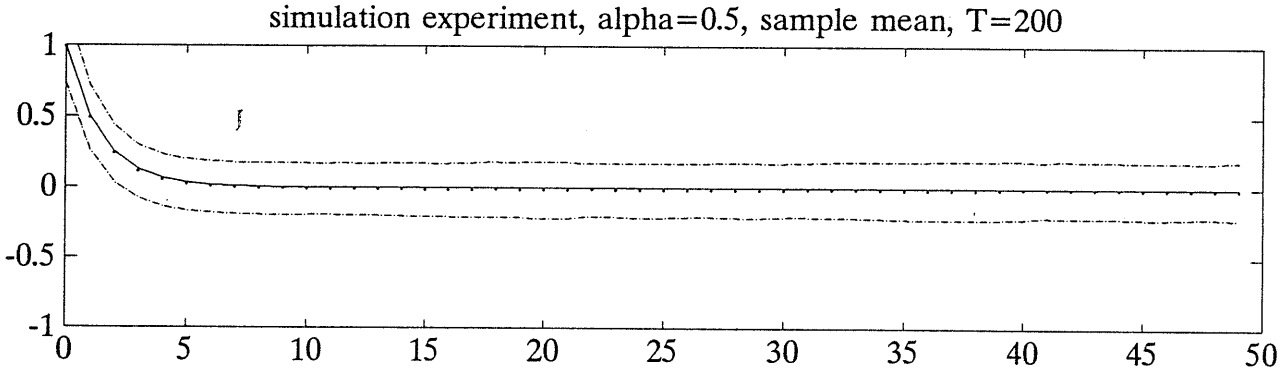
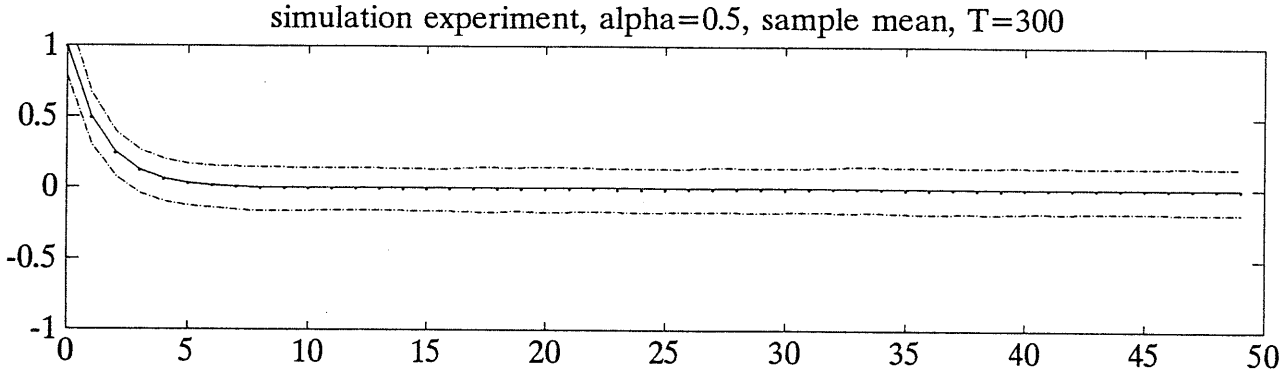

simulation experiment, alpha=0.5, population mean, T=300
1
0.5
0
-0.5
-1
0
5
10
15
20
25
30
35
40
45
50

Figure 2.15


simulation experiment, alpha=0.5, sample mean, T=50
1
0.5
0
-0.5
-1
0
5
10
15
20
25
30
35
40
45
50


simulation experiment, alpha=0.5, sample mean, T=100
1
0.5
0
-0.5
-1
0
5
10
15
20
25
30
35
40
45
50


simulation experiment, alpha=0.5, sample mean, T=200
1
0.5
0
-0.5
-1
0
5
10
15
20
25
30
35
40
45
50


simulation experiment, alpha=0.5, sample mean, T=300
1
0.5
0
-0.5
-1
0
5
10
15
20
25
30
35
40
45
50

3. ESTIMATION OF TIME SERIES MODELS

In this chapter we present four techniques for estimation of the models discussed before. The emphasis will be on the performance of these techniques with respect to models of fractionally integrated processes. In particular, the first three techniques to get an estimate of the fractional order of integration require a specification of the autoregressive and moving-average structure of the process. Application of the fourth technique as presented in Section 3.3 gives a parameter estimate irrespective of this model structure. This makes it possible to direct the estimate of the integration parameter for long-run modeling purposes only.

The first two estimation techniques which will be discussed in Section 1 lead to estimates of autoregressive and/or moving-average parameters only, i.e. the parameter of integration is supposed to be given. In Section 2 a maximum likelihood estimation technique is discussed which takes into account all parameters simultaneously. The Geweke and Porter-Hudak (1983) two-stage estimation technique gives an estimation of the parameter of integration d at the first stage (Section 3). The remaining *ARMA* parameters are estimated at the second stage (Yule-Walker, OLS, ML).

3.1 YULE-WALKER AND ORDINARY LEAST SQUARES ESTIMATION

Given a stationary process y we can estimate an autoregressive model without moving-average parameters by using the Yule-Walker equations (see Section 2.1.3); substitution of the sample autocovariances (or autocorrelations) then gives the estimates of the finite order autoregressive parameters $a_{k,p}$. The same methodology can be applied when we have to estimate an *ARMA* model. Box and Jenkins (1976) present Yule-Walker equations where both autoregressive and moving-average parameters are taken into account. McLeod (1975) presented an algorithm suitable for the machine calculation of the solution to these equations.[29] It should be noted that here the quality of the autoregressive and moving-average estimates depends on the quality of the estimated autocorrelations as well.

---

[29] The definition of b_k in McLeod (1975) should read $b_k=\sum_{i=k}^{q}-\theta_i c_{i-k}$.

An alternative method to estimate autoregressive models is the ordinary least-squares technique, the advantage of which is that using this technique under non-stationarity is justified as well. Unit root tests are not necessary as we can model the realization in levels.

3.2 MAXIMUM LIKELIHOOD ESTIMATION

A third methodology is called maximum likelihood and is based on the joint probability density function (p.d.f.) of the process y, which has to be stationary once more. The joint probability density function of the process y as in equation (2.1.1) can be written as

$$(1) \qquad f(y_1, y_2, \ldots, y_T | \Psi) = \prod_{t=0}^{T-1} f(y_{T-t} | Y_{T-t-1}, \Psi)$$

where $Y_t = (y_t, y_{t-1}, \ldots, y_1)$ is the information available up to period t and $f(y_1 | Y_0) = f(y_1)$ by definition. The vector of parameters is defined as

$$\Psi' = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_\varepsilon^2)$$

in case of *ARMA* processes. If we deal with fractionally integrated *ARMA* processes this vector reads

$$\Psi' = (d, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_\varepsilon^2).$$

If the observations become available we can calculate the value of the function f, given some vector of parameters Ψ. If other parameter values lead to an increase in the value of the function f these values can be seen as more adequate. This is the principle of maximum likelihood: the function f is reinterpreted as the <u>likelihood</u> <u>function</u> $L(\Psi | Y_T)$ and has to be maximized for a unique vector of parameters.

If y is a process that has p autoregressive components, the joint probability density function (1) can be written as

$$f(y_1, y_2, \ldots, y_T | \Psi) = f(Y_p) \cdot \prod_{t=p+1}^{T-1} f(y_{T-t} | Y_{T-t-1})$$

where $f(Y_p)$ is the joint probability density function of the first p observations. Depending on the decision to ignore $f(Y_p)$ or not, the <u>conditional</u> and <u>exact</u> log-likelihood functions follow from

$$lnf(y_1, y_2, \ldots, y_T | \Psi) = \sum_{t=p+1}^{T-1} lnf(y_{T-t} | Y_{T-t-1})$$

and

$$lnf(y_1,y_2,\ldots,y_T|\Psi) \;=\; lnf(Y_p) \;+\; \sum_{t=p+1}^{T-1} lnf(y_{T-t}|Y_{T-t-1}).$$

respectively. As exact maximum likelihood procedures turn out to be superior to conditional ones they are preferred.

Exact maximum likelihood requires specification of the function $f(Y_p)$. Given our earlier experience (see Section 2.4), however, there appears to be a maximum likelihood concept that does not suffer from this problem with starting-up values. Under the normality assumption the p.d.f. of the zero mean process y equals

$$f(y_1,y_2,\ldots,y_T|\Psi) \;=\; (2\pi)^{-T/2}|Q|^{-1/2}exp\{\; -\tfrac{1}{2}y'Q^{-1}y \;\}.$$

where Q is the variance-autocovariance matrix of the process. Normalizing this matrix with respect to the variance of the process, i.e. writing $Q=\gamma_y(0).Q_1$, we obtain

$$f(y_1,y_2,\ldots,y_T|\Psi) \;=\; (2\pi)^{-T/2}|\gamma_y(0)Q_1|^{-1/2}exp\{-\frac{1}{2\gamma_y(0)}y'Q_1^{-1}y\;\}$$

The matrix Q, and thus $\gamma_y(0)$ and the normalized variance-autocovariance matrix Q_1, are determined by the vector of parameters $\boldsymbol{\Psi}$. The inverse and determinant of the matrix Q_1 are obtained by solving the Levinson-Durbin-Whittle recursion formulas (see appendix B) as $Q_1^{-1}=A\Sigma^{-1}A'$ and $|\Sigma^2|$, respectively. The matrices A and Σ are defined as in (2.4.2) and (2.4.3), respectively.

Substitution, taking logarithms and finally multiplying through by -2 we obtain

(2) $$-2.lnf(y_1,y_2,\ldots,y_T|\Psi) \;=$$

$$= \; T.ln(2\pi) \;+\; T.ln(\gamma_y(0)) \;+\; |\Sigma^2| \;+\; y'A\Sigma^{-1}A'y \;/\; \gamma_y(0)$$

$$= \; T.ln(2\pi) \;+\; T.ln(\gamma_y(0)) \;+\; \sum_{t=1}^{T}\sigma_{t-1}^2 \;+\; \sum_{t=1}^{T}(y_t-y_{t-1,1})^2/\;(\sigma_{t-1}^2\gamma_y(0))$$

where $y_{t-1,1}$ equals the right-hand side of equation (2.1.7). The function $f(Y_t|\Psi)$ is reinterpreted as the likelihood function $L(\Psi|Y_T)$ and has to be maximized with respect to $\boldsymbol{\Psi}$ as soon as the observations become available. Given a numerical optimization procedure we can find the optimal vector of parameters $\boldsymbol{\Psi}$.

The second expression of $-2lnf(Y_t|\Psi)$ in (2) is termed the prediction error decomposition form of the likelihood (see e.g. Harvey 1989). Note that the prediction error $(y_t-y_{t-1,1})$, σ_{t-1}^2 and the variance of the process $\gamma_y(0)$ are

determined by the autoregressive and moving-average parameters and the residual variance σ_ε^2.

3.3 TWO-STAGE SEMI PARAMETRIC ESTIMATION

In their article Geweke and Porter-Hudak discuss two so-called 'long memory' models, that is the 'simple fractional Gaussian noise' and the 'simple integrated (or fractionally differenced) series'. They show the equivalence of both types of models. These models are called 'long memory' models because the spectrum is infinite at a frequency (not necessarily zero). Both models are generalized by postmultiplying their spectral densities with $f_u(\lambda)$, the spectral density of a 'short memory' process with *ARMA* representation. A two-stage procedure is proposed to model 'general integrated series': (1) An estimator of the 'long memory' parameter of fractional integration of the process, d, is obtained as the coefficient in the linear regression of the log periodogram of the process on that of a deterministic process. This regression is based on the low-frequency ordinates of the periodogram and the parameter d is an estimate of the slope of the spectral density. (2) Then an *ARMA* representation of the transformed series is estimated at the second stage.

Because the integration parameter d does not depend on the actual *ARMA* representation parameters, the estimator $\hat{d}$ is called semi-parametric. There are however estimation procedures that give more efficient estimators than the two-stage procedure, *if the assumed model structure is right.* These methods are maximum likelihood methods, either applied in the time or in the frequency domain. The estimator of d is no longer semi-parametric, because one specifies the full model in a finite number of parameters.

The most important theorem in the Geweke and Porter-Hudak article is the following one:

Theorem (*Geweke and Porter-Hudak* 1983, theorem 2, p. 227):

Suppose $\{y_t\}$ *is a general integrated linear process, with* $d<0$.[30] *Let* $I_y(\lambda_{j,T})$

---

[30] As shown before, the assumption $d<0$ implies that we are dealing with a short memory stationary process.

denote the periodogram of $\{y_t\}$ *at the harmonic frequencies* $\lambda_{j,T}=2\pi j/T$ *in a sample of size* T. *Let* $b_{1,T}$ *denote the ordinary least squares estimator of* β_1 *in the regression equation*

$$(3) \qquad ln\ I_y(\lambda_{j,T}) \ = \ \beta_0 \ + \ \beta_1.ln\{4sin^2(\lambda_{j,T}/2)\} \ + \ u_{j,T}$$

where $j=1,\ldots,n$.
Then there exists a function $z(T)$, *which will have the properties* $lim_{T\to\infty}z(T)=\infty$ *and* $lim_{T\to\infty}z(T)/T=0$, *such that*

(*i*) *if* $n=z(T)$, *then plim* $b_1=-d$

(*ii*) *if* $lim_{T\to\infty}(lnT)^2/z(T)=0$, *then* $(b_1+d)/\{\hat{var}(b_1)\}^{\frac{1}{2}}\to^D N(0,1)$

where $\hat{var}(b_1)$ *is the usual least squares estimator of* $var(b_1)$. ∎

From the theorem it follows that there exists a function $z(.)$, that depends on the length T of the time-series, such that estimation of equation (3) for $n=z(T)$ periodogram ordinates gives a consistent estimator of d. Asymptotic normality is assured, given an additional regularity condition.

From equation (2.3.3) it can be seen that the second regressor in equation (3) is the log of the pseudo spectrum of a fractionally integrated series y_t. To come from equation (2.3.3) at equation (3), take logarithms of both sides and add and subtract $ln\{g_y(1/2\pi)\}$, the spectrum of the *ARMA* part of the process at zero. Evaluation at the harmonic frequencies $\lambda_{j,T}$ as defined in the theorem gives

$$lnf_y(\lambda_{j,T}) \ = \ lng_y(1)-d.ln\{4sin^2(\lambda_{j,T}/2)\}+ln\{g_y(exp\{-i\lambda_{j,T}\})/g_y(1)\}$$

When we restrict attention to the low-frequency ordinates near zero, say, $\lambda_{j,T}$ where $j\leq K\ll T$, the last term can be dropped as negligible[31]. To make the equation operational, add the log periodogram at ordinate $\lambda_{j,T}$, say, $I_y(\lambda_{j,T})$, to both sides of the equation and rearrange to obtain

$$lnI_y(\lambda_{j,T}) \ = \ lng_y(1) \ - \ d.ln\{4sin^2(\lambda_{j,T}/2)\} \ + \ ln\{I_y(\lambda_{j,T})/f_y(\lambda_{j,T})\}$$

This is equation (3). The residuals $u_{j,T}$ are reflected by the third term at the right-hand side of the above equation. At the periodogram points where $g_y(exp\{-i\lambda_{j,T}\})\approx g_y(1)$ we are left with the spectrum of the long memory component.

---

[31] Geweke and Porter-Hudak (1983) propose to use the rule $z(T)=T^{\alpha}$, where $\alpha=0.55$ or $\alpha=0.6$, based on Monte Carlo simulations.

The two-stage semi-parametric spectral estimation procedure that is proposed by Geweke and Porter-Hudak consists of the following two steps

(a) finding an estimate $\hat{d}$ of the parameter d as a result of the regression as described in (3) and transforming the data by the filter $(1-L)^{\hat{d}}$ in the frequency domain;
(b) estimating the parameters of the remaining *ARMA* part of the process in the usual way in the time domain.

The second-stage *ARMA* parameters are consistent, but their asymptotic distribution is unknown.

Unit root tests

Diebold and Nerlove (1990) conclude that the potential usefulness of the results of the first-stage d estimate as a unit-root test is obvious. First, the first-stage d estimate is asymptotically normal when $d=1$, but also when $0<d<1$. This differs sharply from the nonstandard distribution theory required for least-squares estimates of unit-roots. Second, estimation and testing of d proceeds independently of the (generally infinite dimensional) nuisance "parameter" $B(z)=\phi^{-1}(z)\theta(z)$, freeing the investigator from (frequently dubious) assumptions regarding the form of the data-generating process.

3.4 THE CHOICE OF AN OPTIMAL MODEL

In this thesis, the choice of a criterion function to find an optimal model depends on the structure of the estimated models. If the models are estimated by the same estimation technique but the autoregressive and/or moving-average polynomial orders p and q are different, two optimal models are chosen: one by the Akaike criterion (*AIC*) and one by the Schwartz criterion (*SIC*) which read

$$AIC = 2.lnf \; - \; 2 \; . \; k/T$$
$$SIC = 2.lnf \; - \; lnT \; . \; k/T$$

where k denotes the number of elements of the vector of parameters Ψ. In order to have a maximization problem as object in the empirical part of this thesis we computed the negative values of *AIC* and *SIC*. According to Judge (1985) *AIC* asymptotically overestimates the autoregressive order with positive probability, given a finite order *AR* process; an estimator of the *AR* order based on *AIC* will therefore not be consistent. *SIC* does give consistent estimators in this case.

If the *AIC* and *SIC* criteria lead to different optimal models, we subsequently have to choose between these according to some device. If the optimal model out of several models of the same type is not unique or the models are not of the same type, the following forecasting device is used to make a unique choice possible: the square root of the mean squared in-sample forecast errors, i.e. the minimum sum of squared residuals. The model for which this criterion is minimal is called optimal. This strategy amounts to maximizing the R^2. For example, the forecasting device is applied if we have four different optimal models as a consequence of the four estimation techniques proposed in the previous sections.[32]

[32] Contrary to the procedure presented here, Sowell (1992a) applies the *AIC* and *SIC* criteria functions to make a choice between fractional and non-fractional *ARMA* and fractional *ARMA* models.

4. MODELING ECONOMIC AND NON-ECONOMIC DATA

In this section we give an empirical illustration of the theory presented in the sections before. In the section 4.1 we describe the maximum likelihood computer codes that we applied for estimation of nonfractional and fractional *ARMA* models. In both codes the input data is standardized initially: its estimated mean is subtracted and the result is divided by its estimated standard deviation. The value of the likelihood function is corrected for this standardization afterwards. In section 4.2 the long- and short-memory behavior of quarterly real US gross national product is investigated. In section 4.3 we determine the long- and short memory characteristics of the annual Trier oak tree ring widths series. The length of this series is large as compared to most economic series and is therefore interesting for long-memory studies.

4.1 THE MAXIMUM LIKELIHOOD COMPUTER CODE

To evaluate the exact likelihood function of a nonfractional *ARMA* model the autocovariances have to be expressed in the parameters of the model. The variance-autocovariance matrix Q can be calculated from Ψ by using the Yule-Walker equations corresponding to stationary nonfractional *ARMA* models as given by Box and Jenkins (1976) and McLeod (1975) (see Section 2.1.3), given that y has a stationary *ARMA* representation. Given an estimate of the vector of parameters Ψ we can substitute these values to obtain the variance-autocovariance matrix Q corresponding to the estimated *ARMA* model. Note that the application of the Yule-Walker equations in this way is conceptually different from its application as an estimation technique: there the parameters are unknown and the estimated autocovariances are substituted.

The exact maximum likelihood procedure applied to fractionally integrated *ARMA* processes is presented in Sowell (1992b), based on Sowell (1987). A recursion formula for the calculation of the autocovariance function of processes with an autoregressive and/or moving-average component is obtained which reduces the number of calculations substantially (see Section 2.3.2.1). If there is no *ARMA* structure we simply employ the equation presented in Section 2.3.1.1.

The Vax/VMS Fortran computer code to evaluate the likelihood function of a fractionally integrated *ARMA* process given an estimate of the vector of parameters Ψ was obtained by the author of this thesis from Sowell on request. This Vax/VMS Fortran computer code has been transformed to Convex/Unix Fortran computer code and adjusted to allow for times series of length 1200 instead of 500. Instead of the numerical maximization package GQOPT which is available from R.E. Quandt and applied by Sowell (1992b), we implemented the numerical minimization library called VARM (i.e. a variable metric or quasi-Newton sub- routine), which is IBM/DOS Fortran computer code originally and written by economists of the Econometric Department of the Erasmus University Rotterdam (EUR). Furthermore, the IMSL routine DZPLRC and the function DGAMMA are substituted by the Numerical Algorithms Group Limited routines (NAG) C02ABF and the function S14AAF, respectively. The NAG routines G13ABF and G13ADF have been applied to find initial *ARMA* parameter estimates along the Yule-Walker lines described by Box and Jenkins (1976) and McLeod (1975).

In this thesis we adopt the iterative quasi-Newton Davidon-Fletcher-Powell (DFP) updating formula of the vector of parameters (see Judge et al 1985 and the references therein)

$$\Psi_{n+1} = \Psi_n - t_n P_n \gamma_n$$

where Ψ_n denotes the vector of parameters to be estimated, t_n is the step length and $P_n\gamma_n$ is the step direction, all in iteration n. The direction matrix P_n is an approximation to the inverse Hessian matrix corresponding to the likelihood funcion and γ_n denotes the first derivative of the likelihood function, both at Ψ_n. Thus $P_{n+1}=P_n+M_n$ where M_n is the DFP correction matrix. In this thesis γ_n is computed numerically. Asymptotic standard errors are based on a finite difference computation of the Hessian matrix in Ψ_n^* which gives the minimum of the likelihood function.

Step length t_n is not determined by an exact line minimization method here.[33] Instead we use an interval enclosure method. As this does not guarantee a positive definite P_{n+1} even if P_n is positive definite, an alternative updating formula is the quasi-Newton Broyden-Fletcher-Goldfarb-Shano (BFGS) formula which differs from the DFP updating formula with respect to M_n (Van der Hoek and Kolen 1986, p.188). It is possible to apply the BFGS updating

[33] Exact line minimization means that in each iteration t_n is selected such that $f(\psi_n+\zeta_n)$ is minimal for each step $\zeta_n=-t_nP_n\gamma_n$, given ψ_n, P_n and γ_n.

formula given the Fortran code described above. However, in this thesis we applied the DFP updating formula only.

Nonfractional as well as fractional *ARMA* models will be estimated by maximum likelihood for the autoregressive polynomial orders $p = 0,1,2,3$ and the moving-average polynomial orders $q = 0,1,2,3$. When there are no moving-average parameters present, for each autoregressive order p (1) the initial autoregressive parameters are estimated as the solution of the Yule-Walker equations, given an estimate of the first p autocorrelations; (2) the variance of the error process is put equal to the estimated residual variance of this autoregressive model; and (3) the fractional order of integration d is set to 0.01;[34]

Then for each autoregressive order p the likelihood function of a model containing 0, 1, 2 and 3 moving-average parameters is maximized by maximum likelihood along the lines described above. The initial moving-average parameters are set to the optimum parameter estimate of the previous model if these estimates are present —— given the same p —— and to -0.01 otherwise.

4.2 QUARTERLY REAL US GROSS NATIONAL PRODUCT

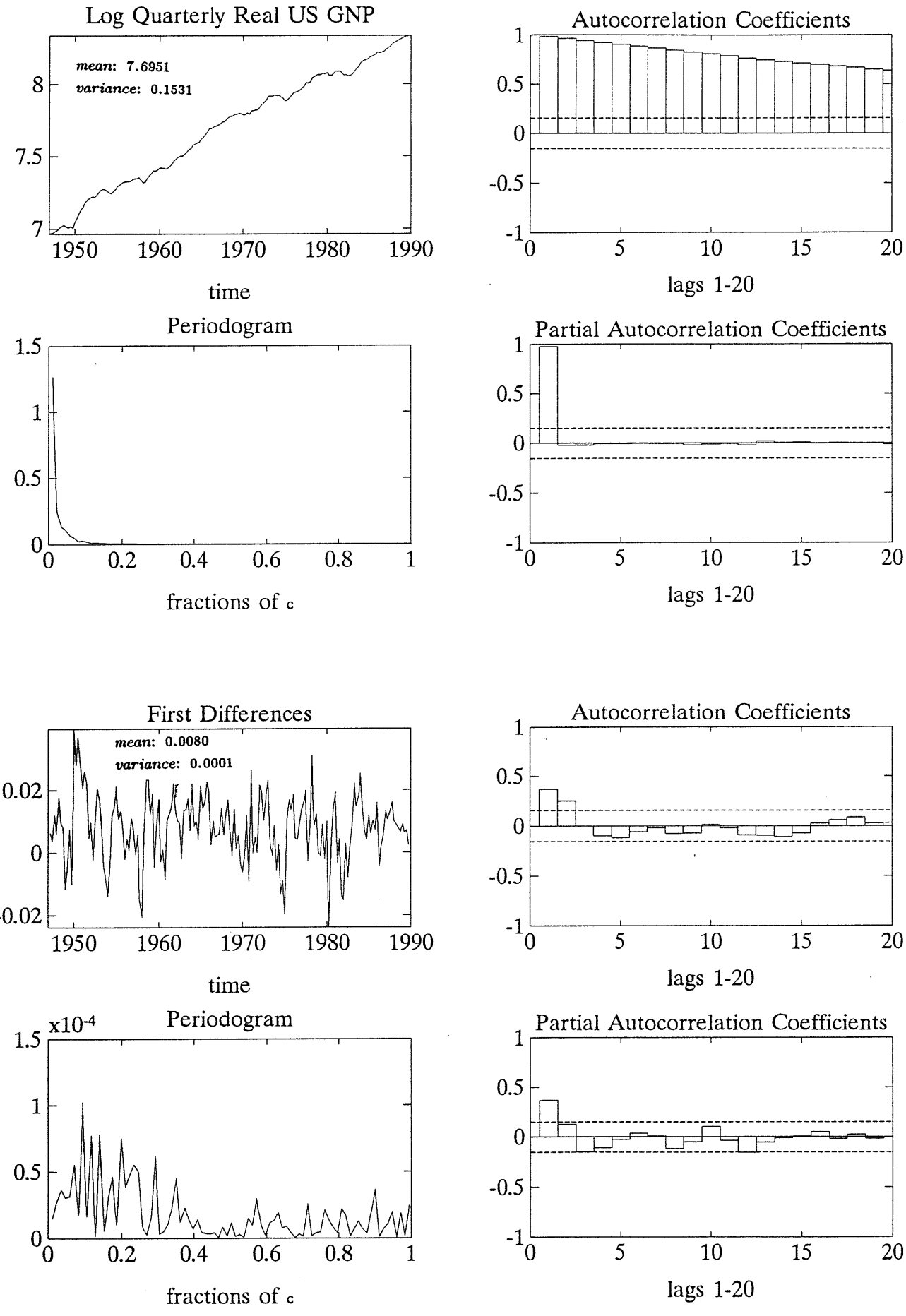
Sowell (1992a) modeled the first differences of the natural logarithm of quarterly real US gross national product (hereafter: log GNP) from 1947:I to 1989:IV (172 data points, seasonally adjusted). As the autocorrelation function of log GNP dies out slowly (see figure 4.1) modeling first differences instead of levels of log GNP is proposed by Box and Jenkins (1976). From the partial autocorrelation function an $AR(1)$ model for the first differences seems appropriate; an estimate of the first order autoregressive coefficient will be near 0.4 (see the autocorrelation function of the first differences).

However, at the frequencies close to zero the shape of the periodogram of these first differences indicates overdifferentiation. Although comparison of the plots of the levels and first differences of log GNP clearly reveals that

---

[34] Contrary to what is stated in Sowell (1992b, footnote 7) the original Vax/VMS Fortran computer code does not allow for a zero parameter of integration d. In this thesis we apply this code (albeit modified) to estimate fractional *ARMA* models only as well and other computer code to estimate stationary nonfractional *ARMA* models only.

Figure 4.1


Log Quarterly Real US GNP
mean: 7.6951
variance: 0.1531
8
7.5
7
1950
1960
1970
1980
1990
time
Autocorrelation Coefficients
1
0.5
0
-0.5
-1
0
5
10
15
20
lags 1-20
Periodogram
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
fractions of c
Partial Autocorrelation Coefficients
1
0.5
0
-0.5
-1
0
5
10
15
20
lags 1-20
First Differences
mean: 0.0080
variance: 0.0001
0.02
0
-0.02
1950
1960
1970
1980
1990
time
Autocorrelation Coefficients
1
0.5
0
-0.5
-1
0
5
10
15
20
lags 1-20
x10-4
Periodogram
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
fractions of c
Partial Autocorrelation Coefficients
1
0.5
0
-0.5
-1
0
5
10
15
20
lags 1-20

the levels are nonstationary whereas the latter are stationary, overdifferentiation indicates that an order of differencing *less than one* is already enough to make a nonstationary series stationary. We will illustrate this later.

Augmented Dickey-Fuller tests

First we perform the ADF test by estimation of equation (2.2.4) including a constant and deterministic linear trend of order one for log GNP, denoted lny_t (i.e. y_t denotes the levels of log GNP). Lagged differences are included up to order 50 ($p=51$). The five percent critical value is -3.45 (Fuller 1976, table 8.5.2, $\hat{\tau}_\tau$, $n=100$).

For lny_t the ADF test statistic not significantly different from zero for all lags up to 50, except at lag 11 (see figure 4.2). As the SIC criterion indicates that only one lag of first differences is appropriate the conclusion to be drawn is that the zero frequency unit root of order one hypothesis for lny_t cannot be rejected given a five percent significance ADF test. For the first differences of log GNP we observe from figure 4.2 that the hypothesis of a unit root in first differences is rejected at a five percent significance level, at least if we do not include lagged differences of too high order. The zero frequency unit root therefore has multiplicity one given these ADF tests results, i.e. log GNP is zero frequency integrated of order one.

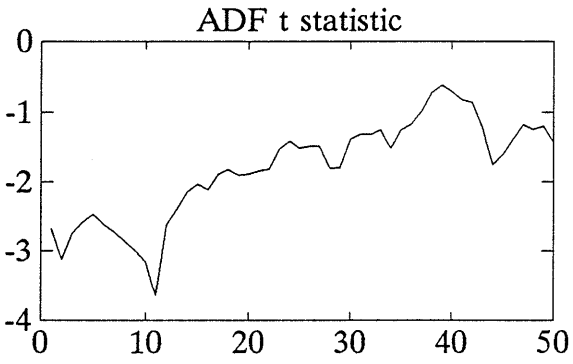
Model estimation

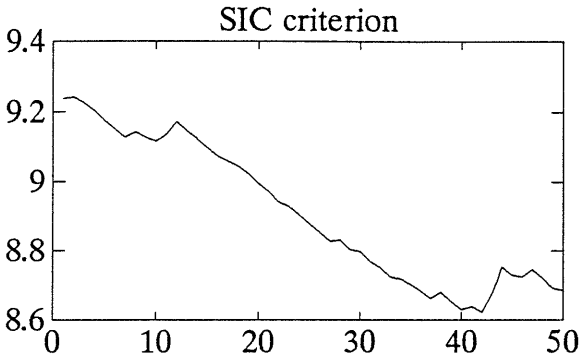
Given the results of the ADF tests we model first differences of log GNP, denoted Δlny_t, by nonfractional *ARMA* models and estimate these for $p=0,1,2,3$ and $q=0,1,2,3$ by maximum likelihood. As we have seen before from the periodogram of Δlny_t modeling this series by some fractional *ARMA* model seems appropriate as well. The corresponding parameter of fractional integration will be negative because of overdifferentiation. The fractional *ARMA* models will be estimated by maximum likelihood as well. A fractional *AR* model for Δlny_t will be estimated following Geweke and Porter-Hudak (1983) and a nonfractional model for Δlny_t will be estimated by solving the Yule-Walker equations. A nonfractional model for the levels of log GNP will be estimated by ordinary least squares.

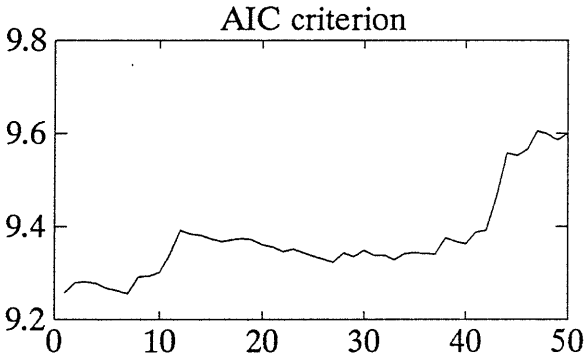
Nonfractional ARMA models estimated by maximum likelihood

Fractional and nonfractional $ARMA(p,q)$ models have been estimated by maximum

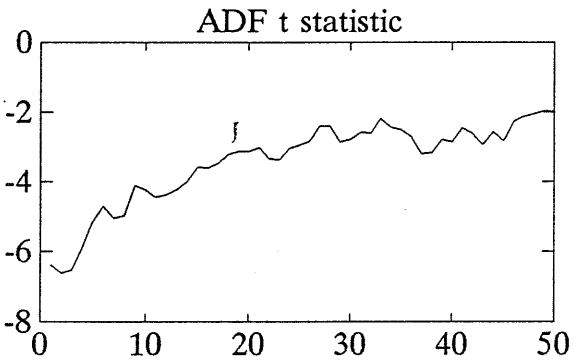
Figure 4.2

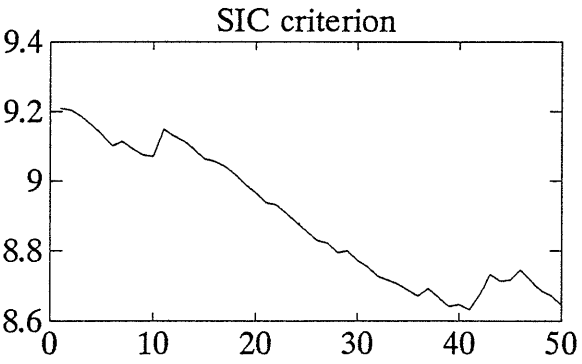


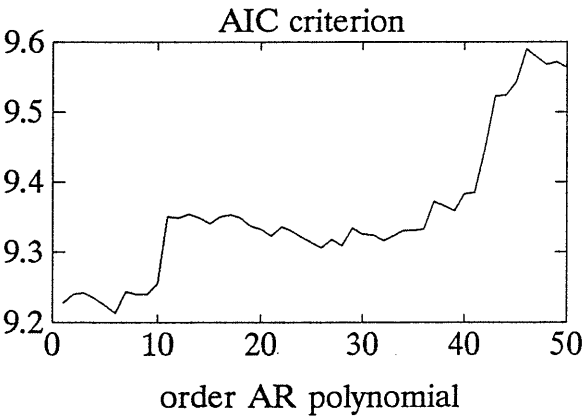




LEVELS







FIRST DIFFERENCES

likelihood for $p=0,1,2,3$ and $q=0,1,2,3$. As Sowell kindly provided the data together with copies of the original computer output corresponding to the estimated fractionally integrated $ARMA(p,q)$ models, we are able to check our results with respect to these models in a replicated experiment. The tables 4.1 and 4.3 presented below can be compared to tabel 1 in Sowell (1992a).

For each model we present in table 4.1 two times the log-likelihood value and the Akaike (AIC) and Schwartz (SIC) information criterion values at the optimal vector of parameters Ψ as defined in section 3.4. In the tables 4.2 and 4.4 we present the parameter estimates of the nonfractional and fractional *ARMA* models, respectively. They can be compared to the tables 3 and 2 of Sowell (1992a), respectively.[35]

Table 4.1

2lnL, AIC and SIC for sixteen nonfractional $ARMA(p,q)$ models of the standardized first differences of log quarterly real GNP (s.a.)

Number of *AR* parameters	Number of *MA* parameters			
	0	1	2	3
	1065.658	1083.242	1095.959	1097.987
0	1065.658	1081.242	1091.959	1091.987
	1065.658	1078.100	1085.675	1082.562
	1090.567	1092.068	1097.256	1098.003
1	1088.567	1088.068	1091.256	1090.003
	1085.425	1081.784	1081.831	1077.436
	1093.472	1094.964	1101.185	1102.362
2	1089.472	1088.964	1093.185	1092.362
	1083.189	1079.539	1080.618	1076.654
	1097.269	1098.936	1102.649	1102.674
3	1091.269	1090.936	1092.649	1090.674
	1081.844	1078.370	1076.941	1071.824

From table 4.1 we observe that the AIC optimal model is $ARMA(2,2)$, whereas the SIC optimal model is $ARMA(0,2)$. The parameter estimates for the sixteen models are presented in table 4.2. The two optimal models for the standardized first differences of log GNP are respectively (t-statistics between parentheses)

[35] The autoregressive parameters in Sowell (1992a) are the negative of the autoregressive parameters in this thesis.

$$(1 - 0.60L + 0.49L^2)\ \Delta lny_t = (1 - 0.30L + 0.64L^2)\ \varepsilon_t$$
$$(-3.91)\quad (3.11)\qquad\qquad (-2.39)\quad (4.12)$$

and

$$\Delta lny_t = (1 + 0.30L + 0.27L^2)\ \varepsilon_t$$
$$(4.07)\quad (3.88)$$

where the corresponding estimated residual variances are 9.3249×10^{-5} and 9.6284×10^{-5}, respectively. A choice between the two models will be made on the basis of their forecasting performances at the end of this section.

Table 4.2

Parameter estimates of sixteen nonfractional ARMA(p,q) models of the standardized first differences of log quarterly real GNP (s.a.); t-statistics between parentheses

Model	ϕ_1	ϕ_2	ϕ_3	θ_1	θ_2	θ_3
(0,1)				0.27		
				(4.43)		
(0,2)				0.30	0.27	
				(4.07)	(3.88)	
(0,3)				0.33	0.34	0.13
				(4.41)	(3.97)	(1.45)
(1,0)	0.37					
	(5.18)					
(1,1)	0.52			-0.17		
	(4.11)			(-1.28)		
(1,2)	0.25			0.07	0.24	
	(1.36)			(0.41)	(2.75)	
(1,3)	-0.04			0.37	0.35	0.14
	(-0.13)			(1.11)	(2.77)	(1.03)
(2,0)	0.32	0.13				
	(4.23)	(1.71)				
(2,1)	-0.07	0.29		0.39		
	(-0.24)	(2.55)		(1.34)		
(2,2)	0.60	-0.49		-0.30	0.64	
	(3.91)	(-3.11)		(-2.39)	(4.12)	
(2,3)	0.40	-0.53		-0.09	0.71	0.14
	(1.85)	(-3.35)		(-0.39)	(6.42)	(1.23)
(3,0)	0.34	0.18	-0.15			
	(4.50)	(2.26)	(-1.96)			
(3,1)	0.83	0.02	-0.22	-0.50		
	(2.29)	(0.15)	(-2.89)	(-1.36)		
(3,2)	0.60	-0.67	0.14	-0.28	0.79	
	(4.05)	(-3.75)	(1.32)	(-2.28)	(6.72)	
(3,3)	0.68	-0.72	0.18	-0.36	0.82	-0.06
	(1.31)	(-2.04)	(0.62)	(-0.69)	(4.00)	(-0.16)

From table 4.2 it appears that the parameter estimates of the nonfractional *ARMA*(1,3) and *ARMA*(3,3) models differ significantly from those presented in table 3 of Sowell (1992a): for the *ARMA*(1,3) model the parameters estimated

by Sowell are respectively 0.93, −0.64, −0.05 and −0.30; for the $ARMA(3,3)$ model the estimated parameters are respectively 1.52, −0.97, 0.41, −1.26, 0.90 and −0.64. Comparison of the corresponding $2lnL$, AIC and SIC values presented in table 4.1 and those presented in Sowell's table 1 reveals that the parameter estimates in our case are suboptimal. A possible explanation is that our norm of gradient convergence criterion in VARM is less strict than the one applied in GQOPT. Furthermore, different starting values could lead to different "optimal" parameter estimates for these two models.

If no starting values are supplied, maximum likelihood estimation of the $ARMA(1,3)$ model in Sun/Unix TSP leads to non-convergence and estimation of the $ARMA(3,3)$ model to parameter estimates close to ours.[36] Moreover, given the same starting values as in our Convex/Unix Fortran code —— these are the same as those applied in Sowell (1992a) —— the TSP parameter estimates converged to our estimates in both cases. Furthermore, the parameter estimates of Sowell (1992a) indicate common zero frequency unit roots.

From table 4.1 it is observed that addition of a third moving-average parameter is redundant: for each autoregressive order the AIC and SIC values decrease as compared to the models with two moving-average parameters.[37] Furthermore, the third moving-average parameter is not significantly different from zero in any case as can be seen in table 4.2.

As the quarterly real US GNP data are seasonally adjusted it is advisable to include a fourth moving-average parameter and check for statistical significance.[38] Fixing the third moving-average parameter value at zero and applying the same starting values strategy as described above only the $ARMA(1,4)$ model resulted in statistically significant parameter values. However, the model estimated is non-stationary and these type of models are excluded from the analysis in this thesis.

Fractional ARMA models estimated by maximum likelihood

From table 4.3 we observe that on the basis of AIC the fractionally

---

[36] In Sun/Unix TSP the Gauss-Marquardt updating formula is applied.

[37] The AIC value slightly increases in case of zero autoregressive order.

[38] In the periodogram of log GNP the dip at the seasonal frequency $\pi/2$ indicates that the series has been seasonally adjusted. Seasonal overdifferentiation is a reason for considering frequency $\pi/2$ integration of fractional order. This will not be performed here.

integrated $ARMA(3,2)$ model is optimal, whereas it is an $ARMA(1,0)$ model if SIC is the chosen criterion.

Table 4.3

2lnL, AIC and SIC for sixteen fractional ARMA(p,q) models of the standardized first differences of log quarterly real GNP (s.a.)

Number of *AR* parameters	Number of *MA* parameters			
	0	1	2	3
	1083.436	1085.865	1096.057	1100.848
0	1081.436	1081.865	1090.057	1092.848
	1078.294	1075.582	1080.632	1080.281
	1095.024	1095.198	1100.911	1102.092
1	1091.024	1089.198	1092.911	1092.092
	1084.741	1079.773	1080.345	1076.384
	1095.468	1097.328	1101.748	1103.053
2	1089.468	1089.328	1091.748	1091.053
	1080.043	1076.761	1076.039	1072.203
	1100.406	1100.467	1105.695	1105.721
3	1092.406	1090.467	1093.695	1091.721
	1079.839	1074.759	1074.845	1069.730

From table 4.4 it appears that the estimated parameters of the fractional *ARMA* model are almost all identical to the corresponding ones in Sowell (1992a).[39] The standard deviations reveal more differences, though never significant. For the optimal models $d = -0.59$ and $d = -0.45$, i.e. log quarterly real US GNP is zero frequency integrated of order 0.41 or 0.55, respectively. Thus, according to the former model log GNP has a long memory stationary model representation whereas it is non-stationary according to the latter model.

$$\underset{(-3.01)}{(1 - 1.18L} + \underset{(2.94)}{0.93L^2} - \underset{(-2.60)}{0.51L^3)}\ \underset{(-1.73)}{(1 - L)^{-0.59}}\ \Delta lny_t = \underset{(-2.23)}{(1 - 0.29L} + \underset{(7.20)}{0.81L^2)}\ \varepsilon_t$$

and

$$\underset{(-6.44)}{(1 - 0.77L)}\ \underset{(-2.86)}{(1 - L)^{-0.45}}\ \Delta lny_t = \varepsilon_t$$

[39] The autoregressive parameters in Sowell (1992a) are the negative of the autoregressive parmaters in this thesis.

where the estimates of the residual variance corresponding to the two nonfractional *ARMA* optimal models are 9.0253×10^{-5} and 9.6508×10^{-5}, respectively. A choice between the two models will be made on the basis of their forecasting performances at the end of this section.

Table 4.4

Parameter estimates of sixteen fractional ARMA(p,q) models of the standardized first differences of log quarterly real GNP (s.a.); t-statistics between parentheses

Model	d	ϕ_1	ϕ_2	ϕ_3	θ_1	θ_2	θ_3
(0,d,0)	0.29						
	(3.98)						
(0,d,1)	0.16				0.16		
	(1.52)				(1.65)		
(0,d,2)	-0.03				0.33	0.28	
	(-0.32)				(2.99)	(3.67)	
(0,d,3)	-0.20				0.48	0.46	0.25
	(-1.92)				(4.66)	(4.70)	(2.37)
(1,d,0)	-0.45	0.77					
	(-2.86)	(6.44)					
(1,d,1)	-0.38	0.74			-0.05		
	(-1.69)	(5.52)			(-0.42)		
(1,d,2)	-0.41	0.65			0.06	0.23	
	(-1.42)	(2.57)			(0.52)	(2.56)	
(1,d,3)	-0.33	0.37			0.26	0.39	0.17
	(-1.73)	(1.04)			(1.17)	(2.61)	(1.17)
(2,d,0)	-0.30	0.60	0.08				
	(-1.27)	(2.44)	(0.79)				
(2,d,1)	-0.29	0.07	0.44		0.52		
	(-1.54)	(0.22)	(2.07)		(1.63)		
(2,d,2)	-0.25	0.78	-0.26		-0.23	0.37	
	(-0.79)	(2.80)	(-1.05)		(-0.93)	(1.45)	
(2,d,3)	-0.13	0.41	-0.39		0.03	0.63	0.19
	(-0.77)	(1.73)	(-1.45)		(0.12)	(3.39)	(1.51)
(3,d,0)	-0.39	0.70	0.17	-0.17			
	(-1.52)	(2.76)	(1.65)	(-2.24)			
(3,d,1)	-0.34	0.75	0.11	-0.18	-0.11		
	(-1.11)	(2.29)	(0.45)	(-2.23)	(-0.25)		
(3,d,2)	-0.59	1.18	-0.93	0.51	-0.29	0.81	
	(-1.73)	(3.01)	(-2.94)	(2.60)	(-2.23)	(7.20)	
(3,d,3)	-0.52	1.16	-0.94	0.50	-0.34	0.83	-0.04
	(-1.10)	(3.04)	(-3.03)	(2.33)	(-1.06)	(5.96)	(-0.17)

In the literature (e.g. Sowell 1992; Diebold and Rudebusch 1989) the zero frequency unit root of order one hypothesis given log GNP is sometimes rejected and sometimes not rejected against the alternative of a deterministic trend by an ADF unit root test. Given our ADF test results this hypothesis could not be rejected. From the maximum likelihood results above

we observe that a zero frequency unit root is present in the autoregressive structure of log GNP, however not of integer order one but of fractional order (0.41 or 0.55). The ADF test for a zero frequency unit root of integer order one therefore restricts the set of possible models for log GNP and seems to be a test for non-stationarity versus stationarity rather than a unit root test.

The modeling strategy of Sowell (1992a) clearly leads to the incorrect impression that the nonfractional $ARMA(0,2)$ model is an important nonfractional alternative to the fractional $ARMA(3,-0.59,2)$ model.[40] From the plot of the spectral densities of the estimated models over the periodogram for the first differences of log GNP it appears that the nonfractional $ARMA(2,2)$ model gives a better description of the frequency domain behavior than the model advocated by Sowell (see figure 4.3).[41] Campbell and Mankiw (1987) selected this model out of a set of nonfractional *ARMA* models. This finding gives support to our proposition that application of the AIC or SIC criterion to find one model out of a set of nonfractional *and* fractional alternatives is a strategy not to be pursued. It is advisable to apply the criteria to the set of nonfractional alternatives only to find an AIC optimal and an SIC optimal model and do the same for a set of fractional alternatives only. A choice between the models can be made by studying the forecasting performances of the set of optimal models.[42]

The second stage of the GPH* (1983) *estimation procedure

The GPH estimation of fractional *ARMA* models procedure consists of two stages: at the first stage the long memory parameter d is estimated by a log periodogram regression (equation (3.3)), and at the second stage the short memory autoregressive parameters are estimated. In order to get a clear picture of the meaning of fractional integration in general and the GPH methodology in particular we first discuss the GPH results given the fractional $ARMA(1,-0.45,0)$ model for Δlny (this model is SIC optimal when estimated by maximum likelihood). Thus in this discussion estimation of the

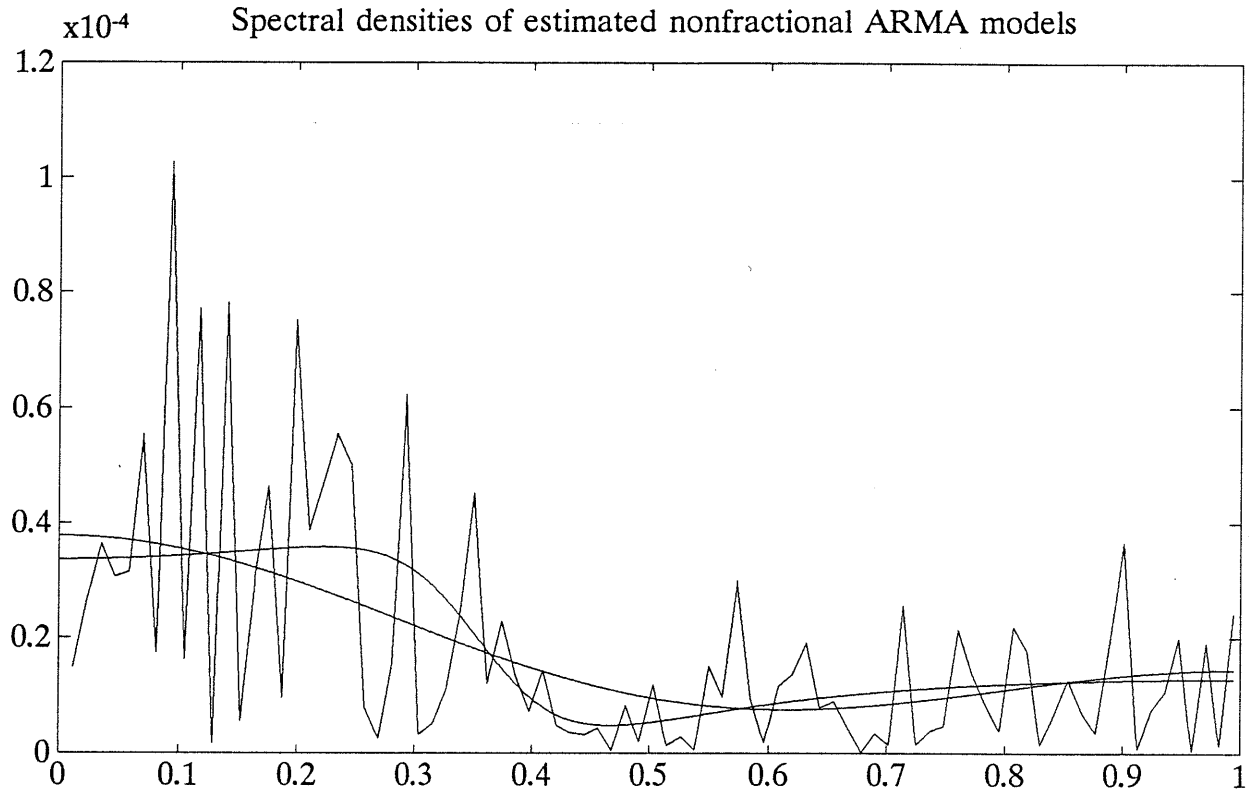
---

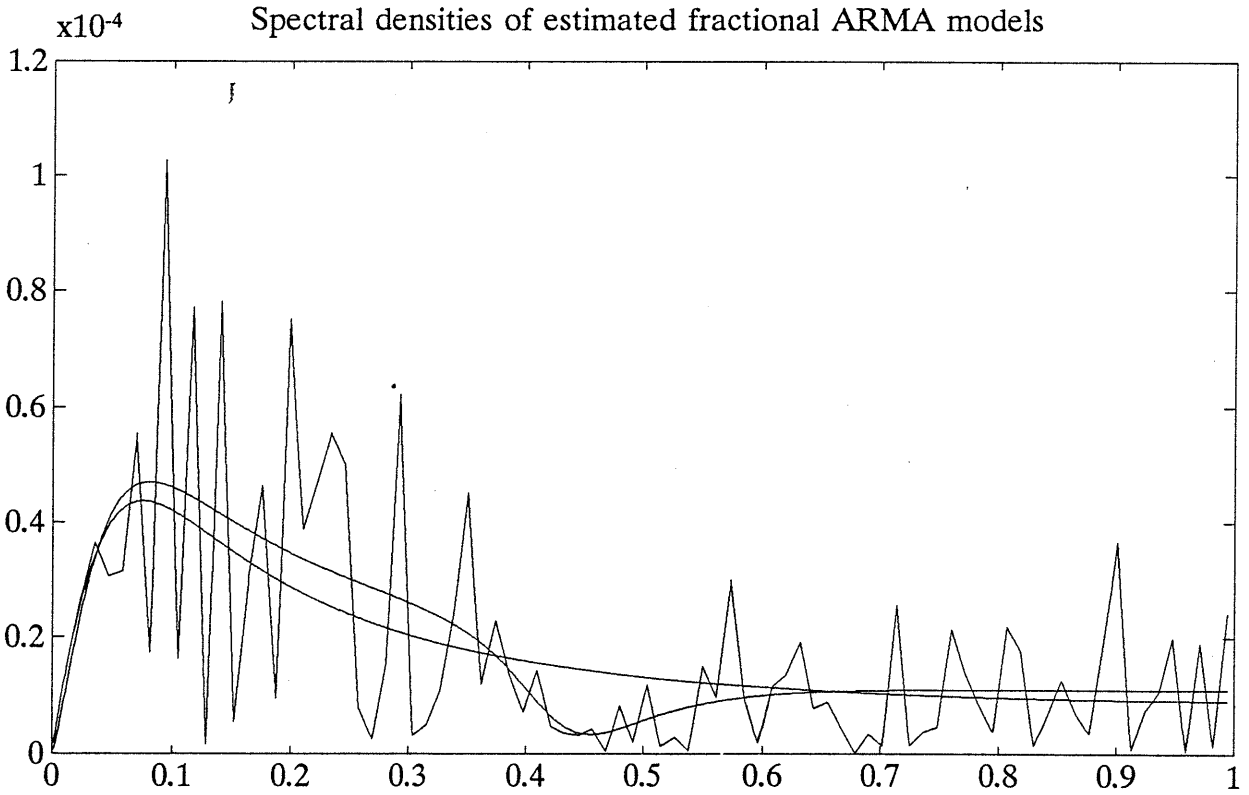
[40] Sowell chooses an optimal model out of nonfractional *and* fractional models according to some criterium (AIC,SIC). In this thesis however we choose an optimal nonfractional model and an optimal fractional model according to some criterium (AIC,SIC).

[41] The shape of the periodogram of Δlny_t at the higher frequencies differs from the one depicted in Sowell (1992a, figures 2, 3 and 11).

[42] Sowell (1992a) remarks the alternative of a nonfractional *ARMA* model at the end of the discussion only.

Figure 4.3


x10-4
Spectral densities of estimated nonfractional ARMA models
1.2
1
0.8
0.6
0.4
0.2
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


x10-4
Spectral densities of estimated fractional ARMA models
1.2
1
0.8
0.6
0.4
0.2
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

parameter of fractional integration at the first stage of the GPH procedure is skipped. The GPH estimation results when performing the first stage log periodogram regression will be discussed hereafter.

Given $d = -0.45$, we can compute the mean-corrected series

$$(1) \qquad u_t = (1-L)^{-0.45} \Delta lny_t$$

by premultiplication of Δlny_t in the frequency domain. In figure 4.4 we plotted the series u_t against Δlny_t where both series run from 1947:II to 1989:IV (171 data points).

The negative sign of the parameter of fractional integration indicates that in order to attain stationarity the series lny_t has been overdifferenced when considering Δlny_t. As $u \sim I(0)$ we observe that lny is already stationary and without long memory when corrected for an integration order of 0.55; it is overdifferenced when corrected for an integration order of 1.00. Notice that $u_t = (1-L)^{0.55} lny_t$.

Yule-Walker estimation of the autoregressive structure of u at the second stage of the GPH procedure gives the following model for the mean-corrected series u_t:

$$(2) \qquad \underset{(-15.70)}{(1 - 0.77L)}\, u_t = \varepsilon_t$$

where the first order autoregressive parameter estimate appears to be equal to the corresponding maximum likelihood estimate (see table 4.4).

Predictions obtained by time domain transforms

In order to construct forecasts of Δlny_t and thereby of lny_t we can perform convolution of the short-memory autoregressive polynomial $(1-0.77L)$ and the long memory autoregressive polynomial (an alternative device will be discussed later). The latter polynomial is approximated by its infinite order autoregressive representation, truncated after lag 50. In figure 4.5 we have illustrated the convolution of these autoregressive polynomials. We have depicted the short and long memory autoregressive coefficients and the coefficients resulting after convolution. Furthermore we have plotted the implied short and long memory estimated spectral densities and the spectral density corresponding to the convoluted polynomial, each estimated density plotted against the periodogram of Δlny_t. It is clear that the shape of the periodogram at the higher frequencies is approximated best by the short memory spectral density whereas the lower frequency behavior is captured best

Figure 4.4

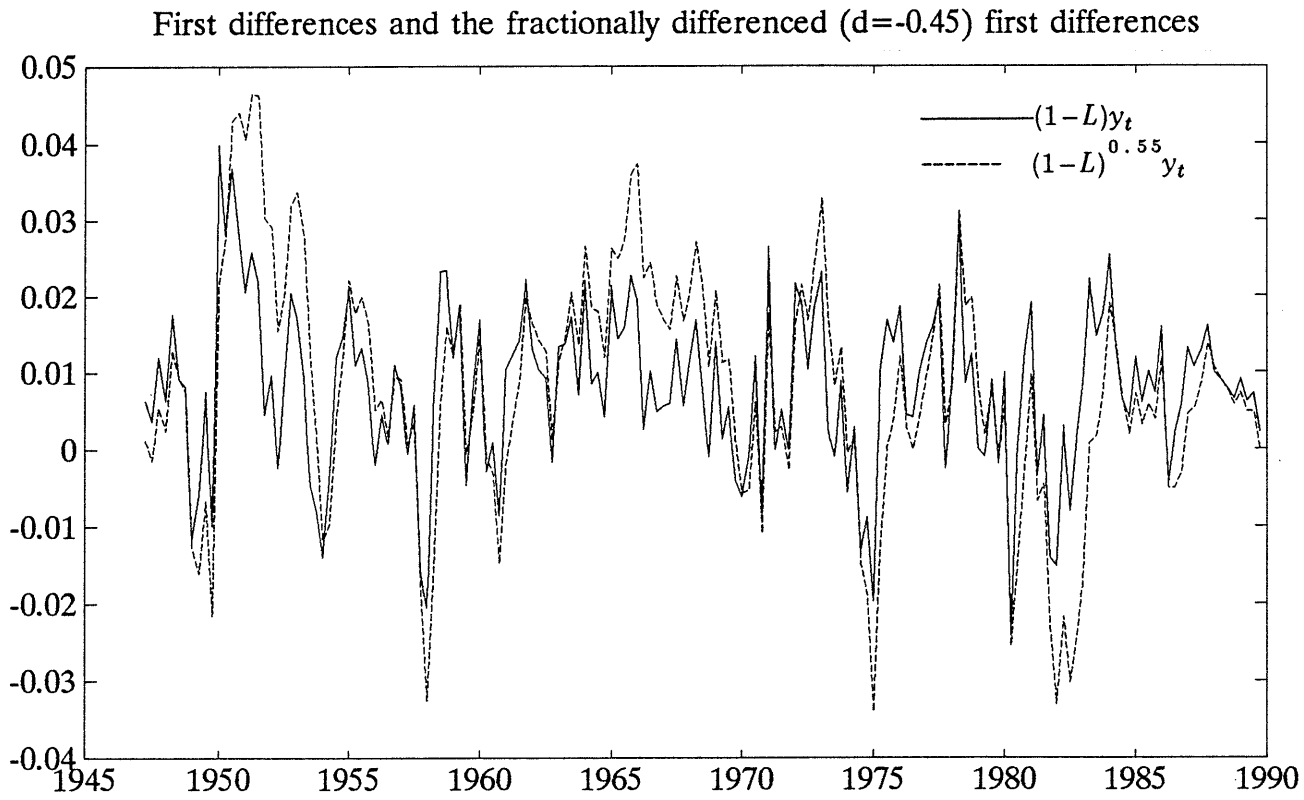
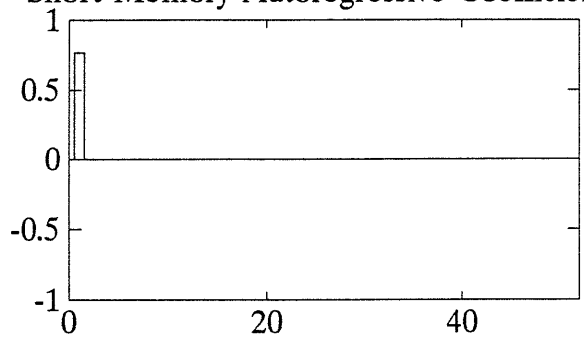

First differences and the fractionally differenced (d=-0.45) first differences
0.05
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
$(1-L)y_t$
$(1-L)^{0.55}y_t$

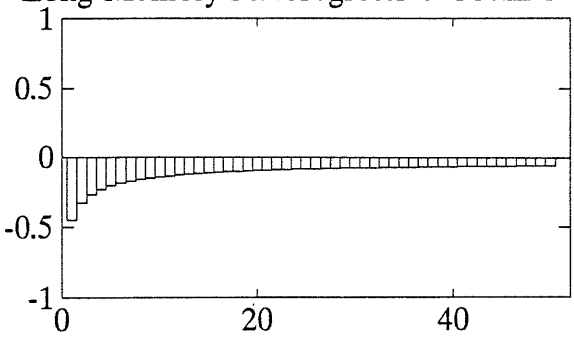
Figure 4.5

Short Memory Autoregressive Coefficients



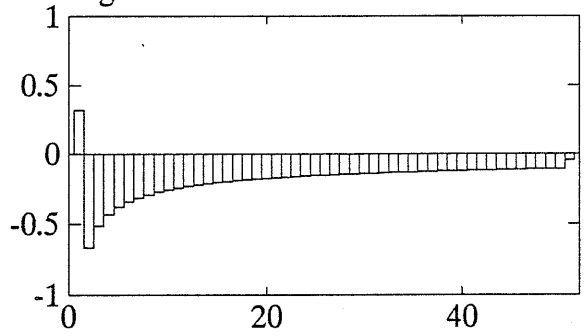
lags 1-51

Long Memory Autoregressive Coefficients



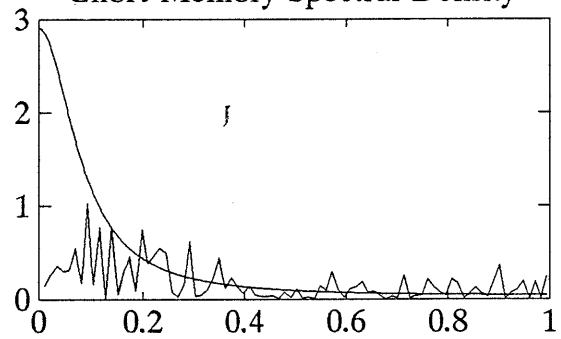
lags 1-51

Autoregressive Coefficients After Convolution



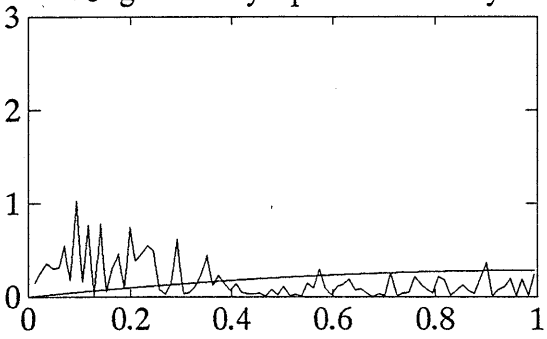
lags 1-51

Short Memory Spectral Density



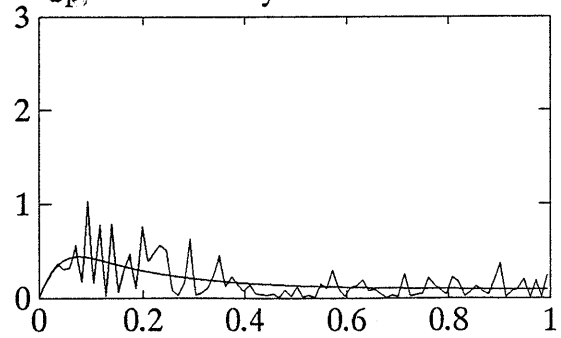
fractions of pi

Long Memory Spectral Density



fractions of pi

Spectral Density After Convolution



fractions of pi

by the long memory spectral density.

In figure 4.6 the one-step ahead predictions $\Delta lny_{t,1}$ are plotted against Δlny_t, given the bases 1959:II up to 1989:III. From the histogram of the one-step ahead prediction errors in figure 4.7 (122 data points) we observe a slight skewness to the left with an extreme prediction error of -0.06 in 1982:III. The square root of the mean squared one-step ahead in-sample prediction errors of 0.0177 can be compared to 0.0097, i.e. the maximum likelihood estimate of the standard deviation of the fractional $ARMA(1,-0.45,0)$ model. Notice however that the former figure is based on 122 data points whereas the latter is based on 172 data points. Compare the results with the predictions obtained by frequency domain transforms.

One-step ahead predictions of lny_t are obtained by integration of the forecasts $\Delta lny_{t,1}$, thereby making use of

$$(3) \qquad lny_{t,1} = lny_t + \Delta lny_{t,1}$$

The one-step ahead in-sample predictions $lny_{t,1}$ are plotted against lny_t in figure 4.8, given the bases 1959:II up to 1989:III. By substitution of the identity $lny_t = lny_{t+1} - \Delta lny_{t+1}$ in (3) it can be derived that

$$\Delta lny_{t+1} - \Delta lny_{t,1} = lny_{t+1} - lny_{t,1}$$

i.e. the one-step ahead prediction errors for the first differences of log GNP equal those of the levels. For the histogram of the one-step ahead prediction errors of lny_t we therefore refer to the one corresponding to Δlny_t. Furthermore, the square root of the mean squared one-step ahead in-sample prediction errors of 0.0177 for lny_t can thus be compared to the maximum likelihood estimate of the standard deviation of the fractional $ARMA(1,-0.45,0)$ model for Δlny_t.

In general h-step ahead predictions of lny_t are obtained by making use of the identity $lny_{t,h} = lny_{t,h-1} + \Delta lny_{t,h}$. Given the basis 1959:II the predictions of log GNP for horizons 1 up to 120 are plotted against the realizations in figure 4.9.

Predictions obtained by frequency domain transforms

An alternative device to generate one-step ahead predictions of Δlny_t given the model in equation (2) is the following. One-step ahead predictions of u_t can be derived from model (2) as

Figure 4.6

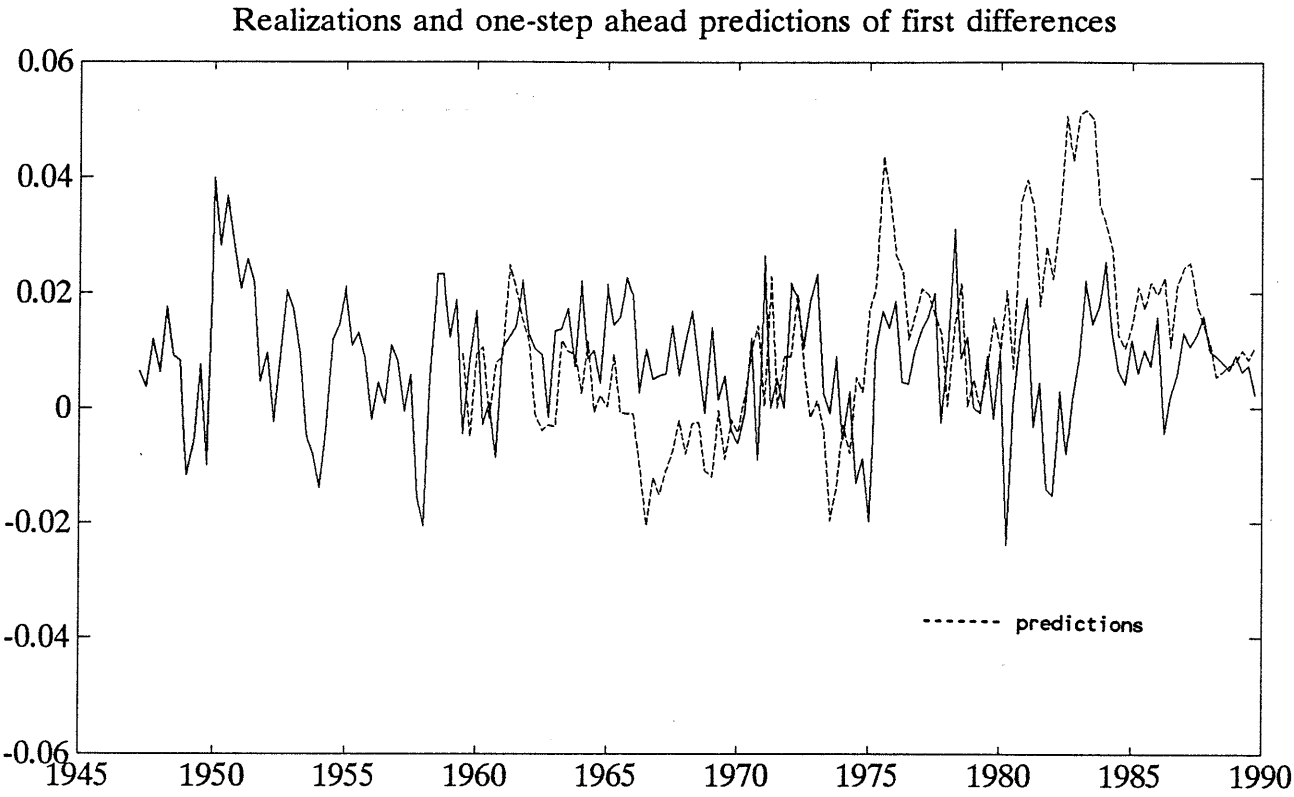

Realizations and one-step ahead predictions of first differences
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
predictions

Figure 4.7

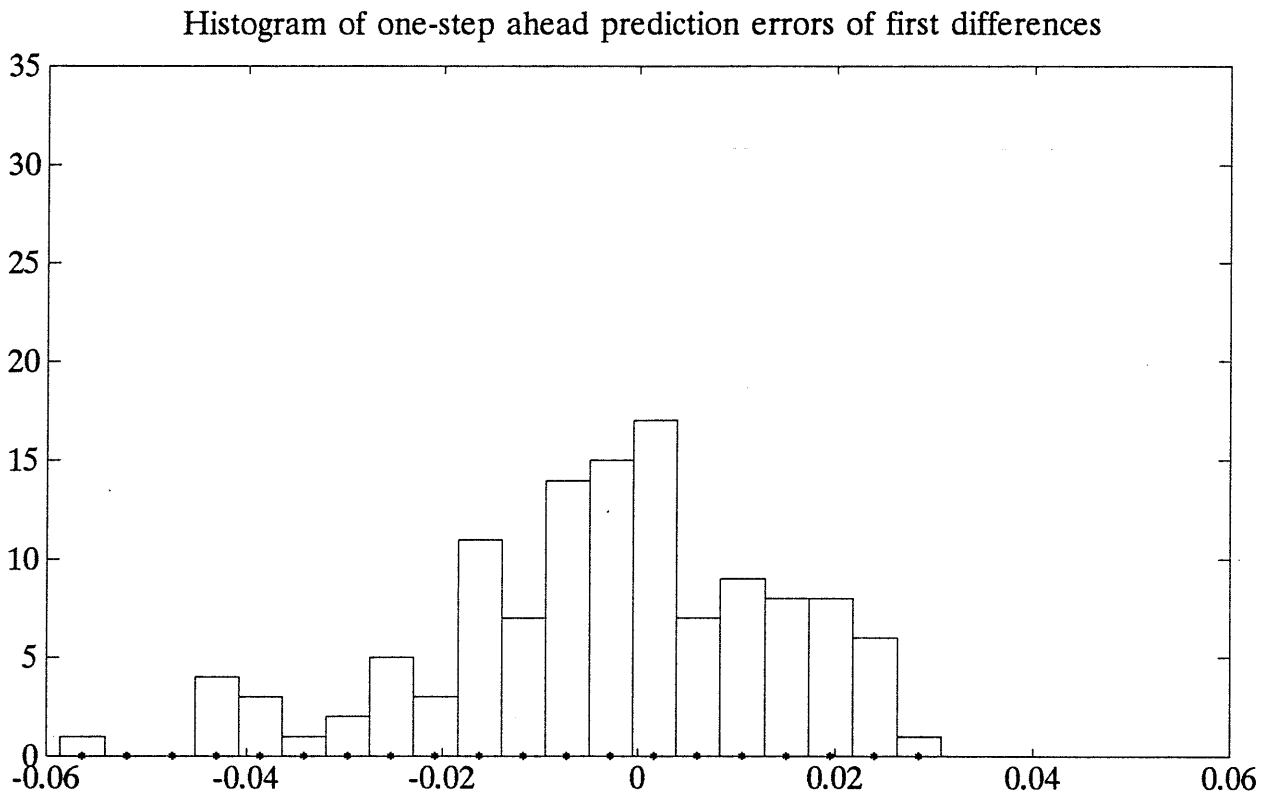

Histogram of one-step ahead prediction errors of first differences
35
30
25
20
15
10
5
0
-0.06
-0.04
-0.02
0
0.02
0.04
0.06

Figure 4.8

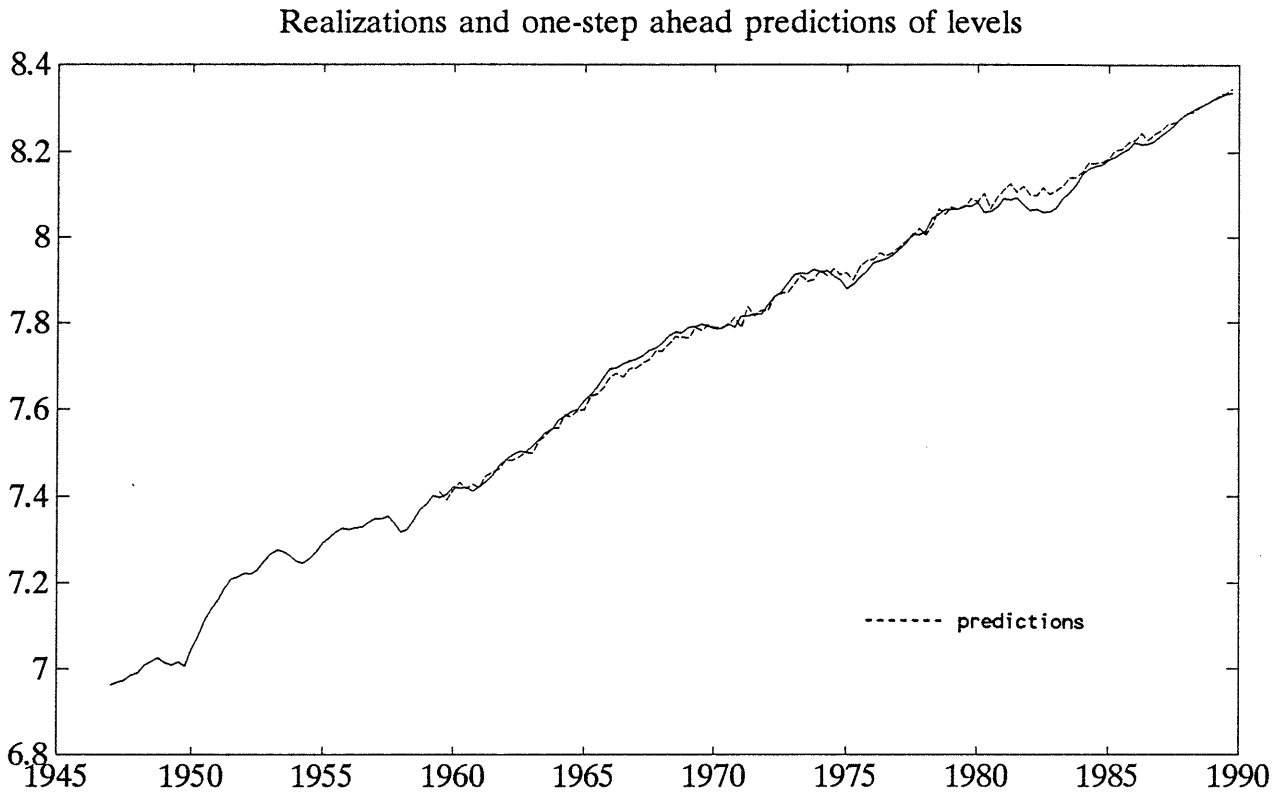
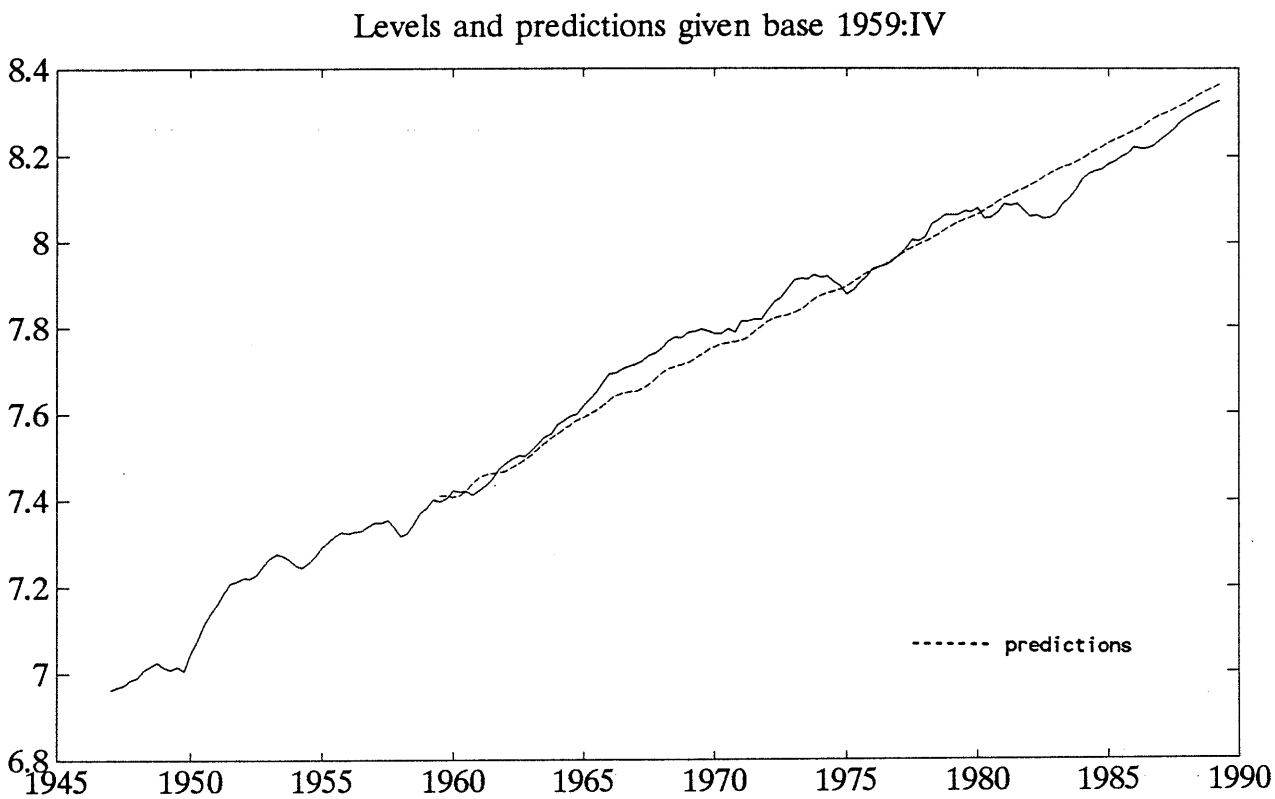

Realizations and one-step ahead predictions of levels
8.4
8.2
8
7.8
7.6
7.4
7.2
7
6.8
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
------- predictions

Figure 4.9


Levels and predictions given base 1959:IV
8.4
8.2
8
7.8
7.6
7.4
7.2
7
6.8
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
------- predictions

(4) $\hat{u}_t = 0.77\ u_{t-1}$

for $t=2,\ldots,171$. Premultiplication of $\hat{u}_t$ by the filter $(1-L)^{0.45}$ in the frequency domain gives one-step ahead predictions of Δlny_t, say $\Delta l\hat{n}y_t$. Notice the difference in the generation of predictions as described before: they were derived after convolution of the short- and long-memory autoregressive polynomials as given in equations (1) and (2). The histogram of the one-step ahead predictions of Δlny_t derived in this way is presented in figure 4.10. The square root of the mean squared in-sample predictions errors based on 170 data-points is 0.0110; based on the period after 1959:II the corresponding figure is 0.0108.

***Fractional AR models estimated according to GPH* (1983)**

At the first stage of the GPH estimation procedure the number of periodogram ordinates n to be used in the log periodogram regression (3.3) is determined by the rule $n=z(T)$ where $z(T)=T^{\alpha}$ as proposed by Geweke and Porter-Hudak (1983) where $0<\alpha\leq 1$ has to be chosen. Although $\alpha=0.55$ or $\alpha=0.6$ is proposed by Geweke and Porter-Hudak (1983), in the literature usually $\alpha=0.5$ is taken (e.g. Diebold and Rudebusch 1989). If the process under consideration is fractional noise Hassler (1993a, 1993b) suggests that the log periodogram regression should be run over the whole range of harmonic frequencies. We estimated the short-memory component of the fractional models for autoregressive orders of one up to 50. According to the SIC criterion an optimal short-memory autoregressive order for the mean-corrected series Δlny_t is determined.

The first differences of log GNP reveal short-memory behavior. The log periodogram regression has been run over the first $n=13$ ($\alpha=0.5$), $n=16$ ($\alpha=0.55$) and $n=21$ ($\alpha=0.6$) nonzero periodogram ordinates.[43] From the first panel of table 4.5 we conclude that the estimates of the parameter of fractional integration for $n=13,16,21$ are not significantly different from zero. This is in sharp contrast with the maximum likelihood results.

---

[43] As application of the rule $n=z(T)$ often leads to a real number n we choose n as the entier of $z(T)$.

Figure 4.10

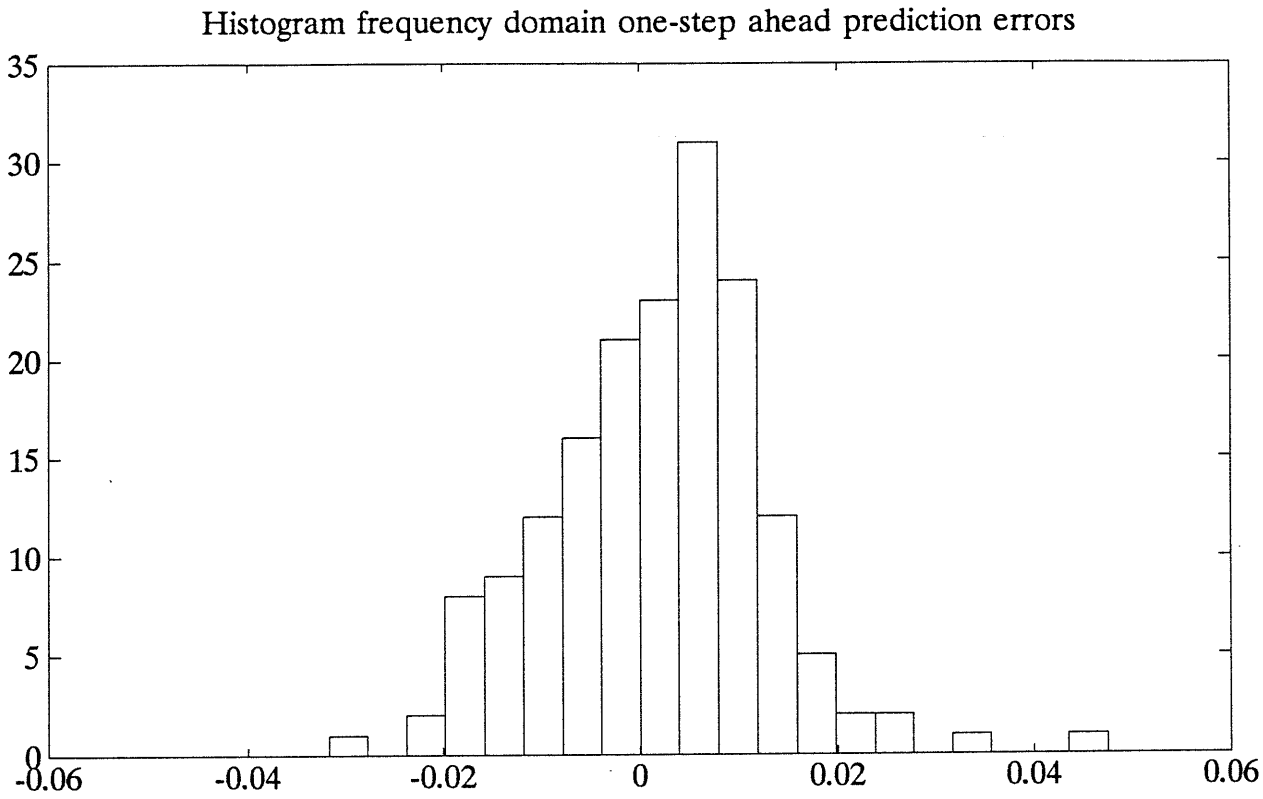

Histogram frequency domain one-step ahead prediction errors
35
30
25
20
15
10
5
0
-0.06
-0.04
-0.02
0
0.02
0.04
0.06

Table 4.5

Parameter estimates of fractional AR(p) models of the first differences of log quarterly real GNP (s.a.); t-statistics between parentheses

periodogram regression				*AR(p)* parameters		
α	n	constant	d	p	ϕ_1	SIC value
0.5	13	-10.79	0.06	1	0.31	9.181
		(-13.81)	(0.25)		(4.29)	
0.55	16	-10.76	0.05	1	0.32	9.181
		(-18.77)	(0.28)		(4.38)	
0.6	21	-10.27	-0.07	1	0.44	9.193
		(-26.56)	(-0.53)		(6.44)	
	6	-9.12	-0.29	1	0.65	9.211
		(-27.50)	(-4.02)		(11.07)	
	7	-9.83	-0.16	1	0.52	9.200
		(-17.64)	(-1.22)		(8.02)	
	8	-9.26	-0.27	1	0.63	9.209
		(-15.61)	(-1.89)		(10.50)	

Note: In the first two columns α and n according to the rule $n=T^{\alpha}$ are presented, where T=171. In the third and fourth column the estimated constant of the log periodogram regression and the estimated parameter of fractional integration are shown. The optimal order of the short memory autoregressive model, the corresponding estimate of the autoregressive parameter(s) and the corresponding SIC value are depicted in the columns five, six and seven, respectively.

The periodogram of Δlny_t increases up to $n=8$ and decreases afterwards. As the increment has to be described by the long-memory part of the fractional *AR* model it is advisable to perform the log periodogram regression for $n=8$. The estimation results are presented in the second panel of table 4.5, together with those for $n=6$ and $n=7$ for comparison. The estimated parameter of integration $d=-0.27$ for $n=8$ is almost significantly different from zero and the SIC value is higher than the ones corresponding to the estimates discussed before. For $n=6$ the estimate $d=-0.29$ is significantly different from zero and corresponds to the overall highest SIC value reported in this table. The model for the mean corrected series Δlny_t reads

$$(5) \qquad \underset{(-11.07)}{(1 - 0.65L)}\ \underset{(-4.02)}{(1 - L)^{-0.29}}\ \Delta lny_t = \varepsilon_t$$

where the estimated variance equals 17.5652×10^{-5}. For $n=7$ the parameter of integration is not significantly different from zero.

The rule $z(T)=T^{\alpha}$, where $\alpha\in\{0.5,0.55,0.6\}$, does lead to a serious bias in the parameter of fractional integration and it is therefore not advisable to apply this rule in case of fractional *AR* modeling of log GNP. In figure 4.11 we have plotted the log periodogram of the first differences of log GNP, the deterministic regressor $ln\{4sin^2(\lambda_{j,T}/2)\}$ and the regressor multiplied by the maximum likelihood estimate $d=-0.45$. It is clear that the latter gives a good description only of the first few low harmonic frequencies. The periodogram at the higher frequencies causes the upward bias in the estimate of the long-run parameter of integration, as these are best described given a positive order of fractional integration.

The bias in d is caused by the presence of strong short memory autoregressive correlation of the first order as can be seen from equation (2). In a Monte Carlo experiment Sowell (1992a) estimates the parameter d by performing the log periodogram regression for the first k nonzero periodogram ordinates where k was varied between 3 and 32. One thousand samples were simulated from the fractional $ARIMA(3,d,2)$ model as estimated by maximum likelihood for the first differences of log GNP. The means of the estimated parameters d all show upward bias from the population value which monotonically increases with the number of ordinates used. Hassler (1993b) derives an analytical representation of the bias in the estimate of the parameter d which depends on the coefficients of the *ARMA* polynomials only. Indeed, for first order autoregressive processes with first order autocorrelation of 0.8 extremely biased estimators are obtained, the bias growing with n.

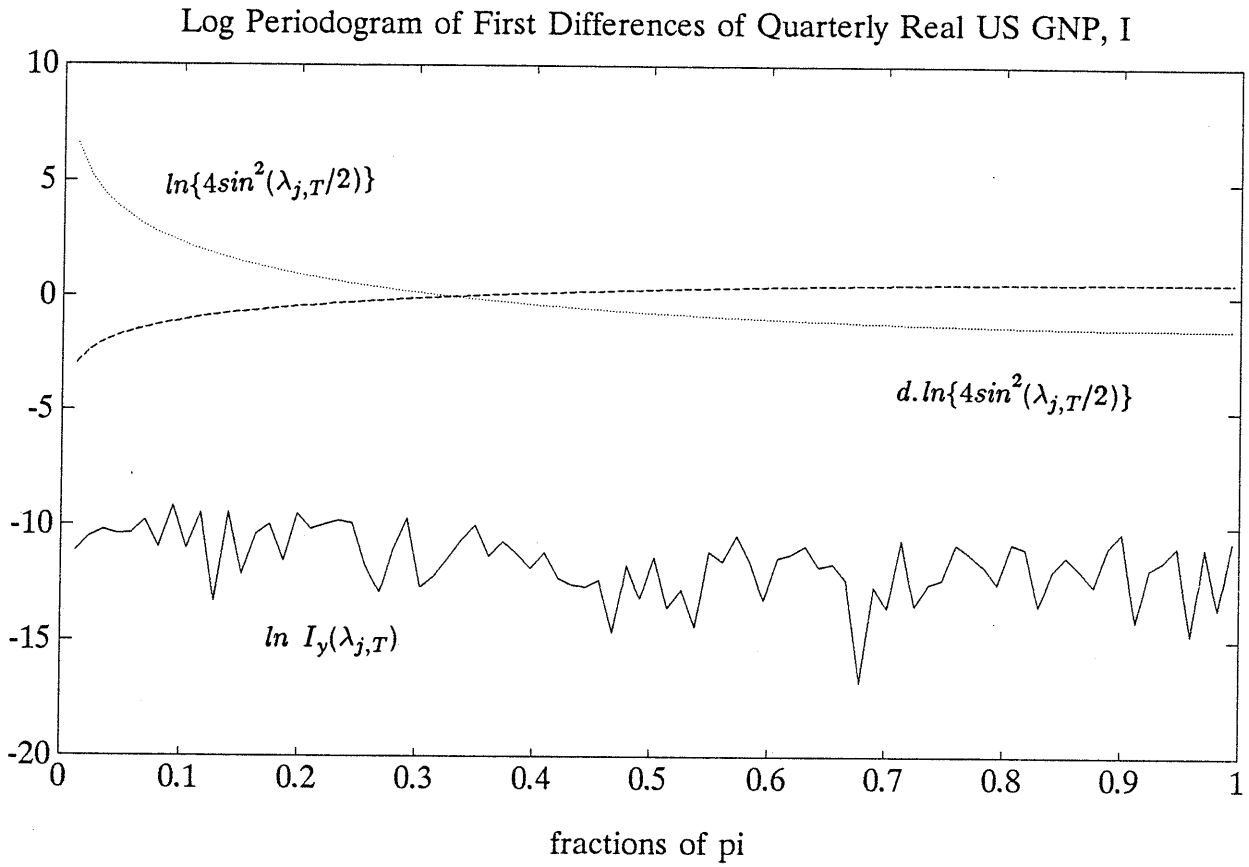
A Yule-Walker nonfractional AR model

To apply the Yule-Walker estimation procedure stationarity is necessary. As we want to estimate a nonfractional *AR* model and given de ADF test results the relevant series should be Δlny_t. We estimated nonfractional *AR* models for an autoregressive order of one up to 50. According to the SIC criterion the following Yule-Walker estimated model for the mean-corrected series Δlny_t is optimal

$$(6) \qquad \underset{(-5.18)}{(1-0.37L)}\,\Delta lny_t = \varepsilon_t$$

where the estimate of the first order autoregressive parameter equals the corresponding maximum likelihood estimate (see table 4.2). The estimated variance equals 10.0018×10^{-5}.

Figure 4.11


Log Periodogram of First Differences of Quarterly Real US GNP, I
10
5
0
-5
-10
-15
-20
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
$ln\{4sin^2(\lambda_{j,T}/2)\}$
$d.ln\{4sin^2(\lambda_{j,T}/2)\}$
$ln\ I_y(\lambda_{j,T})$
fractions of pi

In the discussion of the maximum likelihood estimation of fractional *ARMA* models we noticed that the generation of one-step ahead predictions of Δlny_t given the model (2) could be done in two ways: first convolute the short- and long-memory autoregressive polynomials and then compute the forecasts of Δlny_t, or first generate predictions of the long-memory series u_t and then premultiplicate these predictions by the filter $(1-L)^{0.45}$ in the frequency domain (see equation (4)). It is the latter procedure that makes an interesting comparison possible.

Given model (6) it is possible to derive predictions of Δlny_t as

$$\Delta l\hat{n}y_t = 0.37\ \Delta lny_{t-1}$$

for $t=2,\ldots,171$. If model (6) as well as model (2) give a good description of the underlying process these one-step ahead prediction errors should be similar in distribution as those computed according via the frequency domain. From the histogram in figure 4.12 we observe that this is indeed the case (compare figure 4.10). The square root of the mean squared in-sample predictions errors based on 170 data-points is 0.0112; based on the period after 1959:II the corresponding figure is 0.0106.

An ordinary least squares nonfractional AR model

In order to estimate a nonfractional *AR* model by ordinary least squares stationarity of the series to be modeled is not necessary. According to the SIC criterion the following ordinary least squares estimated model for the lny_t is optimal[44]

$$(7) \qquad \underset{(-18.90)}{(1 - 1.36L} + \underset{(5.06)}{0.36L^2}) lny_t = \underset{(1.43)}{0.02} + \varepsilon_t$$

where the estimated variance equals 10.1026×10^{-5}. The presence of a zero frequency unit root is clearly revealed. Rewriting this model we obtain

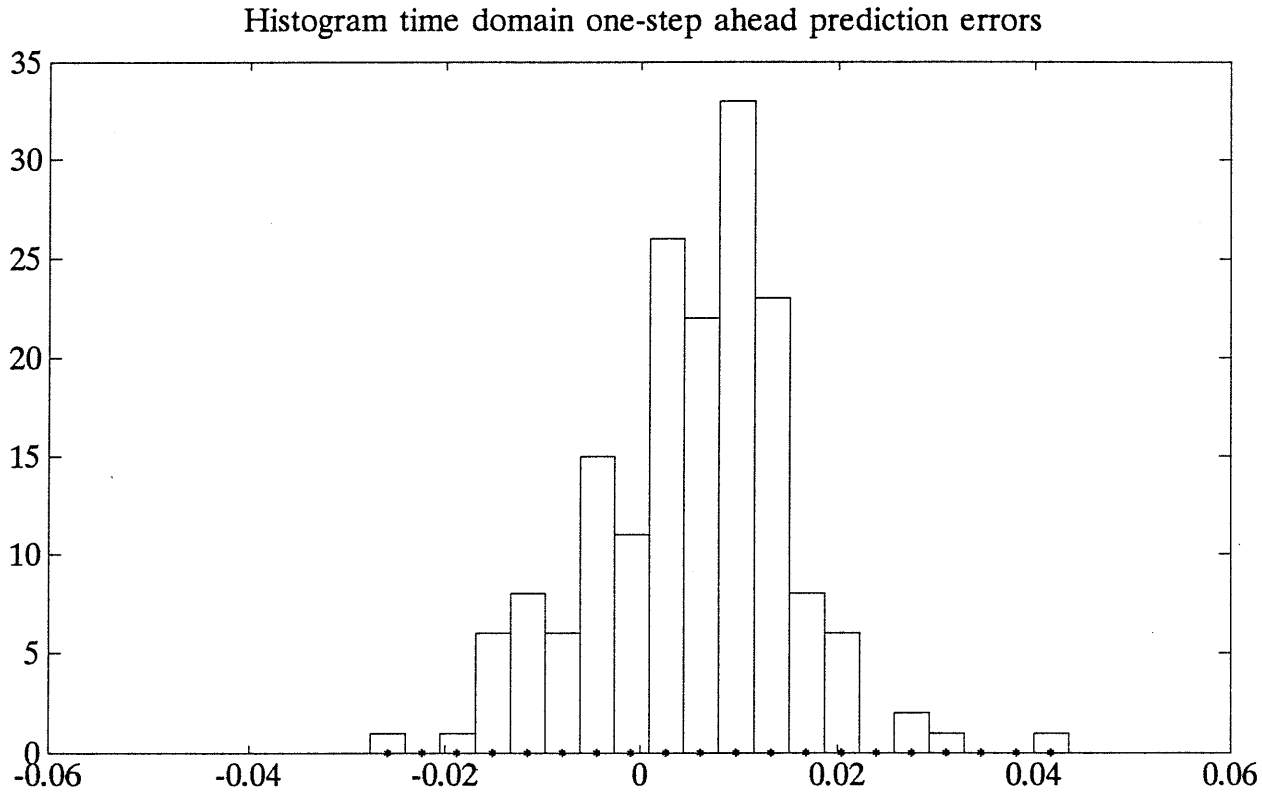
$$(1 - 0.36\ L)\ \Delta lny_t = 0.02 + \varepsilon_t$$

which is very similar to the Yule-Walker estimated model (4) for the mean-corrected series Δlny_t.

---

[44] The series lny_t has not been corrected for its mean.

Figure 4.12


Histogram time domain one-step ahead prediction errors
35
30
25
20
15
10
5
0
-0.06
-0.04
-0.02
0
0.02
0.04
0.06

The choice of an optimal model

From table 4.6 we observe that the nonfractional models are to be preferred in terms of in-sample forecasting performance. Note however that there are only 3 horizon 120 predictions, so that conclusions should be drawn carefully. The ADF test result is in accordance with the optimal model choices.

Table 4.6

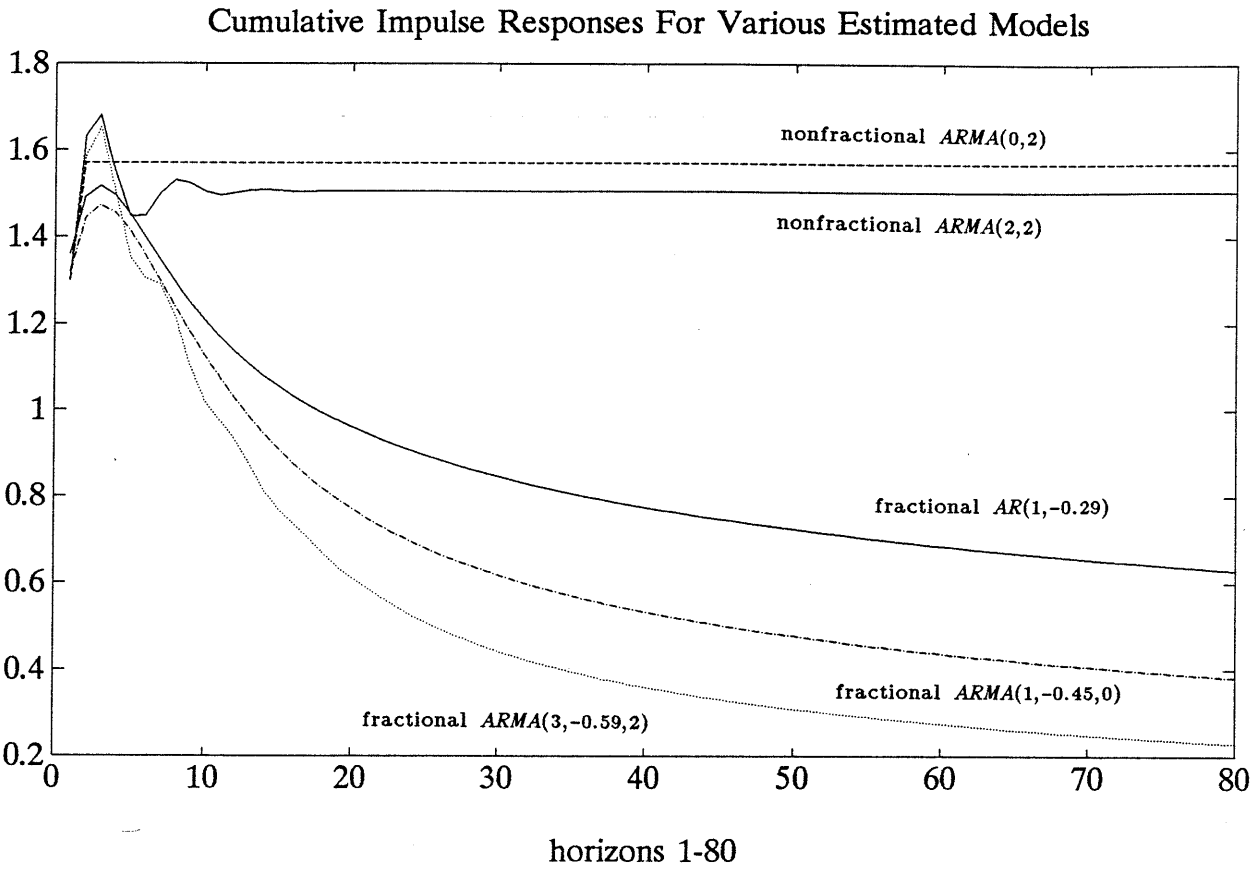
Square roots of mean squared in-sample prediction errors for horizons 1, 5, 20, 40, 60 *and* 120 *corresponding to log quarterly real GNP (s.a.)*

type of model	estimation method	horizons 1	5	20	40	60	120
(models of the levels)							
nonfractional *AR*(2)	OLS	0.009	0.028	0.047	0.043	0.045	0.043
(models of the first differences)							
nonfractional *ARMA*(2,2)	ML	0.012	0.031	0.053	0.064	0.079	0.028
nonfractional *ARMA*(0,2)	ML	0.010	0.029	0.052	0.066	0.081	0.028
fractional *ARMA*(3,-0.59,2)	ML	–	–	–	–	–	–
fractional *ARMA*(1,-0.45,0)	ML	0.018	0.038	0.054	0.066	0.068	0.039
fractional *AR*(1,-0.29)	GPH	0.012	0.035	0.053	0.065	0.071	0.039
nonfractional *AR*(1)	YW	0.009	0.029	0.053	0.069	0.084	0.033

Cumulative impulse responses

From figure 4.13 we see that the cumulative impulse responses are very close for all models. In the longer run the shock persistence corresponding to the the fractional models is much less than that corresponding to the nonfractional models. Given the rather small number of observations available it should be noted that standard errors will be high.

Figure 4.13


Cumulative Impulse Responses For Various Estimated Models
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
10
20
30
40
50
60
70
80
nonfractional ARMA(0,2)
nonfractional ARMA(2,2)
fractional AR(1,-0.29)
fractional ARMA(1,-0.45,0)
fractional ARMA(3,-0.59,2)
horizons 1-80

4.3 ANNUAL TRIER OAK TREE RING WIDTHS

In Lamb (1977, table v21) a series of 1143 data points is tabulated. The data points reflect yearly Trier oak tree ring widths from the west of the Rhine, near Trier and cover the years 822–1964. The data were gathered by felling the oaks and measuring the thickness of the rings, the youngest ring being the one nearest to the bark. Generally in each year there have been between 10 and 36 sample measurements; in the years before 910 there have been only 1–9 measurements and only 4–8 in the period between 1060 and 1129. The unit of measurement is not reported in Lamb (1977) but may well be 0.1mm (see also Steyn 1992).

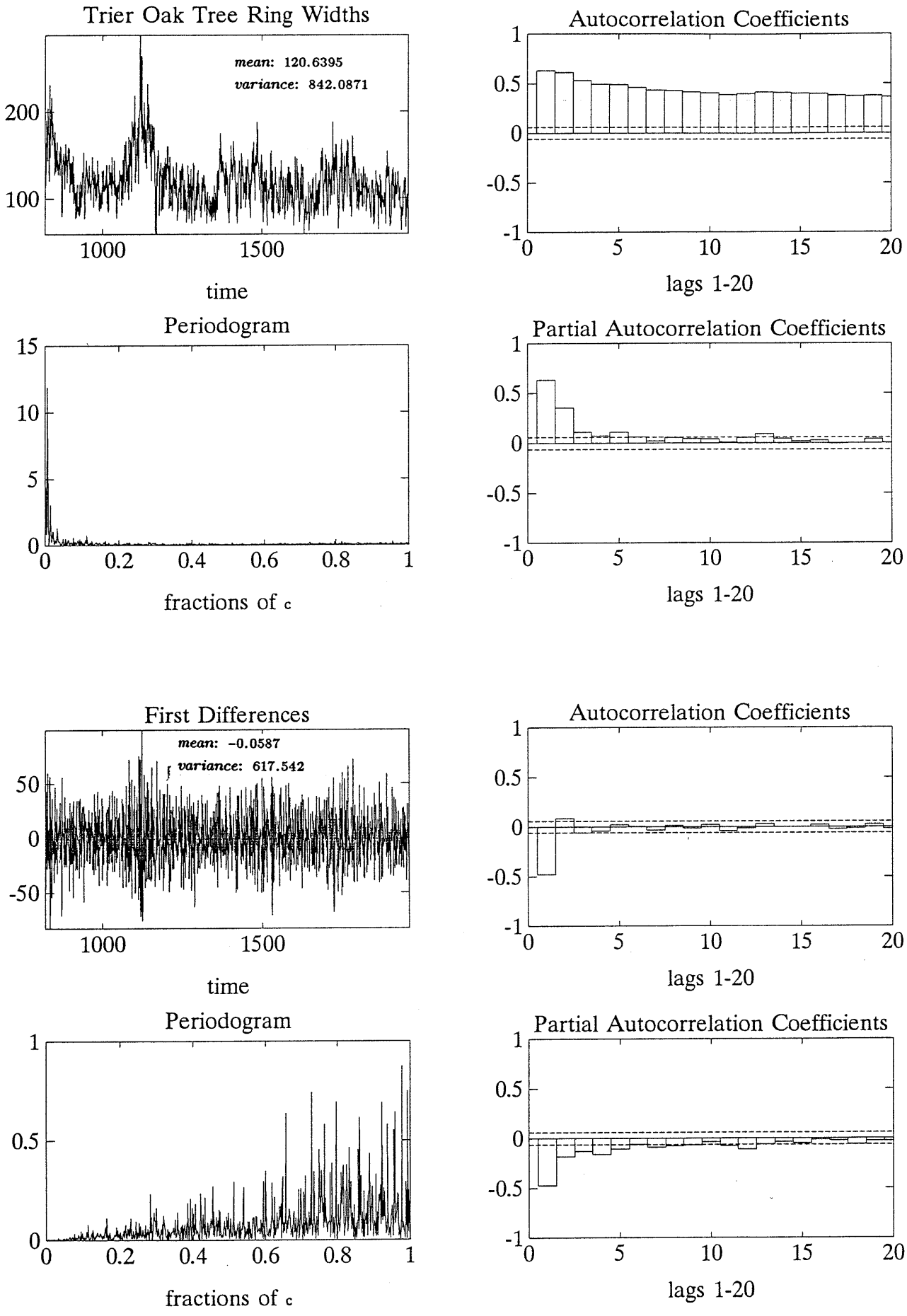
The ring widhts are the subject of dendroclimatology as they provide indirect evidence of weather conditions. The nature of the weather dependence of this series has not been investigated at Lamb's (1977) time of writing. Decades with mild winters and warm summers which also gave plenty of rain seem to have produced the most growth; colder or drier summers or winters probably restricted growth. The ring widths seem to register the effects of more than one year. The Trier oak tree ring widths are plotted in figure 4.14.

As the emphasis in this thesis is on long memory behavior of stochastic processes a series of this length should be considered as very welcome, given economic series which are often of rather short length and are therefore rather inadequate for this purpose. From the periodogram of the Trier oak tree ring widths we observe that long memory behavior is present in this series. The correponding autocorrelation function dies out slowly which is an indication of the presence of a zero frequency unit root as well.

From the figures corresponding to the first differences a surprising phenomenon is manifest. While the levels of the series are primarily dominated by the low frequency behavior, the first differences are primarily dominated by the behavior at the high frequencies. Furthermore, from the estimated autocorrelation function and partial autocorrelation function an appropriate model for the first differences seems to be an $MA(1)$ model with first order autocorrelation of about -0.5. However, the corresponding first order moving-average coefficient then should be about -1, which is an indication of overdifferentiation as well.

Thus the levels of the series are non-stationary whereas taking first

Figure 4.14


Trier Oak Tree Ring Widths
mean: 120.6395
variance: 842.0871
200
100
1000
1500
time
Autocorrelation Coefficients
1
0.5
0
-0.5
-1
0
5
10
15
20
lags 1-20
Periodogram
15
10
5
0
0
0.2
0.4
0.6
0.8
1
fractions of c
Partial Autocorrelation Coefficients
1
0.5
0
-0.5
-1
0
5
10
15
20
lags 1-20
First Differences
mean: -0.0587
variance: 617.542
50
0
-50
1000
1500
time
Autocorrelation Coefficients
1
0.5
0
-0.5
-1
0
5
10
15
20
lags 1-20
Periodogram
1
0.5
0
0
0.2
0.4
0.6
0.8
1
fractions of c
Partial Autocorrelation Coefficients
1
0.5
0
-0.5
-1
0
5
10
15
20
lags 1-20

differences is just too much. This dilemma very strongly indicates the necessity of allowing for fractional integration when modeling this series. Fractional integration seems to be tailor-made for description of the behavior of the Trier oak tree ring widths series.

Augmented Dickey-Fuller tests

The ADF test is performed by estimation of equation (2.2.4) for the Trier oak tree ring widths, denoted y_t, including a constant, a deterministic linear trend of order one and lagged differences of y_t up to order 50 ($p=51$). The five percent critical value is -3.41 (Fuller 1976, table 8.5.2, $\hat{\tau}_\tau$, $n=\infty$). For both y_t and Δy_t the ADF test statistic is significantly different from zero for all lags up to 50: the zero frequency integer unit root hypothesis is rejected against a fractional unit root less than one at a five percent significance level in both cases (see figure 4.15).

Model estimation

Given the results of the ADF tests we model levels of the Trier oak tree ring widths by a nonfractional and a fractional *ARMA* model first. Estimation of these models will be performed by maximum likelihood. A fractional *AR* model for y_t will also be estimated following Geweke and Porter-Hudak (1983). A nonfractional *AR* model for y_t will once be estimated by solving the Yule-Walker equations and once by ordinary least squares.

Nonfractional ARMA models estimated by maximum likelihood

Estimation of nonfractional *ARMA* models for the levels of the Trier oak tree ring widths indicate the presence of a zero frequency autoregressive unit root of order one.[45] Therefore we present in tables 4.7 and 4.8 the results corresponding to the first differences of the Trier series. The optimal model, both according to AIC and SIC, is the nonfractional $ARMA(2,1)$ model

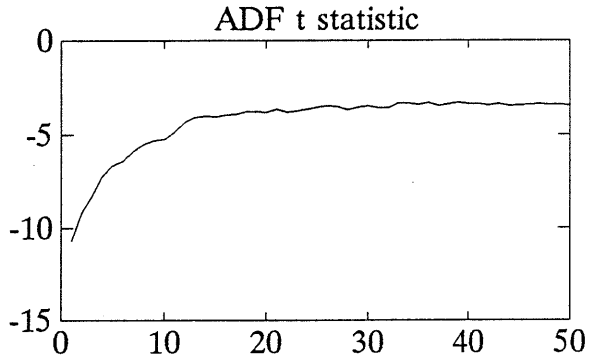
$$\underset{(-7.09)(-6.16)}{(1 - 0.24L - 0.20L^2)}\ \Delta y_t = \underset{(-56.51)}{(1 - 0.92L)}\ \varepsilon_t$$

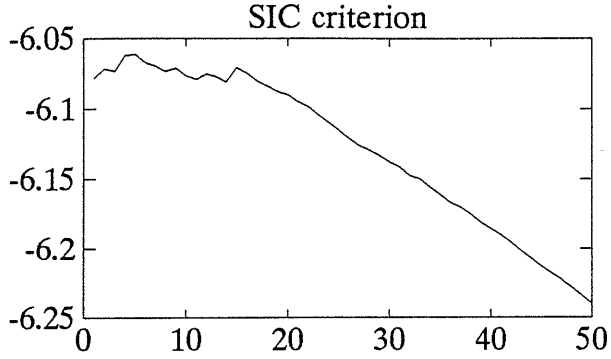
where the corresponding estimated residual variance is 418.0292.

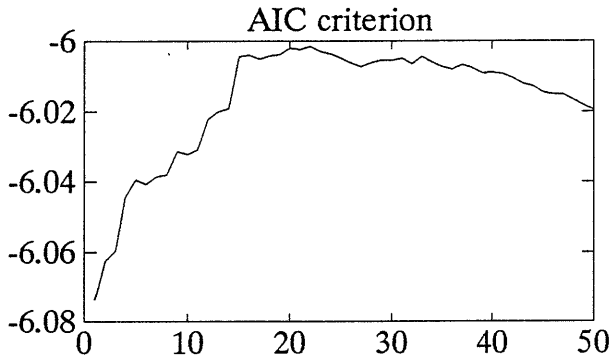
---

[45] These results are not reported here. They can be obtained from the author on request.

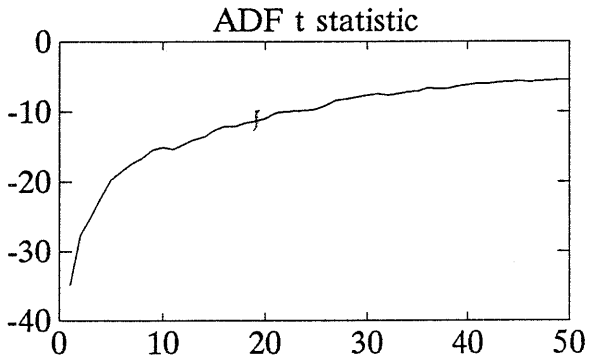
Figure 4.15

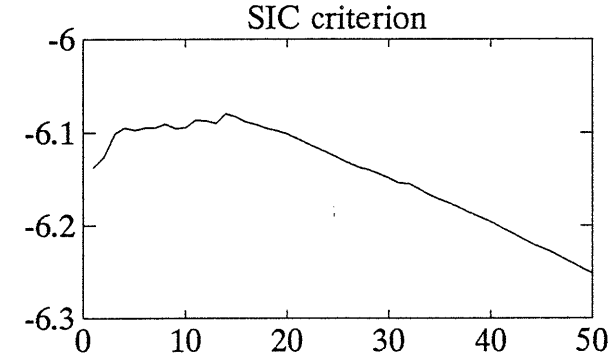


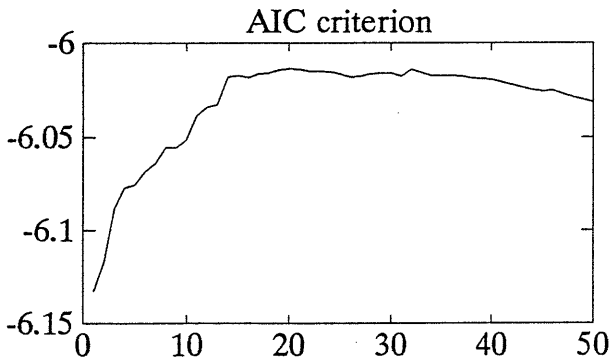




LEVELS







FIRST DIFFERENCES

Table 4.7

2lnL, AIC and SIC for sixteen nonfractional ARMA(p,q) models of the standardized first differences of the Trier oak tree ring widths

Number of *AR* parameters	Number of *MA* parameters			
	0	1	2	3
	-10578.025	-10183.825	-10177.450	-10140.641
0	-10578.025	-10185.825	-10181.450	-10146.641
	-10578.025	-10190.865	-10191.531	-10161.763
	-10285.213	-10170.637	-10145.002	-10134.484
1	-10287.213	-10174.637	-10151.002	-10142.484
	-10292.254	-10184.718	-10166.124	-10162.646
	-10246.630	-10134.488	-10134.481	-10134.079
2	-10250.630	-10140.488	-10142.481	-10144.079
	-10260.711	-10155.610	-10162.643	-10169.281
	-10226.841	-10134.479	-10134.131	-10132.912
3	-10232.841	-10142.479	-10144.131	-10144.912
	-10247.962	-10162.641	-10169.334	-10175.155

Table 4.8

Parameter estimates of sixteen nonfractional ARMA(p,q) models of the standardized first differences of the Trier oak tree ring widths; t-statistics between parentheses

Model	ϕ_1	ϕ_2	ϕ_3	θ_1	θ_2	θ_3
(0,1)				-0.69		
				(-23.43)		
(0,2)				-0.67	-0.08	
				(-23.98)	(-2.56)	
(0,3)				-0.68	0.05	-0.19
				(-23.48)	(1.22)	(-6.25)
(1,0)	-0.48					
	(-18.28)					
(1,1)	0.20			-0.84		
	(4.36)			(-26.93)		
(1,2)	0.72			-1.37	0.41	
	(11.18)			(-18.72)	(6.58)	
(1,3)	0.42			-1.09	0.32	-0.14
	(2.64)			(-6.91)	(3.04)	(-3.36)
(2,0)	-0.56	-0.18				
	(-19.33)	(-6.26)				
(2,1)	0.24	0.20		-0.92		
	(7.09)	(6.16)		(-56.51)		
(2,2)	0.26	0.20		-0.93	0.01	
	(1.74)	(4.08)		(-6.26)	(0.09)	
(2,3)	0.32	0.10		-1.00	0.16	-0.08
	(1.69)	(0.63)		(-5.23)	(0.61)	(-0.66)
(3,0)	-0.59	-0.26	-0.13			
	(-19.98)	(-7.70)	(-4.47)			
(3,1)	0.24	0.20	0.00	-0.92		
	(7.10)	(6.16)	(0.10)	(-52.70)		
(3,2)	-0.35	0.35	0.14	-0.33	-0.55	
	(-0.72)	(2.78)	(1.40)	(-0.67)	(-1.23)	
(3,3)	-0.24	0.10	0.15	-0.43	-0.21	-0.20
	(-0.64)	(0.45)	(1.72)	(-1.15)	(-0.54)	(-1.23)

Fractional ARMA models estimated by maximum likelihood

The tables 4.9 and 4.10 include the results of maximum likelihood estimation of fractional *ARMA* models for the levels of the Trier series, i.e. y_t. According to the AIC criterion a fractional $ARMA(2,0.45,0)$ model is optimal, whereas the SIC criterion results in a fractional $ARMA(1,0.49,0)$ model. From the parameters of integration we conclude that the Trier series has a zero frequency unit root of fractional order about 0.47 which implies stationarity and invertibility. This is in accordance with the results of the ADF test. Once again the ADF test seems to be a test for non-stationarity versus stationarity instead of a unit root test as a zero frequency unit root is present. Integer differencing the series y_t once therefore leads to serious overdifferencing as we have already seen from the periodogram of Δy_t. However, modeling y_t without any correction leaves too much long-memory behavior in this series and will be not justified as well (see estimation of nonfractional *ARMA* models for y_t).

Table 4.9

2lnL, AIC and SIC for sixteen fractional ARMA(p,q) models of the standardized Trier oak tree ring widths

Number of *AR* parameters	Number of *MA* parameters 0	1	2	3
	-10170.996	-10156.470	-10147.821	-10146.455
0	-10172.996	-10160.470	-10153.821	-10154.455
	-10178.037	-10170.553	-10168.945	-10174.621
	-10152.494	-10148.842	-10147.095	-10146.122
1	-10156.494	-10154.842	-10155.095	-10156.122
	-10166.577	-10169.966	-10175.260	-10181.329
	-10147.638	-10147.638	-10144.905	-10143.158
2	-10153.638	-10155.638	-10154.905	-10155.158
	-10168.763	-10175.804	-10180.112	-10185.406
	-10147.638	-10147.355	-10142.892	-10142.827
3	-10155.638	-10157.355	-10154.892	-10156.827
	-10175.804	-10182.562	-10185.140	-10192.117

Table 4.10

Parameter estimates of sixteen fractional ARMA(p,q) models of the standardized Trier oak tree ring widths; t-statistics between parentheses

Model	d	ϕ_1	ϕ_2	ϕ_3	θ_1	θ_2	θ_3
(0,d,0)	0.40						
	(20.15)						
(0,d,1)	0.48				-0.15		
	(23.15)				(-4.27)		
(0,d,2)	0.44				-0.11	0.10	
	(11.67)				(-2.26)	(2.89)	
(0,d,3)	0.46				-0.14	0.10	-0.04
	(12.25)				(-2.78)	(2.92)	(-1.19)
(1,d,0)	0.49	-0.17					
	(28.20)	(-5.08)					
(1,d,1)	0.47	-0.45			0.29		
	(20.15)	(-3.46)			(2.04)		
(1,d,2)	0.45	-0.20			0.08	0.08	
	(12.13)	(-0.90)			(0.37)	(1.53)	
(1,d,3)	0.47	0.26			-0.40	0.13	-0.07
	(13.00)	(0.60)			(-0.91)	(1.99)	(-1.40)
(2,d,0)	0.45	-0.13	0.08				
	(12.39)	(-2.64)	(2.08)				
(2,d,1)	0.45	-0.13	0.08		0.00		
	(11.97)	(-0.45)	(1.33)		(0.01)		
(2,d,2)	0.45	-0.58	-0.38		0.47	0.42	
	(13.55)	(-1.89)	(-2.20)		(1.63)	(2.39)	
(2,d,3)	0.47	-0.27	-0.59		0.12	0.64	-0.10
	(15.60)	(-1.13)	(-3.06)		(0.51)	(3.82)	(-1.79)
(3,d,0)	0.45	-0.13	0.08	-0.00			
	(11.70)	(-2.60)	(1.99)	(-0.01)			
(3,d,1)	0.47	0.48	0.15	-0.06	-0.62		
	(11.83)	(0.91)	(1.87)	(-1.41)	(-1.17)		
(3,d,2)	0.47	-0.41	-0.69	-0.10	0.26	0.73	
	(16.55)	(-1.82)	(-3.61)	(-1.89)	(1.20)	(4.60)	
(3,d,3)	0.46	-0.58	-0.80	-0.20	0.44	0.82	0.10
	(14.69)	(-0.82)	(-1.60)	(-0.50)	(0.61)	(1.96)	(0.25)

The AIC and SIC optimal models are respectively

$$\underset{(2.64)(-2.08)}{(1 + 0.13L - 0.08L^2)}\ \underset{(12.39)}{(1 - L)^{0.45}}\ y_t = \varepsilon_t$$

and

$$\underset{(5.08)}{(1 + 0.17L)}\ \underset{(28.20)}{(1 - L)^{0.49}}\ y_t = \varepsilon_t$$

where the estimates of the residual variance corresponding to the nonfractional *ARMA* optimal models are 418.92 and 420.25, respectively. It should be noted that there is another model with an AIC value slightly less

than that of the fractional $ARMA(2,0.45,0)$ model: the fractional $ARMA(0,0.44,2)$ model

$$\underset{(11.67)}{(1-L)^{0.44}}\, y_t = \underset{(-2.26)}{(1-0.11L} + \underset{(2.89)}{0.10L^2})\, \varepsilon_t$$

where the estimate of the residual variance is 419.10. A choice between the three models will be made on the basis of their forecasting performances.

Table 4.11

2lnL, AIC and SIC for sixteen fractional ARMA(p,q) models of the standardized first differences of the Trier oak tree ring widths

Number of *AR* parameters	Number of *MA* parameters: 0	1	2	3
	-10161.763	-10144.171	-10137.960	-10135.418
0	-10163.763	-10148.171	-10143.960	-10143.418
	-10168.804	-10158.252	-10159.082	-10163.580
	-10139.821	-10137.684	-10136.934	-10132.496
1	-10143.821	-10143.684	-10144.934	-10142.496
	-10153.902	-10158.806	-10165.096	-10167.699
	-10137.309	-10137.280	-10134.820	-10131.936
2	-10143.309	-10145.280	-10144.820	-10143.936
	-10158.430	-10165.442	-10170.022	-10174.179
	-10137.223	-10132.982	-10130.396	-10129.787
3	-10145.223	-10142.982	-10142.396	-10143.787
	-10165.385	-10168.184	-10172.639	-10179.071

For completeness we present the maximum likelihood results for the first differences of the Trier series in case of fractional *ARMA* models as well. From table 4.11 the fractional $ARMA(3,-0.32,2)$ model appears to be AIC optimal whereas the fractional $ARMA(1,-0.48,0)$ model is optimal according to SIC. From table 4.12 the models are respectively

$$\underset{(-9.40)}{(1-1.49L} + \underset{(1.98)}{0.31L^2} + \underset{(6.00)}{0.22L^3})\, \underset{(-2.94)}{(1-L)^{-0.32}}\, \Delta y_t = \underset{(-25.06)}{(1-1.86L} + \underset{(12.58)}{0.88L^2})\, \varepsilon_t$$

and

$$\underset{(4.98)}{(1+0.21L)}\, \underset{(-14.11)}{(1-L)^{-0.48}}\, \Delta y_t = \varepsilon_t$$

where the estimated residual variances are 416.42 and 419.65, respectively. Again there is another model with almost the same AIC value as the optimal

one, i.e. the fractional $ARMA(1,-0.29,3)$ which reads

$$\underset{(-4.07)}{(1-0.57L)}\ \underset{(-1.76)}{(1-L)^{-0.29}}\ \Delta y_t = \underset{(-4.76)\ (2.82)\ (-3.01)}{(1-0.96L+0.25L^2-0.12L^3)}\ \varepsilon_t$$

where the estimated residual variance is 417.15. A choice between the models will be made on the basis of their forecasting performances. It should be noted that the fractional orders of integration indicate that the levels of the Trier series may be non-stationary as the confidence intervals of the fractional $ARMA(1,-0.48,0)$ model and the fractional $ARMA(1,-0.29,3)$ contain $d=0.5$ as a possible estimate of the parameter of integration.

Table 4.12

Parameter estimates of sixteen fractional ARMA(p,q) models of the standardized first differences of the Trier oak tree ring widths; t-statistics between parentheses

Model	d	ϕ_1	ϕ_2	ϕ_3	θ_1	θ_2	θ_3
(0,d,0)	-0.59						
	(-28.92)						
(0,d,1)	-0.47				-0.19		
	(-11.31)				(-3.82)		
(0,d,2)	-0.54				-0.13	0.09	
	(-12.34)				(-2.37)	(2.54)	
(0,d,3)	-0.49				-0.19	0.08	-0.06
	(-7.34)				(-2.59)	(2.16)	(-1.55)
(1,d,0)	-0.48	-0.21					
	(-14.11)	(-4.98)					
(1,d,1)	-0.50	-0.41			0.24		
	(-14.73)	(-3.10)			(1.56)		
(1,d,2)	-0.53	-0.24			0.10	0.06	
	(-11.76)	(-1.06)			(0.47)	(0.99)	
(1,d,3)	-0.29	0.57			-0.96	0.25	-0.12
	(-1.76)	(4.07)			(-4.76)	(2.82)	(-3.01)
(2,d,0)	-0.52	-0.15	0.06				
	(-11.73)	(-2.83)	(1.57)				
(2,d,1)	-0.52	-0.21	0.05		0.05		
	(-11.49)	(-0.64)	(0.74)		(0.17)		
(2,d,2)	-0.54	-0.67	-0.38		0.54	0.39	
	(-13.72)	(-1.69)	(-2.20)		(1.45)	(2.36)	
(2,d,3)	-0.49	-0.21	-0.60		0.03	0.64	-0.14
	(-9.99)	(-0.80)	(-3.23)		(0.10)	(4.10)	(-2.12)
(3,d,0)	-0.52	-0.16	0.06	-0.01			
	(-10.42)	(-2.76)	(1.30)	(-0.29)			
(3,d,1)	-0.24	0.43	0.21	-0.03	-0.86		
	(-1.37)	(3.24)	(5.51)	(-0.66)	(-13.45)		
(3,d,2)	-0.32	1.49	-0.31	-0.22	-1.86	0.88	
	(-2.94)	(9.40)	(-1.98)	(-6.00)	(-25.06)	(12.58)	
(3,d,3)	-0.38	1.44	-0.15	-0.31	-1.74	0.62	0.13
	(-3.37)	(9.00)	(-0.64)	(-2.54)	(-9.97)	(1.86)	(0.80)

***Fractional AR models estimated according to GPH* (1983)**

In table 4.13 we have tabulated the results from the application of the Geweke and Porter-Hudak (1983) procedure to the levels of the Trier series, i.e. y_t. At the first stage of the GPH estimation procedure the number of periodogram ordinates n to be used in the log periodogram regression (3.3) is determined by the rule $n=z(T)$ where $z(T)=T^{\alpha}$ as proposed by Geweke and Porter-Hudak (1983) where $0<\alpha\leq 1$ has to be chosen. In table 4.13 results are presented for $n=33$ ($\alpha=0.5$), $n=48$ ($\alpha=0.55$) and $n=68$ ($\alpha=0.6$). As compared to the parameter of fractional integration as estimated by maximum likelihood before there is an upward bias in the GPH estimator of this parameter d. However, the bias is less than the bias reported for the log GNP series. The reason for this is the much smaller short memory autoregressive component of the Trier series.

According to the SIC criterion an optimal short-memory autoregressive order for the mean-corrected series y_t is determined. The optimal order turned out to be one for $\alpha=0.5$ or $\alpha=0.55$ and two for $\alpha=0.6$. The short-memory behavior of the estimated autoregressive models is not too high. According to the SIC values the fractional $AR(2,0.52)$ model is superior:

$$\underset{(4.35)(-1.13)}{(1+0.18L-0.05L^2)}\ \underset{(5.72)}{(1-L)^{0.52}}\ y_t=\varepsilon_t$$

Table 4.13

Parameter estimates of fractional AR(p) models of the Trier oak tree ring widths; t-statistics between parentheses

periodogram regression				*AR(p)* parameters			
α	n	constant	d	p	ϕ_1	ϕ_1	SIC value
0.5	33	2.62	0.62	1	-0.27		-6.059
		(3.47)	(4.51)		(-9.55)		
0.55	48	2.66	0.61	1	-0.27		-6.058
		(4.94)	(5.56)		(-9.41)		
0.6	68	3.23	0.52	2	-0.18	0.05	-6.055
		(8.33)	(5.72)		(-4.35)	(1.13)	

Note: In the first two columns α and n according to the rule $n=T^{\alpha}$ are presented, where T=1143. In the third and fourth column the estimated constant of the log periodogram regression and the estimated parameter of fractional integration are shown. The optimal order of the short memory autoregressive model, the corresponding estimate of the autoregressive parameter(s) and the corresponding SIC value are depicted in the columns five, six and seven, respectively.

Although not reported here we also performed the GPH procedure on the first differences of the Trier series. The estimated parameters of fractional integration were $d = -0.14$ ($\alpha = 0.5$), $d = -0.28$ ($\alpha = 0.55$) and $d = -0.40$ ($\alpha = 0.6$). The short memory autoregressive models were of orders 15, 15 and 16 respectively.

A Yule-Walker nonfractional AR model

As the ADF tests revealed that the Trier series is stationary we applied the Yule-Walker equations to y_t. The SIC optimal model is

$$\underset{(-4.94)}{(1 - 0.34L} \underset{(-5.12)}{- 0.28L^2} \underset{(-0.96)}{- 0.05L^3} \underset{(-0.34)}{- 0.02L^4} \underset{(-1.81)}{- 0.09L^5} \underset{(-1.32)}{- 0.07L^6)}\ y_t = \varepsilon_t$$

where the estimated residual variance is 425.78. For the first differences an $AR(15)$ model was found to be SIC optimal.

An ordinary least squares nonfractional AR model

As no stationarity is required in order to perform ordinary least squares we we estimated AR models for y_t. The SIC optimal model turned out to be

$$\underset{(-11.39)}{(1 - 0.34L} \underset{(-9.03)}{- 0.28L^2} \underset{(-1.47)}{- 0.05L^3} \underset{(-0.36)}{- 0.01L^4} \underset{(-2.92)}{- 0.09L^5} \underset{(-2.37)}{- 0.07L^6)}\ y_t = \varepsilon_t$$

where the estimated residual variance is 422.96. The coefficients very much resemble those of the Yule-Walker estimated model for y_t. The OLS coefficients are more often significantly different from zero however, at a five percent significance level.

The choice of an optimal model

From table 4.14 we observe that the fractional models are to be preferred over the nonfractional ones. The ADF test rejects integer unit root stationarity against the fractional unit root alternatives of lower order. Stationarity is however not that clear as the parameter of integration is somewhere about 0.5.

Table 4.14

Square roots of mean squared in-sample prediction errors for horizons 1, 5, 20, 40, 60 *and* 120 *corresponding to annual Trier oak tree ring widths*

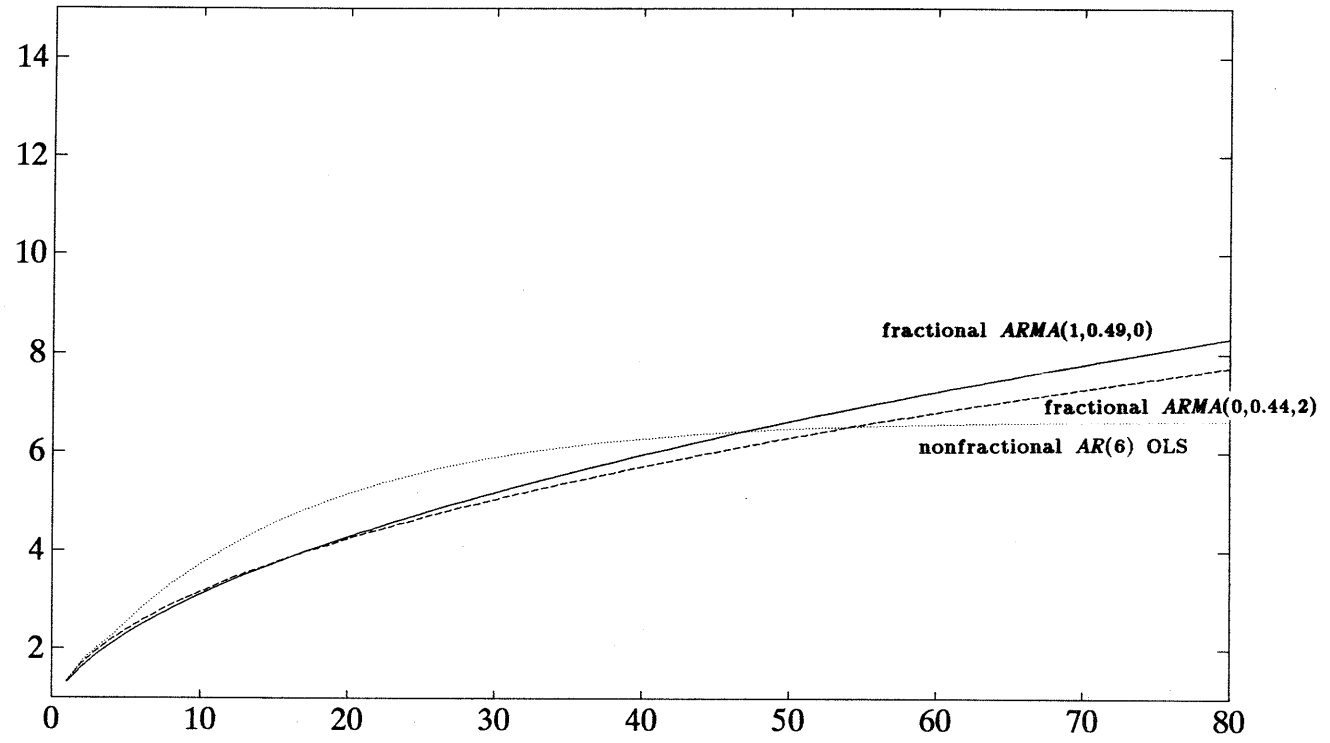
type of model	estimation method	horizons 1	5	20	40	60	120
(models of the levels)							
fractional *ARMA*(2,0.45,0)	ML	20.41	23.52	25.90	27.48	27.91	28.67
fractional *ARMA*(1,0.49,0)	ML	21.45	24.99	27.04	27.86	28.02	28.64
fractional *ARMA*(0,0.44,2)	ML	20.72	23.35	25.38	27.33	27.91	28.85
fractional *AR*(2,0.52)	GPH	20.94	24.44	26.80	27.80	28.01	28.64
nonfractional *AR*(6)	YW	20.35	23.92	26.27	27.79	28.01	28.64
nonfractional *AR*(6)	OLS	20.34	23.91	26.22	27.73	27.96	28.59
(models of the first differences)							
nonfractional *ARMA*(2,1)	ML	32.00	47.20	49.95	52.41	54.17	58.42
fractional *ARMA*(3,-0.32,2)	ML	–	–	–	–	–	–
fractional *ARMA*(1,-0.48,0)	ML	20.67	24.13	27.07	30.59	32.27	36.99
fractional *ARMA*(1,-0.29,3)	ML	33.21	23.08	26.64	30.33	32.35	37.23

Cumulative impulse responses

From figure 4.17 we observe that the cumulative impulse responses do not differ for horizons up to 50. Figure 4.17 shows horizon 120 predictions for the fractional $ARMA(1,0.49,0)$ model: obviously there is mean reversion. Finally, figure 4.18 showes the estimated residuals of the latter model, the autocorrelation coefficients, the periodogram and the partial autocorrelation coefficients. From the log periodograms plotted below we observe that the one corresponding to the fractional differences is more evenly distributed over the frequencies.

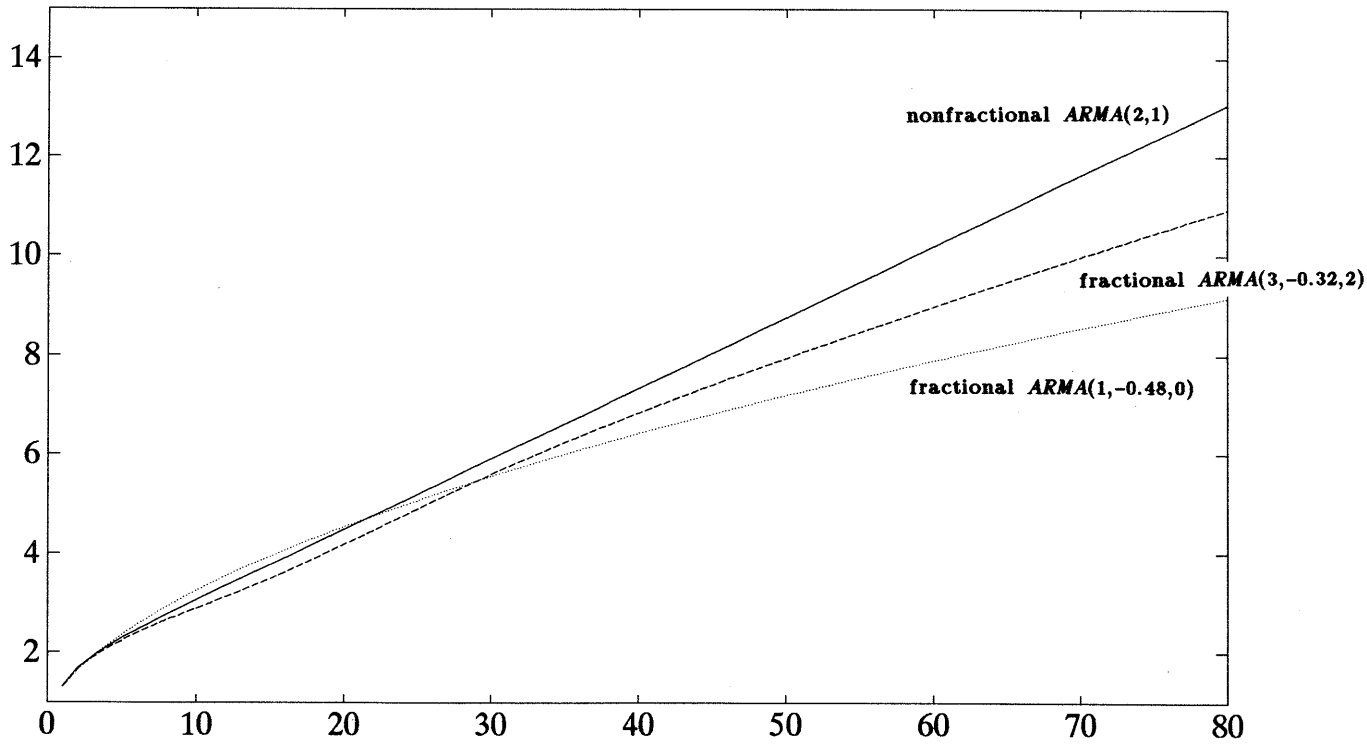
Figure 4.16

Cumulative Impulse Responses - Estimated Models for Levels



horizons 1-80

Cumulative Impulse Responses - Estimated Models for First Differences



horizons 1-80

Figure 4.17

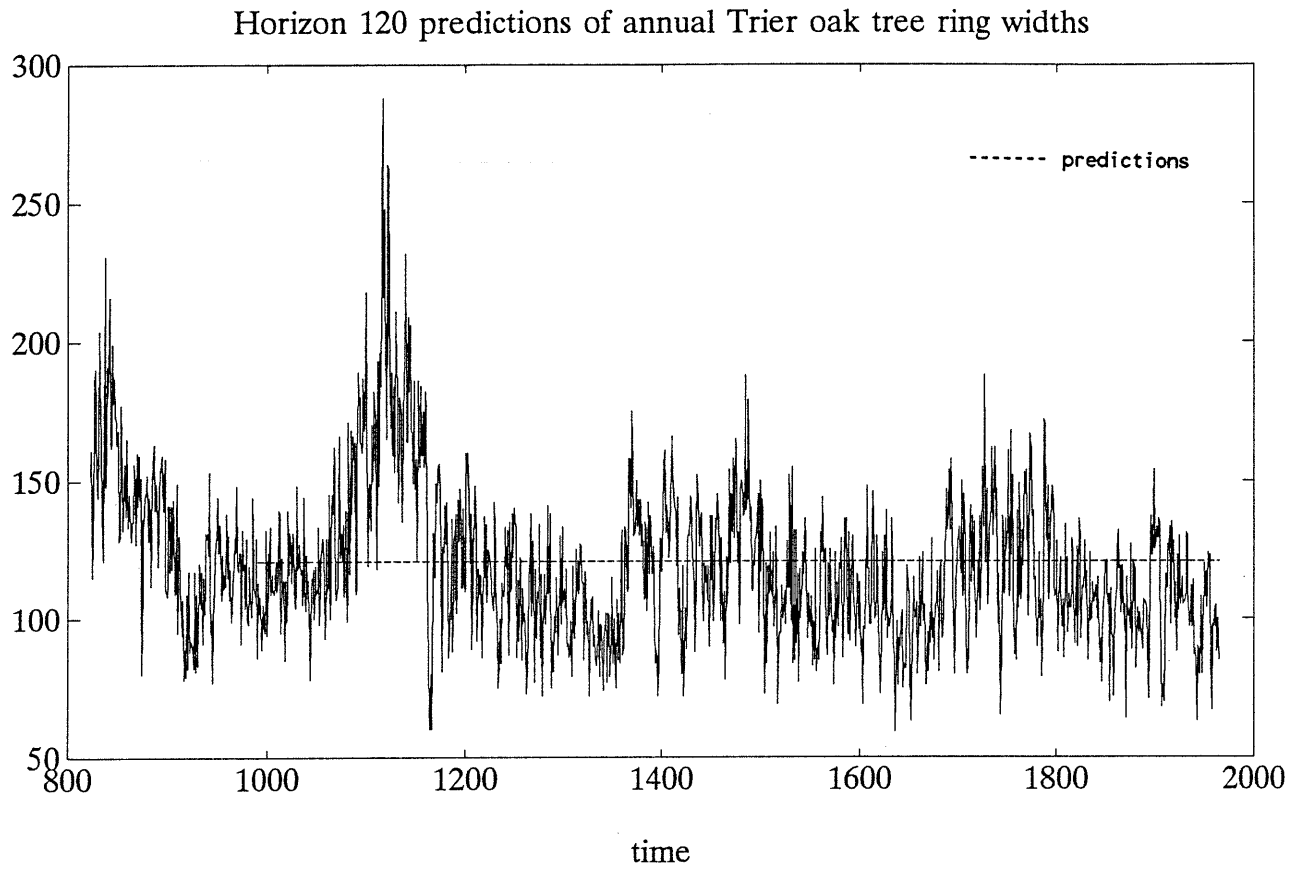
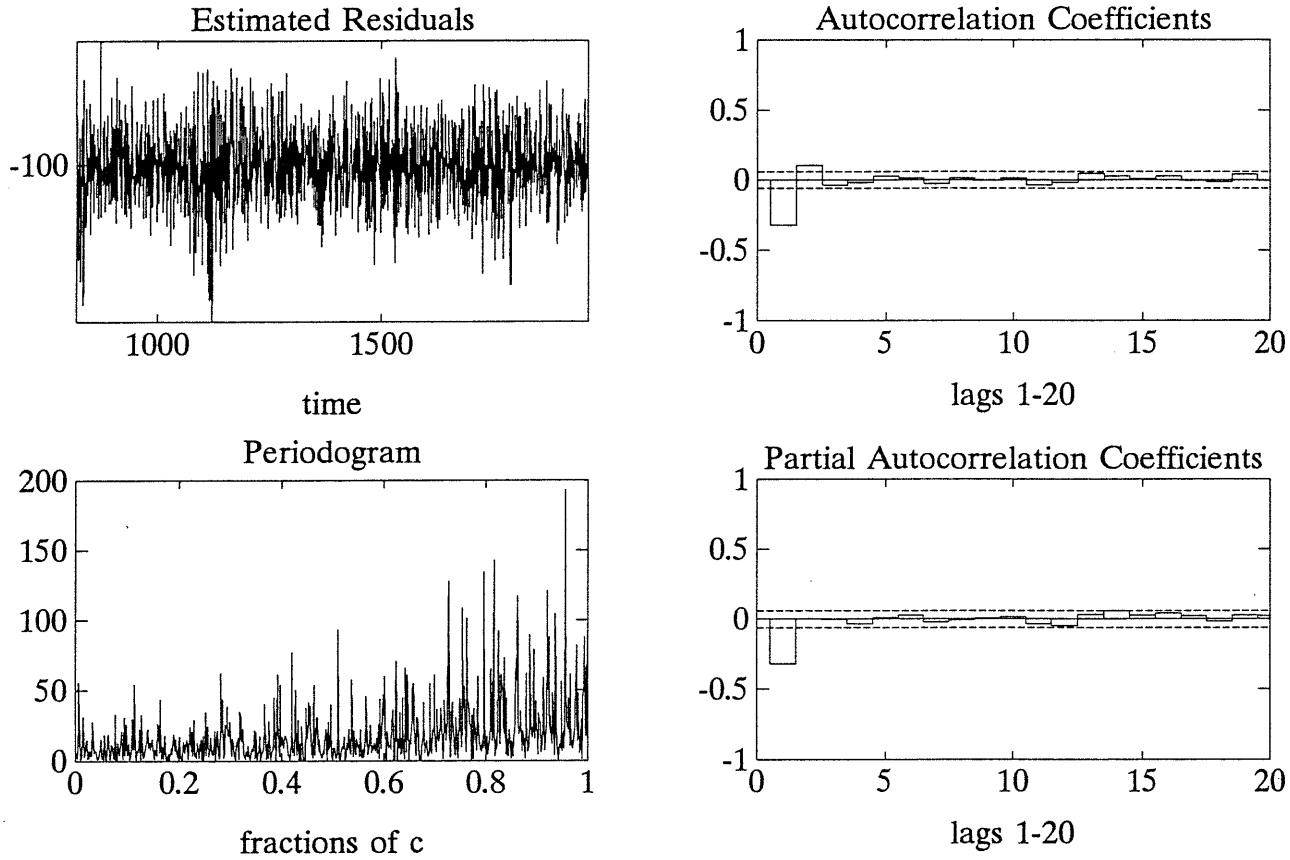

Horizon 120 predictions of annual Trier oak tree ring widths
300
250
200
150
100
50
800
1000
1200
1400
1600
1800
2000
------- predictions
time

Figure 4.18


Estimated Residuals
-100
1000
1500
time
Autocorrelation Coefficients
1
0.5
0
-0.5
-1
0
5
10
15
20
lags 1-20
Periodogram
200
150
100
50
0
0
0.2
0.4
0.6
0.8
1
fractions of c
Partial Autocorrelation Coefficients
1
0.5
0
-0.5
-1
0
5
10
15
20
lags 1-20
Log Periodogram Levels
10
5
0
-5
-10
0
0.2
0.4
0.6
0.8
1
fractions of c
Log Periodogram Fractional Differences
10
5
0
-5
-10
0
0.2
0.4
0.6
0.8
1
fractions of c
Log Periodogram First Differences
10
5
0
-5
0
0.2
0.4
0.6
0.8
1
fractions of c

5. CONCLUSIONS

In this thesis the emphasis was on long memory stochastic processes. Traditional *ARIMA* models were generalized to allow for zero frequency fractional integration instead of zero frequency integer integration only. The fractional parameter describes the long memory behavior of a stochastic process, thereby modeling low frequency behavior more flexible than when trying to model this behavior by e.g. an autoregressive filter. The autoregressive and moving-average components can be applied to describe short-term behavior.

From the simulation experiments we concluded that the estimator of the population mean of fractionally integrated series can be seriously biased. In Geweke and Porter-Hudak (1983) the population mean has been used in the simulation study on the autocorrelation function of such series. Our experiments show that computation of this function given empirical mean leads to underestimation at all autocorrelation lags. One can partially circumvent this problem by modeling first differences.

Whereas (augmented) Dickey-Fuller tests are not developed for fractionally integrated processes, zero frequency unit root tests based on estimation of the parameter of fractional integration can be used in case of fracional as well as integer integrated processes. In the second example Dickey-Fuller rejects the integer unit root against lower order fractional unit root.

Estimation of the fractional parameter of integration is performed by the two-stage semi-parametric Geweke and Porter-Hudak (1983) procedure. Although estimation of this parameter by frequency domain log periodogram regression is independent from the mean of the series, the influence of short-run *ARMA* parameters can lead to serious bias in the estimator. This bias showed up when analyzing quarterly real US GNP (s.a.). Simultaneous estimation of the long and short memory parameters according to Sowell (1992a,b) performs better in this case.

From the frequency domain periodograms it appeared that imposing a zero-frequency unit root by taking first differences of realizations of some process can lead to overdifferentiation. With respect to long run forecasting taking first differences is preferred over the fractional alternatives in

case of quarterly real US GNP (s.a.). Overdifferentiation is present when first differences of Trier oak tree ring widths are taken. However, there is some zero frequency unit root long memory of fractional order. As the fractional alternatives are preferred, mean reversion under the fractional alternatives is present and the series is estimated as stationary although non-stationarity cannot be excluded.

The oak tree data give a very clear illustration of the applicability of time series modeling in empirical work.

REFERENCES

Blough, S. R. (1992), "The Relationship Between Power and Level for Generic Unit Root Tests in Finite Samples", *Journal of Applied Econometrics*, 7, 295-308.

Box, G. E. P., and Jenkins, G. M. (1970), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

———— (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, second edition.

Brigham, E. O. (1974), *The Fast Fourier Transform*, Englewood Cliffs, N.J: Prentice-Hall.

Campbell, J. Y. and Mankiw, N. G. (1987), "Are Output Fluctuations Transitory?", *Quarterly Journal of Economics*, 102, 857-880.

Clark, P. K. (1987), "The Cyclical Component of U.S. Economic Activity", *Quarterly Journal of Economics*, 102, 798-814.

Cochrane, J. H. (1988), "How Big is the Random Walk in GNP?", *Journal of Political Economy*, 96, 893-919.

———— (1991), "A Critique of the Application of Unit Root Tests", *Journal of Economic Dynamics and Control*, 15, 275-284.

DeJong, D. N., Nankervis, J. C., Savin, N. E., and Whiteman, C. H., "Integration Versus Trend Stationarity in Time Series", *Econometrica*, 60, 2, 423-433.

Dickey, D. A. (1976), *Estimation and Testing on Nonstationary Time Series*, Ph.D. Thesis, Iowa State University.

Dickey, D. A., Bell, W. R., and Miller, R. B. (1986), "Unit Roots in Time Series Models: Tests and Implications", *The American Statistician*, 40, 1, 12-26.

Dickey, D. A., and Fuller, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series With a Unit Root", *Journal of the American Statistical Association*, 74, 366.

———— (1981), "Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root", *Econometrica*, 49, 4.

Diebold, F. X., and Nerlove, M. (1990), "Unit Roots in Economic Time Series: A Selective Survey", *Finance and Economic Discussion Series*, Division of Research and Statistics, Federal Reserve Board, Washington, D.C.

Diebold, F. X., and Rudebusch (1989), "Long Memory and Persistence in Aggregate Output", *Journal of Monetary Economics*, 24, 189-209.

———— (1991), "On the Power of Dickey-Fuller Tests Against Fractional Alternatives", *Economics Letters*, 35, 155-160.

Durbin, J. (1960), *see the references in Jonas* (1983).

Engle, R. F., and Granger, C. W. J. (1987), "Co-Integration and Error Correction: Representation, Estimation, and Testing", *Econometrica*, 55, 251-276.

Evans, G. B. A., and Savin, N. E. (1981), "Testing for Unit Roots: I", *Econometrica*, 49, 3.

Fox, R., and Taqqu, M. S. (1986), "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time-Series", *Annals of Statistics*, 14, 517-532.

Franses, P. H. (1990), "Testing for Seasonal Unit Roots in Monthly Data", *Report Econometric Institute*, 9032/A.

——— (1991), *Model Selection and Seasonality in Time Series*, Ph.D. Thesis, Tinbergen Institute Research Series, Erasmus University Rotterdam, Rotterdam.

Fuller, W. A. (1976), *Introduction to Statistical Time Series*, John Wiley & Sons, New York.

Geweke, J. (1992), *Personal Communication*, August 25^{th}.

Geweke, J., and Porter-Hudak, S. (1983), "The Estimation and Application of Long-Memory Time Series Models", *Journal of Time Series Analysis*, 4, 221-237.

Gourieroux, C., and Monfort, A. (1990), *Séries Temporelles et Modèles Dynamiques*, Economica, Paris.

Granger, C. W. J. (1966), "The Typical Spectral Shape of an Economic Variable", *Econometrica*, 34, 150-161.

Granger, C. W. J., and Joyeux, R. (1980), "An Introduction to Long-Memory Time Series Models and Fractional Differencing", *Journal of Time Series Analysis*, 1, 1, 15-29.

Granger, C. W. J., and Newbold, P. (1986), *Forecasting Economic Time Series*, Academic Press, Harcourt Brace Jovanovich, New York, NY, USA, second edition.

Griliches, Z., and Intriligator, M. D. (1983), *Handbook of Econometrics*, North-Holland, Amsterdam, Volumes I, II and III.

Hall, A. (1988), "Intertemporal Substitution in Consumption", *Journal of Political Economy*, 96, 339-357.

Harvey, A. C. (1981), *Time Series Models*, Allan, Oxford.

——— (1989), *Forecasting, Structural Time Series and the Kalman Filter*, Cambridge University Press, Cambridge.

Hassler, U. (1993a), "Regression of Spectral Estimators With Fractionally Integrated Time Series", *Journal of Time Series Analysis*, 14, 4, 369-80.

——— (1993b), "Unit Root Tests: the Autoregressive Approach in Comparison with the Periodogram Regression", *Statistical Papers*, 34, 67-82.

Hosking, J. R. M. (1981), "Fractional Differencing", *Biometrika*, 68, 1, 165-76.

Jonas, A. B. (1981), "Long-Memory Self-Similar Time Series Models", *unpublished manuscript*, Department of Statistics, Harvard University.

——— (1983), *Persistent Memory Random Processes*, Ph.D. thesis, Harvard University, Cambridge, Massachusetts.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H., and Lee, T.-C. (1982), *Introduction to The Theory and Practice of Econometrics*, John Wiley and Sons, New York.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., and Lee, T.-C. (1985), *The Theory and Practice of Econometrics*, John Wiley and Sons, New York, second edition.

Kailath, T. (1980), *Linear Systems*, Englewood Cliffs, N.J: Prentice-Hall.

Künsch, H (1986), "Discrimination Between Monotonic Trends and Long-Range Dependence", *Journal of Applied Probability*, 23, 1025–1030.

Lamb, H. H. (1977), *Climate: Present, Past and Future, vol.2 – Climatic History and the Future*, Methuen & Co.: London.

Levinson, N. (1947), *see the references in Jonas* (1983).

McLeod, A. I. (1975), "Derivation of the Theoretical Autocovariance Function of Autoregressive-Moving Average Time Series", 24, 2, pp. 255–256.

McLeod, A. I., and Hipel, K. W. (1978), "Preservation of the Rescaled Adjusted Range. 1. A Reassessment of the Hurst Phenomenon", *Water Resources Research*, 14, 3, 491–508.

Mills, T. C. (1991), "Nonlinear Time Series in Economics", *Journal of Economic Surveys*, 5, 3, pp. 215–242.

Nelson, C. R., and Plosser, C. I. (1982), "Trends and Random Walks in Macroeconomic Time Series", *Journal of Monetary Economics*, 10, 139–162.

Nerlove, M., Grether, D. M., and Carvalho, J. L. (1979), *Analysis of Economic Time Series: A Synthesis*, Academic Press, New York.

Ooms, M. (1993), *Empirical Vector Autoregressive Modeling*, Ph.D. thesis, Erasmus University, Rotterdam.

Phillips, P. C. B. (1987), "Time Series Regression With a Unit Root", *Econometrica*, 55, 277–301.

Priestley, M. B. (1981), *Spectral Analysis and Time Series*, Academic Press, London, Volumes I and II.

Said, S. E., and Dickey, D. A. (1984), "Testing for Unit Roots in ARMA Models of Unknown Order", *Biometrika*, 71, 599–607.

——— (1985), "Hypothesis Testing in ARIMA(p,1,q) Models", *JASA*, 80, 369–374.

Sowell, F. B. (1987), "Maximum Likelihood Estimation of Fractionally Integrated Time Series Models", *Working Papers in Economics*, Department of Economics, Duke University.

———— (1990), "Fractional Unit Root Distribution", *Econometrica*, 58, 495–506.

———— (1992a), "Modeling Long-Run Behavior With the Fractional ARIMA Model", *Journal of Monetary Economics*, 29, 277–302.

———— (1992b), "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models", *Journal of Econometrics*, 53, 165–188.

Steyn, I. J. (1992), "Censorship and Hidden Error-Correction Mechanisms", *Working Paper*, Department of Econometrics, Free University, Amsterdam.

Van der Hoek, G., and Kolen, A. W. J. (1986), *Inleiding Lineaire en Niet – Lineaire Programmering*, H. E. Stenfert Kroese B.V., Leiden.

Watson, M. W. (1986), "Univariate Detrending Methods With Stochastic Trends", *Journal of Monetary Economics*, 18, 49–75.

Whittle, P. (1983), *Prediction and Regulation by Linear Least – Square Methods*, Basil Blackwell Publisher, Oxford, second edition, revised.

APPENDIX A

Asymptotic and Covariance Stationarity

Suppose the discrete time series y_t can be represented by the following $AR(1)$ model

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t$$

for $t=1,2,\ldots,T$, which after $(k-1)$ substitutions can be written as

$$(A1) \qquad y_t = \alpha_1^k y_{t-k} + \sum_{i=0}^{k-1} \alpha_1^i \varepsilon_{t-i},$$

$\varepsilon_t \sim NID(0,\sigma_\varepsilon^2)$. Furthermore, if we assume that $\varepsilon_0 = 0$ the following equation holds exact, i.e. substitution does not need to be continued:

$$(A2) \qquad y_t = \alpha_1^t y_0 + \sum_{i=0}^{t-1} \alpha_1^i \varepsilon_{t-i}$$

It is clear that the mean and the variance respectively of the series y_t depend on the initial observation y_0:[46]

$$\mathcal{E}(y_t|y_0) = \alpha_1^t . \mathcal{E}(y_0)$$

$$var(y_t|y_0) = \alpha_1^{2t} . var(y_0) + \sigma_\varepsilon^2 . (1-\alpha_1^{2t})/(1-\alpha_1^2)$$

$$cov(y_t, y_{t-k}|y_0) = \alpha_1^k . var(y_t|y_0)$$

where the covariance is derived by making use of $(A1)$.

If t tends to infinity, conditional mean and variance of the process y are zero and $\sigma_\varepsilon^2/(1-\alpha_1)$ respectively and therefore the process y is called asymptotically stationary. To arrive at covariance stationarity we furthermore assume that the initial observation has zero mean and variance $\sigma_\varepsilon^2/(1-\alpha_1^2)$.[47] Then there holds

$$\mathcal{E}(y_t) = \mathcal{E}(y_0) = 0$$

$$var(y_t) = var(y_0) = \sigma_\varepsilon^2/(1-\alpha_1^2)$$

$$cov(y_t, y_{t-k}) = \alpha_1^k \sigma_\varepsilon^2/(1-\alpha_1^2)$$

---

[46] To derive the conditional variance of the series y_t the following additional assumptions are necessary: y_0 is independently distributed from the disturbances $\{\varepsilon_t\}_{t=1}^T$ and $|\alpha_1|<1$. Notice that the necessary condition $|\alpha_1|<1$ boils down to assuming that the zero of the polynomial function $1-\alpha_1 z$ lies outside the unit circle.

[47] Alternatively we can assume zero mean and zero variance of the initial observation to arrive at covariance stationarity.

Notice that from $(A2)$ it can be seen that

$$(A3) \qquad \mathcal{E}(y_t^2|y_0) = \alpha_1^{2t}\mathcal{E}(y_0^2) + (1+\alpha_1^2+\alpha_1^4+\ldots+\alpha_1^{2(t-1)})\,\sigma_\varepsilon^2$$

and so the variance of y_t increases out of bound if $|\alpha_1| \geq 1$.

In the appendix A to their paper Dickey, Bell and Miller (1986) note that the argument above can be extended to higher order models. It can be demonstrated that for y_t following an $AR(p)$ model to be stationary, we need (1) the zeros of the AR-polynomial $\alpha(z)$ to lie outside the unit circle, (2) the p starting values $y_0, y_{-1}, \ldots, y_{1-p}$, say, to be independent of the shocks $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T$, (3) the initial shocks $\varepsilon_0, \varepsilon_{-1}, \varepsilon_{-2}, \ldots$ to be zero and (4) the $y_0, y_{-1}, \ldots, y_{1-p}$ to have the correct variances and covariances. Furthermore they note that the series y_t can equivalently be seen to arise from a particular linear combination of current and past shocks (Wold decomposition):

$$y_t = \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i}$$

This can be viewed as generating starting values $y_{1-p}, \ldots, y_0$ independent of $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T$ and with the correct covariance structure.

APPENDIX B

The Levinson-Durbin-Whittle Recursions

The coefficients $\{a_{1,p}, a_{2,p}, \ldots, a_{p,p}\}$ of the $AR(p)$ model (2.1.7) can be obtained as a solution to the corresponding Yule-Walker equations (2.1.8) for $k=1,2,\ldots,p$, or alternatively to the linear system (2.1.9). In prediction theory this solution leads to an optimal one step ahead prediction $y_{t-1,1}$ of y_t in a mean squared error sense, $y_{t-1,1}$ resulting from equation (2.1.7) by substitution of this solution.[48]

Levinson (1947) and Durbin (1960) independently discovered formulas to solve the linear system (2.1.8) recursively: the $(p+1)$ Yule-Walker optimal estimates $\{a_{1,p+1}, a_{2,p+1}, \ldots, a_{p+1,p+1}\}$ can be derived from the Yule-Walker optimal estimates $\{a_{1,p}, a_{2,p}, \ldots, a_{p,p}\}$. The recursions read

$$a_{p+1,p+1} = \frac{\gamma_y(p+1) - \sum_{k=1}^{p} a_{k,p}\gamma_y(p-k+1)}{\sigma_p^2}$$

$$a_{k,p+1} = a_{k,p} - a_{p+1,p+1}a_{p-k+1,p} \qquad k=1,2,\ldots,p$$

where

$$\sigma_p^2 = \gamma_y(0) - \sum_{k=1}^{p} a_{k,p}\gamma_y(k) = \sigma_{p-1}^2(1-a_{p,p}^2)$$

can be interpreted as the variance of the one step ahead prediction error. The coefficient $a_{p+1,p+1}$ is the partial correlation at lag $(p+1)$. The initial conditions are: $a_{0,0}=1$, $\sigma_0^2=\gamma_y(0)=\gamma_\varepsilon(0)$ and $a_{1,1}=\gamma_y(1)/\sigma_0^2$. Furthermore, $a_{k,p}=0$ if $k>p$ and $a_{0,p}=1$ for all p. Notice that the nominator of the right-hand side of $(B1)$ and the first expression of σ_p^2 in $(B3)$ are Yule-Walker equations (2.1.8) for $k=p+1$ and $k=0$, respectively.[49]

According to Jonas (1983), Whittle (1963) made the algorithm widely known among statisticians by showing the correspondence between the autocovariance matrix of a process and the finite order autoregressive coefficients $a_{k,p}$ as $Q^{-1}=A\Sigma^{-2}A'$ where A and Σ are defined in equations (2.4.2) and (2.4.3) (see also Whittle 1983, Sections 3.4 and 7.1). The recursions are therefore often referred to as the Levinson-Durbin-Whittle recursions.

---

[48] See section 2.1.3, footnote 8.

[49] As equation (2.1.9) is a linear system $Ax=b$ where A is a Toeplitz matrix the Levinson-Durbin-Whittle algorithm can be applied to find a solution x to any such system.

For illustrative purposes we present the first three iterations of the Levinson–Durbin–Whittle algorithm:

Table B1

The first three Levinson – Durbin – Whittle iterations

Iteration	Model	Prediction variance	
0	$y_t=\varepsilon_t$	$\sigma_0^2=\gamma_y(0)=\gamma_\varepsilon(0)$	
1	$y_t=a_{1,1}y_{t-1}+\varepsilon_t$	$\sigma_1^2=\gamma_y(0)-a_{1,1}\gamma_y(1)$	
2	$y_t=a_{1,2}y_{t-1}+a_{2,2}y_{t-2}+\varepsilon_t$	$\sigma_2^2=\gamma_y(0)-a_{1,2}\gamma_y(1)-a_{2,2}\gamma_y(2)$	
Iteration	**Parameter estimates**		
0			$a_{0,0}=1$
1	$a_{1,1}=\gamma_y(1)/\sigma_0^2$		$a_{0,1}=1$
2	$a_{2,2}=(\gamma_y(2)-a_{1,1}\gamma_y(1))/\sigma_1^2$	$a_{1,2}=a_{1,1}-a_{2,2}a_{1,1}$	$a_{0,2}=1$

APPENDIX C
Tables on the Simulation Experiments[50]

Table contents:

The Geweke and Porter-Hudak (1983) *experiment*:

$N = 4000$, $T = 265$, $\mu = 0$, $\gamma_y(0) = 1$, *indirect Cholesky decomposition of* Q:
a) $d = 0.25$, *cpu time* $= 88.78$ *seconds*
b) $d = 0.45$, *cpu time* $= 89.16$ *seconds*
c) $d = 0.25$, *cpu time* $= 93.87$ *seconds*
d) $d = 0.45$, *cpu time* $= 94.23$ *seconds*

The Granger and Joyeux (1980) *experiment*:

$N = 1000$, $T = 400$, $\mu = 0$, $\gamma_y(0) = 1$, *finite order autoregressive polynomial*:
a) $d = 0.25$, *cpu time* $= 17.50$ *seconds*
b) $d = 0.45$, *cpu time* $= 17.37$ *seconds*
c) $d = 0.25$, *cpu time* $= 18.88$ *seconds*
d) $d = 0.45$, *cpu time* $= 18.73$ *seconds*

[50] As the variances equal one autocovariances and autocorrelations mean the same.

A replication of the Geweke and Porter-Hudak (1983) experiment

autocovariances are computed given the population mean

	d=0.25		d=0.45	
lag	theoretical autocovar.	estimated autocovar.	theoretical autocovar.	estimated autocovar.
0	1.00000000	0.99892257	1.00000000	1.01571072
1	0.33333333	0.33145570	0.81818182	0.83368131
2	0.23809524	0.23662850	0.76539589	0.78096778
3	0.19480519	0.19370126	0.73538037	0.75107941
4	0.16883117	0.16894420	0.71466543	0.73071218
5	0.15105947	0.15007588	0.69895850	0.71474245
6	0.13792386	0.13830638	0.68636465	0.70253296
7	0.12770728	0.12777324	0.67588580	0.69198604
8	0.11946810	0.11807165	0.66693367	0.68270921
9	0.11264135	0.11342874	0.65913328	0.67545295
10	0.10686487	0.10707379	0.65223136	0.66843851
11	0.10189441	0.10133280	0.64604907	0.66207036
12	0.09755848	0.09770913	0.64045557	0.65669971
13	0.09373266	0.09440022	0.63535234	0.65177599
14	0.09032420	0.09023845	0.63066339	0.64682324
15	0.08726236	0.08531729	0.62632894	0.64187663
16	0.08449213	0.08337668	0.62230110	0.63798440
17	0.08196997	0.08053059	0.61854097	0.63413879
18	0.07966096	0.08045070	0.61501652	0.63117719
19	0.07753667	0.07666994	0.61170107	0.62743592
20	0.07557371	0.07493702	0.60857216	0.62430527
21	0.07375266	0.07237761	0.60561074	0.62111890
22	0.07205720	0.07067978	0.60280048	0.61829108
23	0.07047352	0.06881413	0.60012731	0.61551146
24	0.06898987	0.06685963	0.59757900	0.61288206
25	0.06759613	0.06512605	0.59514487	0.61042358
26	0.06628359	0.06415841	0.59281553	0.60829986
27	0.06504464	0.06508817	0.59058271	0.60682958
28	0.06387267	0.06379554	0.58843903	0.60480914
29	0.06276184	0.06256522	0.58637795	0.60279390
30	0.06170702	0.06042041	0.58439359	0.60064071
31	0.06070365	0.05990556	0.58248068	0.59897738
32	0.05974769	0.05963907	0.58063447	0.59749820
33	0.05883551	0.05919701	0.57885064	0.59603137
34	0.05796387	0.05946656	0.57712531	0.59476485
35	0.05712986	0.05800298	0.57545490	0.59304377
36	0.05633084	0.05695970	0.57383618	0.59144132
37	0.05556443	0.05652877	0.57226618	0.59004908
38	0.05482848	0.05532967	0.57074216	0.58845330
39	0.05412102	0.05516929	0.56926164	0.58726893
40	0.05344025	0.05559460	0.56782229	0.58626291
41	0.05278454	0.05521454	0.56642199	0.58512492
42	0.05215239	0.05466122	0.56505876	0.58388129
43	0.05154242	0.05388237	0.56373077	0.58256087
44	0.05095337	0.05292751	0.56243633	0.58122124
45	0.05038405	0.05205151	0.56117385	0.57996346
46	0.04983341	0.05283045	0.55994185	0.57918520
47	0.04930043	0.05325471	0.55873897	0.57835123
48	0.04878420	0.05236838	0.55756391	0.57708737
49	0.04828385	0.05077349	0.55641548	0.57561479

An alternative to the Geweke and Porter-Hudak (1983) experiment

autocovariances are computed given the sample mean

	d=0.25		d=0.45	
lag	theoretical autocovar.	estimated autocovar.	theoretical autocovar.	estimated autocovar.
0	1.00000000	0.94604895	1.00000000	0.44985038
1	0.33333333	0.27598257	0.81818182	0.26697599
2	0.23809524	0.18073274	0.76539589	0.21393821
3	0.19480519	0.13758251	0.73538037	0.18381534
4	0.16883117	0.11269060	0.71466543	0.16326978
5	0.15105947	0.09367739	0.69895850	0.14713482
6	0.13792386	0.08178519	0.68636465	0.13472038
7	0.12770728	0.07117068	0.67588580	0.12400657
8	0.11946810	0.06135520	0.66693367	0.11451752
9	0.11264135	0.05666546	0.65913328	0.10716015
10	0.10686487	0.05022890	0.65223136	0.09996686
11	0.10189441	0.04441599	0.64604907	0.09340787
12	0.09755848	0.04074068	0.64045557	0.08785837
13	0.09373266	0.03732841	0.63535234	0.08269932
14	0.09032420	0.03308405	0.63066339	0.07757849
15	0.08726236	0.02808459	0.62632894	0.07250191
16	0.08449213	0.02610286	0.62230110	0.06848359
17	0.08196997	0.02322293	0.61854097	0.06455305
18	0.07966096	0.02313036	0.61501652	0.06147877
19	0.07753667	0.01931147	0.61170107	0.05762145
20	0.07557371	0.01752785	0.60857216	0.05436050
21	0.07375266	0.01493090	0.60561074	0.05107266
22	0.07205720	0.01318040	0.60280048	0.04810657
23	0.07047352	0.01128568	0.60012731	0.04520391
24	0.06898987	0.00926624	0.59757900	0.04238675
25	0.06759613	0.00749712	0.59514487	0.03980632
26	0.06628359	0.00651170	0.59281553	0.03761153
27	0.06504464	0.00739205	0.59058271	0.03599426
28	0.06387267	0.00608603	0.58843903	0.03387924
29	0.06276184	0.00484344	0.58637795	0.03176204
30	0.06170702	0.00264797	0.58439359	0.02950199
31	0.06070365	0.00206742	0.58248068	0.02772400
32	0.05974769	0.00175393	0.58063447	0.02610720
33	0.05883551	0.00125057	0.57885064	0.02449849
34	0.05796387	0.00145320	0.57712531	0.02309721
35	0.05712986	-0.00004609	0.57545490	0.02128191
36	0.05633084	-0.00112508	0.57383618	0.01958590
37	0.05556443	-0.00160231	0.57226618	0.01811179
38	0.05482848	-0.00281043	0.57074216	0.01650421
39	0.05412102	-0.00299914	0.56926164	0.01524086
40	0.05344025	-0.00257302	0.56782229	0.01416867
41	0.05278454	-0.00300844	0.56642199	0.01289347
42	0.05215239	-0.00359686	0.56505876	0.01159972
43	0.05154242	-0.00437931	0.56373077	0.01023275
44	0.05095337	-0.00537438	0.56243633	0.00881546
45	0.05038405	-0.00628302	0.56117385	0.00748112
46	0.04983341	-0.00553285	0.55994185	0.00661437
47	0.04930043	-0.00512134	0.55873897	0.00569030
48	0.04878420	-0.00602178	0.55756391	0.00436651
49	0.04828385	-0.00762075	0.55641548	0.00284630

A replication of the Granger and Joyeux (1980) experiment

autocovariances are computed given the population mean

	d=0.25		d=0.45	
lag	theoretical autocovar.	estimated autocovar.	theoretical autocovar.	estimated autocovar.
0	1.00000000	1.17911916	1.00000000	3.00811533
1	0.33333333	0.39213423	0.81818182	2.34801506
2	0.23809524	0.28038493	0.76539589	2.15764260
3	0.19480519	0.23123892	0.73538037	2.05116768
4	0.16883117	0.20051089	0.71466543	1.97664643
5	0.15105947	0.17881517	0.69895850	1.91961928
6	0.13792386	0.16434088	0.68636465	1.87505276
7	0.12770728	0.15039411	0.67588580	1.83544523
8	0.11946810	0.13743772	0.66693367	1.80039420
9	0.11264135	0.13062982	0.65913328	1.77374692
10	0.10686487	0.12481034	0.65223136	1.75049581
11	0.10189441	0.12144259	0.64604907	1.73095395
12	0.09755848	0.11619966	0.64045557	1.71129826
13	0.09373266	0.11167732	0.63535234	1.69308868
14	0.09032420	0.10602917	0.63066339	1.67457092
15	0.08726236	0.09910784	0.62632894	1.65607762
16	0.08449213	0.09751168	0.62230110	1.64282028
17	0.08196997	0.09296925	0.61854097	1.62856799
18	0.07966096	0.09352323	0.61501652	1.61905002
19	0.07753667	0.09022827	0.61170107	1.60707843
20	0.07557371	0.08771528	0.60857216	1.59586653
21	0.07375266	0.08402612	0.60561074	1.58428972
22	0.07205720	0.08196621	0.60280048	1.57463807
23	0.07047352	0.08049645	0.60012731	1.56609892
24	0.06898987	0.07981561	0.59757900	1.55851891
25	0.06759613	0.07647006	0.59514487	1.54905454
26	0.06628359	0.07483575	0.59281553	1.54158525
27	0.06504464	0.07825569	0.59058271	1.53861452
28	0.06387267	0.07610061	0.58843903	1.53128790
29	0.06276184	0.07254869	0.58637795	1.52267157
30	0.06170702	0.07158961	0.58439359	1.51636921
31	0.06070365	0.07071606	0.58248068	1.51113580
32	0.05974769	0.07323736	0.58063447	1.50901943
33	0.05883551	0.07408256	0.57885064	1.50553045
34	0.05796387	0.07345045	0.57712531	1.50030213
35	0.05712986	0.06897345	0.57545490	1.49195222
36	0.05633084	0.06674320	0.57383618	1.48569869
37	0.05556443	0.07045022	0.57226618	1.48447526
38	0.05482848	0.06633375	0.57074216	1.47727917
39	0.05412102	0.06680779	0.56926164	1.47364709
40	0.05344025	0.06720905	0.56782229	1.47039267
41	0.05278454	0.06557384	0.56642199	1.46570487
42	0.05215239	0.06699446	0.56505876	1.46338926
43	0.05154242	0.06559026	0.56373077	1.45862519
44	0.05095337	0.06521286	0.56243633	1.45458464
45	0.05038405	0.06365026	0.56117385	1.44946856
46	0.04983341	0.06117426	0.55994185	1.44341943
47	0.04930043	0.06020679	0.55873897	1.43894873
48	0.04878420	0.06199947	0.55756391	1.43679953
49	0.04828385	0.06082388	0.55641548	1.43234851

An alternative to the Granger and Joyeux (1980) experiment

autocovariances are computed given the population mean

	$d = 0.25$		$d = 0.45$	
lag	theoretical autocovar.	estimated autocovar.	theoretical autocovar.	estimated autocovar.
0	1.00000000	1.13606434	1.00000000	1.81324962
1	0.33333333	0.30110952	0.81818182	1.14983849
2	0.23809524	0.20217451	0.76539589	0.95755264
3	0.19480519	0.15891194	0.73538037	0.84927140
4	0.16883117	0.13171897	0.71466543	0.77317284
5	0.15105947	0.11261497	0.69895850	0.71466018
6	0.13792386	0.09987352	0.68636465	0.66856173
7	0.12770728	0.08742812	0.67588580	0.62761747
8	0.11946810	0.07576923	0.66693367	0.59121949
9	0.11264135	0.06990991	0.65913328	0.56316444
10	0.10686487	0.06480213	0.65223136	0.53869500
11	0.10189441	0.06183652	0.64604907	0.51806409
12	0.09755848	0.05699240	0.64045557	0.49722525
13	0.09373266	0.05300567	0.63535234	0.47789102
14	0.09032420	0.04821774	0.63066339	0.45837148
15	0.08726236	0.04211403	0.62632894	0.43886084
16	0.08449213	0.04058033	0.62230110	0.42441293
17	0.08196997	0.03675255	0.61854097	0.40883924
18	0.07966096	0.03715211	0.61501652	0.39822251
19	0.07753667	0.03425397	0.61170107	0.38514230
20	0.07557371	0.03215145	0.60857216	0.37275959
21	0.07375266	0.02899242	0.60561074	0.35992361
22	0.07205720	0.02709729	0.60280048	0.34899236
23	0.07047352	0.02573541	0.60012731	0.33918139
24	0.06898987	0.02511706	0.59757900	0.33049533
25	0.06759613	0.02238412	0.59514487	0.32008446
26	0.06628359	0.02107271	0.59281553	0.31165390
27	0.06504464	0.02378724	0.59058271	0.30771938
28	0.06387267	0.02174860	0.58843903	0.29942685
29	0.06276184	0.01882282	0.58637795	0.28977140
30	0.06170702	0.01817588	0.58439359	0.28279290
31	0.06070365	0.01747374	0.58248068	0.27658372
32	0.05974769	0.01933196	0.58063447	0.27354791
33	0.05883551	0.01970032	0.57885064	0.26930958
34	0.05796387	0.01912326	0.57712531	0.26321116
35	0.05712986	0.01539504	0.57545490	0.25391418
36	0.05633084	0.01350941	0.57383618	0.24665616
37	0.05556443	0.01648800	0.57226618	0.24451767
38	0.05482848	0.01311061	0.57074216	0.23632258
39	0.05412102	0.01343742	0.56926164	0.23187373
40	0.05344025	0.01375375	0.56782229	0.22776013
41	0.05278454	0.01231234	0.56642199	0.22217041
42	0.05215239	0.01351158	0.56505876	0.21898340
43	0.05154242	0.01216806	0.56373077	0.21353331
44	0.05095337	0.01176210	0.56243633	0.20877590
45	0.05038405	0.01070373	0.56117385	0.20272393
46	0.04983341	0.00894330	0.55994185	0.19569927
47	0.04930043	0.00807930	0.55873897	0.19047764
48	0.04878420	0.00936004	0.55756391	0.18761257
49	0.04828385	0.00843962	0.55641548	0.18250847

APPENDIX D

More on ADF Unit Root Testing

In Section 2.2.2 Dickey-Fuller zero frequency unit root tests are presented which are useful if the processes under consideration have zero mean. If a process has a non-zero mean, a natural way to model this under the alternative hypothesis is:

$$(D1a) \qquad H_1: (y_t-\mu) = \phi.(y_{t-1}-\mu) + \varepsilon_t, \qquad \phi\neq 1,$$

which is equivalent to

$$(D1b) \qquad H_1: y_t = \alpha + \phi.y_{t-1} + \varepsilon_t,$$

where $\alpha=\mu(1-\phi)$ and $y_0=0$. Alternatively, given a process with a non-zero mean and deterministic trend component the alternative hypothesis is

$$(D2a) \qquad H_1: (y_t-\mu-\gamma t) = \phi.(y_{t-1}-\mu-\gamma.(t-1)) + \varepsilon_t, \quad \phi\neq 1,$$

which is equivalent to

$$(D2b) \qquad H_1: y_t = \alpha + \beta t + \phi.y_{t-1} + \varepsilon_t,$$

where $\alpha=\mu(1-\phi)+\gamma\phi$ and $\beta=\gamma(1-\phi)$.

Limiting and finite-sample distributions of ρ and τ

As mean μ and trend component γt are usually unknown in practice unit root tests are usually based on the models $(D1b)$ and $(D2b)$. Fuller (1976) tabulated the finite-sample and asymptotic critical values under the null hypotheses $(\alpha,\phi)=(0,1)$ for model $(D1b)$ and $(\alpha,\beta,\phi)=(0,0,1)$ for model $(D2b)$ based on Monte Carlo studies. Thus, for the latter model trending behavior under the null is excluded. The estimators of ϕ and corresponding t-statistics are denoted by $\hat{\phi}_\mu$ and $\hat{\tau}_\mu$ for model $(D1b)$ respectively, and by $\hat{\phi}_\tau$ and $\hat{\tau}_\tau$ for model $(D2b)$ respectively (a.o. Fuller (1976)). The subscript μ reflects the inclusion of a constant, whereas the subscript τ reflects the presence of a trend (and a constant).

Dickey and Fuller (1979) derived the analytical expression of the limit distributions of $\hat{\phi}_\mu$ and $\hat{\tau}_\mu$ under the assumption $(\alpha,\phi)=(0,1)$, and of $\hat{\phi}_\tau$ and $\hat{\tau}_\tau$ under the assumption $(\beta,\phi)=(0,1)$. It should be noted that the limit distributions of $\hat{\phi}_\tau$ and $\hat{\tau}_\tau$ are unaffected by the value of α in model $(D2b)$. However, the finite-sample distributions do depend on α. This is the reason why α has been put to a specific value (i.e. zero) in Monte Carlo studies.

For model (*D2a*) a similar result holds under the null hypothesis, as it then reads $y_t = y_{t-1} + \gamma + \varepsilon_t$ for any μ. As, given stationarity, μ represents the mean of the series we say that the mean is not identified under the null. In finite sampling μ can only be related to the initial conditions under the null (Said and Dickey (1985)).

Given the dependence of the limit distributions of $\hat{\phi}_\mu$ and $\hat{\tau}_\mu$ on α, and the dependence of the limit distributions of $\hat{\phi}_\tau$ and $\hat{\tau}_\tau$ on β, an F-test as advocated in Dickey and Fuller (1981) seems to be more appropriate than tests based on ϕ solely. However, F-tests have limited use (Ooms 1993, p. 57);

In zero frequency unit root testing linear trending behavior under the null is excluded for model (*D1b*) and quadratic trending behavior under the null is excluded for model (*D2b*). However, if we allow for such behavior Dickey and Fuller (1979) find that the limiting distributions of $\hat{\tau}_\mu$ in the presence of a linear trend and $\hat{\tau}_\tau$ in the presence of a quadratic trend are normal. So, if (*D1b*) is the maintained model the null hypothesis will be accepted more often than the nominal level.

Limiting and finite-sample distributions of α and β

In Dickey and Fuller (1981) limiting distributions are also derived for the estimator of α in model (*D1b*) and the estimators of α and β in model (*D2b*), together with those of their respective t-statistics. The asymptotic distributions are symmetric and independent of the initial condition y_0. Furthermore, the ε_t should form a sequence of independent identically distributed random variables, not necessarily normal. For model (*D1b*) the statistics are denoted by $\hat{\alpha}_\mu$ and $\hat{\tau}_{\alpha\mu}$; for the model (*D2b*) the statistics are denoted by $\hat{\alpha}_\tau$, $\hat{\beta}_\tau$, $\hat{\tau}_{\alpha\tau}$ and $\hat{\tau}_{\beta\tau}$. Percentage points for the limiting distributions are tabulated.

Dickey and Fuller (1981) also tabulated finite sample critical values for these statistics drawn from Monte Carlo studies. The asymptotic distributions continue to hold for higher order autoregressive processes as well (see also Diebold and Nerlove (1990)).

Testing under joint hypotheses

Dickey and Fuller (1981) derive a likelihood ratio (LR) test of the hypothesis $(\alpha, \phi) = (0, 1)$ for model (2b), of the hypothesis $(\alpha, \beta, \phi) = (0, 0, 1)$ for model (3b) and of the hypothesis $(\alpha, \beta, \phi) = (\alpha, 0, 1)$ for model (3b). Furthermore,

they derive the limiting distributions of these test statistics. The limiting distributions for the latter two hypotheses are monotone transformations of common regression 'F-tests' one would construct. The latter limiting distribution is independent of α.

Evans and Savin (1981) show that the powers of the LM tests are conditional on the value of the parameter c/σ, just as the finite sample distribution of the normalized least squares estimator. However, the asymptotic distribution of this statistic depends on c/σ for values of ϕ greater than one in absolute value only.