Hans Gerhard Strohe

Time Series Analysis

Textbook for Students of Economics and Business Administration

Universität Potsdam
Lehrstuhl für Statistik und Ökonometrie

Potsdam 2004

Hans Gerhard Strohe, Universität Potsdam

Time Series Analysis

1. Deterministic Time Series Modelling

To be subject of a separate textbook**.**

2. Stationary Stochastic Processes

2.1. Definitions

A set of random variables $X(t)$ where $t \in \Theta \subseteq \Re$ (real numbers) is referred to as a stochastic process. A discrete stochastic process is defined as a sequence of random variables $X(t)$ where $t = t_1 , t_2 , ..., t_T , ...$, shortly $..., X_1 , X_2 , ..., X_T ...$ or X_t

The mathematical expectations $\mathrm{E}(X_t)$ can differ from time to time and form a mean function depending on time

$$\mu(t) = \mu_t = \mathrm{E}[X_t] \tag{2.1}$$

The same way the variances $\mathrm{var}(X_t)$ form the variance function, depending on time too:

$$\sigma^2(t) = \sigma_t^2 = \mathrm{E}[(X_t - \mu_t)^2]. \tag{2.2}$$

Generally, there is a certain variance at each point of time. Principally, this is not the same as variability of empirical data during the run of the process over time.

The autocovariance

$$\gamma_{t_1,t_2} = \mathrm{cov}(X_{t_1}, X_{t_2}) = \mathrm{E}[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})] \tag{2.3}$$

generally depends on each t_1 and t_2.

One finite realization $x_1, x_2, ..., x_T$ of a discrete stochastic process … $X_1, X_2, \ldots X_T \ldots$ is called a time series. In this chapter we shall consistently distinguish between stochastic processes and time series generated by them. Processes are marked by capitals. Small letters mark time series. Exceptions are residual processes belonging to models for stochastic processes and not having any independent practical content. They are designed by small letters such as a, u and ε, too. The strict distinction is necessary for correctly deducing properties of time series from those of stochastic processes. When practically modelling empirical time series later on, this distinction can be more or less neglected.

A stochastic process X_t is called *strongly stationary*, if the joint probability distribution of all variables

$X_{t_1}, X_{t_2}, ..., X_{t_n}$ is the same as that

of $X_{t_{1+\tau}}, X_{t_{2+\tau}}, ..., X_{t_{n+\tau}}$.

X_t is *weakly stationary*, if

Mean $\mu_t = \mu = const$
Variance $\sigma_t^2 = \sigma^2 = const$
Autocovariance $\gamma_{t_1, t_2} = \gamma_{t_1 - t_2} = \gamma_\tau$ with $\tau = t_1 - t_2$ (Lag)

Being a function of only the lag τ, the autocovariance

$$\gamma(\tau) = \gamma_\tau = \mathrm{E}[(X_t - \mu)(X_{t-\tau} - \mu)] \qquad (2.4)$$

is called autocovariance function. For $\tau = 0$ it is equal to the variance.

By standardization with $\sigma^2 = \gamma_0$ the autocorrelation function of a stationary stochastic process is obtained:

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0} \quad \text{with} \quad -1 \le \rho_\tau \le 1 \; . \qquad (2.5)$$

A time series $x_1, x_2, ..., x_T$ that is one realization of a stationary stochastic process X_t is called stationary as well.

In practical analytical work stationarity of a time series means
- no trend
- no systematic change of variance
- no strictly periodic fluctuations
- no systematically changing interdependencies between the elements of the time series

Economic time series consisting of data observed in practice such as gross national product for a sequence of years usually are not stationary.

2.2. Ergodicity

A fundamental obstacle to estimating the distribution parameters of a stochastic process is that generally the sample size is $n = 1$, because usually here exists only one time series for a process. Thus a sensible estimation virtually is not possible. The stochastic process to be examined itself is unknown. Its stationarity or nonstationarity can be found only by analyzing this one existing time series. But on the other hand: Many analysis methods for time series assume stationarity. This leads to the sort of circular conclusion, that the property to be found firstly has to be assumed to exist.

A solution can be found by using the notion of *ergodicity*: this is the behaviour of a large class of stationary processes, where the arithmetic mean over time periods converges to the mathematical expectation μ. Ergodicity makes it possible to estimate μ, σ^2, $\gamma(\tau)$ of the underlying process by using one time series only.

Approaches for recognizing the stationarity of a time series are various:

- Graphical representation of the time series and visual check for trend, i.e. for changing mean, increasing or decreasing variance and strong periodicities
- Examination of the empirical autocorrelation
- Tests for a deterministic trend, e.g. *t*-test of a least squares estimation
- Tests for a stochastic trend, e.g. unit root test.

2.3. Special Cases

A process is called a *normal process* if the joint distribution of $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ is an *n*-dimensional normal distribution. In this case, from weak stationarity follows strong stationarity.
White noise is a purely random process i.e. a series of independent identically distributed random variables a_t (iid). The most important properties of white noise are

$$\begin{aligned} \mu_t &= \mathrm{E}(a_t) = const = \mu \\ \sigma_t^2 &= const = \sigma_a^2 \\ \gamma_{t_1,t_2} &= 0 \quad \text{für} \quad t_1 \neq t_2 \end{aligned} \tag{2.6}$$

Stationarity immediately follows from this. White noise plays an important role in modelling for the representation of the error or innovations part in a data generating stochastic process.

Example 2.1:

Let us consider two white noise processes

$X_t = a_t$
and $Y_t = 3+1.5a_t$
where a_t is white noise with zero mean and unit variance. Obviously Y_t has the mean $\mu = 3$ and the variance 1.5.

Figure 2.1 displays two independent realizations for each process generated by normally

In order to find out whether or not a time series x_t represents white noise it is useful to test its empirical autocorrelation r_τ (see section 4.1) by the Box-Pierce Q-statistic:

$$Q = T\sum_{k=1}^{P} r_\tau^2 \tag{2.7}$$

Under the null hypothesis that X_t is white noise Q follows a χ^2 – distribution with p degrees of freedom. In the case of example (2.1) Q assumes values between 1 and 14 for τ = 1 to 16, respectively. The values do not exceed the corresponding 1% critical values of χ^2. Thus the null hypothesis of white noise cannot be rejected distributed random numbers for a_t.

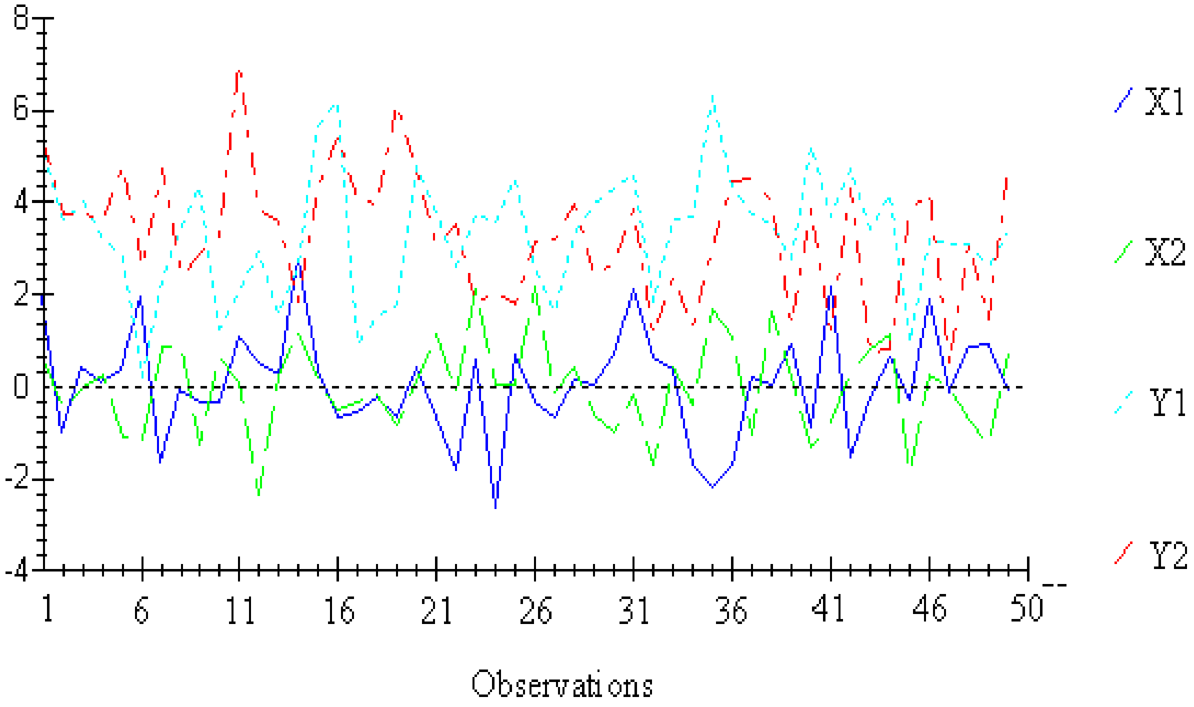

Figure 2.1: Plots of white noise realizations with means 0 and 3, respectively.

.

3. ARMA Processes

3.1. MA Models

Consider a process that is nothing more than a linear combination of two white noise elements following one after the other:

$$X_t = a_t - \theta_1 a_{t-1} \tag{3.1}$$

where a_t is white noise with μ = 0. Then X_t is called a first order moving average process MA(1).

Here the white noise term sometimes is referred to as “innovations” or, more dramatically, as “shocks” because it is the only new, i.e. previously unknown, information entering the process in every point of time.

A moving average process of order q [MA(q)] is a process X_t with

$$X_t = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} , \tag{3.2}$$

where a_t is white noise with $\mu=0$.

By introducing the lag or backshift operator L with

$$\begin{aligned} \mathrm{L}(X_t) &= X_{t-1} \\ \mathrm{L}^2(X_t) &= X_{t-2} \\ \mathrm{L}^k(X_t) &= X_{t-k} \end{aligned} \tag{3.3}$$

an MA(q) process can be written shorter if we substitute

$$a_{t-k} = \mathrm{L}^k(a_t) \tag{3.4}$$

And use the operator function:

$$\Theta_q(\mathrm{L}) = 1 - \theta_1 \mathrm{L} - \theta_2 \mathrm{L}^2 - \ldots - \theta_q \mathrm{L}^q . \tag{3.5}$$

Then the MA(q) process (3.2) is simply defined by

$$X_t = \Theta_q(\mathrm{L}) a_t \tag{3.6}$$

An MA(q) process has the following properties:

$$\begin{aligned} \mathrm{E}[X_t] &= 0 \\ \mathrm{var}[X_t] &= \sigma^2 \sum_{i=0}^{q} \theta_i^2 \end{aligned} \tag{3.7}$$

$$\gamma_{t,t+\tau} = \begin{cases} 0 & \tau > q \\ \sigma^2 \sum_{i=0}^{q-\tau} \theta_i \theta_{i+\tau} & \tau = 0,1,\ldots,q \end{cases} \tag{3.8}$$

The mean, the variance and the covariance do not depend on time. Therefore a MA process is weakly stationary.

Example 3.1:

MA(2) processes.

Figure 3.1 shows two time series generated by the MA(2) processes

$$X_t^1 = a_t + 0{,}75a_{t-1} + 0{,}4a_{t-2},$$
$$X_t^2 = a_t - 0{,}75a_{t-1} - 0{,}4a_{t-2},$$

respectively. The white noise innovations a_t are represented by zero mean normal random numbers. It is visible from the graph that the process X_t^2 with negative coefficients (i.e. $\theta_i > 0$!) is more oscillating than the first one.

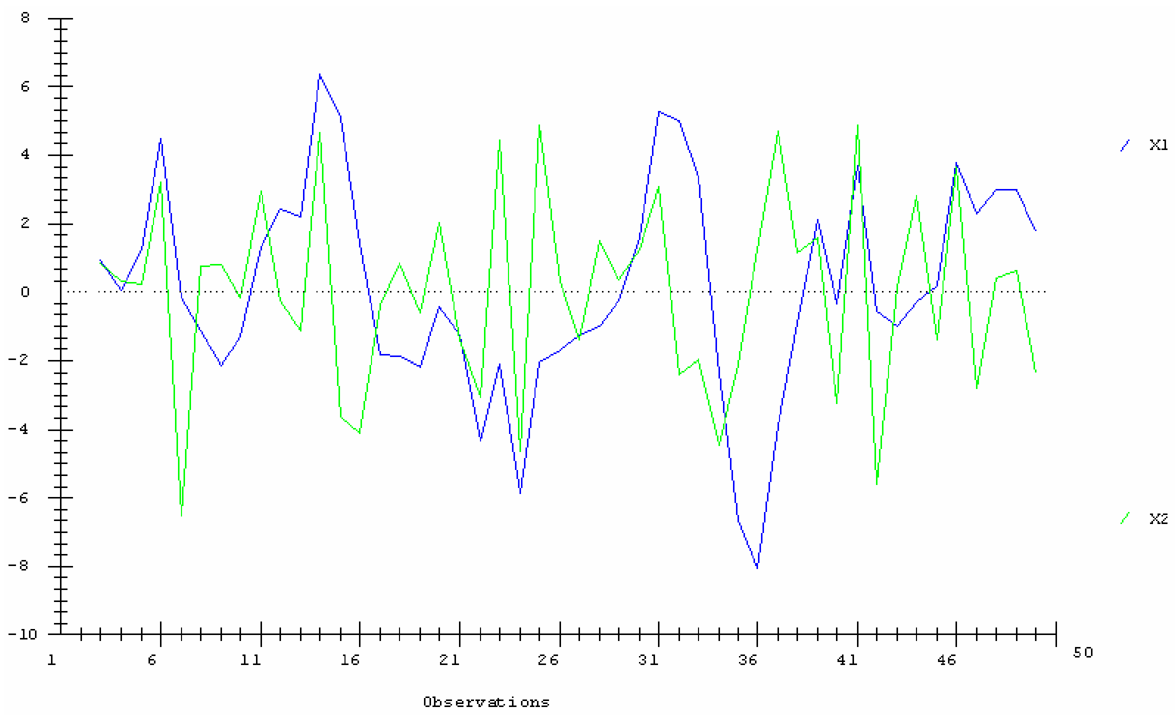

Figure 3.1: MA(2) time series

A second generation by each process with other random numbers a_t delivers time series different from the former ones in detail but similar to them in the general shape (figure 3.2).

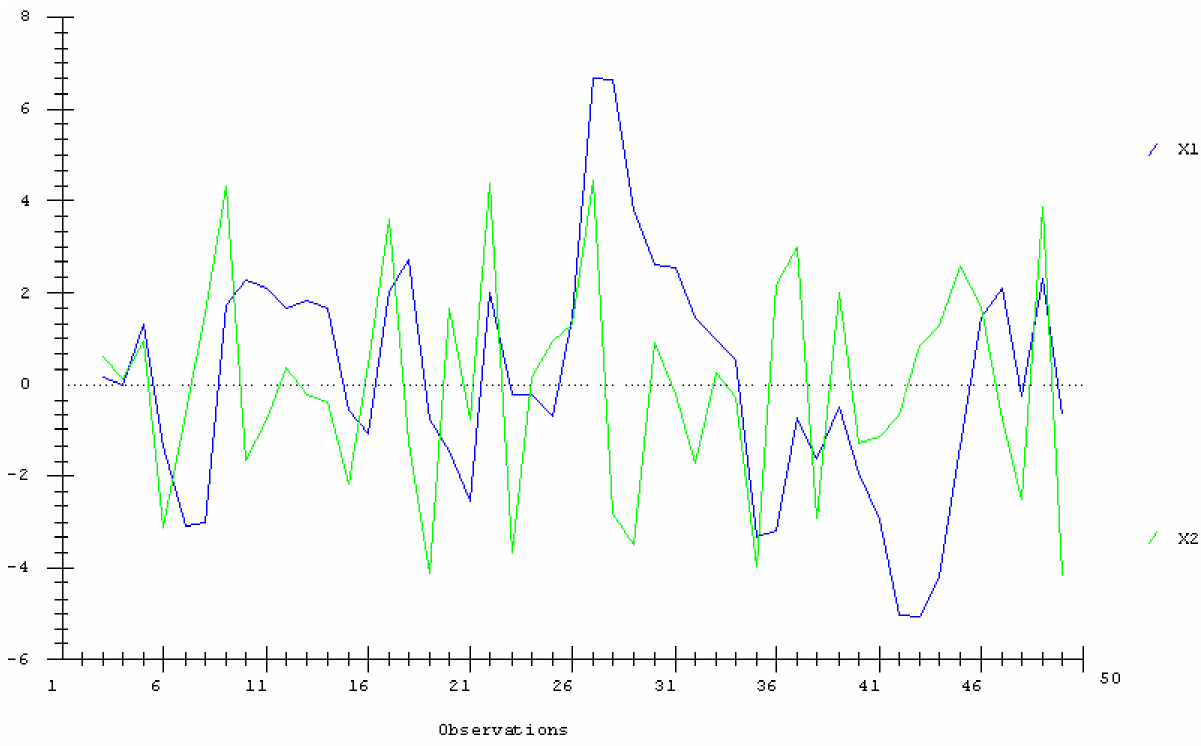

Figure 3.2: MA(2) time series generated by the same processes

3.2. AR Models

An autoregressive process of order p (AR(p)) is a stochastic process X_t with

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + a_t \qquad (3.9)$$

where a_t is white noise with $\mu_a = 0$. The intercept ϕ_0 is often set to zero.
By using the lag operator function

$$\Phi_p(\mathrm{L}) = 1 - \phi_1 \mathrm{L} - \phi_2 \mathrm{L}^2 - ... - \phi_p \mathrm{L}^p \qquad (3.10)$$

it can be shortly written

$$\Phi_p(\mathrm{L}) X_t = \phi_0 + a_t \qquad (3.11)$$

An AR process is not in every case stationary. If we know the representation (3.9) or (3.11) of the process, what is called the characteristic equation is helpful in finding whether the process is stationary or not.

The characteristic equation is defined as

$$1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p = 0$$

$$\text{i.e.} \quad \Phi_p(z) = 0 \qquad (3.12)$$

where z is assumed a complex variable. The following necessary and sufficient condition for stationarity of an AR process can be proved:

If and only if all (complex) solutions (roots) of the characteristic equation lie outside the unit circle, i.e. $|z| > 1$, the AR process is stationary.
Particularly, if $|z| = 1$, what is called a unit root, the process is just nonstationary.

Example 3.2:

Let $X_t = 1.1\, X_{t-1} + a_t$ be an AR(1) process with zero mean white noise a_t.

Its characteristic equation is $1 - 1.1\, z = 0$ with the root $z = 0.91$, which lies inside the unit circle: $|z| < 1$. Thus the process is nonstationary.
This is obvious also without solving the characteristic equation because the coefficient 1.1 generates a permanent increase of the following values. Figure 3.3 displays the graphs of two independent realizations of this process.

But the process $X_t = 0.8\, X_{t-1} + a_t$ has the characteristic equation $1\text{-}0.8\, z = 0$ with the root $z = 1.25$, i.e. $|z| > 1$. It is stationary. Its values move around zero, what is visible from figure 3.4.

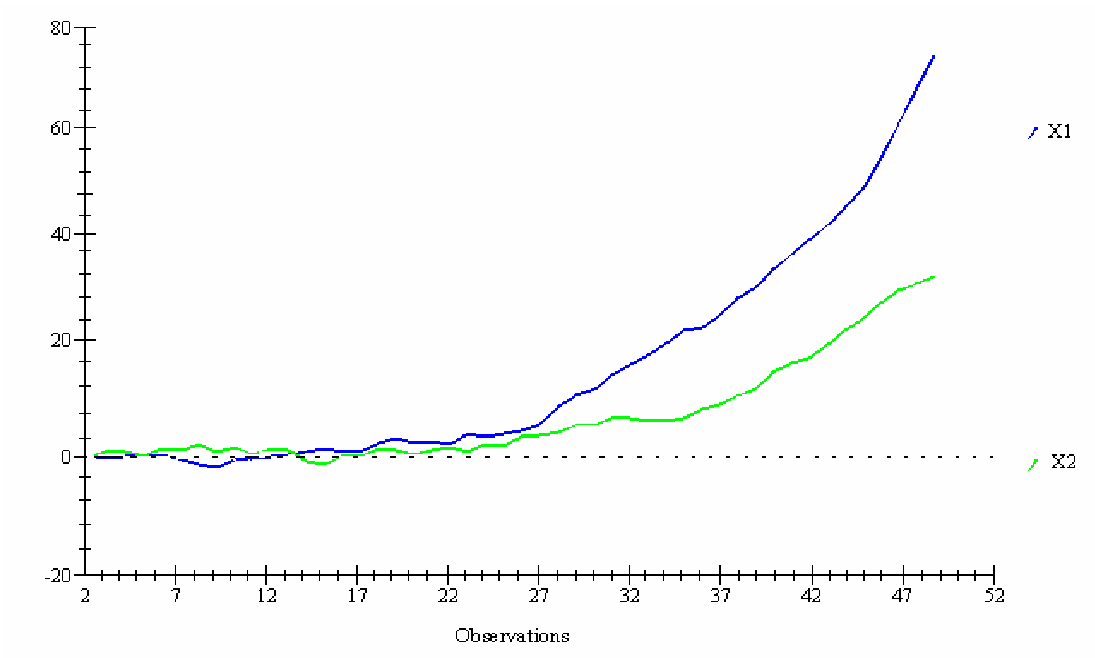

Figure 3.3: Two realizations of one nonstationary process.

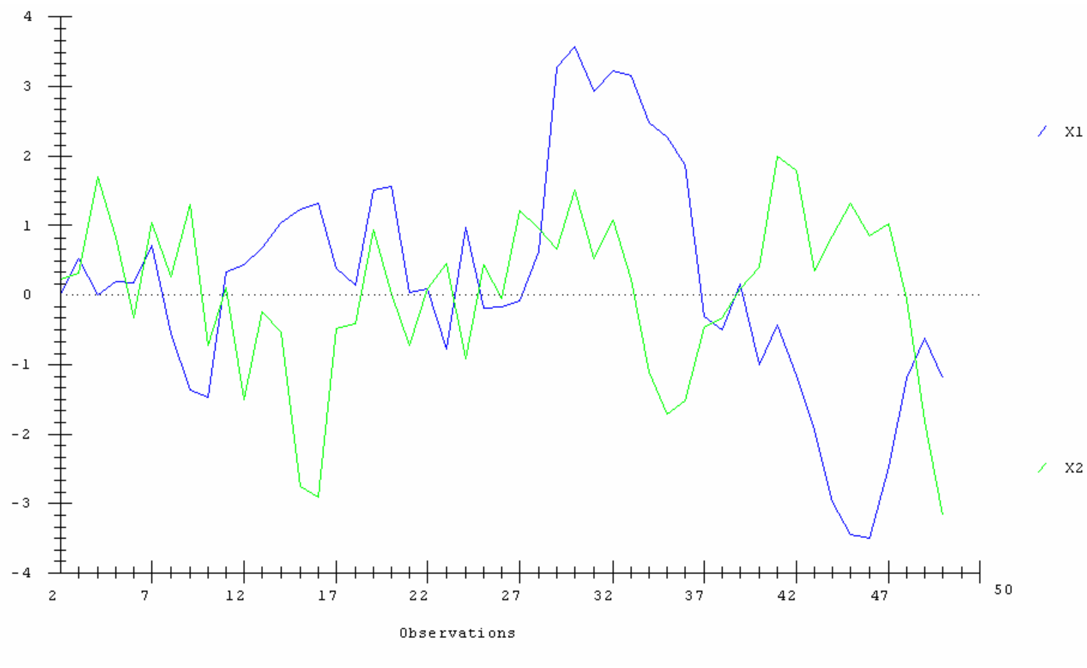

Figure 3.4: Two realizations of one stationary AR process.

Example 3.3:

In order to present a unit root process a random walk is generated**:**

$X_t = X_{t-1} + a_t$

Figure 3.5 shows two realizations. A random walk is nonstationary as the solution of the characteristic equation z-1=0 is z =1 and lies on the unit circle. It is a “unit root”. Later will be proved that despite X_t has a constant expectation its variance would be dependent on time.

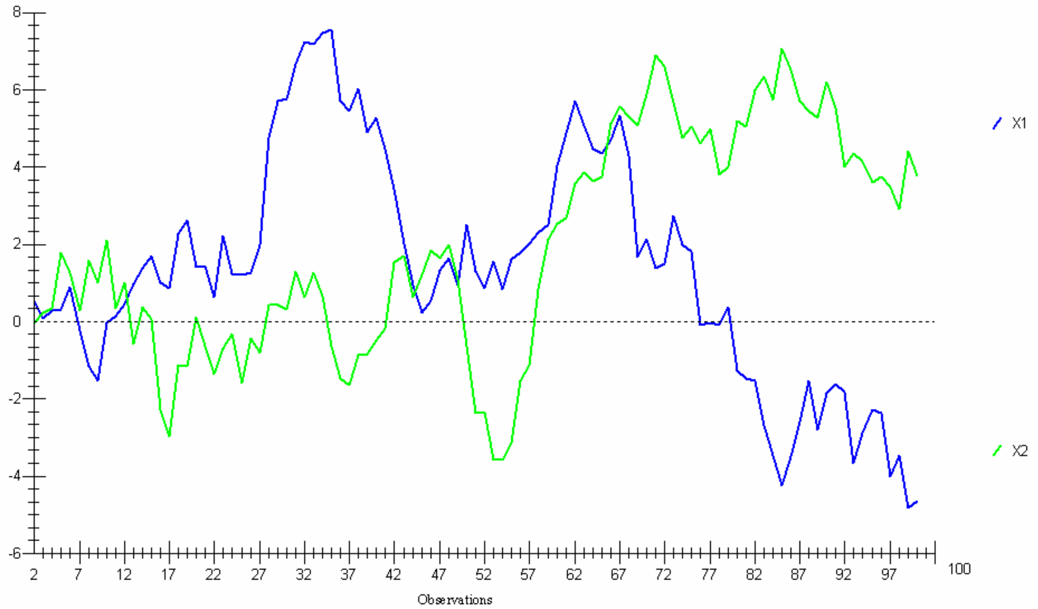

Figure 3.5: Two realizations of one random walk with zero mean.

3.3. ARMA Models

A mixture of AR and MA processes of orders p and q, respectively, is called an autoregressive moving average process [ARMA(p,q)]:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - ... - \theta_q a_{t-q} \quad (3.13)$$

or

$$X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p} = a_t - \theta_1 a_{t-1} - ... - \theta_q a_{t-q} \quad (3.14)$$

Here the single error term a_t of an AR process is substituted by an MA(q) process.

The ARMA(p,q) can be written shortly:

$$\Phi_p(L) X_t = \phi_0 + \Theta_q(L) a_t \quad (3.15)$$

where $\Phi_p(L)$ and $\Theta_q(L)$ are the lag operator functions of the corresponding AR(p) and MA(q) processes, respectively, and ϕ_0 is mostly assumed to be zero.

Example 3.4:

Figure 3.6 displays the graph of a time series. Here the data generating process is ARMA(1,2):

$X_t = 0.8X_{t-1} + a_t + 0.24a_{t-1} + 0.4a_{t-2}$

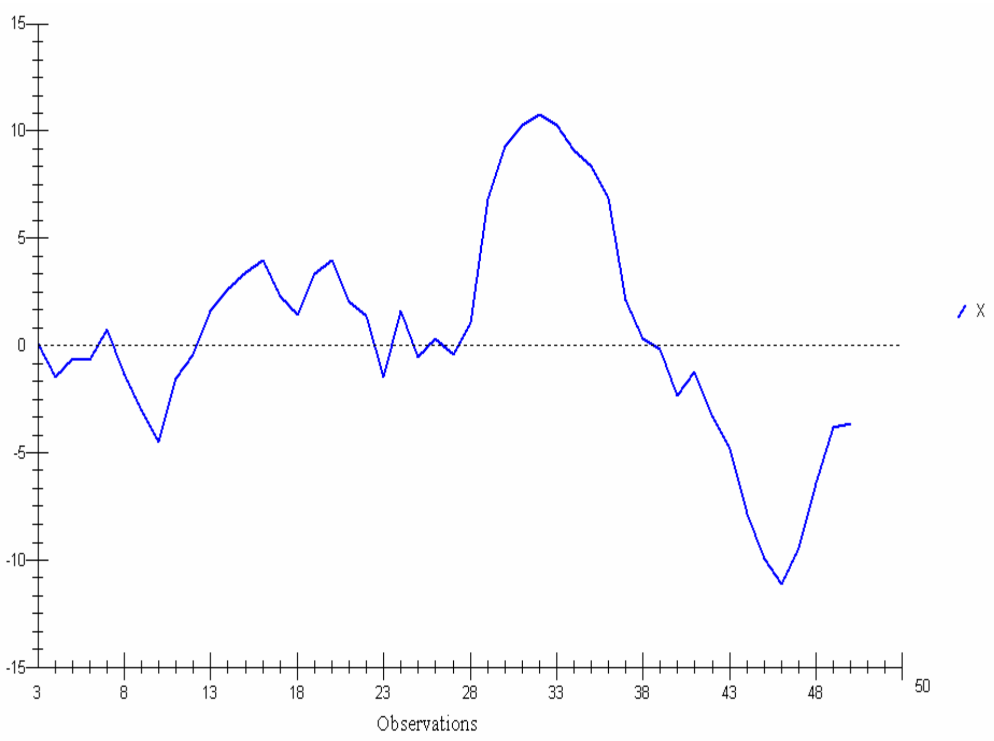

Figure 3.6: Graph of an ARMA(1,2) process

Under very general conditions, a stationary ARMA process $\Phi_p(\mathrm{L})X_t = \phi_0 + \Theta_q(\mathrm{L})a_t$ can be transformed in an infinite AR process as well as in an infinite MA process:

$$X_t = \phi_0 + a_t - \psi_1 a_{t-1} - \psi_2 a_{t-2} - \dots$$

or $X_t = \phi_0 + \Psi(\mathrm{L})a_t$ *(3.16)*

with $\Psi(\mathrm{L}) = 1 - \psi_1\,\mathrm{L} - \psi_2\,\mathrm{L}^2 - ...$ *(3.17)*

The infinite lag polynom $\Psi(\mathrm{L})$ is determined by

$$\Psi(\mathrm{L}) = \frac{\Theta_q(\mathrm{L})}{\Phi_p(\mathrm{L})}.$$

Particularly, stationary AR processes can be represented by infinite MA processes and most MA processes (under an invertibility condition) by infinite AR processes. In practical time series analysis, the representation with as few as possible parameters should be chosen.

Example 3.5:

Consider the MA(1) process

$$X_t = a_t - \theta a_{t-1} \qquad (3.18)$$

From $X_{t-1} = a_{t-1} - \theta a_{t-2}$

follows $a_{t-1} = X_{t-1} + \theta a_{t-2}$

and from $X_{t-2} = a_{t-2} - \theta a_{t-3}$

follows $a_{t-2} = X_{t-2} + \theta a_{t-3}$

and so further. By successive substitution of a_{t-1}, a_{t-2} and so on in (3.18) we obtain:

$$X_t = a_t - \theta\, X_{t-1} - \theta^2 X_{t-2} - ... \qquad (3.19)$$

i.e. an infinite AR process, which converges under the invertibility condition $|\theta|<1$.

Example 3.6:

Let us consider the stationary AR(1) process

$$X_t = \phi X_{t-1} + a_t \quad \text{with} \quad |\phi|<1 \qquad (3.20)$$

By backshifting the whole process we obtain

$$X_{t-1} = \phi X_{t-2} + a_{t-1}$$

and (3.20) becomes $X_t = a_t + \phi a_{t-1} + \phi^2 X_{t-2}$.

Further back shifting gives $X_{t-2} = \phi X_{t-3} + a_{t-2}$.

By successive substitution of these lagged elements of *X*, an infinite moving average process is produced:

$$X_t = a_t + \phi\, a_{t-1} + \phi^2 a_{t-2} + ... \qquad (3.21)$$

Because of $|\phi| < 1$ this representation converges.

ARMA processes have a more complex structure than pure AR or MA processes of the same behaviour, but they have less parameters. Parsimony is one of their advantages compared with fitted AR or MA processes.

4. Autocorrelation and Spectrum

4.1. Autocorrelation Function

According to (2.5) the autocorrelation function of a stationary process X_t is

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0} = \frac{1}{\gamma_0}\mathrm{E}[(X_t - \mu)(X_{t+\tau} - \mu)]. \qquad (4.1)$$

The graph of ρ_τ is called correlogram. The shape of the correlogram and other functions is characteristic for special ARMA processes. Therefore such functions are used in time series analysis for finding the type and order of a process and the corresponding model.
For an AR(p) process the correlogram is a mixture of exponential and sinuous curves.

Example 4.1:

Let X_t be an AR(1) process without constant and $|\phi_1| < 1$

From $X_t = \phi_1 X_{t-1} + a_t$

and $X_t^2 = (\phi_1 X_{t-1} + a_t)^2 = \phi_1^2 X_{t-1}^2 + 2\phi_1 X_{t-1} a_t + a_t^2$

follows $\gamma_0 = \mathrm{E}(X_t^2) = \phi_1^2 \gamma_0 + 0 + \sigma_a^2$

and by re-arranging

$$\gamma_0 = \frac{\sigma_a^2}{1 - \phi_1^2} \qquad (4.2)$$

For calculating γ_1 we consider

$$X_{t-1} = \phi_1 X_{t-2} + a_{t-1}$$

and $X_t X_{t-1} = (\phi_1 X_{t-1} + a_t) X_{t-1} = \phi_1 X_{t-1}^2 + X_{t-1} a_t$

Thus we obtain

$$\gamma_1 = \mathrm{E}(X_t X_{t-1}) = \phi_1\, \mathrm{E}(X_{t-1}^2) + 0 = \frac{\sigma_a^2}{1 - \phi_1^2}\phi_1$$

From this follows $\rho_1 = \phi_1$. In an analogous procedure we find $\rho_2 = \phi_1^2$.
Generally, a geometrically decreasing sequence will be obtained:

$$\rho_k = \phi_1^k. \qquad (4.3)$$

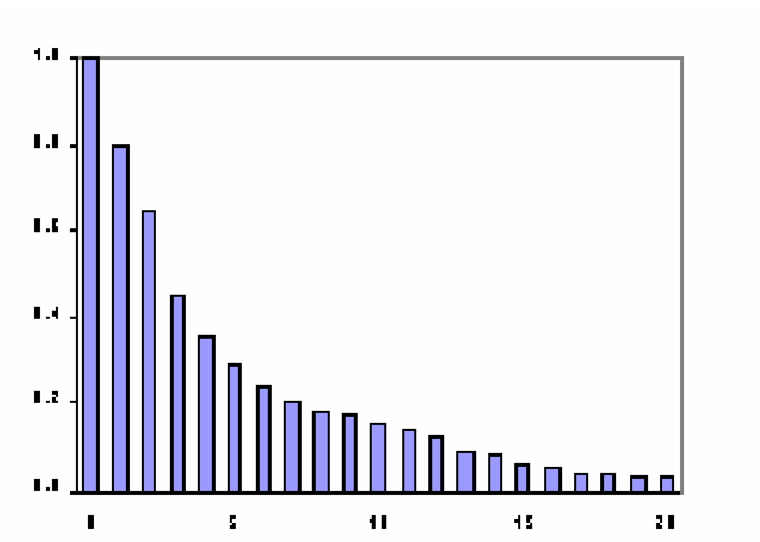

Figure 4.1: Autocorrelation function of an AR(1) process with $\phi_1 > 0$

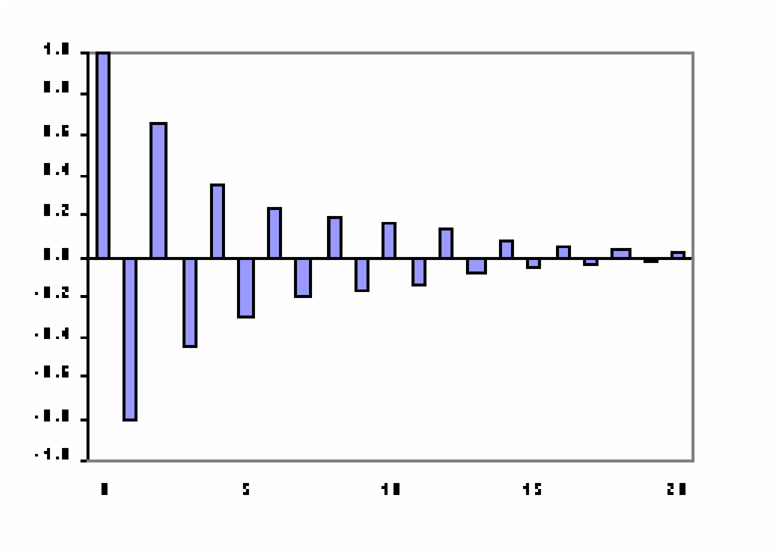

Figure 4.2: Autocorrelation function of an AR(1) process with $\phi_1 < 0$

Example 4.2:

Now, the autocorrelation of the MA(1) process $X_t = a_t - \theta_1 a_{t-1}$ is examined.

From $X_t^2 = a_t^2 - 2\theta_1 a_t + (\theta_1 a_{t-1})^2$
follows

$$\gamma_0 = \mathrm{E}(X_t^2) = (1+\theta_1^2)\sigma_a^2 \quad \text{with} \quad \sigma_a^2 = \mathrm{var}(a_t) = \mathrm{E}(a_t^2) \qquad (4.4)$$

from $X_{t-1} = a_{t-1} - \theta_1 a_{t-2}$

and $X_t X_{t-1} = (a_t - \theta_1 a_{t-1})(a_{t-1} - \theta_1 a_{t-2}) = a_t a_{t-1} - \theta_1 a_t a_{t-2} - \theta_1 a_{t-1}^2 + \theta_1^2 a_{t-1} a_{t-2}$

we obtain $\gamma_1 = \mathrm{E}(X_t X_{t-1}) = 0 - 0 - \theta_1 \sigma_a^2 + 0$

and dividing this by the variance γ_0:

$$\rho_1 = \frac{-\theta_1}{(1+\theta_1^2)} \tag{4.5}$$

From $X_t X_{t-2} = (a_t - \theta_1 a_{t-1})(a_{t-2} - \theta_1 a_{t-3}) = a_t a_{t-2} - \theta_1 a_t a_{t-3} - \theta_1 a_{t-1} a_{t-2} + \theta_1^2 a_{t-1} a_{t-3}$
follows because of the independence of the a_t, a_{t-1}…:
$\gamma_2 = \mathrm{E}(X_t X_{t-2}) = 0$,
All following autocovariances γ_3, γ_4,… and autocorrelations are equal zero, too.

The generalization of this result to processes of higher order gives us the opportunity to recognise the order q of an MA(q): it is determined by the highest number q of autocorrelation coefficients significantly differing from zero, while all following values are zero or close to it.

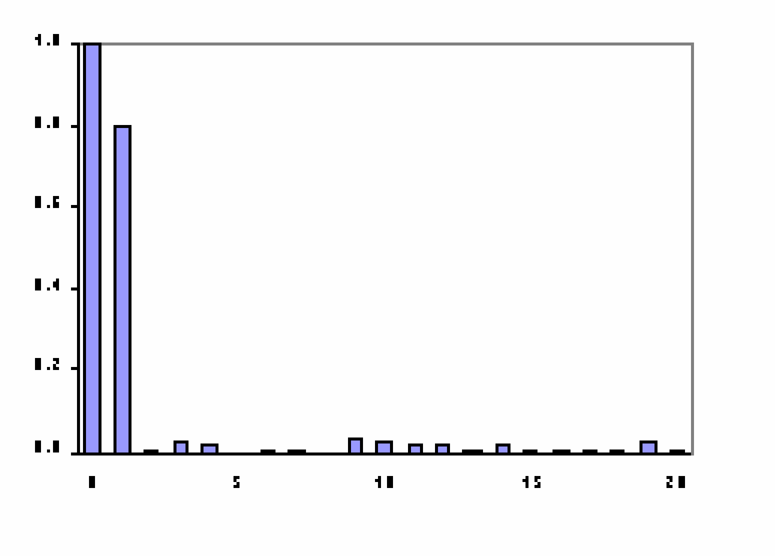

Figure 4.3: Correlogram of an MA(1) process with $\theta_1<0$

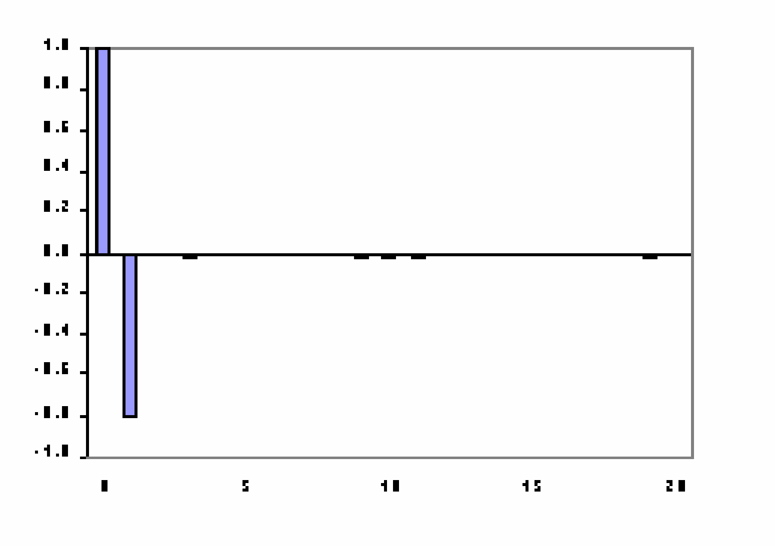

Figure 4.4: Correlogram of an MA(1) process with $\theta_1>0$

Suitable estimators for the autocovariance function of an ergodic process on the base of one time series x_t are

$$c_\tau = \hat{\gamma}_\tau = \frac{\sum_{t=1}^{T-\tau}(x_t - \bar{x})(x_{t+\tau} - \bar{x})}{T} \qquad (4.6)$$

or

$$c_\tau = \hat{\gamma}_\tau = \frac{\sum_{t=1}^{T-\tau}(x_t - \bar{x})(x_{t+\tau} - \bar{x})}{T - \tau} \qquad (4.7)$$

Both estimators are mentioned here because they are used in different text books and computer packages and have slightly different properties for short time series only.
The estimator for the autocorrelation function is the sample autocorrelation function

$$r_\tau = \hat{\rho}_\tau = \frac{\hat{\gamma}_\tau}{\hat{\gamma}_0} = \frac{c_\tau}{s_x^2} \qquad (4.8)$$

with s_x being the sample standard deviation of the time series x_t.

Example 4.3:

Let x_t be a time series generated by the process

$X_t = 0.8\, X_{t-1} + a_t$ with $a_t \sim N(0,1)$.

Figure 4.5 shows the sample autocorrelation function of x_t.

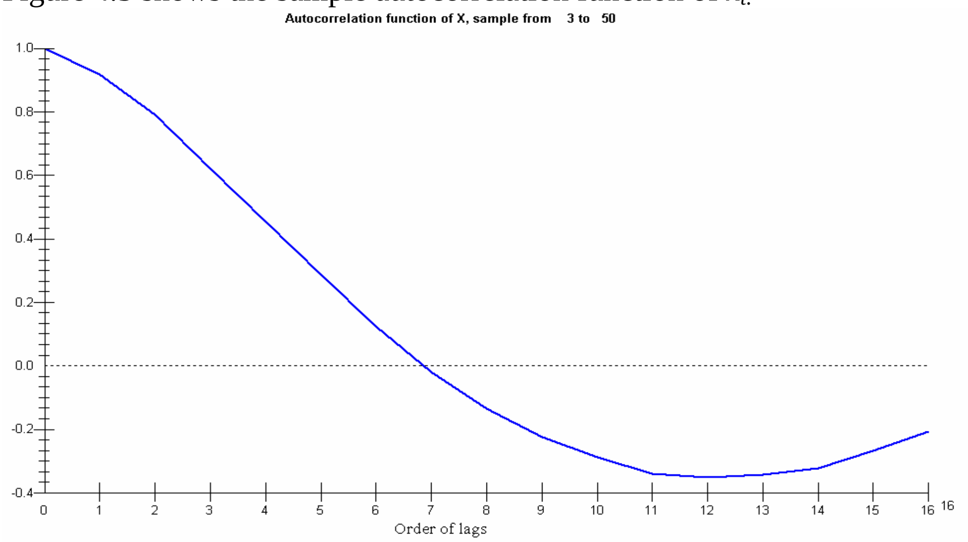

Figure 4.5: Sample autocorrelation function of one realisation of the AR(1) process

The Box-Pierce statistic (2.7) results in values from 35 to 114 for τ =1 to 16, respectively. This allows to reject the null hypothesis of white noise on the 1% significance level.

Example 4.4:

Now, z_t is a time series generated by the MA(2) process

$Z_t = a_t + 0.25a_{t-1} + 0.4a_{t-2}$ with $a_t \sim N(0,1)$.

Figure 4.6 displays the sample autocorrelation function. The values of the function beyond the lag τ =2 are close to zero. This indicates the MA(2) process.

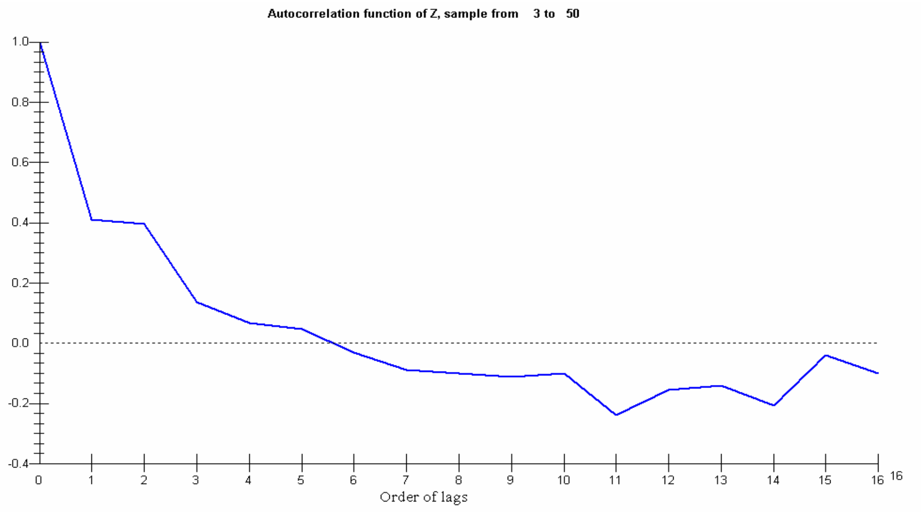

Figure 4.6: Sample correlogram of an MA(2) process.

4.2. Partial Autocorrelation Function

Another diagnostic function is the partial autocorrelation function (PAC) of a stationary stochastic process.
For calculating this function we assume X_t approximated by an AR(τ) process:

$$X_t^{(\tau)} = \phi_{1\tau}\, X_{t-1}^{(\tau)} + \ldots + \phi_{\tau\tau}\, X_{t-\tau}^{(\tau)} \qquad (4.9)$$

Then the last coefficient $\phi_{\tau\tau}$ is referred to as *partial autocorrelation coefficient* of X_t for the lag τ.
The series $\rho_{part}(\tau) = \phi_{\tau\tau}$ with varying τ is called *partial autocorrelation function* (PAC).

For an AR(p) process $\rho_{part}(\tau)$ is equal zero beyond the lag $\tau = p$.

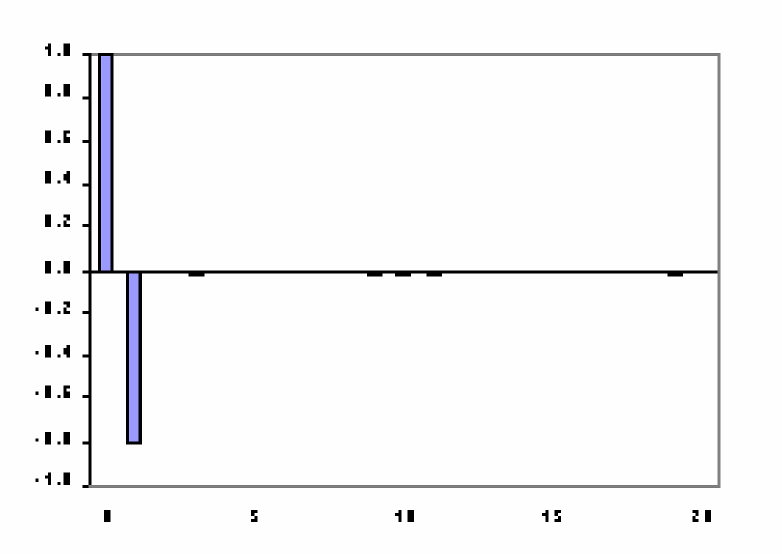

Figure 4.7: Partial autocorrelation function of an AR(1) process with $\phi_1 < 0$

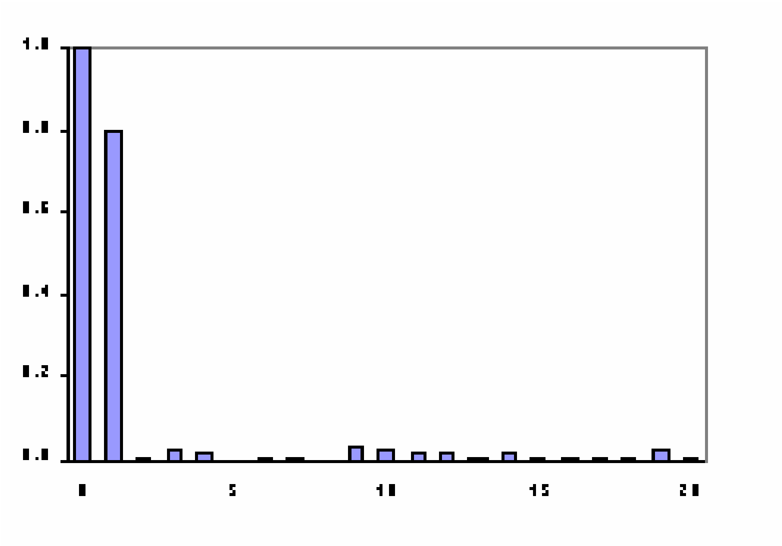

Figure 4.8: Partial autocorrelation function of an AR(1) process with $\phi_1 > 0$

On the other hand, for an MA(q) process the PAC is an exponentially decreasing sequence.

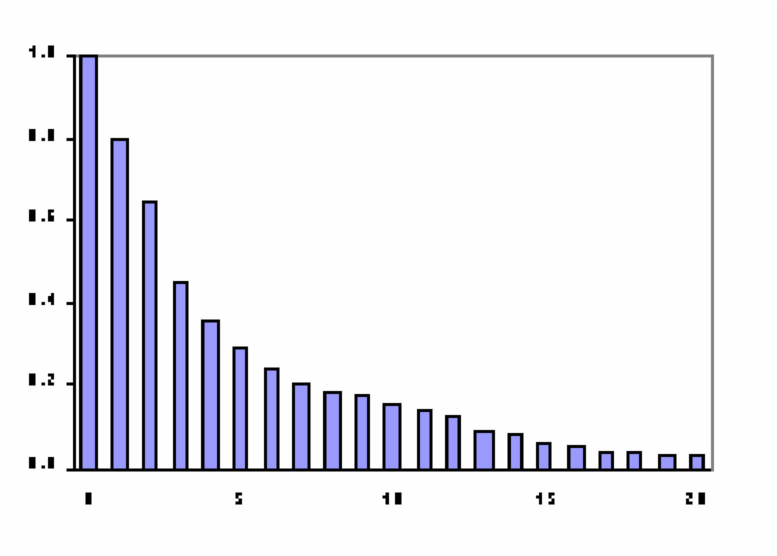

Figure 4.9: Partial autocorrelation function of an MA(1) process with $\theta_1 < 0$

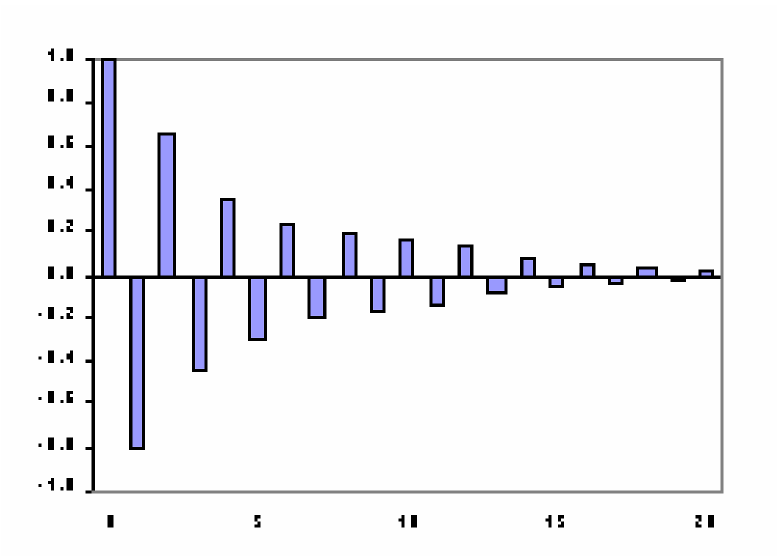

Figure 4.10: Partial autocorrelation function of an MA(1) process with $\theta_1 > 0$

The value of the PAC for a time series given at a selected lag τ can be estimated by OLS fitting an AR(τ) model and taking the estimated highest order coefficient $\hat{\phi}_{\tau\tau}$.

Example 4.5:

Let x_t be a time series generated by the AR(1) process $X_t = 0.8X_{t-1} + a_t$ with unit variance white noise a_t. The linear regression of x_t on x_{t-1} gives the estimated coefficient 0.791 with the standard error 0.023. OLS regression on x_{t-1} and x_{t-2} produces the estimation of ϕ_{22} in table 4.1. Then the regression of x_t on x_{t-1}, x_{t-2}, x_{t-3} results in table 4.2 and so on.

Table 4.1: OLS regression of x_t

****

Regressor	Coefficient	Standard Error
x_{t-1}	.89002	.15171
x_{t-2}	**-.057163**	.15181

****

Table 4.2: OLS regression of x_t

****

Regressor	Coefficient	Standard Error
x_{t-1}	.88283	.15609
x_{t-2}	.0029261	.21856
x_{t-3}	**-.063269**	.16234

****

Thus the first three values of the PAC estimated are

$\rho_{part}(1) = 0.791$

$\rho_{part}(2) = -0.057$

$\rho_{part}(3) = -0.063$

It is recognizable that the graph of ρ_{part} would drop down to approximately zero after the lag 1. This is characteristic for AR(1) processes.

4.3. Spectral Density

The spectral density or the power spectrum is the Fourier transform of the autocovariance function or of the autocorrelation function, e.g.

$$p(f) = 2(\gamma_o + 2\sum_{\tau=1}^{\infty} \gamma_\tau \cos 2\pi f \tau); \tag{4.10}$$

where $f = \frac{1}{P}$ is the frequency and P the period length of a supposed periodic component within the process.

The value $p(f)$ can be interpreted as the amplitude of this periodic cycle with $0 \leq f \leq \frac{1}{2}$.

The lower limit f=0 means an infinite period length, e.g. the trend component, and the maximum frequency f=0.5 means extremely short oscillations.

The function $p(f)$ as a whole distributes the variance to variations with frequencies between 0 and $^1/_2$:

$$\sigma_x^2 = \gamma_0 = \int_0^{1/2} p(f)df \,. \tag{4.11}$$

On the base of one empirical time series the estimator for the spectral density is the sample spectral density:

$$\hat{p}(f) = 2(c_0 + 2\sum_{\tau}^{T} g_\tau c_\tau \cos 2\pi f \tau) \tag{4.12}$$

where the empirical sample autocovariances are weighted by a suitable 'window' g_τ in order to obtain consistent estimations. Computer programs offer several spectral windows, mostly those by Parzen, Hannen or Bartlett.
Some programs use the „circular frequency" $\omega=2\pi f$ instead of the frequency f. Then the special density is defined for $\omega=0$ to $\omega=3.14$.

If the spectrogram, i.e. the plot of the spectral density over all frequencies under consideration, shows for a special frequency f or ω a high pique then the process contains a periodical component of the period length $P=\frac{1}{f}$ and the share of the variance covered by this component totals to the share of the area concentrated under this pique. Therefore an important application field of spectral analysis is the analysis of cyclical variations.

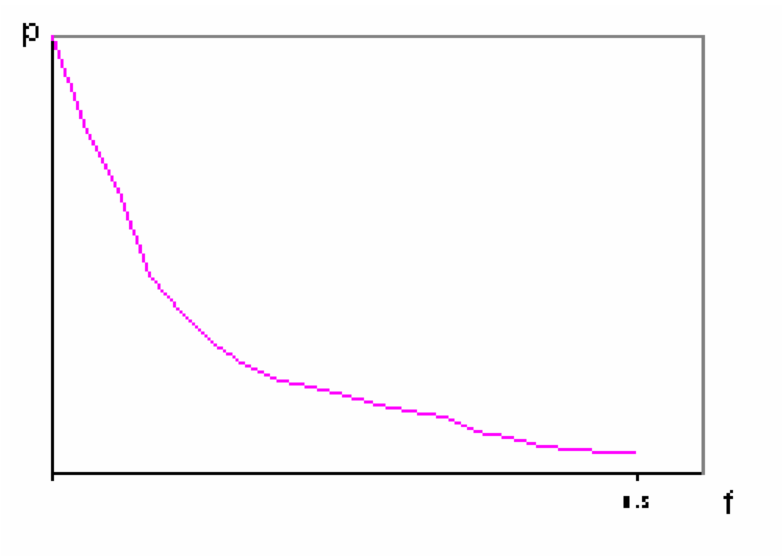

Figure 4.11: Spectral density of an AR(1) process with $\phi_1>0$

Figure 4.11 shows the typical spectral density function of an AR(1) process with positive coefficient. The spectral power is concentrated at the frequency zero, i.e. the behaviour of the process is interpreted by the spectrum as an infinitely long periodic movement. In the same way, the oscillations of a process with negative ϕ_1 occur in the spectrogram as a pique at the frequency $f=\frac{1}{2}$ or $\omega = \pi$ what means the dominance of periodic oscillations with the minimum length of 2 time units (see figure 4.12).

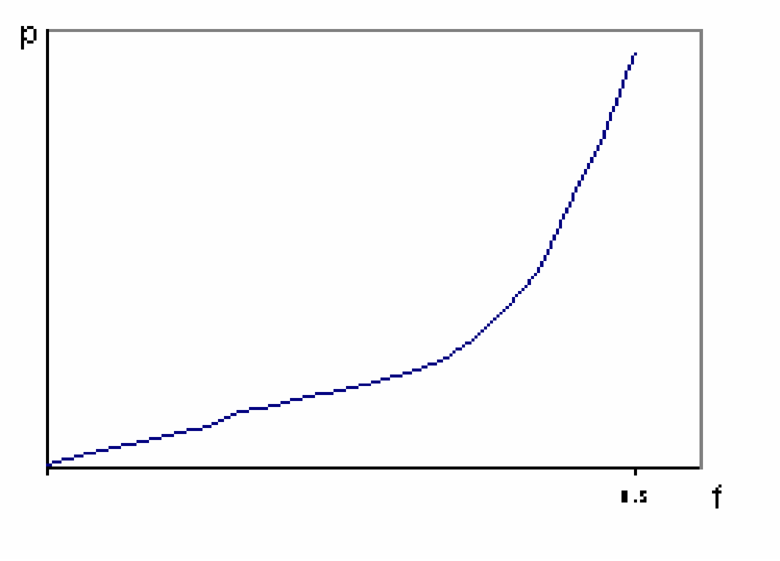

Figure 4.12: Spectral density of an AR(1) process with $\phi_1 < 0$

The spectral shapes of the corresponding MA(1) processes are similar, but the spectral power is not so sharply concentrated at the points 0 or $^1/_2$ respectively. That means the variations concentrated in MA processes are distributed over a broader what is called band width than in AR processes.

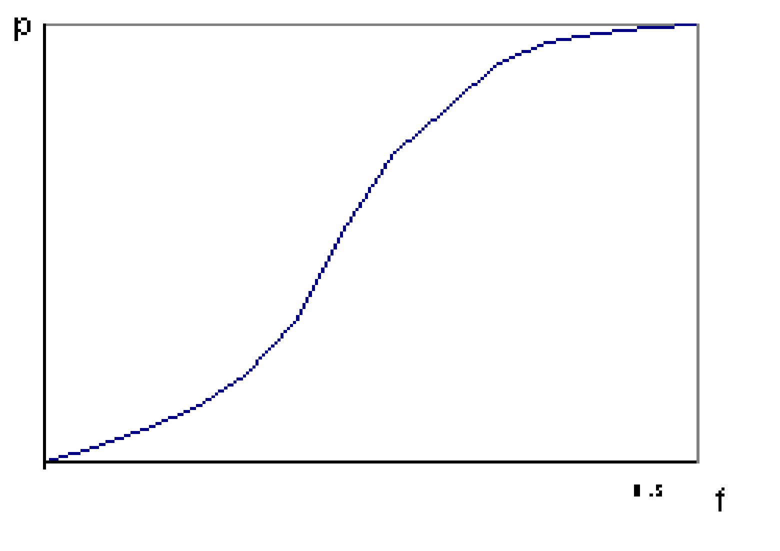

Figure 4.13: Spectral density of an MA(1) process with $\theta_1 > 0$.

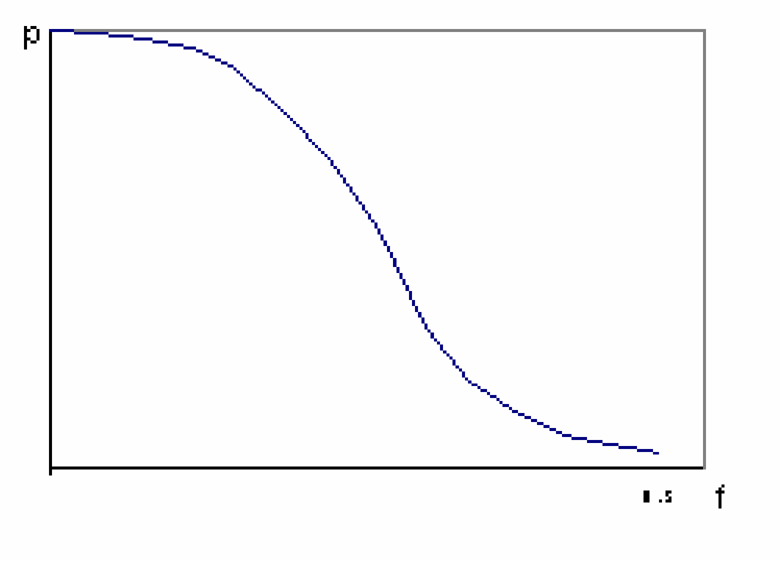

Figure 4.14: Spectral density of an MA(1) process with $\theta_1 < 0$.

While the curves at figures 4.11 to 4.14 display the theoretical spectral density of AR(1) and MA(1) processes the examples 4.6 and 4.7 present sample spectral densities of time series.

Example 4.6:

Let z_t be a time series generated by the MA(2) process $Z_t = a_t + 0.25a_{t-1} + 0.4a_{t-2}$ where a_t is zero mean white noise with variance 2. Two of the spectral density estimators indicate a pique at $\omega = 0.2$. But because of the large variation between the different estimators in the area around zero, this can be assumed as a random sample effect and it is rather a signal for a very flat extremum at $\omega = 0$.

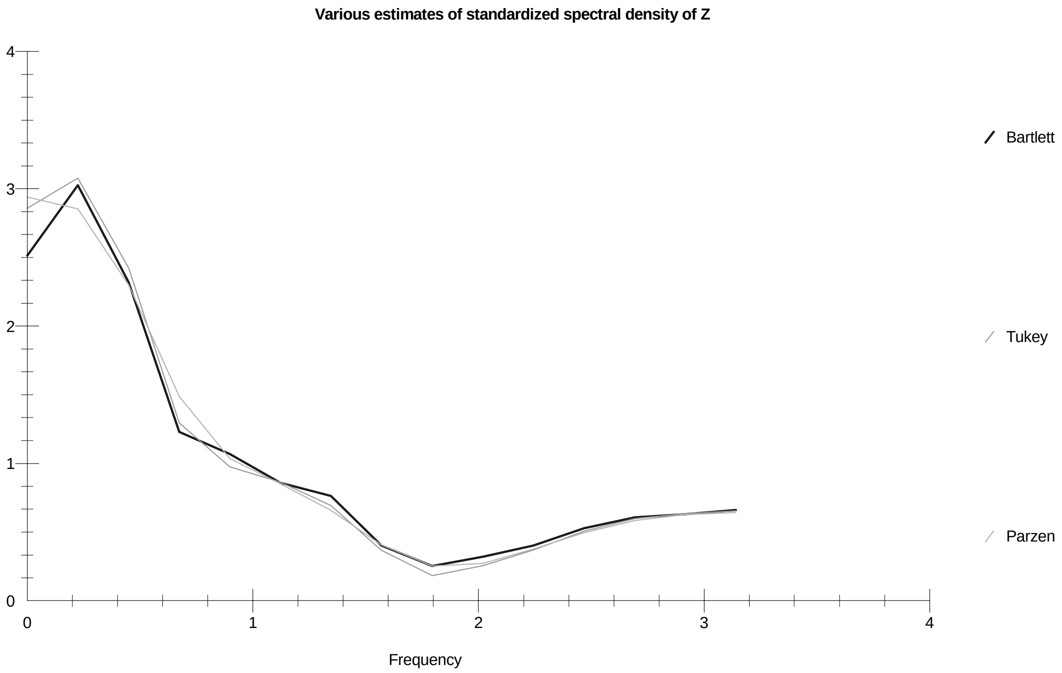

Figure 4.15: Spectral density of an MA(2) process

Example 4.7:

For a stationary time series of quarterly data x_t with a dominant seasonality the spectral density has been estimated. We find the highest peak of the graph in figure 4.16 in the very centre of the frequency range i.e. at circular frequency $\omega=\pi/2$. It indicates f=¼ and a period length P= 4 quarters i.e. one year as previously expected for a seasonal component.

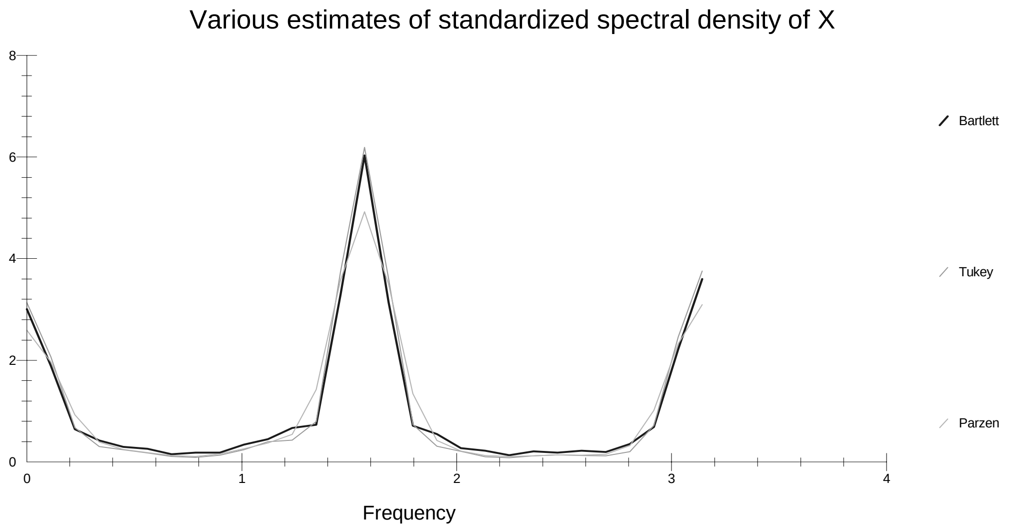

Figure 4.16: Spectral density of quarterly seasonal time series

Example 4.8:

Figure 4.17 displays the development of the total value *OR* of monthly incoming orders for the construction industry in East Germany after the unification of Germany. Besides the changing trend there seem to be several periodicities. The graph of the spectrum in figure 4.18 confirms the existence of very long waves or a trend by a peak near frequency 0. Furthermore one can find minor peaks at circular frequencies ω equal 0.52 and 2.1 what indicates periodicities with length 12 and 3 month, respectively.

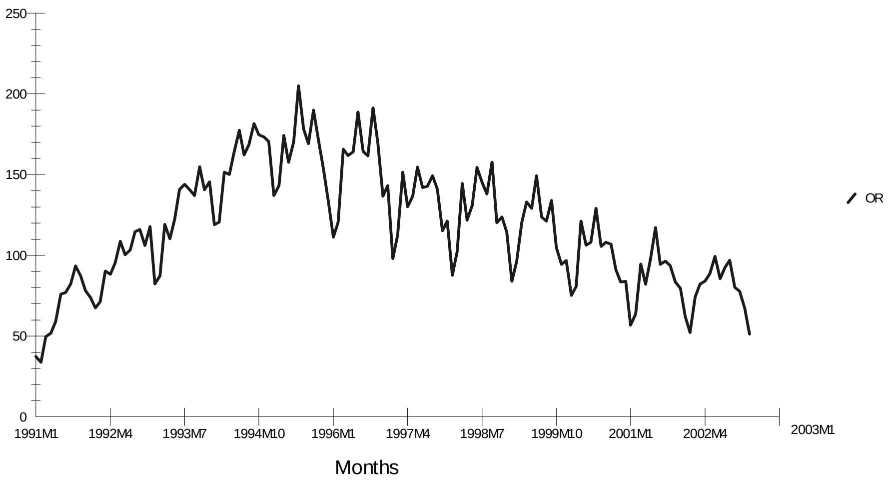

Figure 4.17: Orders to East German construction industries

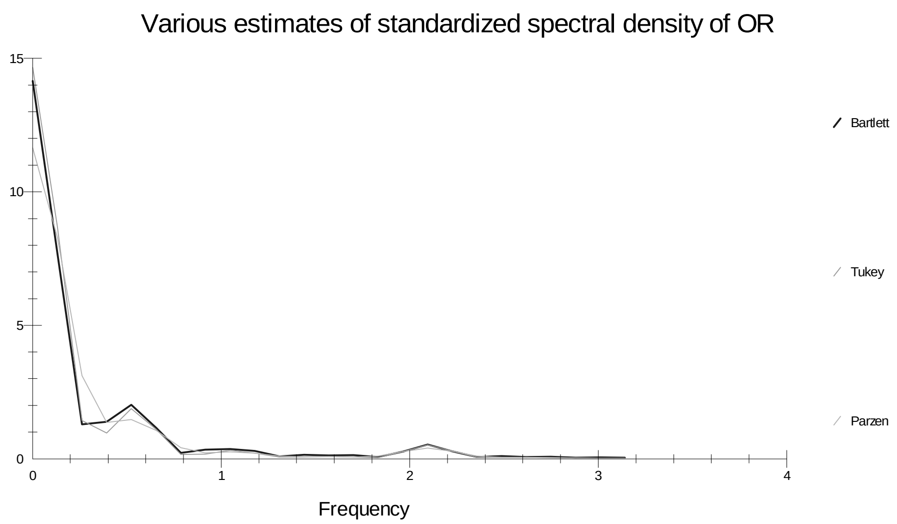

Figure 4.18: Spectral density of the orders to East German construction industries

In order to examine periodic fluctuations more detailed it can be useful to consider the increase rate of a variable. The increase rate of a time series x_t is

$$\frac{x_t - x_{t-1}}{x_{t-1}} \qquad (4.13)$$

that can be approximated by the logarithmic increase rate for comparatively small changes:

$$\Delta \ln x_t = \ln x_t - \ln x_{t-1} = \ln\left(\frac{x_t}{x_{t-1}}\right). \qquad (4.14)$$

The next figure shows the increase rate Δln*OR* of the order value *OR*. The trend is eliminated and the periodicities seem to be amplified.

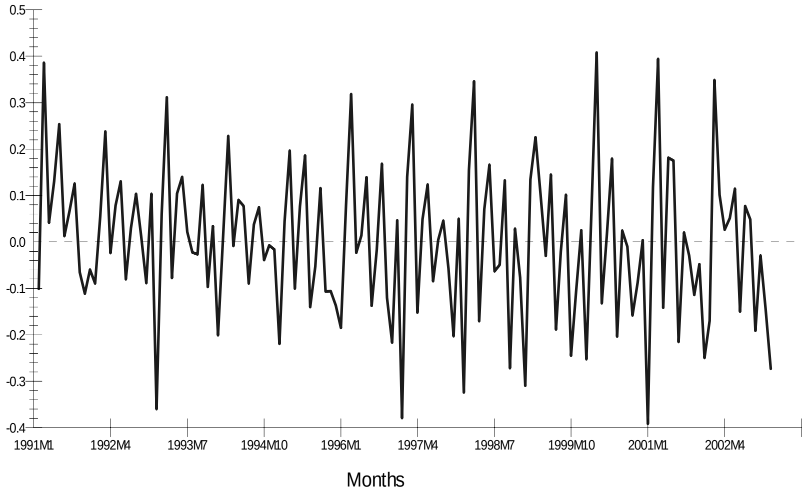

Figure 4.19: Growth rate of the orders to the construction industries

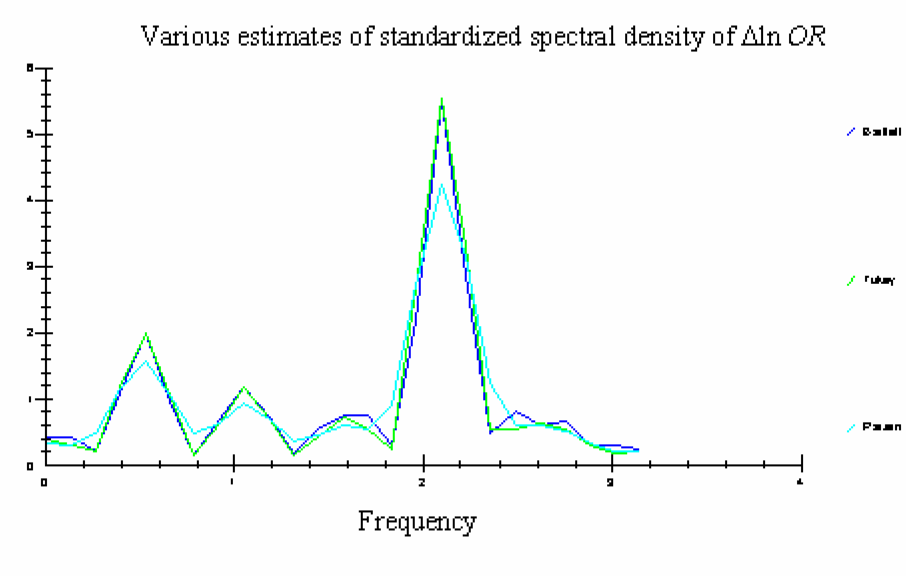

Figure 4.20: Spectral density of the growth rate of the orders

Figure 4.20 and table 4.3 show the spectral density of the growth rates. Surprisingly, now the highest peak occurs at the circular frequency ω =2.1 i.e. the three month periodicity. That means the biggest part of the increase variance is produced by regular fluctuations within the quarters. We find a lower peak at ω =5.2 what indicates an additional 12-month seasonality with a minor share at the variance.

Table 4.3:

Standardized spectral density functions of $\Delta \ln OR$, sample 1991M2 to 2003M1

Circular frequency	Period	Bartlett	Tukey	Parzen
0.00	Inf.	.41414	.40327	.34745
.13090	48.0000	.42359	.28945	.31183
.26180	24.0000	.21619	.22185	.47551
.39270	16.0000	1.0765	1.1668	1.1293
.52360	**12.0000**	**1.9989**	**2.0018**	**1.5528**
.65450	9.6000	.97483	1.0439	1.0502
.78540	8.0000	.15450	.13615	.48733
.91630	6.8571	.68276	.63810	.65174
1.0472	**6.0000**	**1.1892**	**1.1760**	**.91865**
1.1781	5.3333	.70727	.68282	.67360
1.3090	4.8000	.16923	.15761	.36567
...				
1.8326	3.4286	.30370	.24614	.91571
1.9635	3.2000	2.1926	2.6362	2.7224
2.0944	**3.0000**	**5.5038**	**5.5131**	**4.2373**
2.2253	2.8235	2.8631	3.2895	3.1299
2.3562	2.6667	.49168	.56624	1.2268
...				
3.1416	2.0000	.24909	.21564	.20655

5. Integrated Processes

5.1. Nonstationary time series

If a time series contains a development tendency, the assumption of constant mean or constant variance is violated. Then this time series would be considered a realisation of a nonstationary process.
Practically, a time series x_t should be subjectively judged for nonstationarity by means of its graph and its correlograms. If there are found

- trend
- Strongly deterministic periodicities
- Systematically varying variance
- Changing autocorrelation,

then there would be good reason to assume the underlying process to be nonstationary. Theoretically it is nonstationary if the mean or the variance or the covariance of the generating process change by time. As a rule, most processes representing real economic phenomena prove nonstationary because economics grow or decrease or change in some other way.

5.2. Differentiation and Integration

On the one hand, most economic time series are nonstationary. On the other hand, many methods and models demand stationary time series.
In many cases differentiation of the time series is a successful approach to obtaining stationary time series.

The first differences of a stochastic process are
$(1-\mathrm{L})X_t = \Delta X_t = X_t - X_{t-1}$

Or for a seasonal process with period length *s*:
$(1-\mathrm{L}^s)X_t = \Delta_s X_t = X_t - X_{t-s}$

If the first differences of X_t are stationary, than X_t is termed integrated of first order.
Else further differentiation will lead to the second differences
$(1-\mathrm{L})^2 = \Delta^2 X_t = \Delta X_t - \Delta X_{t-1}$

If this is stationary, than X_t is called integrated of second order. If we obtain the first stationary result after *k*-fold differentiation, the process is said to be integrated of *k*-th order.
A time series generated by a *k*-th order integrated process is said to be integrated of *k*-th order as well.

Example 5.1:

Let $X_t=a_t$ be white noise. Obviously it is at least weakly stationary, because $\mathrm{E}(a_t) = \mu$ and $\mathrm{var}(a_t)=\sigma_a^2$ are constant. As the elements a_t, $a_{t-1},\dots$ of the process are defined to be independent, the covariance

$$\operatorname{cov}(a_t, a_{t-\tau}) = \mathrm{E}([a_t - \mu][a_{t-\tau} - \mu]) \tag{5.1}$$

would be constantly zero, i.e. not depending on time.

5.3. Random Walk

The process $X_t = X_{t-1} + a_t$ *(5.2)*

with white noise a_t is referred to as random walk. It is the most elentary case of a nonstationary process. Its characteristic function has the unit root $z=1$.

The mean of x_t $\mathrm{E}(x_t) = \mathrm{E}(x_{t-1}) + \mathrm{E}(a_t) = \mu$ is constant.

The nonstationarity can be proved only by considering the variance:

$$\begin{aligned} &\operatorname{var}(X_t) = \operatorname{var}(X_{t-1}) + \operatorname{var}(a_t) \\ &= \operatorname{var}(X_{t-2}) + \operatorname{var}(a_t) + \operatorname{var}(a_t) \\ &= \ldots \\ &= t\sigma_a^2 \end{aligned} \tag{5.3}$$

That means $\operatorname{var}(X_t)$ depends on time t. This property of X_t is what is called variance nonstationarity.

The first differences of X_t are white noise a_t and stationary: $\Delta x_t = x_t - x_{t-1} = a_t$. Thus the random walk is integrated of first order.

We know: the seemingly very similar process $Y_t = 0{,}999Y_{t-1} + a_t$ with the root $z = 1{,}001$ outside the unit circle is stationary. Thus one of the most serious problems of time series analyse is the distinction between time series x_t and y_t as random realizations of an unit root process X_t and a stationary process Y_t, respectively. Helpful instruments for this purpose are unit root tests.

5.4. Unit Root Tests

Dickey-Fuller-Test

The most widespread method for checking whether or not a process is stationary was the unit root test developed by Dickey and Fuller til 1979. The basic idea consists in the assumption that the process is a random walk, i.e. nonstationary, and the possible rejection of this hypothesis. For the Dickey-Fuller-Test the process X_t is approximated by an AR(1) process:

$$X_t = (\phi_0 +) \quad \phi_1 X_{t-1} + a_t \tag{5.4}$$

The intercept (ϕ_0) is set in brackets because it is often assumed zero.

The nontstatitonarity hypothesis to be tested

H_0: $\phi_1 = 1$

means X_t is a unit root, i.e. nonstationary, whereas the alternative

H_1: $\phi_1 < 1$

means stationarity.

For easier handling the test the process can be transformed into its 1st differences:

$$\Delta X_t = X_t - X_{t-1} = (\phi_0 +)(\phi_1 - 1) X_{t-1} + a_t$$
$$\Delta X = (\phi_0 +)\gamma X_{t-1} + a_t \qquad (5.5)$$
$$\text{where } (\phi_1 - 1) = \gamma$$

The intercept ϕ_o is often assumed zero
Now the hypotheses show the more common shape of a one-side t-test of a regression coefficient:

$$H_0 : \gamma = 0$$
$$H_1 : \gamma < 0$$

Practically, the following steps have to be performed:

1. Backshifting x_t to the lagged series x_{t-1}
2. Calculation of the 1st differences Δx_{t-1}
3. OLS regression of Δx_t (dependent variable) on x_{t-1} (independent variable):

$$\Delta x_t = (f_0 +) \quad g x_{t-1} + e_t \qquad (5.6)$$

with the estimates f_0 and g and the residuals e_t.

4. Calculating the empirical value of the test variable:

$$t^{emp} = \frac{g}{sdv(g)} \qquad (5.7)$$

If the true intercept $\phi_0 = 0$ then t^{emp} is asymptotically standard normal distributed, else

5. calculating the critical value: If there is no deterministic trend, the 5% critical values are approximately

$$t_{0.95}^{DF} = -2{,}86 - \frac{2{,}74}{T} - \frac{8{,}36}{T^2} \qquad (5.8)$$

(it is always negative). Otherwise they can be obtained out of tables (e.g. Eckey).

6. If $t^{\text{emp}} < {}^{DF}{}_{1-\alpha}$ then the null hypothesis can be rejected on the α significance level and X would be stationary.

Example 5.2:

Let y_t be a realisation of a white noise process with $\sigma_y^2 = 2$ and $\mu = 0$

For performing a DF-Test the first differences Δy_t of the time series are put in a linear regression relationship on the lagged time series y_{t-1} without an intercept (table 5.1).

Table 5.1: Ordinary Least Squares Estimation

****

Dependent variable is Δy_t

49 observations used for estimation from 2 to 50

****

Regressor	Coefficient	Standard Error	*t*-Value
y_{t-1}	-1.1353	.14202	-7.99

****

Because of the lack of an intercept the empirical *t*-value can be compared with the left-sided 5% quantile of the normal distribution, i.e. $z_{0.05}$ = -1.65. The *t*-value –7.99 indicates rejection of the zero (nonstationarity) hypothesis. That means the process is to be considered stationary.

While in the last example the intercept was dropped the next test takes the existence of a constant a_0 into consideration:

Example 5.3:

Let us consider the same time series but without having the a-priori knowledge about the zero intercept. Then we are to include an intercept term in the regression (table 5.2).

Table 5.2: Ordinary Least Squares Estimation

****

Dependent variable is Δy_t

49 observations used for estimation from 2 to 50

****

Regressor	Coefficient	Standard Error	*t*-Value
C	.49606	.25257	1.964
y_{t-1}	-1.1353	.14202	-8.46

****

Because of the intercept the normal distribution is not applicable. Formula 5.9 gives the 5% critical value.

$$t_{0.95}^{DF} = -2{,}86 - \frac{2{,}74}{50} - \frac{8{,}36}{2500} = -2{,}918$$

The test results in favour for stationarity, too.

In the following example the assumption on nonstationarity should not be rejected because we have chosen and generated by simulation a unit root process:

Example 5.4:

A random walk is to be examined. Let x_t be a realization of the process

$X_t = X_{t-1} + a_t$ where a_t is normal with $E(a_t)=0$ and $\sigma_a^2 = 1$. For the DF-Test again a regression of Δx_t on x_{t-1} without an intercept is estimated:

Table 5.3: Ordinary Least Squares estimation

****

Dependent variable is Δx_t

49 observations used for estimation from 2 to 50

****

Regressor	Coefficient	Standard Error
x_{t-1}	.026941	.023797

****

The coefficient is positive. Therefore the *t*-statistic cannot be smaller than any (always negative) critical value. Thus, nonstationarity cannot be rejected at any significance level.

Augmented Dickey-Fuller Test

In the above examples, fitting the processes by AR(1) was absolutely correct according to our knowledge of the generating processes. But in praxi, assuming an AR(1) process can be a gross simplification. Better and more general would be to allow for an AR(p) representation of the errors in (5.9):

$$\Delta X_t = (\phi_0 +)\ \gamma X_{t-1} + \phi_1 \Delta X_{t-1} + \phi_2 \Delta X_{t-2} + \ldots + \phi_p \Delta X_{t-p} + \varepsilon_t \qquad (5.10)$$

with white noise e_t . The hypotheses to be tested are the same as for the DF-Test

$H_0 : \gamma = 0$ (for a unit root)

$H_1 : \gamma < 0$ (for stationarity)

Then γ will be estimated by OLS simultaneously with the ϕ's. The t-test of γ is performed in the same way as in the DF-Test. The critical values for the t-value of the estimate g for γ are equal to those of Dickey-Fuller. This improved unit root test is called Augmented Dickey-Fuller Test (ADF).

The order p of the AR(p) process in (5.13) can be found by the Akaike Information Criterion in the error variance form

$$AIC_\sigma = \{1 + \ln(2 \cdot \pi)\} + \ln\left(\frac{\sum \hat{\varepsilon}^2}{T}\right) + \frac{2 \cdot p}{T} \qquad (5.11)$$

or in the likelihood form

$$AIC_L = l_{max}(T,p) - p \qquad (5.12)$$

where $l_{max}(T,p)$ is the logarithmic likelihood of the model estimated.

The value of p with minimum AIC_σ or maximum AIC_L is to be taken.

As an alternative, the Schwarz Criterion in its two forms SC_σ or SC_L can be used in an analogous way

$$SC_\sigma = \{1+\ln(2\cdot\pi)\} + \ln\left(\frac{\sum\hat{\varepsilon}^2}{T}\right) + \frac{p\cdot\ln T}{T} \qquad (5.13)$$

or

$$SC_L = l_{\max}(T,p) - \frac{p}{2}\ln T\,. \qquad (5.14)$$

Example 5.5:

Let y_t be a realization of the ARMA(2,1) process:

$Y_t = 0.9Y_{t-1} - 0.3Y_{t-1} + a_t + 0.25a_{t-1}$

with $E(a_t) = 0$ and $\sigma_a^2 = 4$.

Table 5.4: Unit root tests for variable *Y*.

The Dickey-Fuller Regressions include an intercept but not a trend
****
46 observations used in the estimation of all ADF regressions
Sample period from 5 to 50
****

	Test Statistic	LL	AIC_L	SC_L
DF	-2.1556	-132.3946	-134.3946	-136.2233
ADF(1)	**-3.3536**	-127.1521	**-130.1521**	**-132.8951**
ADF(2)	-2.5808	-126.7647	-130.7647	-134.4220
ADF(3)	-2.0708	-126.4142	-131.4142	-135.9858

****
95% critical value for the augmented Dickey-Fuller statistic = -2.9256
LL = Maximized log-likelihood AIC_L = Akaike Information Criterion
SC_L = Schwarz Bayesian Criterion

According to the higher values of both the ACI_L and the SC_L, the augmented Dickey-Fuller Test ADF(1) on the base of an AR(1) model for the first differences is to be preferred to the others. The corresponding value of the test statistic is smaller than the critical value. That means the zero hypothesis (nonstationarity) is to be rejected on the 5% level. The simple Dickey-Fuller Test would not allow for this decision.

Example 5.6:

Figure 5.1 presents the graph of the daily closing value of the Russian share price index Moscow Times from 1st July 1997 by 6th May 2003. It can be guessed from the growing curve that the time series is not stationary. This assumption would be confirmed by an augmented Dickey-Fuller Test as shown in table 5.5. The H_0 of nonstationarity can neither by DF nor by ADF be rejected. Because of the unanimity of the decision of the tests for nonstationarity there is no need for model choice according to ACI_L or SC_L.

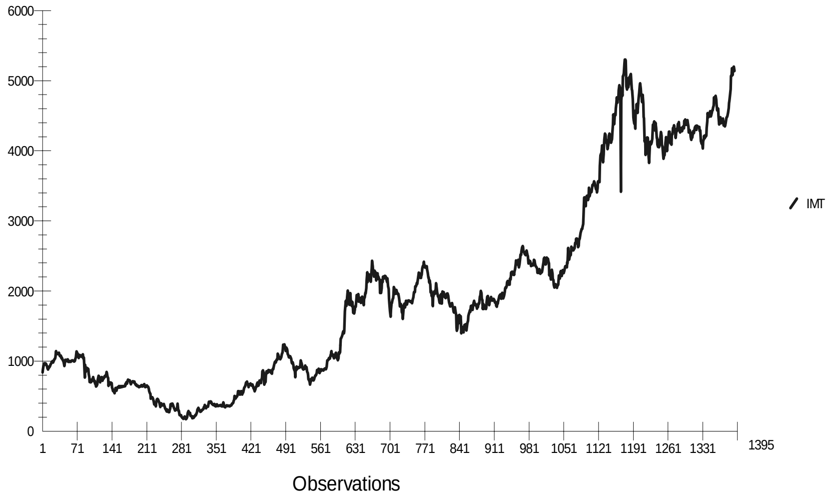

Figure 5.1: The Moscow Times share price index from 1st July 1997 by 6th May 2003

Table 5.5: Unit root tests for variable *IMT*

The Dickey-Fuller Regressions include an intercept but not a trend
****
1391 observations used in the estimation of all ADF regressions.
Sample period from 5 July 1997 by 6 May 2003
****

	Test Statistic	LL	AIC_L	SC_L
DF	.12822	-7999.2	-8001.2	-8006.4
ADF(1)	**.58275**	-7969.7	**-7972.7**	**-7980.6**
ADF(2)	.63345	-7969.3	-7973.3	-7983.8
ADF(3)	.61590	-7969.3	-7974.3	-7987.4

****
95% critical value for the augmented Dickey-Fuller statistic = -2.8641
LL = Maximized log-likelihood AIC_L = Akaike Information Criterion
SC_L = Schwarz Criterion

Example 5.7:

It can be proved the same way as in example 5.6 that the natural logarithms of *IMT* are , too. But the finance market is more interested in the return drawn out of a share than in the share price level. The return usually is measured by the increase rate of the price, particularly by its logarithmic form (4.14). Therefore figure 5.2 displays the daily rate of return $\Delta \ln IMT$ of the Moscow Times Index from 2nd July 1997 by 6th May 2003. Obviously, the graph does not show any trend and the time series can be expected to be stationary. This is verified by the test results shown in table 5.6. The test statistics of all test variants considered lie far beyond the critical value. Thus independently from any model choice, the null hypothesis of nonstationarity can be rejected and it can be stated on a high level of significance that the time series of return rates is stationary. From this follows that lnIMT_t is integrated of order 1 or I(1).

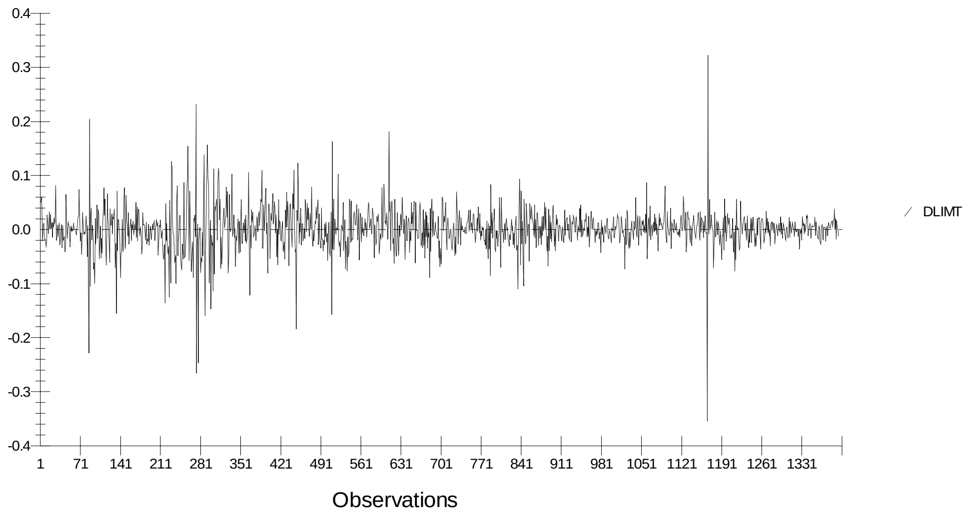

Figure 5.2: The daily return rate of the Moscow Times Index

Table 5.6: Unit root tests for variable Δln*IMT*

The Dickey-Fuller Regressions include an intercept but not a trend
****
1390 observations used in the estimation of all ADF regressions.
Sample period from July 1997 by 6 May 2003
****

	Test Statistic	LL	AIC_L	SC_L
DF	**-38.1617**	2510.2	2508.2	**2503.0**
ADF(1)	-25.2673	2512.2	2509.2	2501.3
ADF(2)	-20.4274	2512.6	2508.6	2498.1
ADF(3)	**-19.1745**	2515.8	**2510.8**	2497.7

****
95% critical value for the augmented Dickey-Fuller statistic = -2.8641
LL = Maximized log-likelihood AIC_L = Akaike Information Criterion
SC_L = Schwarz Criterion

,

6. ARIMA models

6.1. Definition

Let X_t be a nonstationary process with stationary *d*th differences, i.e. $Y_t = (1\text{-L})^d X_t = \Delta^d X_t$ is a stationary process but $\Delta^{d-1}X_t$ is nonstationary. That means X_t is integrated of *d*th order.

If Y_t is an ARMA(*p*,*q*) process, i.e.

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - ... - \theta_q a_{t-q} \qquad (6.1)$$

then X_t is said to be an ARIMA(p,d,q) process. Often the mean or constant ϕ_0 is dropped i.e. set to zero.

Most empirical time series can be considered as realizations of ARIMA processes. In other words: for most time series can be found an ARIMA process, the ARIMA model, that can be thought as data generating process having generated this special time series. The main task of time series analysis is to specify the order of the ARIMA(*p,d,q*) model according to the properties of one time series und to estimate by statistical means the parameters of the model equation and the variance of the error term. As already mentioned, the problem is that generally there is only this one realization of the process.

6.2. Model identification and parameter estimation.

Modelling a time series usually consists of the following steps:

a) Diagnosis, i.e.
  – Checking the time series for stationarity, a precondition of ergodicity
    • Graphical inspection of the time series
    • Unit root test
  – In the case of nonstationarity differentiation and repeated testing these differences
  – Estimation of diagnostic functions such as autocorrelation and inspection of their graphs

b) Choice of a set of process types, what is called the identification of the model.
  In the result, three primary parameters are to be obtained: *d* - the order of integration, *p* and *q* – the orders of the AR and MA components, respectively.
  During the process of diagnosis the parameter *d* is easily found as the number of differentiations necessary for stationarity. For economic time series, *d* is typically 1 but sometimes 0 or two. More difficult is the search for *p* and *q*. Inspection of autocorrelation function (ACF), partial autocorrelation function () and inverse autocorrelation function (IAC) would be helpful. Parameter parsimony should be the principle in the case of doubt.

c) Estimation of the parameters for all versions by suitable methods such as
    • Ordinary Least Squares (OLS)
    • Maximum Likelihood (ML)
    • Minimum squared forecast errors
    • Marquardt algorithm

d) Choice of the most suitable model among a number fitted ones
    • Model check
    • Analysis of residuals. They should be white noise
    • Consideration of the best fit and the most parsimonious representation. Again on the base of the residuals the Akaike Information Criterion or the Schwarz Criterion are to be calculated for each model and compared in order to find the optimum.

Diagnostic functions can give useful hints to the type of underlying process but they are not unambiguous. In example 6.1 for time a series x_t and its 1st and 2nd differences ACF and PAC are estimated in order to answer the question for the order of the underlying ARIMA process.

Example 6.1:
Let us try to answer the question: What type of model is being indicated by the following diagnostics of a time series?

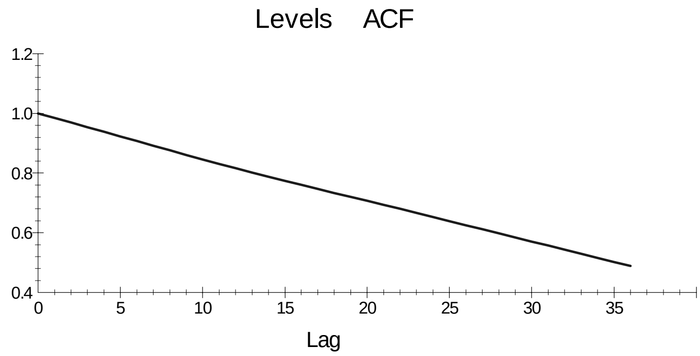

Figure 6.1: Sample autocorrelation function of the original time series

Figure 6.1 suggest an AR(1) process with ϕ_1 close to 1 or a nonstationary process because the ACF is decreasing very slowly. Figure 6.2 indicates stationary AR(1) first differences because there is an almost exponentially decreasing autocorrelation (compare figure 4.5). Figure 6.3 let us guess MA(0) i.e. white noise second differences because ACF is sharply falling down after lag $\tau = 0$ and then moving around zero. But Box-Pierce statistics Q reject the hypothesis of white noise (table 6.1).

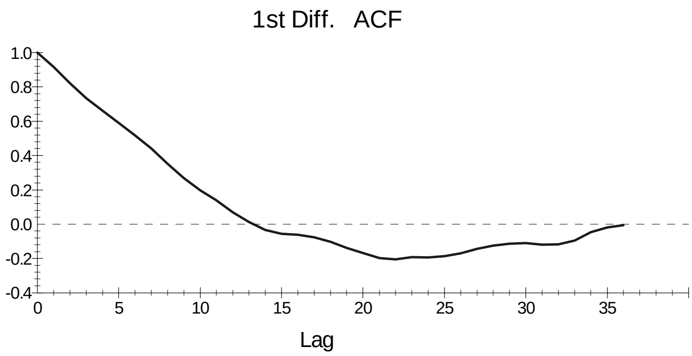

Figure 6.2: Sample autocorrelation function of the 1st differences

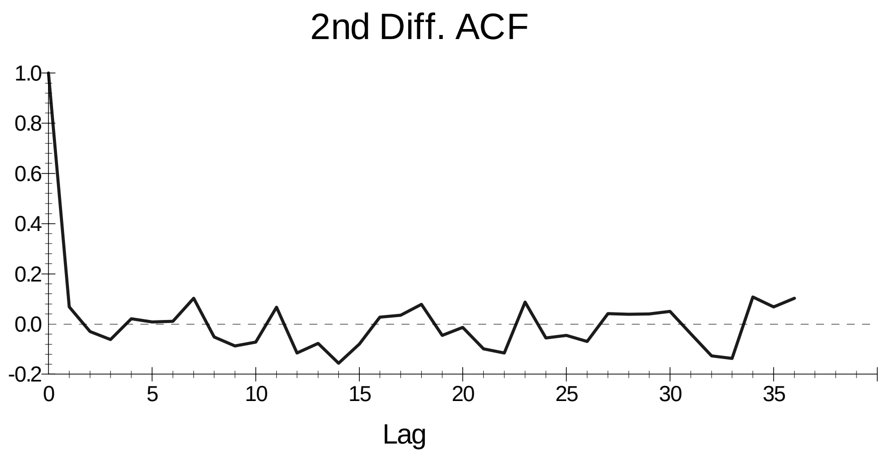

Figure 6.3: Sample autocorrelation function of the 2nd differences

Table 6.1: Box-Pierce statistic of the 2nd differences

T	Q-Stat	Prob	τ	Q-Stat	Prob
1	63.087	0.000	19	99.236	0.000
2	71.862	0.000	20	99.370	0.000
3	78.373	0.000	21	100.57	0.000
4	78.448	0.000	22	100.88	0.000
5	78.682	0.000	23	100.98	0.000
6	78.951	0.000	24	101.76	0.000
7	79.169	0.000	25	102.91	0.000
8	79.278	0.000	26	103.54	0.000
9	81.352	0.000	27	106.32	0.000
10	86.908	0.000	28	110.14	0.000
11	91.571	0.000	29	113.31	0.000
12	95.272	0.000	30	116.30	0.000
13	95.852	0.000	31	117.49	0.000
14	95.873	0.000	32	117.55	0.000
15	96.190	0.000	33	117.56	0.000
16	97.872	0.000	34	117.56	0.000
17	97.878	0.000	35	119.12	0.000
18	97.959	0.000	36	121.40	0.000

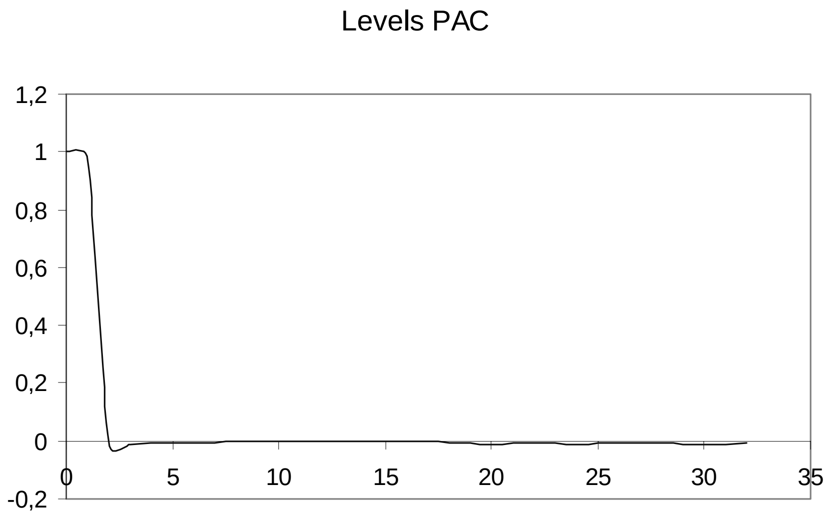

Figure 6.4: Sample partial autocorrelation functions of the original time series

From figure 6.4 could be concluded that the levels of X_t are AR(1) (or AR(2)) because their PAC disappears beyond τ=1 (or 2). An AR(1) process can be taken for the 1st differences despite the negative pique at τ=2. In the same way PAC of the 2nd differences reflect MA(0) i.e. white noise. Figure 6.5 confirms the assumption of AR(1) first differences.

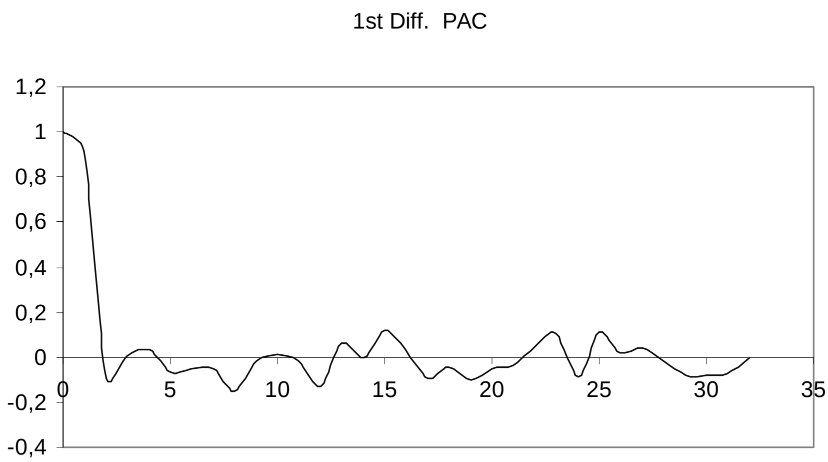

Figure 6.5: Sample partial autocorrelation function of the 1st differences

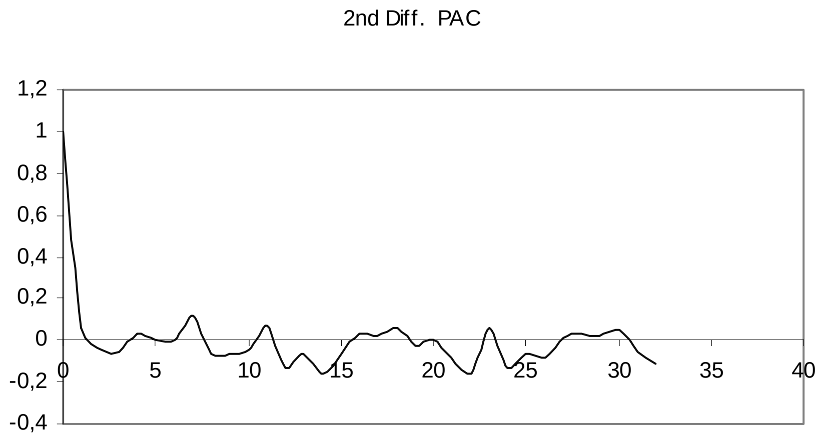

Figure 6.6: Sample partial autocorrelation function of the 2nd differences

To sum up, there are three options for model choice:
1. An nonstationary ARIMA(1,0,0) with $\phi_1 \geq 1$ or perhaps ARIMA(2,0,0).
2. An ARIMA(1,1,0) because of AR(1) first differences.
3. An ARIMA(0,2,0) because of possible white noise second differences.

Decision between the cases can be done by unit root tests in practice. But here we know the data generating process: it is an ARIMA(1,1,0), namely $Y_t = 0.97Y_{t-1} + a_t$ where $Y_t = \Delta X_t$ and $a_t \sim N(0,1)$, i.e. case 2. On the other hand, this means that also the following relationship is valid: X_t=1.97X_{t-1} -0.97X_{t-2}+a_t, what corresponds with case1. And finally by rounding up the coefficient of $Y_t = 0.97Y_{t-1} + a_t$ to unity we obtain the case 3.

If the underlying process is a pure AR(q) or MA(p) with small p or q then this can be easily recognised by inspecting ACF and PAC. For mixed ARIMA(p,q) processes or AR(p) and MA(q) with high orders p and q it would be difficult to give an reliable rule for model identification on the base of a time series and its sample diagnostic functions.

Anyway, after the difficult choice of suitable *p* and *q* the estimation of the more specific parameters ϕ_i, θ_i and σ_a^2 is the next complex problem.

In the case of an AR process, there are among others the following options for estimating the parameters of an AR(p) model:

1. Ordinary last squares regression of x_t depending on x_{t-1}, x_{t-2},…x_{t-p} with certain deteriorated properties of the test statistics because of the lagged regressors.

2. Maximisation of the log-likelihood function (ML estimation).
   The difference to OLS is for long time series negligible.

3. Solving the Yule-Walker equations**.**

The Yule-Walker equations are an equations system describing linear relationships between AR coefficients and the autocovariances of the zero-mean AR(p) process

$$\begin{aligned} \gamma_0 &= \phi_1\gamma_1 + \phi_2\gamma_2 + \ldots + \phi_p\gamma_p + \sigma_a^2 \\ \gamma_1 &= \phi_1\gamma_0 + \phi_2\gamma_1 + \ldots + \phi_p\gamma_{p-1} \\ \vdots \quad & \quad \vdots \qquad\qquad \vdots \end{aligned} \tag{6.2}$$

When substituting sample covariances for the γ_i the solutions $\hat{\phi}_j$ of (6.2) are the Yule-Walker estimators for the ϕ_j. They are consistent.

In the case of a pure MA(q) process software packages usually offer nonlinear least squares methods of parameter estimation. Such as the Conditional-Sum-of-Squares estimation (CSS): Here the missing white noise data a_t are generated as ex-post forecasting errors under the condition of minimum sum of squares.

Example 6.2:
For an MA(1) process $X_t = a_t - \theta_1 a_{t-1}$ error data $\hat{a}_t = x_t + \hat{\theta}_1\hat{a}_{t-1}$ are generated with a coefficient $\hat{\theta}_1$ to be estimated under the condition $S(\hat{\theta}_1) = \sum_1^T \hat{a}_t^{\,2} \overset{!}{=} \text{Min}$ that is a nonlinear function of the parameter θ_1 and is to minimize in an iterative process.

If there is a moving average term in the ARMA model, an Ordinary Least Squares estimation such as in the AR case is not possible. Therefore the parameter estimation for ARMA(p,q) parameters mostly follows the same principle as that for MA processes but is extremely more complex, e.g. there is the problem of the choice of initial values for x_t because of the lagged regressors. Analogously to the example 6.2 the conditional sum of squares

$$S(\hat{\phi}_0,\ldots,\hat{\phi}_p;\hat{\theta}_1,\ldots,\hat{\theta}_q) = \sum_{t=p+q+1}^{T} \hat{a}_t^2 \tag{6.3}$$

is to be minimized, where the error data $\hat{a}_t$ are estimated by the provisional model itself. What is called nonlinear least squares estimation is the iterative procedure

$$\begin{bmatrix} \hat{\mathbf{\Phi}}_{k+1} \\ \hat{\mathbf{\Theta}}_{k+1} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{\Phi}}_{k} \\ \hat{\mathbf{\Theta}}_{k} \end{bmatrix} - \left[\mathbf{D}_k'\mathbf{D}_k\right]^{-1}\mathbf{D}_k'\hat{\mathbf{a}}_k \tag{6.4}$$

where $\hat{\mathbf{\Phi}}_k$ and $\hat{\mathbf{\Theta}}_k$ are the vectors of the kth iteration of the estimated AR and MA coefficients respectively. $\mathbf{D}_k$ denotes the (T-p-q)×(1+p+q) matrix of the derivatives of the estimated disturbances $\hat{\mathbf{a}}_k$ with respect to the parameters ϕ_i and θ_j. Computer programs offer several options for the choice of initial values $\hat{\mathbf{\Phi}}_\mathbf{0}$ and $\hat{\mathbf{\Theta}}_\mathbf{0}$.

Another broadly used option offered by computer program packages is the maximum likelihood (ML) estimation. Here the autoregression coefficients of an AR(p) model are iteratively calculated by ML estimation under the condition that the model errors are MA(q). Initial values can be given or estimated such as with the nonlinear LS estimation.

Example 6.3:

We try to find suitable models for the daily Moscow Times share price index (IMT).
As shown in example 5.6, this time series is nonstationary. By augmented Dickey-Fuller-Test it can easily be demonstrated that the first differences are stationary (table 6.2). Consequently, the Moscow Times Index itself is first order integrated.

Table 6.2: Unit root tests for variable ΔIMT

The Dickey-Fuller Regressions include an intercept but not a trend

****

1390 observations used in the estimation of all ADF regressions.
Sample period from 6 to 1395

****

	Test Statistic	LL	AIC_L	SC_L
DF	-45.7675	-7964.6	-7966.6	-7971.8
ADF(1)	-29.5211	-7964.3	-7967.3	-7975.1
ADF(2)	-22.9344	-7964.2	-7968.2	-7978.7
ADF(3)	-20.3820	-7962.9	-7967.9	-7981.0

****

95% critical value for the augmented Dickey-Fuller statistic = -2.8641
LL = Maximized log-likelihood AIC_L = Akaike Information Criterion
SC_L = Schwarz Criterion

In the following search for an ARIMA(p,**1**,q) Model of IMT we concentrate on the recent time, i. e. only on the last period from 3rd August 2002 by 6th May 2003. The sample autocorrelations function for this period gives some hints about the possible types of the model. According to the steep decrease after lag two it could be an MA(2) model for Δ*IMT*. But the sample PAC (figure 6.8) has almost the same shape. Thus an AR(2) model could be as good. If we take into consideration also minor peaks of the diagnostic functions one could also try models of order 10 and mixed
mode

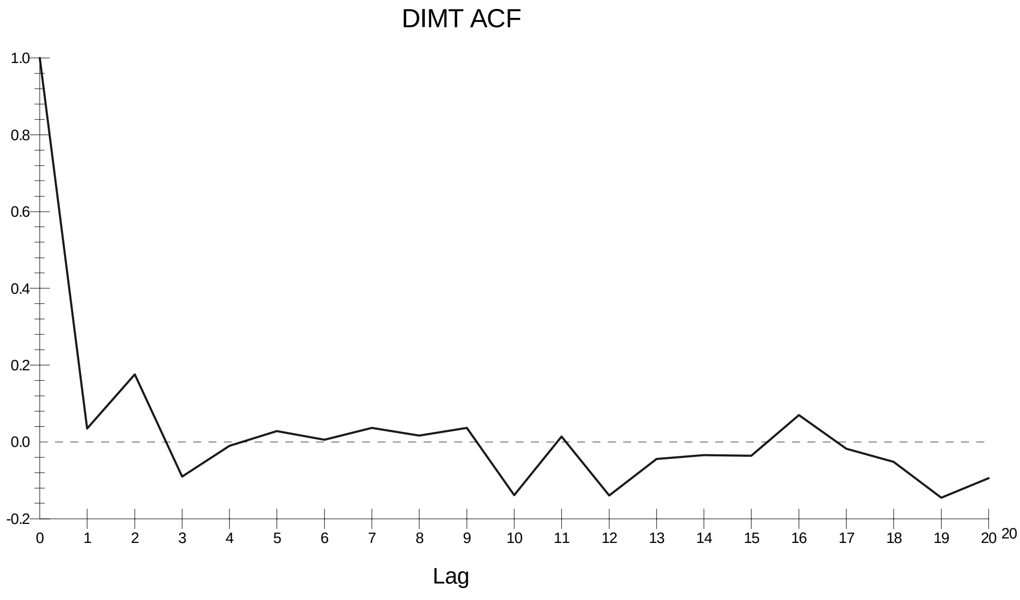

Figure 6.7: Autocorrelation function of the first differences of the Moscow Times Index

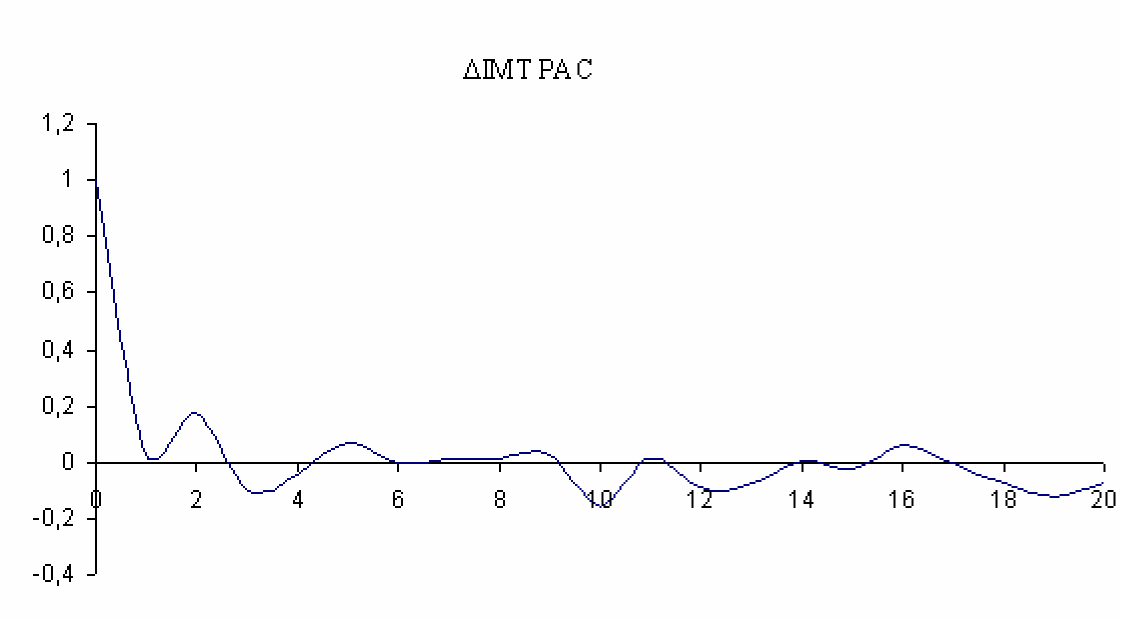

Figure 6.8: Partial autocorrelation function of the first differences of the Moscow Times Index

Our first trial is to deal with such a mixed one, namely an ARIMA(2,1,2) for IMT. The constant term can be dropped because of missing significance. The symbols AR(1), MA(2) etc. in table 6.3 to table 6.9 does not mean the model with this denomination but only the corresponding coefficient of the indicated order. Considering the values of the t-statistic in table 6.3 and the following we can decide about the significance of the corresponding model coefficient: shortly, it is called significant on the 5% level if the absolute value of the t-statistic exceeds the two sided 5% critical value of the standard normal distribution, i.e. 1.96. We have tried to obtain significant coefficients only.

Table 6.3: ARIMA(2,1,2)

Dependent Variable: ΔIMT
Method: Least Squares
Date: 11/17/03 Time: 21:54
Sample: 1200 1395
Included observations: 196
Convergence achieved after 21 iterations
Backcast: 1198 1199

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.727445	0.040086	-18.14692	0.0000
AR(2)	-0.831707	0.046215	-17.99662	0.0000
MA(1)	0.778122	0.009785	79.52184	0.0000
MA(2)	0.983501	0.019035	51.66710	0.0000

R-squared	0.102555	Mean dependent var	2.931173
Adjusted R-squared	0.088533	S.D. dependent var	75.95919
S.E. of regression	72.51885	Akaike info criterion (σ)	11.42577
Sum squared resid	1009725.	Schwarz Criterion (σ)	11.49267
Log likelihood	-1115.725	Durbin-Watson stat	1.938888

The next trial is dedicated to a simpler type, the ARIMA(0,1,2) where the constant and θ_1 have been restricted to zero because of nonsignificance.

Table 6.4: ARIMA(0,1,2), restricted

Convergence achieved after 5 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(2)	0.193886	0.070392	2.754370	0.0064
R-squared	0.032978	Mean dependent var		2.931173
S.E. of regression	74.69618	S.D. dependent var		75.95919
Akaike info criterion (σ)	11.46982	Schwarz Criterion (σ)		11.48655
Log likelihood	-1123.043	Durbin-Watson stat		1.900061

For completeness and because of the similarity between ACF and PAC, the ARIMA(2,1,0) model type will be analysed as well. After two steps of zero restricting nonsignificant coefficients we obtain the also very parsimonious model in table 6.5.

Table 6.5: ARIMA(2,1,0)

Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(2)	0.159330	0.070053	2.274417	0.0240
R-squared	0.024384	Mean dependent var		2.931173
S.E. of regression	75.02736	S.D. dependent var		75.95919
Akaike info criterion (σ)	11.47867	Schwarz Criterion (σ)		11.49540
Log likelihood	-1123.910	Durbin-Watson stat		1.903325

As mentioned above, there is a certain chance of improving the models estimated by extending them to higher orders corresponding to peaks in the ACF and PAC right from τ=2, particularly at τ=10. For this purpose the model is to subject to some restrictions, here to zero restrictions for the coefficient most remote from significance as is indicated by smallest absolute t-value. This way we can set to zero or exclude AR(3) to AR(9) and, after the next estimations not shown here, MA(3) to MA(9). We skip the procedure of numerous estimations and excluding nonsignificant coefficients one after the other and present only the final highly restricted results in table 6.6 to table 6.9. The coefficients in all four variants are significant at the 5% level.

Table 6.6: ARIMA(10,1,0), restricted

Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(2)	0.167751	0.069113	2.427189	0.0161
AR(10)	-0.171843	0.065995	-2.603864	0.0099
R-squared	0.057330	Mean dependent var		2.931173
S.E. of regression	73.93952	S.D. dependent var		75.95919
Akaike info criterion (σ)	11.45452	Schwarz Criterion (σ)		11.48797
Log likelihood	-1120.543	Durbin-Watson stat		1.908686

Table 6.7: ARIMA(0,1,10), restricted

Convergence achieved after 11 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(2)	0.160845	0.069857	2.302493	0.0224
MA(10)	-0.165091	0.071934	-2.295052	0.0228
R-squared	0.049384	Mean dependent var		2.931173
S.E. of regression	74.25048	S.D. dependent var		75.95919
Akaike info criterion (σ)	11.46292	Schwarz Criterion (σ)		11.49637
Log likelihood	-1121.366	Durbin-Watson stat		1.889381

Table 6.8: ARIMA(10,1,2), restricted

Convergence achieved after 6 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(10)	-0.156193	0.067173	-2.325237	0.0211
MA(2)	0.186053	0.070843	2.626275	0.0093
R-squared	0.059239	Mean dependent var		2.931173
S.E. of regression	73.86460	S.D. dependent var		75.95919
Akaike info criterion (σ)	11.45250	Schwarz Criterion (σ)		11.48595
Log likelihood	-1120.345	Durbin-Watson stat		1.899561

Table 6.9: ARIMA(2, 1, 10)

Convergence achieved after 10 iterations

Variable	Coefficien t	Std. Error	t-Statistic	Prob.
AR(2)	0.144618	0.070392	2.054453	0.0413
MA(10)	-0.209273	0.072653	-2.880442	0.0044
R-squared	0.050544	Mean dependent var		2.931173
S.E. of regression	74.20519	S.D. dependent var		75.95919
Akaike info criterion (σ)	11.46170	Schwarz Criterion (σ)		11.49515
Log likelihood	-1121.246	Durbin-Watson stat		1.888897

Before continuing the search for the best of the seven models estimated the decision criteria should be considered more detailed.

When comparing different models for the same time series we are to target competing aims:
- in the case of any kind of least square estimation minimizing the error variance among the set of models estimated *and* at the same time minimizing the number of model parameters,
- or in the case of ML estimation maximizing the likelihood among the set of models *and* at the same time minimizing the number of model parameters.

Usually we would obtain a better fit of the model to be estimated if we choose higher orders p and q of the ARMA model. The price for this seemingly gain of accuracy is a loss of simplicity and parsimony. Therefore usually it is impossible to reach both aims by the same model selected. A compromise between best fit and lowest number of parameters is to be found. As known from section 5.4 useful instruments for finding such a compromise are the Akaike Information Criterion and the Schwarz Criterion (often called Schwarz-Bayes Criterion SBC).

Again here the **Akaike** Information Criterion has got two forms

$$AIC_\sigma = \{1+\ln 2\pi\} + \ln \hat{\sigma}_a^2 + 2\frac{p+q}{T} \qquad \text{to be minimized} \qquad (6.5)$$

or

$$AIC_L = l_{max}(T,p,q) - p - q \qquad \text{to be maximized} \qquad (6.6)$$

with σ_a^2 being the error variance and $l_{max}(T,p,q)$ the logarithmic likelihood of the ARMA model with p and q coefficients (not necessarily equal to the real order of the model!) estimated for time series with length T.

The corresponding two shapes of the **Schwarz** Criterion are:

$$SC_\sigma = \{1+\ln 2\pi\} + \ln \hat{\sigma}^2 + \frac{p+q}{T}\ln T \qquad \text{to be minimized} \qquad (6.7)$$

$$SC_L = l_{\max}(T,p,q) - \frac{p+q}{2}\ln T \qquad \text{to be maximized} \qquad (6.8)$$

The Schwarz Criterion is more parsimonious concerning the number of parameters

Example 6.3: (continued)

Table 6.10 assembles the values of the error variance oriented Akaike Information Criterion AIC_σ of the models for the Moscow Times Index estimated in the first part of this example. Here p and q indicate the order of the model independently from the number of effectively estimated coefficients. Table 6.11 shows the corresponding values of the Schwarz Criterion SC_σ.

Table 6.10: AIC_σ

q \ p	0	2	10
0	...	11.479...	11.455
2	*11.470*	**11.426**	11.453
10	11.463	11.462	...

Table 6.11: SC_σ

q \ p	0	2	10
0	...	11.495	11.488
2	*11.487*	11.493	**11.486**
10	11.496	11.495	...

The ARIMA(10,1,2) model for ΔIMT proved best concerning the smallest values of both AIC and SC criteria. Its model equation is

$$\Delta IMT_t = -0.156193\, \Delta IMT_{t-10} + a_t + 0.186053\, a_{t-2} \qquad (6.9)$$

or for the levels of the original time series:

$$IMT_t = IMT_{t-1} - 0.156193\, IMT_{t-10} + 0.156193\, IMT_{t-11} + a_t + 0.186053\, a_{t-2} \qquad (6.10)$$

But because it would be difficult to explain the meaning of the AR(10) term it could be sufficient to choose the best model among the ARIMA(2,1,2) and smaller models. By both criteria then ARIMA(0,1,2) would be chosen:

$$\Delta IMT_t = a_t + 0.193886 a_{t-2} \qquad (6.11)$$

or $$IMT_t = IMT_{t-1} + a_t + 0.193886 a_{t-2} \qquad (6.12)$$

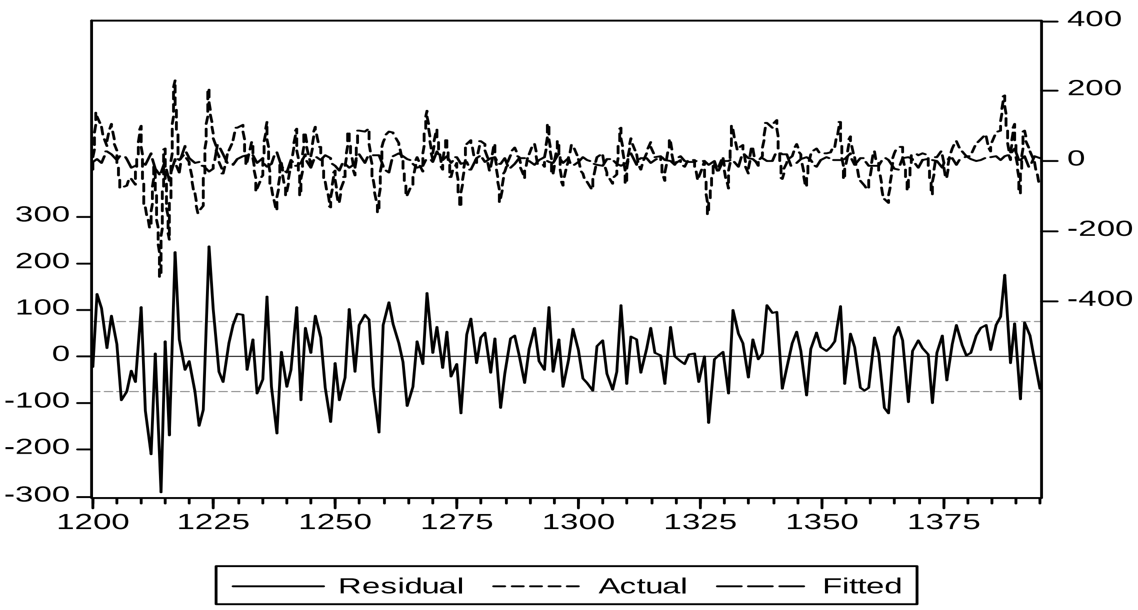

Figure 6.9: MA(2) model for the stationary increase ΔIMT_t of the Moscow Times Index

6.3. Multiplicative ARIMA models – Seasonality

Seasonalities in time series can be dealt with by seasonal differences:
e.g. for seasonal period length s=12, i.e. for monthly data:

$$X_t = \Delta_{12} Z_t = (1 - \mathrm{L}^{12}) Z_t = Z_t - Z_{t-12} \qquad (6.13)$$

or by seasonal lags in ARMA models.
Simple examples are the seasonal AR(1) model

$$X_t = \varphi_1^* X_{t-12} + a_t \qquad (6.14)$$

and the seasonal MA(1) model

$$X_t = a_t - \theta_1^* a_{t-12} \qquad (6.15)$$

Model (6.14) can be written as

$$(1 - \phi_1^* \mathrm{L}^{12}) X_t = a_t \,. \qquad (6.16)$$

If for a periodical monthly time series x_t the residuals $\hat{a}_t$ prove free of seasonal effects the assumed underlying stationary process a_t itself can be subject a second ARMA(p,q) modelling:

$\Phi_\mathrm{p}(\mathrm{L}) a_t = \Theta_\mathrm{q}(\mathrm{L}) \varepsilon_t$ where ε_t is white noise
or, substituting (6.16),

$$\Phi_\mathrm{p}(\mathrm{L})(1 - \phi_1^* \mathrm{L}^{12}) X_t = \Theta_\mathrm{q}(\mathrm{L}) \varepsilon_t \qquad (6.17)$$

If we take as examples for $\Phi_\mathrm{p}(\mathrm{L})$ and $\Theta_\mathrm{q}(\mathrm{L})$ the functions

$\Phi_2(\mathrm{L}) = \mathrm{L}^0 - \phi_1 \mathrm{L} - \phi_2 \mathrm{L}^2$
$\Theta_1(\mathrm{L}) = \mathrm{L}^0 - \theta_1 \mathrm{L}$
then (6.17) assumes the form
$\Phi_2(\mathrm{L})(1 - \phi_1^* \mathrm{L}^{12}) X_t = \Theta_1(\mathrm{L}) \varepsilon_t$
or
$(\mathrm{L}^0 - \phi_1 \mathrm{L} - \phi_2 \mathrm{L}^2)(1 - \phi_1^* \mathrm{L}^{12}) X_t = (\mathrm{L}^0 - \theta_1 \mathrm{L}) \varepsilon_t$

This results after multiplying the operator terms to
$(\mathrm{L}^0 - \phi_1 \mathrm{L} - \phi_2 \mathrm{L}^2 - \phi_1^* \mathrm{L}^{12} + \phi_1 \phi_1^* \mathrm{L}^{13} + \phi_2 \phi_1^* \mathrm{L}^{14}) X_t = (\mathrm{L}^0 - \theta_1 \mathrm{L}) \varepsilon_t$.
or explicitly to
$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-1} + \phi_1^* X_{t-12} - \phi_1 \phi_1^* X_{t-13} - \phi_2 \phi_1^* X_{t-14} = \varepsilon_t - \theta_1 \varepsilon_{t-1}$

Because the operator functions $\Phi_p(\mathrm{L})$ and $\Phi_1^*(\mathrm{L}^{12}) = 1 - \phi^* \mathrm{L}^{12}$ to be executed one after the other can be formally multiplied like arithmetical terms this kind of model is called multiplicative.

More generally, both parts of the "multiplication" can be full ARIMA models. Then the combined model is termed as a seasonal ARIMA(p,d,q)×(P,D,Q)$_s$ model, where P, D, Q denote the orders of the seasonal model for the period length s.

Then the differenced process $\Delta^d \Delta_s^D X_t = (1-\mathrm{L})^d (1-\mathrm{L}^s)^D X_t = Y_t$ is assumed to be a stationary ARMA process

$$\Phi_p(\mathrm{L})\Phi_P^*(\mathrm{L}^s)Y_t = \Theta_q(\mathrm{L})\Theta_Q^*(\mathrm{L}^s)\varepsilon_t \qquad (6.18)$$

where

$$\Phi_p(\mathrm{L}) = 1-\phi_1 \mathrm{L} - \ldots - \phi_p \mathrm{L}^p$$

$$\Phi_P^*(\mathrm{L}^s) = 1-\phi_1^* \mathrm{L}^s - \ldots - \phi_P^* \mathrm{L}^{sP}$$

$$\Theta_q(\mathrm{L}) = 1-\theta_1 \mathrm{L} - \ldots - \theta_q \mathrm{L}^q$$

$$\Theta_Q^*(\mathrm{L}^s) = 1-\theta_1^* \mathrm{L}^s - \ldots - \theta_Q^* \mathrm{L}^{sQ}$$

If we use instead of the symbol Δ again the operator 1-L and introduce again the original process under consideration x_t we find for the general multiplicative seasonal ARIMA(p,d,q)×(P,D,Q) process X_t with the seasonality s the representation

$$\Phi_p(\mathrm{L})\Phi_P^*(\mathrm{L}^s)(1-\mathrm{L}^s)^D(1-\mathrm{L})^d X_t = \Theta_q(\mathrm{L})\Theta_Q(\mathrm{L}^s)a_t \qquad (6.19)$$

Example 6.4:

In example 4.8 we introduced the nonstationary time series of monthly incoming orders for the construction industries in East Germany *OR*. This time series is first order integrated I(1). For the stationary first differences Δ*OR* a seasonal ARMA(3,3)×$(0,1)_{12}$ is estimated. The coefficients and the standard error can be found in table 6.2. The model for the levels *OR* is a seasonal ARIMA(3,1,1)×$(0,0,1)_{12}$

Table 6.12:

Dependent Variable: Δ*OR*
Method: Least Squares
Sample(adjusted): 1991:05 2003:01
Included observations: 141 after adjusting endpoints
Convergence achieved after 11 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(3)	0.984586	0.018048	54.55340	0.0000
MA(1)	-0.087409	0.044901	-1.946723	0.0536
MA(3)	-0.875578	0.044068	-19.86861	0.0000
SMA(1)	0.223344	0.092243	2.421250	0.0168

R-squared	0.473244	Durbin-Watson stat	1.996023
Adjusted R-squared	0.461709	S.D. dependent var	17.90197
S.E. of regression	13.13438	Akaike info criterion (σ)	8.016302
Log likelihood	-561.1493	Schwarz Criterion (σ)	8.099955

Thus we have found for the representation (6.19) of *OR* the following details with s=12, d=1, D=0, p=3, P=0, q=3 and Q=1:

$\Phi_3(\mathrm{L}) = 1 - 0.985\,\mathrm{L}^3$

$\Phi_p^* = 1$

$\Theta_3(\mathrm{L}) = 1 + 0.087\,\mathrm{L}^1 + 0.876\,\mathrm{L}^3$

$\Theta_1^*(\mathrm{L}) = 1 - 0.223\,\mathrm{L}^{12}$.

Then (6.19) is

$(1 - 0.985\,\mathrm{L}^3)(1 - \mathrm{L})OR_t = (1 + 0.087\,\mathrm{L} + 0.876\,\mathrm{L}^3)(1 - 0.223\,\mathrm{L}^{12})a_t$

or writing it explicitly:

$$OR_t = OR_{t-1} + 0.895OR_{t-3} - 0.985OR_{t-4} + a_t + 0.087a_{t-1} + 0.876a_{t-3} - 0.223a_{t-12} - \\ - 0.087 \cdot 0.223a_{t-13} - 0.876 \cdot 0.223a_{t-15}$$

The last equation can be easily used as a forecast formula.

7. Forecasting

7.1. Forecasting ARMA processes

Let us consider the stationary ARMA model

$$\Phi(\mathrm{L})X_t = \Theta(\mathrm{L})a_t \qquad (7.1)$$

transformed into the infinite MA representation (random shock model)

$$X_t = \Psi(\mathrm{L})a_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots \qquad (7.2)$$

or for $t=T+h$

$$X_{T+h} = \sum_{i=0}^{\infty} \psi_i a_{T+h-i} \qquad (7.3)$$

Let T now be the origin for forecasting over a time horizon h taking into consideration that there is no information for the time points $T+1, T+2, \ldots T+h$. The forecasting formula can be reduced to

$$\hat{X}_{T(+h)} = \psi_h a_T + \psi_{h+1} a_{T-1} + \psi_{h+2} a_{T-2} + \ldots = \mathrm{E}(X_{T+h} | X_T, X_{T-1}, X_{T-2} \ldots) \qquad (7.4)$$

where $\hat{X}_{T(+h)}$ designs the h-step forecast on the base of the knowledge of the process till $t=T$.

The corresponding forecast error is

$$e_{T(+h)} = X_{T+h} - \hat{X}_{T(+h)} = \sum_{i=0}^{h-1} \psi_i a_{T+h-i}$$

Because of $\mathrm{E}(a_{T+h-i}) = 0$ for $i = 0, 1, \ldots h-1$ the conditional expectation of the forecast error is

$$\mathrm{E}(e_{T(+h)} | X_T, X_{T-1}, X_{T-2} \ldots) = 0,$$

therefore the estimator $\hat{X}_{T(+h)}$ being unbiased.

The forecast error variance

$$\mathrm{var}(e_{T(+h)}) = \sigma_a^2 \sum_{i=0}^{h-1} \psi_i^2 \qquad (7.5)$$

allows the calculation of limits for forecast intervals, if a special distribution of the white noise a_t is assumed, e.g. a normal distribution.

In practical forecasting the true ARMA parameters ϕ_k and θ_j are substituted by estimated values $\hat{\phi}_k$ and $\hat{\theta}_j$ and the random shocks a_t by the residuals $\hat{a}_t$ of the model fitted or the error e_{T+h-i} of previous forecasts.

Example 7.1:

The AR(1) model $X_t = \phi_0 + \phi_1 X_{t-1} + a_t, \quad |\phi_1| < 1$

can be transformed for $t = T+h$ to $X_{T+h} = \phi_0 + \phi_1 X_{T+h-1} + a_{T+h-1}$,

what gives the recursive forecast formula

$$\hat{X}_{T(+h)} = \phi_0 + \phi_1 \hat{X}_{T(+h-1)}$$

If we substitute

$$\hat{X}_{T(+h-1)} = \phi_0 + \phi_1 \hat{X}_{T(+h-2)}$$

and so further, we obtain the final formula

$$\hat{X}_{T(+h)} = \phi_0 (1 + \phi_1 + \phi_1^2 + \ldots + \phi_1^{h-1}) + \phi_1^h X_T \qquad (7.6)$$

In the long run this expression asymptotically tends to the expectation of *X*:

$$\lim_{n\to\infty} \hat{X}_{T(+h)} = \frac{\phi_0}{1-\phi_1} = E(X) = \mu_x$$

Furthermore it is easy to proof that in this example the coefficient of the representation (7.2) are (compare example 3.6)

$$\psi_j = \phi_1^i$$

Therefore the forecast error variance (7.5) is

$$\text{var}(e_{T(+h)}) = \sigma_a^2(1 + \phi_1^2 + \phi_1^4 + \ldots + \phi_1^{2(h-1)}) = \sigma_a^2 \frac{1-\phi_1^{2h}}{1-\phi_1^2}$$

In the long run this variance tends to the variance of the process *X*, namely σ_X^2.

Example 7.2:

For the AR(2) model $X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2}$

the equivalent forecast formulae immediately follow:

one-step forecast: $\hat{X}_{T(+1)} = \phi_0 + \phi_1 X_T + \phi_2 X_{T-1}$

two-step forecast: $\hat{X}_{T(+2)} = \phi_0 + \phi_1 \hat{X}_{T(+1)} + \phi_2 X_T$

h-step forecast: $\hat{X}_{T(+h)} = \phi_0 + \phi_1 \hat{X}_{T(+h-1)} + \phi_2 \hat{X}_{T(+h-2)}$ for $h \geq 3$

Equally to the result in example 7.1, the expectation of this forecast tends to the process mean:

$$\lim_{n\to\infty} X_{T(+h)} = \frac{\phi_0}{1-\phi_1-\phi_2} = \mathrm{E}(X_t) = \mu_X$$

and

$$\lim_{n\to\infty} \text{var}(e_{T(+h)}) = \sigma_X^2.$$

Example 7.3:

For the MA(1) model $X_t = a_t - \theta_1 a_{t-1}$

the forecast formulae obviously are

$$\hat{X}_{T(+1)} = \theta_1 a_T$$

and $\hat{X}_{T(+h)} = 0$ for $h \geq 2$.

Because of

$$\psi_0 = 1$$
$$\psi_1 = -\theta_1$$
$$\psi_i = 0 \text{ for } i \geq 2$$

the forecast error variance according to (7.5) results in

$$\operatorname{var}(e_{T(+1)}) = \sigma_a^2$$

and $\operatorname{var}(e_{T(+h)}) = \sigma_a^2(1+\theta_1^2) = \sigma_X^2$ for $h \geq 2$.

That means the forecast intervals have a constant width independent on the forecast horizon *h*.

In the same way, we obtain for the MA(2) process

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

the forecast formulae

$$\hat{X}_{T(+1)} = -\theta_1 a_T - \theta_2 a_{T-1}$$
$$\hat{X}_{T(+2)} = -\theta_2 a_T$$
$$\hat{X}_{T(+h)} = 0 \text{ for } h \geq 3$$

and the forecast error variance

$$\operatorname{var}(e_{T(+1)}) = \sigma_a^2$$
$$\operatorname{var}(e_{T(+2)}) = \sigma_a^2(1+\theta_1^2)$$

and for $h \geq 3$ $\operatorname{var}(e_{T(+2)}) = \sigma_a^2(1+\theta_1^2+\theta_2^2) = \sigma_X^2$.

Example 7.4:

For the ARMA (1,1) model $X_t = \phi_0 + \phi_1 X_{t-1} - \theta_1 a_{t-1}$

again we easily find the forecast formulae

$$\hat{X}_{T(+1)} = \phi_0 + \phi_1 X_T - \theta_1 a_T$$

and $\hat{X}_{T(+h)} = \phi_0 + \phi_1 \hat{X}_{T(+h-1)}$ for $h \geq 2$

with $\lim_{h\to\infty} X_{T(+h)} = \dfrac{\phi_0}{1-\phi_1} = \mathrm{E}(X) = \mu_x$

and the forecast error variance

$$\operatorname{var}(e_{T(+h)}) = \sigma_a^2\left(1+\sum_{i=1}^{h-1}\left[\phi_1^{i-1}(\phi_1+\theta_1)\right]^2\right)$$

with $$\lim_{h\to\infty}\operatorname{var}(e_{T(+h)}) = \sigma_a^2\left(1+\frac{(\phi_1+\theta_1)^2}{1-\phi_1^2}\right) = \sigma_X^2$$

As we have seen in all examples considered in this section, the forecast error of these ARMA models is limited by the value of the process variance σ_X in the long run.

This will change in the case of nonstationary processes, that means here: ARIMA processes.

7.2. Forecasting ARIMA processes

As defined in section 6.1, the nonstationary process X_t is called an ARIMA*(p,d,q)* process if the *d*-th differences $Y_t = \Delta^d X_t = (1-\mathrm{L})^d X_t$ is a stationary ARMA(*p,q)* process:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}$$

or $$\Phi(\mathrm{L})Y_t = \phi_0 + \Theta(\mathrm{L})a_t \qquad (7.7)$$

with $\Phi(\mathrm{L}) = \mathrm{L}^0 - \phi_1\,\mathrm{L}^1 - \ldots - \phi_p\,\mathrm{L}^p$

and $\Theta(\mathrm{L}) = \mathrm{L}^0 - \theta_1\,\mathrm{L}^1 - \ldots - \theta_q\,\mathrm{L}^q$

The forecast of this ARIMA process Y_t can be carried out as a two-stage procedure:
First, the stationary ARMA process X_t is extrapolated the way shown in section 7.1.
Secondly the differentiation is reversed into integration i.e. summation of the forecasted increments $\hat{Y}_{T(+h)} = \Delta^d \hat{X}_{T(+h)}$ in order to obtain firstly $\Delta^{d-1}\hat{X}_{T(+h)}$, then in the same way $\Delta^{d-2}\hat{X}_{T(+h)}$ and finally $\hat{X}_{T(+h)}$.
The estimation of the forecast error variance and consequently the width of the forecast interval is to be performed analogously by repeated summation of the error variances of the ARMA process X_t.

Another option is the construction of individual one-stage forecast formulae.
For this purpose the equation (7.7) is modified by substituting the differences
$\Delta^d X_t = (1-\mathrm{L})^d X_t$ for Y_t:

$$\Phi(\mathrm{L})(1-\mathrm{L})^d X_t = \phi_0 + \Theta(\mathrm{L})a_t \qquad (7.8)$$

By multiplying the operator functions on the left hand side and solving the equation for X_t we obtain a model formula that can be extrapolated for $t=T+h$ and in this way transformed into an *h*-step forecast formula for $\hat{X}_{T(+h)}$ with the origin *T*.

Example 7.5:

ARIMA (0,1,0) models

If X_t is a random walk without drift (constant)

where $\Delta X_t = a_t$,
i.e. $(1-\mathrm{L})X_t = a_t$
or $X_t = X_{t-1} + a_t$,

then the extrapolation formula can be written as

$$X_{T+h} = X_{T+h-1} + a_t$$

That means on the base of the last known realisation at t=T the forecast formula is reduced to the simple constant relationship

$$\hat{X}_{T(+h)} = X_T$$

for all forecast horizons $h \geq 1$ but with increasing error variance:

$$\mathrm{var}(e_{T(+h)}) = h\sigma_a^2$$

If X_t is a random walk with shift

$$X_t = X_{t-1} + \phi_0 + a_t$$

then the forecast formula

$$\hat{X}_{T(+h)} = X_T + h\phi_0$$

corresponds to a simple linear trend line.

The error variance is the same as in the above case of $\phi_0 = 0$, that means in both cases, the width of the forecast interval increases proportionally with $\sqrt{h}$.

Example 7.6:

The ARIMA (0,1,1) model $(1-\mathrm{L})X_t = \phi_0 - \theta_1 a_{t-1} + a_t$

This becomes for $t = T+h$

$$X_{T+h} = \phi_0 - X_{T+h-1} - \theta_1 a_{T+h-1} + a_t$$

or in the shape of a h-step forecast with origin T on the base of information prior $T+1$

$$\hat{X}_{T(+h)} = X_T + h\phi_0 - \theta_1 a_T$$

i.e. the forecast curve is linear in h.

From this forecast error variance

$$\mathrm{var}(e_{T(+h)}) = \sigma_z^2 (1 + (h-1)(1-\theta_1)^2)$$

can be derived.

Example 7.7:

The ARIMA (1,1,0) model $(1-\phi_1 \mathrm{L})(1-\mathrm{L})X_t = \phi_0 + a_t$

gives explicitly for $t = T+h$

$$X_{T+h} = \phi_0 + (1+\phi_1)X_{T+h-1} - \phi_1 X_{T+h-2} + a_{T+h}$$

and the iterative forecast for $h \geq 3$

$$\hat{X}_{T(+h)} = \phi_0 + (1+\phi_1)\hat{X}_{T(+h-1)} - \phi_1 \hat{X}_{T(+h-2)}$$

with the starting forecasts

$$\hat{X}_{T(+1)} = \phi_0 + (1+\phi_1)X_T - \phi_1 X_{T-1}$$

$$\hat{X}_{T(+2)} = \phi_0 + (1+\phi_1)\hat{X}_{T(+1)} - \phi_1 X_T \, .$$

Without further details should only be mentioned that the forecast error variance is a rather complex function of ϕ_1 that tends to infinity in the long run.

Example 7.8:

From the ARIMA (1,1,1) model $(1-\phi_1 \mathrm{L})(1-\mathrm{L})X_t = \phi_0 + (1-\theta_1 \mathrm{L})a_t$

with the presentation for $t = T+h$,

$$X_{T+h} = \phi_0 + (1+\phi_1)X_{T+h-1} - \phi_1 X_{T+h-2} + a_{T+h} - \theta_1 a_{T+h-1} \, ,$$

the forecast formula for $h \geq 3$ follows:

$$\hat{X}_{T(+h)} = \phi_0 + (1+\phi_1)\hat{X}_{T(+h-1)} - \phi_1 \hat{X}_{T(+h-2)}$$

with $\quad \hat{X}_{T(+1)} = \phi_0 + (1+\phi_1)X_T - \phi_1 X_{T-1} - \theta_1 a_T$

and $\quad \hat{X}_{T(+2)} = \phi_0 + (1+\phi_1)\hat{X}_{T(+1)} - \phi_1 X_T \, .$

Example 7.9:

For the Moscow Times Index (*IMT*) we had specified and estimated in example 6.3 and equation (6.12) among others the ARIMA(0,1,2) model

$$IMT_t = IMT_{t-1} + a_t + 0.194 a_{t-2} \, . \qquad (7.9)$$

That means for $t = T+h$

$$IMT_{T+h} = IMT_{T+h-1} + a_{T+h} + 0.194 a_{T+h-2}$$

or for a one-step forecast from the origin T

$$\hat{IMT}_{T(+1)} = IMT_T + 0.194 a_{T-1} \qquad (7.10)$$

and for a two-step forecast

$$\hat{IMT}_{T(+2)} = \hat{IMT}_{T(+1)} + 0.194a_T = IMT_T + 0.194(a_T + a_{T-1})$$

The latter forecast will not vary for any h>2:

$$\hat{IMT}_{T(+h)} = IMT_T + 0.194(a_T + a_{T-1}) \qquad (7.11)$$

While the forecast equals to a constant for $h \geq 2$ the forecast error variance increases to infinity with growing forecast horizon h:

$$\mathrm{var}(e_{T(+h)}) = \sigma_a^2 (2 + (h-2)(1+0.194)^2) \qquad (7.12)$$

This relationship follows from the general formula for the forecast error variance

$$\mathrm{var}(e_{T(+h)}) = \sigma_a^2 \sum_{i=0}^{h-1} \psi_i^2 \qquad (7.13)$$

with ψ_i being the coefficients of the infinite MA representations of a process with the lag polynom Ψ(L). In the case of this example Ψ(L) must be equal to the model operator $\frac{\Theta(\mathrm{L})}{(1-\mathrm{L})}$

or $\Psi(\mathrm{L})(1-\mathrm{L}) = \Theta(\mathrm{L})$

i.e. $(\psi_0 \mathrm{L}^0 + \psi_1 \mathrm{L}^1 + \psi_2 \mathrm{L}^2 + \ldots)(1-\mathrm{L}) = 1\mathrm{L}^0 - \theta_2 \mathrm{L}^2$

By multiplying the terms in brackets and equalling the coefficients of the same powers of L on both sides of the equation we obtain

$$\psi_0 = 1$$
$$\psi_1 = 1$$
$$\psi_2 = 1 - \theta_2$$
$$\vdots$$
$$\psi_{h-1} = 1 - \theta_2 \quad \text{with} \quad \theta_2 = -0.194$$

Substituting ψ_i in (7.13) we obtain (7.12).

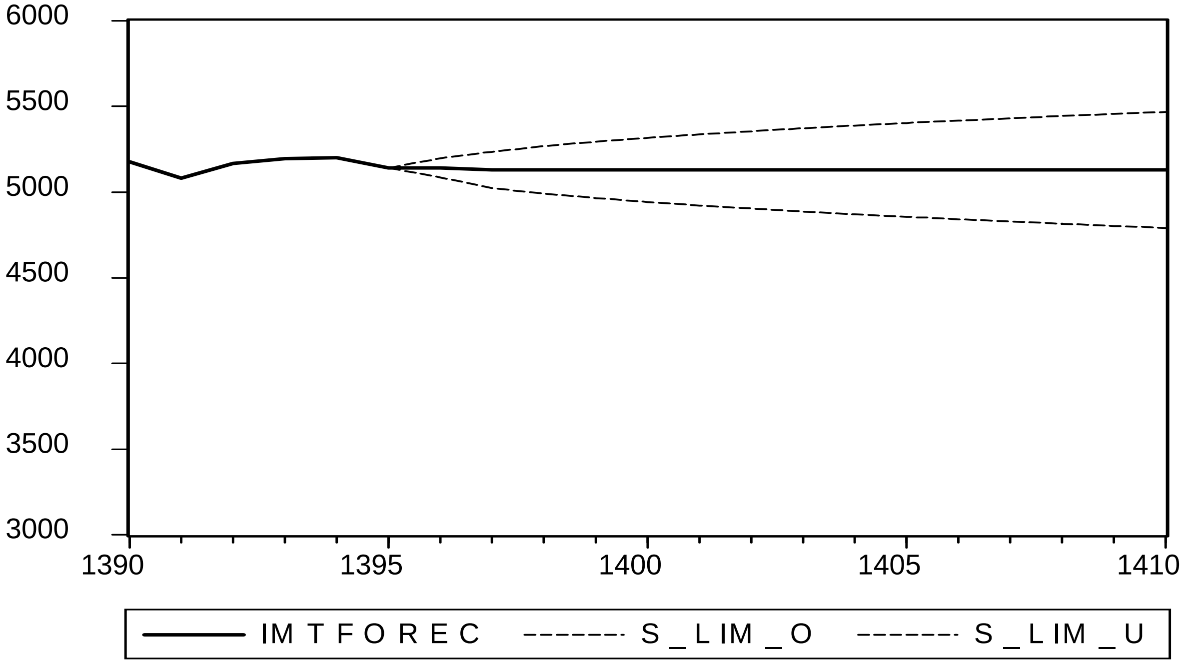

Figure 7.1: Forecast of the Moscow Times share price index.

Figure 7.1 displays the forecasts of the levels $I\hat{M}T_{T(+h)} = IMTFOREC$ of the Moscow Times Index on the base of the data till T=1395 for 6th May 2003 with forecast horizons h=1,2,…,15. The dotted lines below and above the forecast line indicate the one-sigma forecast limits corresponding to (7.12). The forecast values themselves remain constant after the first step of h while the width of the forecast interval increases. As expected, this model would not be a particular efficient tool for long-term forecasting the Russian finance market. But a repeated one-step calculation according to (7.10) on the base of the newest daily data would improve the forecast quality effectively.

8. ARCH and GARCH Processes

Usually econometricians assume the autocorrelations of model disturbances to be zero. But in the last decade the interest of researchers increasingly focused on systematically changing errors and error variances because in time series of exchange rates and stock market return had been found sections of small error changing with sections of large errors or low with high volatility, respectively.

8.1. Conditional Heteroscedasticity

Volatility is usually measured by the variance σ_t^2 of a time series or a stochastic process.

Homoskedasticity of a model such as an AR model means that the error or disturbance term has a constant variance. The antonym is *Heteroskedasticity,* i.e. variability of the error variance.
Conditional Heteroskedasticity (CH) means that the conditional error variance i.e. the variance under the condition of information given depends on time. It can occur in spite of general homoskedasticity (unconditional).

The variance of the model disturbances a_t is

$$\operatorname{var}(a_t) = \mathrm{E}(a_t^2) = \sigma_a^2 \tag{8.1}$$

The corresponding conditional variance on the base of the knowledge of the last value is defined as

$$\operatorname{var}(a_t | a_{t-1}) = \mathrm{E}(a_t^2 | a_{t-1}) \tag{8.2}$$

8.2. The ARCH/GARCH Model

The ARCH(1) model is the simplest example for an ARCH process, i.e. an autoregressive conditional heteroscedasticity process.

Let $X_t = \phi_0 + \phi_1 X_{t-1} + u_t$ (8.3)

be an AR(1) process with an error term u_t and the properties

$$\mathrm{E}(u_t) = 0$$

$$\mathrm{E}(u_t \mid u_{t-1}) = 0$$

$$u_t^2 = (\lambda_0 + \lambda_1 u_{t-1}^2) + a_t^2$$

with λ_1<1 and a_t being white noise. Then the conditional variance of the error is

$$h_t^2 = \operatorname{var}(u_t \mid u_{t-1}) = \mathrm{E}(u_t^2 \mid u_{t-1}) = \lambda_0 + \lambda_1 u_{t-1}^2 \tag{8.4}$$

Obviously it depends on the last value of *u* and is not constant. That means varying conditional variance, i.e. conditional Heteroskedasticity occurs.

But the unconditional variance is constant:

$$\begin{aligned}\mathrm{E}(u_t^2) &= \lambda_0 + \lambda_1\,\mathrm{E}(u_{t-1}^2)\\ &\vdots\\ &= \lambda_0 \sum \lambda_1^i \qquad (\lambda_1 < 1)\\ &= \frac{\lambda_0}{1-\lambda_1} = \text{const}\end{aligned} \tag{8.5}$$

That means the process (8.3) proves homoscedastic despite its conditional heteroscedasticity.

An AR model can be tested for ARCH(1) in the following way

- Fit X_t by an AR model with the error term u_t.
- Calculate residuals $\hat{u}_t$ as estimates for u_t.
- Calculate a linear regression for $\hat{u}_t^2$ with regressor $\hat{u}_{t-1}^2$ and the coefficient λ_1.
- Test the coefficient λ_1 by t-, F-, χ^2-test with the null hypothesis H_0: $\lambda_1=0$.

If λ_1 significantly differs from zero the model is ARCH(1).

Let be again

$X_t = \phi_0 + \phi_1 X_{t-1} + u_t$

But now we assume

$$h_t^2 = \mathrm{var}(u_t \mid u_{t-1,\dots,t-q}) = \lambda_0 + \lambda_1 u_{t-1}^2 + \ldots + \lambda_q u_{t-q}^2 \tag{8.6}$$

Then X_t is considered as being an ARCH(q) process.
A time series can be tested for ARCH(q) by extending the regression in the above described test to a multiple one.

The generalized autoregressive conditional heteroskedasticity model (GARCH (p,q))describes a process where the conditional error variance on all information Ω_{t-1} available at time t

$$h_t^2 = \mathrm{var}(u_t | \Omega_{t-1})$$

is assumed to obey an ARMA(p,q) model:

$$h_t^2 = \alpha_0 + \alpha_1 h_{t-1}^2 + \ldots + \alpha_p h_{t-p}^2 + \beta_1 u_{t-1}^2 + \beta_2 u_{t-2}^2 + \ldots + \beta_q u_{t-q}^2 \tag{8.7}$$

Example 8.1:

Here we try to model the return of the Moscow Times Index over the whole period from 1997 to 2003. In example 5.3 this return was defined as $\Delta\ln IMT$. Table 8.1 shows the result of estimating of an ARCH (1) which coincides with a GARCH(0,1)

Table 8.1:

GARCH(0,1) assuming a normal distribution
converged after 59 iterations

****

Dependent variable is $\Delta \ln_t$
1393 observations used for estimation from 3 to 1395

****

Regressor	Coefficient	Standard Error	T-Ratio[Prob]
$\Delta \ln IMT_{t-1}$	.12158	.034727	3.5012[.000]

****

R-Squared	-.020937	DW-statistic	2.2945
S.E. of Regression	.040219 F-stat.	F(1,1391)	*NONE*
Mean of Dependent Variable	.0012655	S.D. of Dependent Variable	.039790
Residual Sum of Squares	2.2500	Equation Log-likelihood	2634.8
AIC_L	2632.8	SC_L	2627.6

****

Parameters of the Conditional Heteroscedastic Model

****

Dependent variable is the squared error e_t^2

	Coefficient	Asymptotic Standard Error
Constant	.0010248	.4987E-4
e^2_{t-1}	.34542	.052786

****

In contrast to the last examples, here the *likelihood* versions of the Akaike Information and Schwarz Criteria that are to be maximized were used.

An interesting modification of ARCH or GARCH models is *the ARCH/GARCH model in mean* (ARCH-M or GARCH-M)
Here the conditional variance h_t^2 of (8.8) is enclosed as an explicate term in the general model equation for X_t

e.g. $$X_t = \varphi_0 + \varphi_1 X_{t-1} + \gamma h_t^2 + u_t \qquad (8.8)$$

In the case of asset return modelling, the GARCH-M model gives many opportunities to study the influence of the volatility of the process, represented by h_t^2, and the risk of the assets under condition.

Subject index