CHAPTER 2

PRELIMINARIES

In this chapter, I survey some concepts in time series analysis which will be relevant to the discussions in the next chapters. The first part is concerned with univariate time series models and the second part deals with multivariate time series models. I have decided to follow the terminology of Priestley (1988). For more detailed and sometimes alternative accounts, the reader is referred to, e.g., Anderson (1971), Box and Jenkins (1970), Fuller (1976), Granger and Newbold (1986), Harvey (1981), and to parts of Judge et al. (1985).

2.1 UNIVARIATE TIME SERIES

A crucial element in time series modeling is the concept of a white noise process $\{\varepsilon_t\}$, $t = 0,\pm1,\pm2,...$ The definition to be used in the sequel is that $\{\varepsilon_t\}$ is a sequence of uncorrelated random variables with zero mean and constant variance, or

$$\begin{aligned} E\{\varepsilon_t\} &= 0 && \forall t \\ E\{\varepsilon_t^2\} &= \sigma_\varepsilon^2 && \forall t \\ E\{\varepsilon_s\varepsilon_t\} &= 0 && \forall s,t,\ s\neq t, \end{aligned}$$

where E denotes the expectation operator.

Consider the class of linear models

$$\sum_{u=0}^{\infty} h_u y_{t-u} = \varepsilon_t, \tag{2.1}$$

where $\{y_t\}$ is a discrete time series with zero mean, $t = 1,2,..,n$, and where h is a linear function that transforms the series y_t into a white noise process. The mean of y_t is set equal to zero only for notational convenience, but (2.1) may be extended to include a mean μ. The class of models in (2.1) can be written as

$$H(B)y_t = \varepsilon_t, \tag{2.2}$$

with

$$H(z) = \sum_{u=0}^{\infty} h_u z^u,$$

where z can be complex, and where B denotes the backward shift operator, defined by $B^k y_t = y_{t-k}$, for $k = 0,1,2,..$, with $B^0 = 1$. The algebra of the linear operator B is discussed e.g., in Dhrymes (1981, pp. 19-24). It can be shown that the algebra of polynomial operators $R(B)$, where $R(B)$ is defined as the set of all finite linear combinations of elements of $\{B^k;\ k = 0,1,2,..\}$, is isomorphic to the algebra of polynomial functions $R(z)$. This implies for example that $1/(1-\delta B)$ can be written as $\sum_{k=0}^{\infty}\delta^k B^k$.

Under certain conditions it is possible to express y_t as a linear function of present and past values of ε_t, instead of past values of y_t. Then expression (2.2) can be inverted, or

$$y_t = H^{-1}(B)\varepsilon_t. \tag{2.3}$$

If $H^{-1}(z)$ can be expanded as a convergent series for $|z| < 1$, then it can be written as

$$H^{-1}(z) = \sum_{u=0}^{\infty} g_u z^u,$$

and hence (2.3) becomes

$$y_t = \sum_{u=0}^{\infty} g_u \varepsilon_{t-u}. \tag{2.4}$$

This expression is in fact more general than (2.2), a result which is formulated in the Wold decomposition theorem, see e.g., Anderson (1971, p.420).

The requirement $|z| < 1$ means that $H(z)$ has no zeros inside the unit circle. For example, suppose that $H(z)$ is $1+h_1 z$. The root of this equation is found by solving $1+h_1 z = 0$, or $z = -1/h_1$, and no zeros inside the unit circle implies that $|h_1| < 1$. Another convention is to replace z by $1/z$, or to consider $H(z)$ equal to $z+h_1$. In that case $|z| < 1$ means that $H(z)$ has no roots outside the unit circle. In the sequel I will regularly use the latter convention.

The models (2.1) and (2.4) involve an infinite number of parameters,

and to fit them to data it is necessary to make additional assumptions. One such assumption is that $H(z)$ can be approximated by a finite order polynomial of order p, or that (2.1) reduces to

$$y_t + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} = \varepsilon_t, \tag{2.5}$$

which is called an autoregressive model of order p, AR(p), or

$$\phi_p(B) y_t = \varepsilon_t, \tag{2.6}$$

where

$$\phi_p(B) = 1 + \phi_1 B + \ldots + \phi_p B^p.$$

An assumption implying that (2.4) can be written in a similar concise manner is that $H^{-1}(z)$ can be approximated by a finite order polynomial of order q, i.e.

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}, \tag{2.7}$$

or

$$y_t = \theta_q(B) \varepsilon_t, \tag{2.8}$$

where

$$\theta_q(B) = 1 + \theta_1 B + \ldots + \theta_q B^q,$$

which is called a moving average model of order q, MA(q). Another reasonable assumption is that the polynomial $H^{-1}(z)$ can be approximated by a rational function $\theta_q(z)/\phi_p(z)$, where the elements of this ratio are given by the expressions above, such that (2.4) can be written as

$$\phi_p(B) y_t = \theta_q(B) \varepsilon_t, \tag{2.9}$$

which is called a mixed autoregressive–moving average model of order (p,q), or ARMA(p,q). It generally applies that, given a certain accuracy of approximation, the order p and q in (2.9) for a given series y_t are smaller than those in (2.6) or (2.8).

In (2.1) it has been assumed that $\mu = 0$. In case $\mu \neq 0$, y_t can be

replaced by $y_t-\mu$, which implies that its mean is substracted before any modeling analysis. Another possibility is to estimate a model with a constant, from which an estimate for μ can be calculated. For example, for an AR(p) model it applies that this constant equals $-\phi_p(1)\mu$, where $\phi_p(z)$ is defined by (2.6). Henceforth, it will be assumed that all models include a constant term in the estimation phase, but for notational convenience continuing use will be made of zero mean variables.

Stationarity

A useful tool for the identification of ARMA models is the autocorrelation function (ACF). It measures the correlations between current and previous observations. To facilitate an interpretation of this function, it is convenient to assume the stationarity of the time series process y_t. The process y_t is defined to be covariance stationary, or briefly, stationary, when for all t and k it holds that

$$\begin{aligned} &E(y_t) = \mu < \infty && |\mu| < \infty \\ &E(y_t-\mu)^2 = \gamma_0 < \infty \\ &E(y_t-\mu)(y_{t-k}-\mu) = \gamma_k && k = 0,1,2,.. \end{aligned}$$

The γ_k defines the autocovariance of order k. The y_t is called strictly stationary if the multivariate distribution of $(y_t,..y_{t+i})$ is identical to that of the time shifted set $(y_{t+s},..,y_{t+s+i})$. An example of such a strictly stationary process is a Gaussian process where the y_t are assumed to be normally distributed.

An example of a nonstationary process is the random walk model

$$y_t = y_{t-1} + \varepsilon_t, \tag{2.10}$$

which can be written as

$$y_t = y_{t-n} + \varepsilon_{t-n+1} + \varepsilon_{t-n+2} + \ldots + \varepsilon_t,$$

for which it is easy to see that

$$E(y_t^2|y_0) = E(y_{t-n}^2|y_0) + n\sigma_\varepsilon^2.$$

Hence, any new observation causes that the variance of y_t increases with the order n. Note that in this nonstationary AR(1) case where $\phi_1 = -1$ in (2.6), the inverse of the $H(z)$ in (2.2) can not be expanded as a convergent series contrary to (2.4). Hence, there does not exist a useful moving average approximation to this model.

This random walk model belongs to a class of processes which are called integrated processes. A variable y_t is defined to be integrated of order d, or briefly $I(d)$, when $(1-B)^d y_t$ is a stationary variable while $(1-B)^{d-1} y_t$ is not. So, one way to transform the y_t into a stationary series x_t is to apply a difference filter d times, or $x_t = \Delta_k^d y_t$, where Δ_k is defined by $\Delta_k y_t = y_t - y_{t-k}$. For example, $x_t = \Delta_1^2 y_t$ implies the transformation of y_t into $y_t - 2y_{t-1} + y_{t-2}$. A further extension of model (2.9) is now given by an ARIMA(p,d,q) model.

The application of the differencing filter Δ_1 assumes that the model for y_t possesses a unit root. One way to check for such a unit root in the AR(1) case is to estimate

$$y_t = \mu + \delta T_t + \phi_1 y_{t-1} + \varepsilon_t, \tag{2.11}$$

where T_t represents a linear deterministic trend, and to check whether ϕ_1 is equal to 1. This T_t is often included since many economic time series show upward moving patterns which might be explained by such a trend. It is common practice to take the equality $\phi_1 = 1$ as the null hypothesis, and $|\phi_1| < 1$ as the alternative. Unfortunately, most standard statistical techniques do not apply under this nonstationary null hypothesis. One implication of this is that critical values of any test statistics have to be generated by simulations. Given that y_t not only may follow all kinds of ARMA processes, but that the alternative hypothesis may contain, e.g., constant or trend, and that also the distribution of the test statistics may depend on the starting values and on the ratio of μ and σ_ε, one can imagine that the literature on testing for unit roots is rather large. A survey is therefore beyond the scope of this thesis. The testing strategy I will adopt is the regularly applied Dickey and Fuller (1981) approach, which uses the auxiliary regression

$$\Delta_1 y_t = \mu + \delta T_t + \rho y_{t-1} + \psi_1 \Delta_1 y_{t-1} + \ldots + \psi_p \Delta_1 y_{t-p} + \varepsilon_t, \tag{2.12}$$

and hence which assumes that y_t can be approximated by an autoregressive

model of order p+1. The test statistic of interest for the null hypothesis $\rho = 0$ is $\hat{\tau}$, which is the t-ratio for the parameter ρ. Under the alternative hypothesis $\rho < 0$, (2.12) may also be restricted by assuming the absence of deterministic elements, i.e. (μ=0,δ=0) and (δ=0). The critical values of several one-sided test statistics are given in Fuller (1976), but for convenience some relevant values are displayed in table 2.1.

Table 2.1

Some critical values for Dickey-Fuller test statistics

Deterministic elements	Test statistic	Sample size	0.05	0.10
none	$\hat{\tau}$	25	-1.95	-1.60
		100	-1.95	-1.61
constant	$\hat{\tau}_\mu$	25	-3.00	-2.63
		100	-2.89	-2.58
constant and trend	$\hat{\tau}_\tau$	25	-3.60	-3.24
		100	-3.45	-3.15

Source: Fuller (1976, pp. 371-373). The test equation is (2.12).

In case the p in (2.12) equals 0, the test statistic $n\hat{\rho}$ can be used. Since the simulation results in Dickey, Bell and Miller (1986) indicate that the power of this statistic is higher than that of $\hat{\tau}$, it is advised to apply it when y_t can be described by an AR(1) model. Otherwise, this $n\hat{\rho}$ has to be corrected for higher order dynamics, see e.g., Said and Dickey (1984).

Autocorrelation function

Stationarity is assumed for the concepts to be discussed below. The function γ_k establishes the autocorrelation function (ACF) by

$$\rho_k = \gamma_k/\gamma_0, \tag{2.13}$$

for which it applies that $\rho_0 = 1$, that $\rho_{-k} = \rho_k$ and that $-1 < \rho_k < 1$. From the definition of the white noise process it is clear that for ε_t it applies that $\rho_k = 0$ for all $k \neq 0$. For the AR(1) model $y_t = \phi_1 y_{t-1} + \varepsilon_t$ it is easy to derive that $\rho_k = \phi_1^k$ for all $k > 0$. This derivation uses the

result that the covariance of ε_t and y_{t-k} equals zero for all $k > 0$. A formal proof of this is given in Anderson (1971, p.170). An illustration of the determination of the autocorrelation coefficients for higher order autoregressive models is given by considering the AR(2) model,

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = \varepsilon_t. \tag{2.14}$$

Multiplying both sides by y_{t-1}, taking expectations, and deviding them by γ_0 results in

$$\rho_1 - \phi_1 \rho_0 - \phi_2 \rho_1 = 0 \tag{2.15}$$

or

$$\rho_1 = \phi_1/(1-\phi_2).$$

To determine an expression for ρ_2, analogous operations are carried out to (2.14), yielding

$$\rho_2 - \phi_1 \rho_1 - \phi_2 \rho_0 = 0. \tag{2.16}$$

Substituting (2.15) into (2.16) gives that

$$\rho_2 = \phi_1^2/(1-\phi_2) + \phi_2. \tag{2.17}$$

Analogously it can be established that for larger values of k the ρ_k can be solved recursively using $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$.

Partial autocorrelation function

The partial autocorrelation function (PACF), the elements of which will be denoted by ψ_{KK}, can easily be illustrated by considering the AR(p) process, for which it can be shown that

$$\rho_k = \sum_{j=1}^{p} \phi_j \rho_{k-j}, \tag{2.18}$$

for $k > 0$. The partial autocorrelation of order K, ψ_{KK}, is given by the solution of the set of the linear equations

$$\rho_k = \sum_{j=1}^{K} \psi_{Kj} \rho_{k-j}, \tag{2.19}$$

for $k = 1,..,K$. For an AR(1) process with $\phi_1 \neq 0$, and for K is 1, (2.19) reduces to $\rho_1 = \psi_{11}$. For K is 2, this set is given by

$$\rho_1 = \psi_{21} + \psi_{22}\rho_1$$
$$\rho_2 = \psi_{21}\rho_1 + \psi_{22},$$

and since $\rho_2 = \rho_1^2 \neq 0$, it follows that $\psi_{22} = 0$. In general it can be shown that $\psi_{p+k,p+k} = 0$, $k = 1,2,..,$ for an AR(p) process. This provides a tool for the identification of AR time series processes.

The correlation functions are usually unknown in practice and have to be estimated. The ACF can be estimated by means of

$$r_k = C_k/C_0, \tag{2.20}$$

where

$$C_k = \sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t-k} - \bar{y}),$$

where $\bar{y}$ denotes the sample mean of y_t. The sample equivalents of ψ_{KK}, to be denoted as a_{KK}, can be obtained by applying ordinary least squares to

$$y_t - \bar{y} = a_{K1}(y_{t-1} - \bar{y}) + \ldots + a_{KK}(y_{t-K} - \bar{y}) + \nu_t.$$

The issue of estimation of the time series models themselves will however not be discussed in detail here. The interested reader is referred to Anderson (1971, ch.5) and Granger and Newbold (1986, ch.3).

The identification of p of an AR(p) process may be done using the estimated PACF. Analogous results apply to the identification of q of MA(q) processes, where now the ACF turns out to be useful. Consider again (2.7) for which it can be derived that

$$\gamma_k = \sigma_\varepsilon^2 \sum_{i=0}^{q-k} \theta_i\theta_{i+k} \qquad \text{for } k = 0,1,..,q \tag{2.21}$$

$$= 0 \qquad \text{for } k > q,$$

because of the uncorrelated ε_t variates, and where $\theta_0 = 1$.

From (2.3) it can be seen that an implication of stationarity is that a certain $H(z)$ has an inverse $H(z)^{-1}$, which converges. This means that a stationary AR model can be written as a moving average model. The concept

of invertibility of an MA model is dual to this stationarity concept. It is defined by the existence of a converging inverse $G(z)^{-1}$ of an MA polynomial $G(z)$. For example, consider a first order MA process,

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}, \tag{2.22}$$

which can be rewritten as

$$y_t - \theta_1 y_{t-1} + \theta_1^2 y_{t-2} - \theta_1^3 y_{t-3} + .. = \varepsilon_t. \tag{2.23}$$

From this expression it can be seen that for $|\theta_1| \geq 1$ the corresponding $G(z)^{-1}$ for the process on the left hand side does not converge. In that case the MA(1) model is said to be noninvertible. This MA(1) model is invertible when $|\theta_1| < 1$, and it can then be written as an AR(∞) process. Given that the parameters approach zero at higher lags, the PACF of this MA model dies out at high lags.

The patterns of the autocorrelation function and of the partial autocorrelation function of a stationary and invertible ARMA(p,q) model do not cut off after a certain point. A simple example is an ARMA(1,1) model,

$$y_t - \phi_1 y_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}, \tag{2.24}$$

for which it is relatively straightforward to show that

$$\rho_k = \phi_1^{k-1}((1+\phi_1\theta_1)(\phi_1+\theta_1))/(1+2\phi_1\theta_1+\theta_1^2). \tag{2.25}$$

From this expression it can be seen that for distinct choices of ϕ_1 and θ_1, several types of patterns for the ACF can emerge. Together with eventual biases that can occur because of estimation, it may be evident that the identification of an ARMA time series model from the patterns of the (P)ACF alone may be rather difficult. One suggestion is to tentatively specify a model, to estimate it, and check the estimated residuals for the eventual presence of autocorrelation. The type of the detected autocorrelation may then indicate a possible route for modification. This and several more suggestions for time series model selection will be discussed in detail in the next chapter.

Forecasting

An important use of ARMA models is in forecasting exercises. Denote $f_{t,h}$ as the linear h-step forecast of y_{t+h}. First, consider an MA(∞) model as in (2.4) for which this forecast equals

$$f_{t,h} = \sum_{u=0}^{\infty} g_{u+h}\varepsilon_{t-u}, \tag{2.26}$$

with an h-step error

$$e_{t,h} = y_{t+h} - f_{t,h} = \sum_{u=0}^{h-1} g_u\varepsilon_{t+h-u}. \tag{2.27}$$

Given the white noise assumption on ε_t it can be derived that

$$E(e_{t,h}) = 0$$

$$V(h) = E(e_{t,h}^2) = \sigma_\varepsilon^2 \sum_{u=0}^{h-1} g_u^2. \tag{2.28}$$

From (2.26) one can see that for an MA(q) process it applies that $f_{t,h} = 0$ when $h > q$.

The forecasts from an ARMA(p,q) model as in (2.9) are found in a similar way, see Granger and Newbold (1986, p.133), by

$$f_{t,h} - \sum_{j=0}^{p} \phi_j f_{t,h-j} = \sum_{i=0}^{q} \theta_{i+h}\varepsilon_{t-i}, \tag{2.29}$$

where $f_{n,k} = y_{t+k}$ for $k \leq 0$. For an AR(p) the h-step forecasts can be derived from (2.29) by setting the θ_i equal to zero. The variance of these point forecasts is established analogous to (2.28). Denoting

$$\eta(B) = \theta_q(B)/\phi_p(B),$$

then

$$V(h) = \sigma_\varepsilon^2 \sum_{u=0}^{h-1} \eta_u^2, \tag{2.30}$$

for the general ARMA(p,q) model.

2.2 MULTIVARIATE TIME SERIES

One extension of univariate time series models is given by multivariate time series models. Consider an (m×1) zero mean vector $\boldsymbol{y}_t$ containing the observations at time t for m distinct time series. A multiple ARMA model, or vector autoregressive-moving average model, for this $\boldsymbol{y}_t$ is written as

$$\boldsymbol{y}_t - \boldsymbol{\Pi}_1\boldsymbol{y}_{t-1} - \ldots - \boldsymbol{\Pi}_p\boldsymbol{y}_{t-p} = \boldsymbol{\epsilon}_t + \boldsymbol{\Xi}_1\boldsymbol{\epsilon}_{t-1} + \ldots + \boldsymbol{\Xi}_q\boldsymbol{\epsilon}_{t-q}, \tag{2.31}$$

or more compactly

$$\boldsymbol{\Pi}_p(B)\boldsymbol{y}_t = \boldsymbol{\Xi}_q(B)\boldsymbol{\epsilon}_t, \tag{2.32}$$

of which a vector autoregressive model of order p, a VAR(p), and a vector moving average model of order q, VMA(q), are special cases. The $\boldsymbol{\epsilon}_t$ is defined to be a multivariate white noise process, $t = 0,\pm1,+2$, or

$$\begin{aligned} E\{\boldsymbol{\epsilon}_t\} &= \mathbf{0} && \forall t \\ E\{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\} &= \boldsymbol{\Sigma}_\epsilon && \forall t \\ E\{\boldsymbol{\epsilon}_s\boldsymbol{\epsilon}_t'\} &= \mathbf{0} && \forall s,t,\ s\neq t. \end{aligned}$$

The $\boldsymbol{\Pi}_p(B)$ and $\boldsymbol{\Xi}_q(B)$ are matrices, the elements of which are polynomials in the backward shift operator B. For example, consider a simple bivariate ARMA(1,1) model given by

$$\begin{bmatrix} 1-\pi_{11}B & -\pi_{12}B \\ -\pi_{21}B & 1-\pi_{22}B \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_t = \begin{bmatrix} 1+\alpha_{11}B & \alpha_{12}B \\ \alpha_{21}B & 1+\alpha_{22}B \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}_t, \tag{2.33}$$

of which the first equation can be written as

$$y_{1t} - \pi_{11}y_{1t-1} - \pi_{12}y_{2t-1} = \varepsilon_{1t} + \alpha_{11}\varepsilon_{1t-1} + \alpha_{12}\varepsilon_{2t-1}.$$

There are several restrictions on the elements of the matrices $\boldsymbol{\Pi}_p(B)$ and $\boldsymbol{\Xi}_q(B)$ to ensure the uniqueness of multivariate time series models, see for details, e.g., Hannan (1970) and Granger and Newbold (1986).

Model (2.32) is called stationary in case the roots of $|\boldsymbol{\Pi}(z)| = 0$ all lie outside the unit circle. For (2.33) this implies that the absolute

values of the roots of

$$1 - (\pi_{11}+\pi_{22})z + (\pi_{11}\pi_{22} - \pi_{12}\pi_{21})z^2 = 0, \tag{2.34}$$

should exceed unity. Analogously, model (2.32) is said to be invertible in case the roots of $|\Xi(z)| = 0$ are all outside the unit circle.

Identification of a stationary multivariate time series model may proceed along similar lines as for the univariate models. One can define a measure of multiple autocorrelations $E(\boldsymbol{y}_t\boldsymbol{y}_{t-k}')$, of which it can be shown to have similar characteristics as its univariate analogue. For example, for a VMA(q) model it applies that $E(\boldsymbol{y}_t\boldsymbol{y}_{t-q-i}') = \mathbf{0}$ for $i > 0$. One can imagine however that in case of large m, this strategy of identification may not be that easy. Suppose a tentative model is fitted using one of the estimation methods referenced in Granger and Newbold (1986), then a test for residual autocorrelation may be a first check on the appropriateness of the model.

Cointegration

In case the vector $\boldsymbol{y}_t$ contains nonstationary elements, e.g., some or all of the y_{it} are $I(1)$ variables, then one might consider a multiple time series model for the vector containing all variables in first differences. It is however conceivable that there may be linear combinations of some or all of the y_{it} which are $I(0)$. Such combinations might be viewed as equilibrium relationships. If there are such equilibria, it can be said that the y_{it} are cointegrated. And when there is cointegration, one can argue that a model containing only the first differenced variables lacks the information which is available in the equilibrium relations, see Engle and Yoo (1987).

To formalize, it is convenient to consider a VAR(p) model

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\Pi}_1\boldsymbol{y}_{t-1} + \ldots + \boldsymbol{\Pi}_p\boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t, \tag{2.35}$$

see Johansen (1988), and to rewrite this model as

$$\Delta_1\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\Gamma}_1\Delta_1\boldsymbol{y}_{t-1} + \ldots + \boldsymbol{\Gamma}_{p-1}\Delta_1\boldsymbol{y}_{t-p+1} + \boldsymbol{\Pi}\boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t, \tag{2.36}$$

where $\boldsymbol{\Gamma}_i = -\mathbf{I} + \boldsymbol{\Pi}_1 + .. + \boldsymbol{\Pi}_i$ for $i = 1,..,p-1$ and

$$\boldsymbol{\Pi} = -\mathbf{I} + \boldsymbol{\Pi}_1 + .. + \boldsymbol{\Pi}_p. \tag{2.37}$$

This matrix Π is of special interest for it contains information on the cointegration relations. One can distinguish several cases with respect to the rank of Π, say r. The first is that r is equal to m, which indicates that the process $\boldsymbol{y}_t$ is stationary. The second is that r equals 0, i.e. the Π is the null matrix. Model (2.36) now reduces to a model for the vector with the variables in first differences. The third case in which $0 < r < m$ is the most interesting one for it implies the presence of r cointegration relationships, which in essence is the content of the so-called Granger representation theorem (cf. Engle and Granger 1987). The matrix Π can now be written as the product of two $(m \times r)$ matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, or $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$. The interpretation of these matrices is that $\boldsymbol{\beta}'\boldsymbol{y}_t$ is stationary, which in model (2.36) is called an error correction term, and that $\boldsymbol{\alpha}$ contains the coefficients which measure the adjustment of the elements in $\boldsymbol{y}_t$ to disequilibrium errors. For more extensive treatments of the concepts and alternative representations, I refer the reader to Engle and Granger (1987) and Hylleberg and Mizon (1989).

Given certain sets of assumptions, there are several methods to estimate the elements of the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ such as those in Stock and Watson (1989), Engle and Granger (1987), and Boswijk (1991). Most of these methods assume either the value of r or some type of exogeneity for several of the y_{it}. The method in Johansen (1988,1989) however necessitates neither of these assumptions. In fact, this method starts with the estimation of r. An additional advantage is then that a test for integration and cointegration can be carried out simultaneously. The idea is to choose the r linear combinations of the elements of $\boldsymbol{y}_t$ which have the largest partial correlations with the stationary process $\Delta_1\boldsymbol{y}_t$ after corrections for lagged values of $\Delta_1\boldsymbol{y}_t$ (Johansen 1988) and, additionally for eventual mean and trend (Johansen and Juselius 1990). One useful statistical technique is canonical correlation analysis. The partial correlations of interest are the eigenvalues λ_i, where $\lambda_i \geq \lambda_{i+1}$, of the canonical correlation matrix of $\boldsymbol{\beta}'\boldsymbol{y}_t$ and $\Delta_1\boldsymbol{y}_t$ after the above corrections. The columns of $\boldsymbol{\beta}$ are now given by the corresponding eigenvectors. Johansen (1988,1989) has developed test statistics for the number of cointegration relations, one of which is given by the likelihood ratio test statistic

$$Tr(r) = -n \sum_{i=r+1}^{m} \log(1 - \hat{\lambda}_i), \tag{2.38}$$

for the null hypothesis of r cointegration relations, where log denotes the

natural logarithm. Another test is given by testing the significance of the estimated eigenvalues themselves, or

$$Ev(r) = -n\log(1 - \hat{\lambda}_{r+1}), \tag{2.39}$$

which in fact tests r against r+1. Unfortunately, standard distributional results are not valid since this null hypothesis involves nonstationary variables, which provides that the critical values of the test statistics have to be tabulated. This Johansen procedure will often be applied in later sections, so for convenience several relevant critical values will be displayed in table 2.2.

Table 2.2

Critical values for the Johansen cointegration test statistics[1]

	Constant and trend[2]		Constant[2]	
$l=m-r$[3]	90%	95%	90%	95%
	Maximal eigenvalue, $Ev(r)$			
1	6.691	8.083	7.563	9.094
2	12.783	14.595	13.781	15.752
3	18.959	21.279	19.796	21.894
4	24.917	27.341	25.611	28.167
5	30.818	33.262	31.592	34.397
	Trace, $Tr(r)$			
1	6.691	8.083	7.563	9.094
2	15.583	17.844	17.957	20.168
3	28.436	31.256	32.093	35.068
4	45.248	48.419	49.925	53.347
5	65.956	69.977	71.472	75.328

Source: Johansen and Juselius (1990), Tables A.2 and A.3.

(1) The statistics are given in (2.38) and (2.39).

(2) The deterministic elements assumed for the data generating process.

(3) The m is the number of elements in the vector $\boldsymbol{y}_t$, r is the number of cointegration relationships.

This Johansen procedure is applied in a variety of recent occasions with apparent success, see Hendry and Mizon (1990) and Kunst and Neusser (1990) for macroeconomic applications, and Baillie and Bollerslev (1989), and Vogelvang (1990) for financial studies. In Franses and Kofman (1991) the method is used for an empirical test of parities between metal prices at the London Metal Exchange. From simulation evidence reported in Boswijk and Franses (1991) it emerges that empirical size and power properties of this cointegration method are quite reasonable assuming that one does not select too high or too low the order p in (2.35). It appears that in general, too few lags in the model increases the size, leading to spurious cointegration, whereas too many lags in the model decreases the power of the tests. This indicates that there is some optimal lag length. Therefore it seems a sensible strategy to test for cointegration in a range of candidate models.