7

Estimation of Vector Error Correction Models

In this chapter, estimation of VECMs is discussed. The asymptotic properties of estimators for nonstationary models differ in important ways from those of stationary processes. Therefore, in the first section, a simple special case model with no lagged differences and no deterministic terms is considered and different estimation methods for the parameters of the error correction term are treated. For this simple case, the asymptotic properties can be derived with a reasonable amount of effort and the difference to estimation in stationary models can be seen fairly easily. Therefore it is useful to treat this case in some detail. The results can then be extended to more general VECMs which are considered in Section 7.2. In Section 7.3, Bayesian estimation including the Minnesota or Litterman prior for integrated processes is discussed and forecasting and structural analysis based on estimated processes are considered in Sections 7.4–7.6.

7.1 Estimation of a Simple Special Case VECM

In this section, a simple VECM without lagged differences and deterministic terms is considered. More precisely, the model of interest is

$$\Delta y_t = \mathbf{\Pi} y_{t-1} + u_t = \alpha\beta' y_{t-1} + u_t, \quad t = 1, 2, \ldots, \tag{7.1.1}$$

where y_t is K-dimensional, $\mathbf{\Pi}$ is a $(K \times K)$ matrix of rank r, $0 < r < K$, α and β are $(K \times r)$ with rank r, and u_t is K-dimensional white noise with mean zero and nonsingular covariance matrix Σ_u. For simplicity, we assume that u_t is standard white noise so that certain limiting results hold which will be discussed and used in the following. For the time being, the initial vector y_0 is arbitrary with some fixed distribution. We also assume that y_t is an $I(1)$ vector so that we know from Section 6.3 that the $((K-r) \times (K-r))$ matrix

$$\alpha'_\perp \beta_\perp$$

is invertible (see Eq. (6.3.12)). Here $\alpha_\perp$ and $\beta_\perp$ are, as usual, orthogonal complements of α and β, respectively.

The cointegration rank r is assumed to be known and it is strictly between 0 and K. For $r = 0$, Δy_t is stable and for $r = K$, y_t is stable. For the present purposes, these two boundary cases are of limited interest because they can be treated in the stationary framework considered in Part I. If r is not known, however, it may be of interest to consider the case $r = 0$. The matrix $\mathbf{\Pi}$ is then zero, of course. We will comment on this case at the end of this section.

We will discuss different estimators of the matrix $\mathbf{\Pi}$, assuming that a sample $y_1, \dots, y_T$ and a presample vector y_0 are available. Our first estimator is the unrestricted LS estimator,

$$\widehat{\mathbf{\Pi}} = \left(\sum_{t=1}^{T} \Delta y_t y_{t-1}'\right)\left(\sum_{t=1}^{T} y_{t-1} y_{t-1}'\right)^{-1}. \tag{7.1.2}$$

Substituting $\mathbf{\Pi} y_{t-1} + u_t$ for Δy_t gives

$$\widehat{\mathbf{\Pi}} - \mathbf{\Pi} = \left(\sum_{t=1}^{T} u_t y_{t-1}'\right)\left(\sum_{t=1}^{T} y_{t-1} y_{t-1}'\right)^{-1}. \tag{7.1.3}$$

To derive the asymptotic distribution of this quantity, we multiply from the left with the $(K \times K)$ matrix

$$Q := \begin{bmatrix} \beta' \\ \alpha_\perp' \end{bmatrix}$$

and from the right by

$$Q^{-1} = [\alpha(\beta'\alpha)^{-1} : \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}]$$

which yields

$$\begin{aligned} Q(\widehat{\mathbf{\Pi}} - \mathbf{\Pi})Q^{-1} &= Q\left(\sum_{t=1}^{T} u_t y_{t-1}'\right) Q'Q^{-1\prime}\left(\sum_{t=1}^{T} y_{t-1} y_{t-1}'\right)^{-1} Q^{-1} \\ &= \left(\sum_{t=1}^{T} v_t z_{t-1}'\right)\left(\sum_{t=1}^{T} z_{t-1} z_{t-1}'\right)^{-1}, \end{aligned} \tag{7.1.4}$$

where $v_t := Qu_t$ and $z_t := Qy_t$. Notice that invertibility of $\alpha_\perp'\beta_\perp$ follows from our assumption of an $I(1)$ system, as mentioned earlier, and it implies that the inverse of Q exists because

$$\begin{bmatrix} \beta' \\ \alpha_\perp' \end{bmatrix} [\beta : \beta_\perp] = \begin{bmatrix} \beta'\beta & 0 \\ \alpha_\perp'\beta & \alpha_\perp'\beta_\perp \end{bmatrix}$$

is invertible if $\alpha_\perp'\beta_\perp$ is nonsingular. Hence, Q must be invertible and, thus, $\beta'\alpha$ is also nonsingular.

Premultiplying the VECM (7.1.1) by Q shows that

$$\Delta z_t = Q\mathbf{\Pi}Q^{-1}z_{t-1} + v_t = \begin{bmatrix} \beta'\alpha & 0 \\ 0 & 0 \end{bmatrix} z_{t-1} + v_t.$$

Hence, denoting the first r components of z_t by $z_t^{(1)}$, we know that $z_t^{(1)} = \beta' y_t$ consists of the cointegrating relations and is therefore stationary while the last $K-r$ components of z_t, denoted by $z_t^{(2)}$, constitute a $(K-r)$-dimensional random walk because $\Delta z_t^{(2)}$ is white noise. Thus, stationary and nonstationary components are separated in z_t. To derive the asymptotic properties of the LS estimator, it is useful to write

$$Q(\widehat{\mathbf{\Pi}} - \mathbf{\Pi})Q^{-1}$$
$$= \left[\sum_{t=1}^{T} v_t z_{t-1}^{(1)\prime} : \sum_{t=1}^{T} v_t z_{t-1}^{(2)\prime}\right] \begin{bmatrix} \sum_t z_{t-1}^{(1)} z_{t-1}^{(1)\prime} & \sum_t z_{t-1}^{(1)} z_{t-1}^{(2)\prime} \\ \sum_t z_{t-1}^{(2)} z_{t-1}^{(1)\prime} & \sum_t z_{t-1}^{(2)} z_{t-1}^{(2)\prime} \end{bmatrix}^{-1}. \quad (7.1.5)$$

For the cross product terms in this relation, we have the following special case results from Ahn & Reinsel (1990).

Lemma 7.1

(1) $T^{-1}\sum_{t=1}^{T} z_{t-1}^{(1)} z_{t-1}^{(1)\prime} = T^{-1}\sum_{t=1}^{T} \beta' y_{t-1} y_{t-1}' \beta \xrightarrow{p} \Gamma_z^{(1)}$.

(2) $T^{-1/2}\text{vec}\left(\sum_{t=1}^{T} v_t z_{t-1}^{(1)\prime}\right) \xrightarrow{d} \mathcal{N}(0, \Gamma_z^{(1)} \otimes \Sigma_v)$,
where $\Sigma_v := Q\Sigma_u Q'$ is the covariance matrix of v_t.

(3) $T^{-1}\sum_{t=1}^{T} v_t z_{t-1}^{(2)\prime} \xrightarrow{d} \Sigma_v^{1/2}\left(\int_0^1 \mathbf{W}_K d\mathbf{W}_K'\right)' \Sigma_v^{1/2}\begin{bmatrix} 0 \\ I_{K-r} \end{bmatrix}$,
where $\mathbf{W}_K$ abbreviates a standard Wiener process $\mathbf{W}_K(s)$ of dimension K (see Appendix C.8.2).

(4) $T^{-3/2}\sum_{t=1}^{T} z_{t-1}^{(1)} z_{t-1}^{(2)\prime} \xrightarrow{p} 0$.

(5) $T^{-2}\sum_{t=1}^{T} z_{t-1}^{(2)} z_{t-1}^{(2)\prime} \xrightarrow{d} [0 : I_{K-r}]\Sigma_v^{1/2}\left(\int_0^1 \mathbf{W}_K \mathbf{W}_K' ds\right)\Sigma_v^{1/2}\begin{bmatrix} 0 \\ I_{K-r} \end{bmatrix}$.

The quantities in (2), (3), and (5) converge jointly. ■

In this lemma we encounter asymptotic distributions of random matrices. As in Appendix C.8.2, these are understood as the limits in distribution of the vectorized quantities. Because the asymptotic distributions are also conveniently stated in matrix form, not using vectorization here is a useful simplification. Moreover, in the lemma as well as in the following analysis we denote the square root of a positive definite matrix Σ by $\Sigma^{1/2}$, that is, $\Sigma^{1/2}$ is the positive definite symmetric matrix for which $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$ (see Appendix A.9.2).

Proof: The proof follows Ahn & Reinsel (1990). Lemma 7.1(1) is implied by a standard weak law of large numbers (see, e.g., Proposition C.12(7)) because $z_{t-1}^{(1)}$ contains stationary components only.

The second result also involves stationary processes only. Therefore it follows from a martingale difference central limit theorem for stationary processes. Notice that $\text{vec}(v_t z_{t-1}^{(1)\prime})$ is a martingale difference sequence and, hence, a martingale difference array which satisfies the conditions of Proposition C.13(2). Thus, the result follows from that proposition.

To show Lemma 7.1(3), we define a random walk

$$z_t^* = \begin{bmatrix} z_t^{*(1)} \\ z_t^{(2)} \end{bmatrix} = z_{t-1}^* + v_t, \quad t = 1, 2, \ldots,$$

with $z_0^{*(1)} = 0$ and notice that the second part of z_t^* is identical to the last $K - r$ components of z_t. Hence, it follows from Proposition C.18(6) that

$$T^{-1}\sum_{t=1}^{T} v_t z_{t-1}^{*\prime} \xrightarrow{d} \Sigma_v^{1/2}\left(\int_0^1 \mathbf{W}_K d\mathbf{W}_K'\right)' \Sigma_v^{1/2}.$$

Considering the last $K - r$ columns only gives the desired result.

Part (4) of the lemma can be shown by defining

$$z_t^+ = \begin{bmatrix} z_t^{+(1)} \\ z_t^{(2)} \end{bmatrix} = z_{t-1}^+ + v_t^+, \quad t = 1, 2, \ldots,$$

with $z_0^{+(1)} = 0$ and

$$v_t^+ = \begin{bmatrix} z_t^{(1)} \\ v_t^{(2)} \end{bmatrix}.$$

Thus, v_t^+ is an $I(0)$ process. By Proposition C.18(5), we have

$$\sum_{t=1}^{T} z_{t-1}^+ v_t^{+\prime} = \begin{bmatrix} \sum_t z_{t-1}^{+(1)} z_t^{(1)\prime} & \sum_t z_{t-1}^{+(1)} v_t^{(2)\prime} \\ \sum_t z_{t-1}^{(2)} z_t^{(1)\prime} & \sum_t z_{t-1}^{(2)} v_t^{(2)\prime} \end{bmatrix} = O_p(T),$$

which implies the desired result.

Lemma 7.1(5) is just a special case of Proposition C.18(9) because z_t^* is a random walk and the last $K - r$ components of z_t^* are just $z_t^{(2)}$.

Finally, the joint convergence of the quantities in Lemma 7.1(2), (3), and (5) follows because all quantities are eventually made up of the same u_t's. ■

The lemma implies the following limiting result for the LS estimator $\widehat{\mathbf{\Pi}}$.

Result 1
Let

$$D = \begin{bmatrix} T^{1/2} & 0 \\ 0 & T \end{bmatrix}.$$

Then

$$\begin{aligned}
&\text{vec}[Q(\widehat{\mathbf{\Pi}} - \mathbf{\Pi})Q^{-1}D] \\
&\xrightarrow{d} \begin{bmatrix} \mathcal{N}(0, (\Gamma_z^{(1)})^{-1} \otimes \Sigma_v) \\ \text{vec}\Bigg\{ \Sigma_v^{1/2} \left(\int_0^1 \mathbf{W}_K d\mathbf{W}_K' \right)' \Sigma_v^{1/2} \begin{bmatrix} 0 \\ I_{K-r} \end{bmatrix} \\ \qquad \times \left([0 : I_{K-r}] \Sigma_v^{1/2} \left(\int_0^1 \mathbf{W}_K \mathbf{W}_K' ds \right) \Sigma_v^{1/2} \begin{bmatrix} 0 \\ I_{K-r} \end{bmatrix} \right)^{-1} \Bigg\} \end{bmatrix}.
\end{aligned} \quad (7.1.6)$$

■

Proof:

$$\begin{aligned}
&Q(\widehat{\mathbf{\Pi}} - \mathbf{\Pi})Q^{-1}D \\
&= \left[T^{-1/2} \sum_{t=1}^{T} v_t z_{t-1}^{(1)\prime} : T^{-1} \sum_{t=1}^{T} v_t z_{t-1}^{(2)\prime} \right] \\
&\quad \times D \begin{bmatrix} \sum_t z_{t-1}^{(1)} z_{t-1}^{(1)\prime} & \sum_t z_{t-1}^{(1)} z_{t-1}^{(2)\prime} \\ \sum_t z_{t-1}^{(2)} z_{t-1}^{(1)\prime} & \sum_t z_{t-1}^{(2)} z_{t-1}^{(2)\prime} \end{bmatrix}^{-1} D \\
&= \Bigg[\left(T^{-1/2} \sum_{t=1}^{T} v_t z_{t-1}^{(1)\prime} \right) \left(T^{-1} \sum_{t=1}^{T} z_{t-1}^{(1)} z_{t-1}^{(1)\prime} \right)^{-1} \\
&\quad : \left(T^{-1} \sum_{t=1}^{T} v_t z_{t-1}^{(2)\prime} \right) \left(T^{-2} \sum_{t=1}^{T} z_{t-1}^{(2)} z_{t-1}^{(2)\prime} \right)^{-1} \Bigg] + o_p(1).
\end{aligned}$$

The last equality follows from Lemma 7.1(4). The result in (7.1.6) is obtained by vectorizing this matrix and applying Lemma 7.1(2), (3), and (5) and the continuous mapping theorem (see Appendix C.8). ■

An immediate implication of Result 1 follows.

Result 2
The estimator $\widehat{\mathbf{\Pi}}$ is asymptotically normal,

$$\sqrt{T}\text{vec}(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}) \xrightarrow{d} \mathcal{N}\left(0, \beta(\Gamma_z^{(1)})^{-1}\beta' \otimes \Sigma_u\right), \tag{7.1.7}$$

and $\beta(\Gamma_z^{(1)})^{-1}\beta'$ can be estimated consistently by

$$\left(T^{-1}\sum_{t=1}^{T} y_{t-1}y'_{t-1}\right)^{-1}.$$

■

Proof:

$$\begin{aligned}
&\sqrt{T}Q(\widehat{\mathbf{\Pi}} - \mathbf{\Pi})Q^{-1} \\
&= Q(\widehat{\mathbf{\Pi}} - \mathbf{\Pi})Q^{-1}D\begin{bmatrix} 1 & 0 \\ 0 & T^{-1/2} \end{bmatrix} \\
&= \Bigg[\left(T^{-1/2}\sum_{t=1}^{T} v_t z_{t-1}^{(1)\prime}\right)\left(T^{-1}\sum_{t=1}^{T} z_{t-1}^{(1)} z_{t-1}^{(1)\prime}\right)^{-1} \\
&\qquad : T^{-1/2}\left(T^{-1}\sum_{t=1}^{T} v_t z_{t-1}^{(2)\prime}\right)\left(T^{-2}\sum_{t=1}^{T} z_{t-1}^{(2)} z_{t-1}^{(2)\prime}\right)^{-1}\Bigg] + o_p(1)
\end{aligned}$$

from the proof of Result 1 and, hence,

$$\begin{aligned}
&\sqrt{T}\text{vec}[Q(\widehat{\mathbf{\Pi}} - \mathbf{\Pi})Q^{-1}] = (Q^{-1\prime} \otimes Q)\sqrt{T}\text{vec}(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}) \\
&\xrightarrow{d} \begin{bmatrix} \mathcal{N}\left(0, (\Gamma_z^{(1)})^{-1} \otimes \Sigma_v\right) \\ 0 \end{bmatrix}.
\end{aligned}$$

Premultiplying by $Q' \otimes Q^{-1}$ and recalling the definition of Q, gives a multivariate normal limiting distribution with covariance matrix

$$(Q' \otimes Q^{-1})\left(\begin{bmatrix} (\Gamma_z^{(1)})^{-1} & 0 \\ 0 & 0 \end{bmatrix} \otimes \Sigma_v\right)(Q \otimes Q^{-1\prime})$$

or

$$[\beta : \alpha_\perp]\begin{bmatrix} (\Gamma_z^{(1)})^{-1} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \beta' \\ \alpha'_\perp \end{bmatrix} \otimes Q^{-1}\Sigma_v Q^{-1\prime}$$

which implies (7.1.7) because $\Sigma_v = Q\Sigma_u Q'$.

Now consider

$$\left(T^{-1}\sum_{t=1}^{T} y_{t-1}y'_{t-1}\right)^{-1} = Q'\left(T^{-1}\sum_{t=1}^{T} z_{t-1}z'_{t-1}\right)^{-1} Q$$

$$= Q' \begin{bmatrix} T^{-1}\sum_t z_{t-1}^{(1)} z_{t-1}^{(1)\prime} & T^{-1}\sum_t z_{t-1}^{(1)} z_{t-1}^{(2)\prime} \\ T^{-1}\sum_t z_{t-1}^{(2)} z_{t-1}^{(1)\prime} & T^{-1}\sum_t z_{t-1}^{(2)} z_{t-1}^{(2)\prime} \end{bmatrix}^{-1} Q$$

$$= Q' \begin{bmatrix} S_{11}^{-1} + S_{11}^{-1} S_{12} S^* S_{21} S_{11}^{-1} & -S_{11}^{-1} S_{12} S^* \\ -S^* S_{21} S_{11}^{-1} & S^* \end{bmatrix} Q,$$

where the rules for the partitioned inverse from Appendix A.10 have been used and $S^* := (S_{22}^{-1} - S_{21} S_{11}^{-1} S_{12})^{-1}$. Moreover,

$$S_{11} := T^{-1} \sum_t z_{t-1}^{(1)} z_{t-1}^{(1)\prime} \xrightarrow{p} \Gamma_z^{(1)}$$

by Lemma 7.1(1),

$$S_{12} = S_{21}' := T^{-1} \sum_t z_{t-1}^{(1)} z_{t-1}^{(2)\prime} = o_p(T^{1/2})$$

by Lemma 7.1(4), and

$$S_{22} := T^{-1} \sum_t z_{t-1}^{(2)} z_{t-1}^{(2)\prime}.$$

By Lemma 7.1(5) and the continuous mapping theorem, $S_{22}^{-1} = O_p(T^{-1})$. Using again the rules for the partitioned inverse from Appendix A.10,

$$\begin{aligned} S^* &= S_{22}^{-1} + S_{22}^{-1} S_{21} (S_{11} - S_{12} S_{22}^{-1} S_{21})^{-1} S_{12} S_{22}^{-1} \\ &= O_p(T^{-1}) + O_p(T^{-1}) o_p(T^{1/2}) O_p(1) o_p(T^{1/2}) O_p(T^{-1}) \\ &= O_p(T^{-1}), \end{aligned}$$

because

$$S_{11} - S_{12} S_{22}^{-1} S_{21} = S_{11} - o_p(T^{1/2}) O_p(T^{-1}) o_p(T^{1/2}) = S_{11} + o_p(1)$$

so that

$$(S_{11} - S_{12} S_{22}^{-1} S_{21})^{-1} = O_p(1).$$

Hence, we get

$$\begin{aligned} S_{11}^{-1} + S_{11}^{-1} S_{12} S^* S_{21} S_{11}^{-1} &= (\Gamma_z^{(1)})^{-1} \\ &\quad + O_p(1) o_p(T^{1/2}) O_p(T^{-1}) o_p(T^{1/2}) O_p(1) \\ &= (\Gamma_z^{(1)})^{-1} + o_p(1) \end{aligned}$$

and

$$-S_{11}^{-1} S_{12} S^* = O_p(1) o_p(T^{1/2}) O_p(T^{-1}) = o_p(1).$$

Thus,

$$\left(T^{-1}\sum_{t=1}^{T} y_{t-1}y'_{t-1}\right)^{-1} = Q'\begin{bmatrix} (\Gamma_z^{(1)})^{-1} + o_p(1) & o_p(1) \\ o_p(1) & o_p(1) \end{bmatrix} Q$$
$$= \beta(\Gamma_z^{(1)})^{-1}\beta' + o_p(1),$$

which proves Result 2. ■

Thus, the limiting distribution of $\sqrt{T}\text{vec}(\widehat{\mathbf{\Pi}} - \mathbf{\Pi})$ is singular because $\Gamma_z^{(1)}$ is an $(r \times r)$ matrix. Still, we can use the usual estimator of the covariance matrix based on the regressor matrix. Thus, t-ratios can be set up in the standard way and have their usual asymptotic standard normal distributions, if a consistent estimator of Σ_u is used. In Result 8, we will see that the usual residual covariance matrix is in fact a consistent estimator for Σ_u, as in the stationary case. On the other hand, it is not difficult to see that the covariance matrix in the limiting distribution (7.1.7) has rank rK. Therefore, setting up a Wald test for more general restrictions may be problematic. As explained in Appendix C.7, a nonsingular weighting matrix is needed for the Wald test to have its usual limiting χ^2-distribution under the null hypothesis. Thus, if we want to test, for example,

$$H_0 : \mathbf{\Pi} = 0 \quad \text{versus} \quad H_1 : \mathbf{\Pi} \neq 0,$$

the corresponding Wald statistic is

$$\lambda_W = T\text{vec}(\widehat{\mathbf{\Pi}})' \left(\left(T^{-1}\sum_{t=1}^{T} y_{t-1}y'_{t-1} \right) \otimes \widehat{\Sigma}_u^{-1} \right) \text{vec}(\widehat{\mathbf{\Pi}}).$$

Under H_0, the arguments in the proof of Result 2 can be used to show that $T^{-1}\sum_{t=1}^{T} y_{t-1}y'_{t-1}$ converges to zero in probability and, hence, the limit of the weighting matrix in the Wald statistic is singular. Thus, λ_W will not have an asymptotic $\chi^2(K^2)$-distribution. Therefore, caution is necessary in setting up F-tests, for example. In the nonstationary case, they may not have an asymptotic justification. We will provide more discussion of this problem in Section 7.6 in the context of testing for Granger-causality.

It is interesting to note that the asymptotic distribution in (7.1.7) is the same one that is obtained if the cointegration matrix β is known and only α is estimated by LS. To see this result, we consider the LS estimator

$$\widehat{\alpha} = \left(\sum_{t=1}^{T} \Delta y_t y'_{t-1}\beta\right)\left(\sum_{t=1}^{T} \beta' y_{t-1}y'_{t-1}\beta\right)^{-1}. \tag{7.1.8}$$

This estimator has the following properties.

Result 3

$$\sqrt{T}\text{vec}(\widehat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}(0, (\Gamma_z^{(1)})^{-1} \otimes \Sigma_u) \tag{7.1.9}$$

and, thus,

$$\sqrt{T}\text{vec}(\widehat{\alpha}\beta' - \mathbf{\Pi}) \xrightarrow{d} \mathcal{N}(0, \beta(\Gamma_z^{(1)})^{-1}\beta' \otimes \Sigma_u).$$

■

Proof: Substituting $\alpha\beta' y_{t-1} + u_t$ for Δy_t in (7.1.8) and rearranging terms gives

$$\widehat{\alpha} - \alpha = \left(\sum_{t=1}^{T} u_t y_{t-1}'\beta\right)\left(\sum_{t=1}^{T} \beta' y_{t-1} y_{t-1}'\beta\right)^{-1}$$

from which we get (7.1.9) by similar arguments as in the proof of Lemma 7.1. Noting that $\text{vec}(\widehat{\alpha}\beta' - \mathbf{\Pi}) = (\beta \otimes I_K)\text{vec}(\widehat{\alpha} - \alpha)$, gives the stated asymptotic distribution of $\sqrt{T}\text{vec}(\widehat{\alpha}\beta - \mathbf{\Pi})$. ■

Clearly, this result may seem a bit surprising because it means that knowledge of β does not improve our estimator for $\mathbf{\Pi}$, at least asymptotically. In turn, not knowing β does not lead to a reduction in asymptotic precision of our estimator. This is a consequence of the fact that β can be estimated with a better convergence rate than $\sqrt{T}$. To see this fact, suppose for the moment that α is known and that β is normalized as in (6.3.9) such that

$$\beta = \begin{bmatrix} I_r \\ \beta_{(K-r)} \end{bmatrix}. \tag{7.1.10}$$

We know from the discussion in Section 6.3 that this normalization is always possible if the variables are arranged appropriately. Thus, upon normalization, the only unknown elements of β are in the $((K-r) \times r)$ matrix $\beta_{(K-r)}$. This matrix can be estimated from

$$\Delta y_t - \alpha y_{t-1}^{(1)} = \alpha\beta_{(K-r)}' y_{t-1}^{(2)} + u_t = (y_{t-1}^{(2)\prime} \otimes \alpha)\text{vec}(\beta_{(K-r)}') + u_t, \tag{7.1.11}$$

where $y_{t-1}^{(1)}$ and $y_{t-1}^{(2)}$ consist of the first r and the last $K-r$ elements of y_{t-1}, respectively. Because this is a multivariate regression model where the regressors are not identical in the different equations, we assume for the moment that Σ_u is also known and consider the GLS estimator

$$\begin{aligned}\text{vec}(\widehat{\beta}_{(K-r)}') &= \left[\left(\sum_{t=1}^{T} y_{t-1}^{(2)} y_{t-1}^{(2)\prime}\right)^{-1} \otimes (\alpha'\Sigma_u^{-1}\alpha)^{-1}\right] \\ &\quad \times (I_T \otimes \alpha'\Sigma_u^{-1})\text{vec}\left(\sum_{t=1}^{T} (\Delta y_t - \alpha y_{t-1}^{(1)}) y_{t-1}^{(2)\prime}\right).\end{aligned}$$

or

$$\widehat{\beta}'_{(K-r)} = (\alpha'\Sigma_u^{-1}\alpha)^{-1}\alpha'\Sigma_u^{-1} \times \left(\sum_{t=1}^{T}(\Delta y_t - \alpha y_{t-1}^{(1)})y_{t-1}^{(2)\prime}\right)\left(\sum_{t=1}^{T} y_{t-1}^{(2)}y_{t-1}^{(2)\prime}\right)^{-1}. \tag{7.1.12}$$

This estimator has the following asymptotic distribution.

Result 4

$$T(\widehat{\beta}'_{(K-r)} - \beta'_{(K-r)}) \xrightarrow{d} \left(\int_0^1 \mathbf{W}^*_{K-r} d\mathbf{W}^{*\prime}_r\right)'\left(\int_0^1 \mathbf{W}^*_{K-r}\mathbf{W}^{*\prime}_{K-r} ds\right)^{-1}, \tag{7.1.13}$$

where

$$\mathbf{W}^*_{K-r} := Q^{22}[0 : I_{K-r}]\Sigma_v^{1/2}\mathbf{W}_K,$$

Q^{22} denotes the lower right-hand $((K-r) \times (K-r))$ block of Q^{-1} and

$$\mathbf{W}^*_r := (\alpha'\Sigma_u^{-1}\alpha)^{-1}\alpha'\Sigma_u^{-1}Q^{-1}\Sigma_v^{1/2}\mathbf{W}_K.$$

Thus, the asymptotic distribution depends on functionals of a standard Wiener process. ■

Proof: Replacing $\Delta y_t - \alpha y_{t-1}^{(1)}$ in (7.1.12) with $\alpha\beta'_{(K-r)}y_{t-1}^{(2)} + u_t$ and rearranging terms gives

$$\widehat{\beta}'_{(K-r)} - \beta'_{(K-r)} = (\alpha'\Sigma_u^{-1}\alpha)^{-1}\alpha'\Sigma_u^{-1}\left(\sum_{t=1}^{T} u_t y_{t-1}^{(2)\prime}\right)\left(\sum_{t=1}^{T} y_{t-1}^{(2)}y_{t-1}^{(2)\prime}\right)^{-1}. \tag{7.1.14}$$

Thus, we have to consider the quantity

$$T\left(\sum_{t=1}^{T} u_t y_{t-1}^{(2)\prime}\right)\left(\sum_{t=1}^{T} y_{t-1}^{(2)}y_{t-1}^{(2)\prime}\right)^{-1}$$

$$= \left(T^{-1}\sum_{t=1}^{T} u_t y_{t-1}^{(2)\prime}\right)\left(T^{-2}\sum_{t=1}^{T} y_{t-1}^{(2)}y_{t-1}^{(2)\prime}\right)^{-1}.$$

For the first matrix on the right-hand side we have

$$T^{-1}\sum_{t=1}^{T} u_t y_{t-1}^{(2)\prime}$$

$$= \left(T^{-1} \sum_{t=1}^{T} u_t y_{t-1}' \right) \begin{bmatrix} 0 \\ I_{K-r} \end{bmatrix}$$

$$= Q^{-1} \left(T^{-1} \sum_{t=1}^{T} v_t z_{t-1}' \right) Q^{-1\prime} \begin{bmatrix} 0 \\ I_{K-r} \end{bmatrix}$$

$$= Q^{-1} \left[o_p(1) : T^{-1} \sum_{t=1}^{T} v_t z_{t-1}^{(2)\prime} \right] Q^{-1\prime} \begin{bmatrix} 0 \\ I_{K-r} \end{bmatrix}$$

$$\xrightarrow{d} Q^{-1} \Sigma_v^{1/2} \left(\int_0^1 \mathbf{W}_K d\mathbf{W}_K' \right)' \Sigma_v^{1/2} \begin{bmatrix} 0 \\ I_{K-r} \end{bmatrix} Q^{22\prime},$$

where Lemma 7.1(2) and (3) have been used for the last equality and the limiting result, respectively. Thus,

$$T^{-1} \sum_{t=1}^{T} y_{t-1}^{(2)} u_t' \Sigma_u^{-1} \alpha (\alpha' \Sigma_u^{-1} \alpha)^{-1} \xrightarrow{d} \int_0^1 \mathbf{W}_{K-r}^* d\mathbf{W}_r^{*\prime}. \tag{7.1.15}$$

The matrix

$$T^{-2} \sum_{t=1}^{T} y_{t-1}^{(2)} y_{t-1}^{(2)\prime}$$

$$= [0 : I_{K-r}] \left(T^{-2} \sum_{t=1}^{T} y_{t-1} y_{t-1}' \right) \begin{bmatrix} 0 \\ I_{K-r} \end{bmatrix}$$

$$= [0 : I_{K-r}] Q^{-1} \left(T^{-2} \sum_{t=1}^{T} z_{t-1} z_{t-1}' \right) Q^{-1\prime} \begin{bmatrix} 0 \\ I_{K-r} \end{bmatrix}$$

$$= [0 : I_{K-r}] Q^{-1} \begin{bmatrix} o_p(1) & o_p(1) \\ o_p(1) & T^{-2} \sum_{t=1}^{T} z_{t-1}^{(2)} z_{t-1}^{(2)\prime} \end{bmatrix} Q^{-1\prime} \begin{bmatrix} 0 \\ I_{K-r} \end{bmatrix}$$

$$= Q^{22} \left(T^{-2} \sum_{t=1}^{T} z_{t-1}^{(2)} z_{t-1}^{(2)\prime} \right) Q^{22\prime} + o_p(1)$$

$$\xrightarrow{d} Q^{22} [0 : I_{K-r}] \Sigma_v^{1/2} \left(\int_0^1 \mathbf{W}_K \mathbf{W}_K' ds \right) \Sigma_v^{1/2} \begin{bmatrix} 0 \\ I_{K-r} \end{bmatrix} Q^{22\prime}$$

$$= \int_0^1 \mathbf{W}_{K-r}^* \mathbf{W}_{K-r}^{*\prime} ds, \tag{7.1.16}$$

where Lemma 7.1(5) has been applied. Using (7.1.14) and combining (7.1.15) and (7.1.16), gives the result in (7.1.13). ■

Clearly, in the present model setup, the GLS estimator of $\beta'_{(K-r)}$ does not have the usual normal limiting distribution. In fact, it converges with rate T rather than the usual rate $\sqrt{T}$, at least under our present rather restrictive assumptions. The asymptotic distribution consists of functionals of a standard Wiener process. It is also interesting to note that the two Wiener processes $\mathbf{W}_r^*$ and $\mathbf{W}_{K-r}^*$ are independent because their cross-covariance matrix is

$$\begin{aligned} & Q^{22}[0 : I_{K-r}]\Sigma_v Q^{-1\prime}\Sigma_u^{-1}\alpha(\alpha'\Sigma_u^{-1}\alpha)^{-1} \\ &\quad = Q^{22}[0 : I_{K-r}]Q\alpha(\alpha'\Sigma_u^{-1}\alpha)^{-1} \\ &\quad = Q^{22}\alpha'_\perp\alpha(\alpha'\Sigma_u^{-1}\alpha)^{-1} \\ &\quad = 0, \end{aligned}$$

where $\Sigma_v = Q\Sigma_u Q'$ has been used to obtain the first equality. The independence of the two Wiener processes implies that the conditional distribution of

$$\mathrm{vec}\left(\int_0^1 \mathbf{W}_{K-r}^* d\mathbf{W}_r^{*\prime}\right)'$$

given $\mathbf{W}_{K-r}^*$ is

$$\mathcal{N}\left(0, \int_0^1 \mathbf{W}_{K-r}^*\mathbf{W}_{K-r}^{*\prime} ds \otimes (\alpha'\Sigma_u^{-1}\alpha)^{-1}\right)$$

(see Ahn & Reinsel (1990), Phillips & Park (1988) or Johansen (1995)). This reasoning leads to the following interesting result.

Result 5

$$\begin{aligned} & \mathrm{vec}\left[(\widehat{\beta}'_{(K-r)} - \beta'_{(K-r)})\left(\sum_{t=1}^T y_{t-1}^{(2)} y_{t-1}^{(2)\prime}\right)^{1/2}\right] \\ &\qquad \xrightarrow{d} \mathcal{N}\left(0, I_{K-r} \otimes (\alpha'\Sigma_u^{-1}\alpha)^{-1}\right). \end{aligned} \tag{7.1.17}$$

■

Proof: From (7.1.16) we have

$$T^{-2}\sum_{t=1}^T y_{t-1}^{(2)} y_{t-1}^{(2)\prime} \xrightarrow{d} \int_0^1 \mathbf{W}_{K-r}^*\mathbf{W}_{K-r}^{*\prime} ds.$$

Hence, Result 5 follows because

$$\begin{aligned} & \mathrm{vec}\left[(\widehat{\beta}'_{(K-r)} - \beta'_{(K-r)})\left(\sum_{t=1}^T y_{t-1}^{(2)} y_{t-1}^{(2)\prime}\right)^{1/2}\right] \\ &\qquad = \left(\left(\sum_{t=1}^T y_{t-1}^{(2)} y_{t-1}^{(2)\prime}\right)^{1/2} \otimes I_K\right) \mathrm{vec}(\widehat{\beta}'_{(K-r)} - \beta'_{(K-r)}). \end{aligned}$$

■

Result 5 means that, although the GLS estimator $\widehat{\beta}'_{(K-r)}$ has a nonstandard limiting distribution, a transformation is asymptotically normal and can, for example, be used to construct hypothesis tests with standard limiting distributions. For example, t-ratios can be constructed in the usual way by considering an element of $\widehat{\beta}'_{(K-r)}$ and dividing by its asymptotic standard deviation obtained from

$$\left(\sum_{t=1}^{T} y_{t-1}^{(2)} y_{t-1}^{(2)\prime}\right)^{-1} \otimes (\alpha' \Sigma_u^{-1} \alpha)^{-1}.$$

Also Wald tests can be constructed as usual (see Appendix C.7).

Of course, the GLS estimator is only available under the very restrictive assumption that both α and Σ_u are known. It turns out, however, that the same asymptotic distribution is obtained for the corresponding EGLS estimator,

$$\begin{aligned}&\widehat{\widehat{\beta}}'_{(K-r)} = \\ &(\widehat{\alpha}' \widehat{\Sigma}_u^{-1} \widehat{\alpha})^{-1} \widehat{\alpha}' \widehat{\Sigma}_u^{-1} \left(\sum_{t=1}^{T} (\Delta y_t - \widehat{\alpha} y_{t-1}^{(1)}) y_{t-1}^{(2)\prime}\right) \left(\sum_{t=1}^{T} y_{t-1}^{(2)} y_{t-1}^{(2)\prime}\right)^{-1},\end{aligned} \tag{7.1.18}$$

where $\widehat{\alpha}$ and $\widehat{\Sigma}_u$ are consistent estimators of α and Σ_u, respectively. Fortunately, such estimators are available in the present case. A consistent estimator $\widehat{\alpha}$ follows from Result 2. If β is normalized as in (7.1.10), the first r columns of $\mathbf{\Pi}$ are equal to α. Hence, the first r columns of $\widehat{\mathbf{\Pi}}$ are a consistent estimator of α and the usual white noise covariance matrix estimator from the unrestricted LS estimation can be shown to be a consistent estimator of Σ_u, as we will demonstrate later (see Result 8). The following result can be established.

Result 6

$$T(\widehat{\widehat{\beta}}'_{(K-r)} - \widehat{\beta}'_{(K-r)}) = o_p(1). \tag{7.1.19}$$

■

Proof: Defining $u_t^* = \Delta y_t - \widehat{\alpha}\beta' y_{t-1}$ and substituting $\widehat{\alpha}\beta'_{(K-r)} y_{t-1}^{(2)} + u_t^*$ for $\Delta y_t - \widehat{\alpha} y_{t-1}^{(1)}$ in (7.1.18) gives, after rearrangement of terms,

$$\widehat{\widehat{\beta}}'_{(K-r)} - \beta'_{(K-r)} = (\widehat{\alpha}' \widehat{\Sigma}_u^{-1} \widehat{\alpha})^{-1} \widehat{\alpha}' \widehat{\Sigma}_u^{-1} \left(\sum_{t=1}^{T} u_t^* y_{t-1}^{(2)\prime}\right) \left(\sum_{t=1}^{T} y_{t-1}^{(2)} y_{t-1}^{(2)\prime}\right)^{-1}.$$

Hence,

$$T(\widehat{\widehat{\beta}}'_{(K-r)} - \widehat{\beta}'_{(K-r)})$$

$$= \left[(\widehat{\alpha}'\widehat{\Sigma}_u^{-1}\widehat{\alpha})^{-1}\widehat{\alpha}'\widehat{\Sigma}_u^{-1} - (\alpha'\Sigma_u^{-1}\alpha)^{-1}\alpha'\Sigma_u^{-1} \right]$$
$$\times \left(T^{-1}\sum_{t=1}^{T} u_t y_{t-1}^{(2)\prime} \right) \left(T^{-2}\sum_{t=1}^{T} y_{t-1}^{(2)} y_{t-1}^{(2)\prime} \right)^{-1}$$
$$+ (\widehat{\alpha}'\widehat{\Sigma}_u^{-1}\widehat{\alpha})^{-1}\widehat{\alpha}'\widehat{\Sigma}_u^{-1}$$
$$\times \left(T^{-1}\sum_{t=1}^{T} (u_t^* - u_t) y_{t-1}^{(2)\prime} \right) \left(T^{-2}\sum_{t=1}^{T} y_{t-1}^{(2)} y_{t-1}^{(2)\prime} \right)^{-1}.$$

The term in brackets is $o_p(1)$ because $\widehat{\alpha}$ and $\widehat{\Sigma}_u$ are consistent estimators by assumption. Moreover, $T^{-1}\sum_{t=1}^{T}(u_t^* - u_t)y_{t-1}^{(2)\prime} = o_p(1)$ (see Problem 7.1). Thus, the desired result follows because all other terms converge as established previously. ■

If the process is assumed to be Gaussian, ML estimation may be used alternatively. In case α and Σ_u are known, the ML estimator is identical to the GLS estimator for $\beta'_{(K-r)}$ and, hence, $\widehat{\beta}'_{(K-r)}$ is also the ML estimator. If α and Σ_u are unknown, ML estimation under the constraint $\text{rk}(\mathbf{\Pi}) = r$ may be used. The log-likelihood function is

$$\ln l = -\frac{KT}{2}\ln 2\pi - \frac{T}{2}\ln|\Sigma_u| - \frac{1}{2}\sum_{t=1}^{T}(\Delta y_t - \mathbf{\Pi} y_{t-1})'\Sigma_u^{-1}(\Delta y_t - \mathbf{\Pi} y_{t-1}). \tag{7.1.20}$$

From Chapter 3, we know that maximizing this function is equivalent to minimizing the determinant

$$\left| T^{-1}\sum_{t=1}^{T}(\Delta y_t - \mathbf{\Pi} y_{t-1})(\Delta y_t - \mathbf{\Pi} y_{t-1})' \right|.$$

To impose the rank restriction $\text{rk}(\mathbf{\Pi}) = r$, we write $\mathbf{\Pi} = \alpha\beta'$, where α and β are $(K \times r)$ matrices with rank r. For the moment we do not impose any normalization restrictions and consider minimization of the determinant

$$\left| T^{-1}\sum_{t=1}^{T}(\Delta y_t - \alpha\beta' y_{t-1})(\Delta y_t - \alpha\beta' y_{t-1})' \right|$$

with respect to α and β. This minimization problem is solved in Proposition A.7 in Appendix A.14 and the solution is obtained by considering the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_K$ and the associated orthonormal eigenvectors $v_1, \dots, v_K$ of the matrix

$$\left(\sum_{t=1}^{T} y_{t-1}y_{t-1}'\right)^{-1/2} \left(\sum_{t=1}^{T} y_{t-1}\Delta y_t'\right) \left(\sum_{t=1}^{T} \Delta y_t \Delta y_t'\right)$$

$$\times \left(\sum_{t=1}^{T} \Delta y_t y'_{t-1} \right) \left(\sum_{t=1}^{T} y_{t-1} y'_{t-1} \right)^{-1/2}.$$

The minimum of the determinant is attained for

$$\widetilde{\beta} = [\nu_1, \dots, \nu_r]' \left(\sum_{t=1}^{T} y_{t-1} y'_{t-1} \right)^{-1/2} \tag{7.1.21}$$

and

$$\widetilde{\alpha} = \left(\sum_{t=1}^{T} \Delta y_t y'_{t-1} \widetilde{\beta} \right) \left(\sum_{t=1}^{T} \widetilde{\beta}' y_{t-1} y'_{t-1} \widetilde{\beta} \right)^{-1}. \tag{7.1.22}$$

Clearly, the resulting ML estimator $\widetilde{\mathbf{\Pi}} = \widetilde{\alpha}\widetilde{\beta}'$ for $\mathbf{\Pi}$ must have the same asymptotic properties as the unrestricted LS estimator of $\mathbf{\Pi}$ because even the estimator in Result 3, which is based on a known β does not have better properties. Notice that, for a Gaussian model, the LS estimator based on a known β is equal to the ML estimator because the same regressors appear in all equations. Thus, we can draw the following conclusion.

Result 7

$$\sqrt{T}\text{vec}(\widetilde{\alpha}\widetilde{\beta}' - \mathbf{\Pi}) \xrightarrow{d} \mathcal{N}(0, \beta(\Gamma_z^{(1)})^{-1}\beta' \otimes \Sigma_u). \tag{7.1.23}$$

■

This result was derived by Johansen (1995) and other authors for more general models. It is also interesting to note that we can, of course, normalize the ML estimator for β as in (7.1.10), that is, we postmultiply the estimator in (7.1.21) by the inverse of the upper $(r \times r)$ submatrix. Denoting the normalized estimator by $\breve{\beta}$ and using the corresponding estimator for α from (7.1.22),

$$\breve{\alpha} = \left(\sum_{t=1}^{T} \Delta y_t y'_{t-1} \breve{\beta} \right) \left(\sum_{t=1}^{T} \breve{\beta}' y_{t-1} y'_{t-1} \breve{\beta} \right)^{-1},$$

gives an estimator $\breve{\alpha}\breve{\beta}'$ of $\mathbf{\Pi}$ which is identical to $\widetilde{\alpha}\widetilde{\beta}'$. Thus, the asymptotic properties must also be identical. It follows that $\breve{\alpha}$ has the same asymptotic distribution as the LS estimator in (7.1.9). Moreover, the asymptotic distribution of the lower $((K-r) \times r)$ part of $\breve{\beta}$ is the same as that of the GLS estimator in Result 4 because

$$\breve{\beta}'_{(K-r)} = (\breve{\alpha}'\widetilde{\Sigma}_u^{-1}\breve{\alpha})^{-1}\breve{\alpha}'\widetilde{\Sigma}_u^{-1} \left(\sum_{t=1}^{T} (\Delta y_t - \breve{\alpha} y_{t-1}^{(1)}) y_{t-1}^{(2)\prime} \right) \left(\sum_{t=1}^{T} y_{t-1}^{(2)} y_{t-1}^{(2)\prime} \right)^{-1},$$

where the ML estimator $\widetilde{\Sigma}_u$ is substituted for Σ_u. Thus, the asymptotic distribution of $\breve{\beta}_{(K-r)}$ follows from Result 6 and the consistency of the ML estimators $\breve{\alpha}$ and $\widetilde{\Sigma}_u$.

In fact, any of the estimators for $\mathbf{\Pi}$ which we have considered so far, leads to a consistent estimator of the white noise covariance matrix of the form

$$\widetilde{\Sigma}_u = T^{-1}\sum_{t=1}^{T}(\Delta y_t - \widehat{\mathbf{\Pi}}y_{t-1})(\Delta y_t - \widehat{\mathbf{\Pi}}y_{t-1})'. \tag{7.1.24}$$

Here $\widehat{\mathbf{\Pi}}$ can be any of the estimators for $\mathbf{\Pi}$ considered so far, because they are all asymptotically equivalent. The following result can be established.

Result 8

$$\text{plim } \widetilde{\Sigma}_u = \Sigma_u. \tag{7.1.25}$$

■

Proof: Notice that

$$\begin{aligned}
\widetilde{\Sigma}_u &= T^{-1}\sum_{t=1}^{T}(\mathbf{\Pi}y_{t-1} - \widehat{\mathbf{\Pi}}y_{t-1} + u_t)(\mathbf{\Pi}y_{t-1} - \widehat{\mathbf{\Pi}}y_{t-1} + u_t)' \\
&= T^{-1}\sum_{t=1}^{T}u_t u_t' + (\mathbf{\Pi} - \widehat{\mathbf{\Pi}})\left(T^{-1}\sum_{t=1}^{T}y_{t-1}y_{t-1}'\right)(\mathbf{\Pi} - \widehat{\mathbf{\Pi}})' \\
&\quad + \left(T^{-1}\sum_{t=1}^{T}u_t y_{t-1}'\right)(\mathbf{\Pi} - \widehat{\mathbf{\Pi}})' \\
&\quad + (\mathbf{\Pi} - \widehat{\mathbf{\Pi}})\left(T^{-1}\sum_{t=1}^{T}y_{t-1}u_t'\right).
\end{aligned} \tag{7.1.26}$$

Using a standard law of large numbers,

$$\text{plim } T^{-1}\sum_{t=1}^{T}u_t u_t' = \Sigma_u.$$

Thus, it suffices to show that all other terms are $o_p(1)$. This property follows because from Lemma 7.1 we have

$$T^{-1}\sum_{t=1}^{T}y_{t-1}u_t' = O_p(1)$$

and

$$T^{-1}\sum_{t=1}^{T}\beta' y_{t-1}y_{t-1}'\beta = O_p(1).$$

Using the estimator $\widehat{\alpha}\beta'$ for $\mathbf{\Pi}$, it is easily seen that all terms but the first on the right-hand side of the last equality sign in (7.1.26) converge to zero in probability. The argument is easily extended to the other estimators by noting that their difference to the previously treated estimator is $o_p(T^{-1/2})$. ■

So far we have assumed that $r \neq 0$ and, hence, $\mathbf{\Pi} \neq 0$. This assumption is of obvious importance for some of the results to hold and some of the proofs to work. If $\mathbf{\Pi} = 0$, the analysis becomes even simpler in some respects. In that case, y_t is a multivariate random walk and we can apply Proposition C.18 directly to evaluate the asymptotic properties of the term

$$T(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}) = \left(T^{-1}\sum_{t=1}^{T} u_t y_{t-1}'\right)\left(T^{-2}\sum_{t=1}^{T} y_{t-1}y_{t-1}'\right)^{-1},$$

where $\widehat{\mathbf{\Pi}}$ is again the LS estimator. Using Proposition C.18(6) and (9) gives the following result.

Result 9

If the cointegrating rank $r = 0$,

$$T(\widehat{\mathbf{\Pi}}-\mathbf{\Pi}) \xrightarrow{d} \Sigma_u^{1/2}\left(\int_0^1 \mathbf{W}_K d\mathbf{W}_K'\right)'\left(\int_0^1 \mathbf{W}_K\mathbf{W}_K' ds\right)^{-1}\Sigma_u^{-1/2}. \quad (7.1.27)$$

■

The LS estimator is again identical to the ML estimator and, hence, the same result is obtained for the latter. On the other hand, the GLS estimator is not applicable here. Now we cannot even use the usual t-ratios anymore in a standard way because they do not have a limiting standard normal distribution in this case. For the special case of a univariate model this can be seen from Appendix C.8.1. Notice that for $K = 1$, $\widehat{\mathbf{\Pi}} = \widehat{\rho} - 1$ in Proposition C.17 and, thus, the asymptotic distribution of $T\widehat{\mathbf{\Pi}} = T(\widehat{\rho} - 1)$ is clearly different from the standard normal in this case.

The results for the estimator of the VECM imply analogous results for the parameters of the corresponding levels VAR form $y_t = A_1 y_{t-1} + u_t$. Notice that $A_1 = \mathbf{\Pi} + I_K$. Consequently, we have for the LS estimator, for example,

$$\widehat{A}_1 - A_1 = \widehat{\mathbf{\Pi}} - \mathbf{\Pi}. \quad (7.1.28)$$

Hence, the asymptotic properties of $\widehat{A}_1$ follow immediately from those of $\widehat{\mathbf{\Pi}}$.

The simple model we have discussed in this section shows the main differences to the stationary case. All the results can be extended to richer models with short-term dynamics and deterministic terms. Estimation of such models will be considered in the next section.

7.2 Estimation of General VECMs

We first consider a model without deterministic terms,

$$\Delta y_t = \mathbf{\Pi} y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t, \qquad (7.2.1)$$

where y_t is a process of dimension K, $\text{rk}(\mathbf{\Pi}) = r$ with $0 < r < K$ so that $\mathbf{\Pi} = \alpha\beta'$, where α and β are $(K \times r)$ matrices with $\text{rk}(\alpha) = \text{rk}(\beta) = r$. All other symbols have their conventional meanings, that is, the $\mathbf{\Gamma}_j$ $(j = 1, \ldots, p-1)$ are $(K \times K)$ parameter matrices and $u_t \sim (0, \Sigma_u)$ is standard white noise. Also, y_t is assumed to be an $I(1)$ process so that

$$\alpha'_\perp \left(I_K - \sum_{i=1}^{p-1} \mathbf{\Gamma}_i \right) \beta_\perp \qquad (7.2.2)$$

is nonsingular (see Section 6.3, Eq. (6.3.12)). These conditions are always assumed to hold without further notice when the VECM (7.2.1) is considered in this chapter.

For estimation purposes, we assume that a sample $y_1, \ldots, y_T$ and the needed presample values are available. It is then often convenient to write the VECM (7.2.1), for $t = 1, \ldots, T$, in matrix notation as

$$\Delta Y = \mathbf{\Pi} Y_{-1} + \mathbf{\Gamma} \Delta X + U, \qquad (7.2.3)$$

where

$$\Delta Y := [\Delta y_1, \ldots, \Delta y_T],$$

$$Y_{-1} := [y_0, \ldots, y_{T-1}],$$

$$\mathbf{\Gamma} := [\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_{p-1}],$$

$$\Delta X := [\Delta X_0, \ldots, \Delta X_{T-1}] \quad \text{with} \quad \Delta X_{t-1} := \begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \end{bmatrix}$$

and

$$U := [u_1, \ldots, u_T].$$

We will now consider LS, EGLS, and ML estimation of the parameters of this model. Estimation of the parameters of the corresponding levels VAR form will also be discussed and, moreover, we comment on the implications of including deterministic terms.

7.2.1 LS Estimation

From the matrix version (7.2.3) of our VECM, the LS estimator is seen to be

$$[\widehat{\mathbf{\Pi}} : \widehat{\mathbf{\Gamma}}] = [\Delta Y Y_{-1}' : \Delta Y \Delta X'] \begin{bmatrix} Y_{-1}Y_{-1}' & Y_{-1}\Delta X' \\ \Delta X Y_{-1}' & \Delta X \Delta X' \end{bmatrix}^{-1}, \tag{7.2.4}$$

using the usual formulas from Chapter 3. The corresponding white noise covariance matrix estimator is

$$\widehat{\Sigma}_u := (T - Kp)^{-1}(\Delta Y - \widehat{\mathbf{\Pi}}Y_{-1} - \widehat{\mathbf{\Gamma}}\Delta X)(\Delta Y - \widehat{\mathbf{\Pi}}Y_{-1} - \widehat{\mathbf{\Gamma}}\Delta X)'. \tag{7.2.5}$$

The asymptotic properties of these estimators are given in the next proposition.

Proposition 7.1 (*Asymptotic Properties of the LS Estimator for a VECM*)
Consider the VECM (7.2.1). The LS estimator given in (7.2.4) is consistent and

$$\sqrt{T}\,\mathrm{vec}([\widehat{\mathbf{\Pi}} : \widehat{\mathbf{\Gamma}}] - [\mathbf{\Pi} : \mathbf{\Gamma}]) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\mathrm{co}}), \tag{7.2.6}$$

where

$$\Sigma_{\mathrm{co}} = \left(\begin{bmatrix} \beta & 0 \\ 0 & I_{Kp-K} \end{bmatrix} \Omega^{-1} \begin{bmatrix} \beta' & 0 \\ 0 & I_{Kp-K} \end{bmatrix} \right) \otimes \Sigma_u$$

and

$$\Omega = \mathrm{plim}\,\frac{1}{T} \begin{bmatrix} \beta' Y_{-1}Y_{-1}'\beta & \beta' Y_{-1}\Delta X' \\ \Delta X Y_{-1}'\beta & \Delta X \Delta X' \end{bmatrix}.$$

The matrix

$$\begin{bmatrix} \beta & 0 \\ 0 & I_{Kp-K} \end{bmatrix} \Omega^{-1} \begin{bmatrix} \beta' & 0 \\ 0 & I_{Kp-K} \end{bmatrix}$$

is consistently estimated by

$$T \begin{bmatrix} Y_{-1}Y_{-1}' & Y_{-1}\Delta X' \\ \Delta X Y_{-1}' & \Delta X \Delta X' \end{bmatrix}^{-1}$$

and $\widehat{\Sigma}_u$ is a consistent estimator for Σ_u. ■

This proposition generalizes Result 2 of Section 7.1. Therefore similar remarks can be made.

Remark 1 The covariance matrix Σ_{co} is singular. This property is easily seen by noting that Ω is a $[(Kp - K + r) \times (Kp - K + r)]$ matrix. Thus, the rank of the $(K^2p \times K^2p)$ matrix Σ_{co} cannot be greater than $K(Kp - K + r)$ which is smaller than K^2p under our assumption that $r < K$. Still, t-ratios can be set up and interpreted in the usual way because they have standard normal limiting distributions under our assumptions. In contrast, Wald tests and the corresponding F-tests of linear restrictions on the parameters may not have the usual asymptotic χ^2- or approximate F-distributions that are obtained for stationary processes. A more detailed discussion of this issue will be given in Section 7.6. ∎

Remark 2 If β is known, the LS estimator

$$[\hat{\alpha} : \hat{\mathbf{\Gamma}}] = [\Delta Y Y_{-1}'\beta : \Delta Y \Delta X'] \begin{bmatrix} \beta' Y_{-1} Y_{-1}' \beta & \beta' Y_{-1} \Delta X' \\ \Delta X Y_{-1}' \beta & \Delta X \Delta X' \end{bmatrix}^{-1} \tag{7.2.7}$$

of $[\alpha : \mathbf{\Gamma}]$ may be considered. Using standard arguments for stationary processes, its asymptotic distribution is seen to be

$$\sqrt{T}\,\mathrm{vec}([\hat{\alpha} : \hat{\mathbf{\Gamma}}] - [\alpha : \mathbf{\Gamma}]) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\alpha,\mathbf{\Gamma}}), \tag{7.2.8}$$

where

$$\Sigma_{\alpha,\mathbf{\Gamma}} = \Omega^{-1} \otimes \Sigma_u = \mathrm{plim}\, T \begin{bmatrix} \beta' Y_{-1} Y_{-1}' \beta & \beta' Y_{-1} \Delta X' \\ \Delta X Y_{-1}' \beta & \Delta X \Delta X' \end{bmatrix}^{-1} \otimes \Sigma_u.$$

The asymptotic distribution in (7.2.8) is nonsingular so that, for given β, asymptotic inference for α and $\mathbf{\Gamma}$ is standard. Noting that

$$[\hat{\alpha}\beta' : \hat{\mathbf{\Gamma}}] - [\mathbf{\Pi} : \mathbf{\Gamma}] = ([\hat{\alpha} : \hat{\mathbf{\Gamma}}] - [\alpha : \mathbf{\Gamma}]) \begin{bmatrix} \beta' & 0 \\ 0 & I_{Kp-K} \end{bmatrix},$$

it is easy to see that

$$\mathrm{vec}([\hat{\alpha}\beta' : \hat{\mathbf{\Gamma}}] - [\mathbf{\Pi} : \mathbf{\Gamma}])$$

has the same asymptotic distribution as the LS estimator in Proposition 7.1. This finding corresponds to Result 3 in Section 7.1. It means that, whether the cointegrating matrix β is known or estimated is of no consequence for the asymptotic distribution of the LS estimators of $\mathbf{\Pi}$ and $\mathbf{\Gamma}$. The reason is that β is estimated "superconsistently" even if LS estimation is used. This point will be discussed further in Section 7.2.2. ∎

Remark 3 If the cointegrating rank $r = 0$ and, thus, $\mathbf{\Pi} = 0$,

$$\sqrt{T}[\hat{\mathbf{\Pi}} - \mathbf{\Pi}] = o_p(1),$$

that is, the LS estimator of $\mathbf{\Pi}$ converges faster than with the usual rate $\sqrt{T}$. Therefore, Proposition 7.1 remains valid in the sense that all parts of the asymptotic covariance matrix in (7.2.6) related to $\mathbf{\Pi}$ have to be set to zero. In other words, the first K^2 rows and columns of Σ_{co} are zero. ∎

Remark 4 From Proposition 7.1 it is also easy to derive the asymptotic distribution of the LS estimator for the parameters of the levels VAR form corresponding to our VECM,

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t. \tag{7.2.9}$$

The A_i's are related to the VECM parameters by

$$\begin{aligned} A_1 &= \mathbf{\Pi} + I_K + \mathbf{\Gamma}_1 \\ A_i &= \mathbf{\Gamma}_i - \mathbf{\Gamma}_{i-1}, \quad i = 2, \dots, p-1, \\ A_p &= -\mathbf{\Gamma}_{p-1} \end{aligned} \tag{7.2.10}$$

(see also (6.3.7)). Hence, they are obtained by a linear transformation,

$$A := [A_1 : \cdots : A_p] = [\mathbf{\Pi} : \mathbf{\Gamma}]W + J, \tag{7.2.11}$$

where

$$J := [I_K : 0 : \cdots : 0] \quad (K \times Kp)$$

and

$$W := \begin{bmatrix} I_K & 0 & 0 & \dots & 0 & 0 \\ I_K & -I_K & 0 & \dots & 0 & 0 \\ 0 & I_K & -I_K & & 0 & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & I_K & -I_K \end{bmatrix} \quad (Kp \times Kp).$$

Consequently, using

$$\operatorname{vec}([\mathbf{\Pi} : \mathbf{\Gamma}]W) = (W' \otimes I_K) \operatorname{vec}[\mathbf{\Pi} : \mathbf{\Gamma}],$$

we get the following implication of Proposition 7.1 (see also Sims, Stock & Watson (1990)).

Corollary 7.1.1
Under the conditions of Proposition 7.1,

$$\sqrt{T} \operatorname{vec}(\widehat{A} - A) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\boldsymbol{\alpha}}^{co}),$$

where $\widehat{A}$ is the LS estimator of A and

$$\begin{aligned} \Sigma_{\boldsymbol{\alpha}}^{co} &:= \left(W' \begin{bmatrix} \beta & 0 \\ 0 & I_{Kp-K} \end{bmatrix} \Omega^{-1} \begin{bmatrix} \beta' & 0 \\ 0 & I_{Kp-K} \end{bmatrix} W \right) \otimes \Sigma_u \\ &= (W' \otimes I_K) \Sigma_{co} (W \otimes I_K). \end{aligned}$$

Furthermore,

$$\widehat{\Sigma}_{\boldsymbol{\alpha}}^{co} = (XX')^{-1} \otimes [(Y - \widehat{A}X)(Y - \widehat{A}X)']$$

is a consistent estimator of $\Sigma_{\boldsymbol{\alpha}}^{co}$. Here $Y := [y_1, \ldots, y_T]$ and

$$X := [Y_0, \ldots, Y_{T-1}] \quad \text{with} \quad Y_{t-1} := \begin{bmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{bmatrix}.$$

■

Because $\Sigma_{\boldsymbol{\alpha}}^{co}$ is singular, $\widehat{A}$ also has a singular asymptotic distribution. The distribution in Corollary 7.1.1 remains, in fact, valid if $r = 0$. ■

Discussion of the Proof of Proposition 7.1

The proof of Proposition 7.1 is a generalization of that of Result 2 in Section 7.1. Multiplying

$$\begin{bmatrix} y_t \\ \Delta X_t \end{bmatrix}$$

by

$$Q^* := \begin{bmatrix} \beta' & 0 \\ 0 & I_{K(p-1)} \\ \alpha'_\perp & 0 \end{bmatrix}$$

gives a process

$$z_t = \begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \end{bmatrix} := Q^* \begin{bmatrix} y_t \\ \Delta X_t \end{bmatrix}, \tag{7.2.12}$$

where

$$z_t^{(1)} := \begin{bmatrix} \beta' y_t \\ \Delta X_t \end{bmatrix}$$

contains $I(0)$ components only and $z_t^{(2)} := \alpha'_\perp y_t$ consists of $I(1)$ components (see Proposition 6.1). Therefore, a lemma analogous to Lemma 7.1 can be established and used to prove Proposition 7.1. We leave the details as an exercise (see Problem 7.2).

In fact, via the process z_t, we can get the following useful lemma from standard weak laws of large numbers and central limit theorems for stationary processes (see Appendix C.4) as well as Proposition C.18 of Appendix C. It summarizes a number of convergence results for variables generated by the VECM (7.2.1). Some of these or similar results were derived by different authors including Phillips & Durlauf (1986), Johansen (1988), Ahn & Reinsel (1990) and Park & Phillips (1989).

Lemma 7.2

(1) $\Delta X \Delta X' = O_p(T)$ and $(T^{-1}\Delta X \Delta X')^{-1} = O_p(1)$.
(2) $\beta' Y_{-1} \Delta X' = O_p(T)$.
(3) $\beta' Y_{-1} Y_{-1}' \beta = O_p(T)$ and $(T^{-1}\beta' Y_{-1} Y_{-1}' \beta)^{-1} = O_p(1)$.
(4) $\beta' Y_{-1} U' = O_p(T^{1/2})$.
(5) $\beta' Y_{-1} \Delta Y' = O_p(T^{1/2})$.
(6) $Y_{-1} U' = O_p(T)$.
(7) $Y_{-1} \Delta X' = O_p(T)$.
(8) $\beta' Y_{-1} Y_{-1}' = O_p(T)$.
(9) $Y_{-1} Y_{-1}' = O_p(T^2)$.

■

Some of these results are helpful in deriving Proposition 7.1 and they are also useful in proving the next propositions. Because ΔY, $\beta' Y_{-1}$, and ΔX contain $I(0)$ variables only, essentially the same results as in the stable case hold for these quantities. This is reflected in Lemma 7.2(1)–(5). On the other hand, Y_{-1} contains $I(1)$ variables that behave differently from $I(0)$ variables. For instance, for a stable process, $Y_{-1}Y_{-1}'/T$ has a fixed probability limit (see Chapter 3). Now the corresponding quantity $Y_{-1}Y_{-1}'$ is $O_p(T^2)$. Intuitively, the reason is that integrated variables do not fluctuate around a constant mean but are trending. Thus, the sums of products and cross-products go to infinity (or minus infinity) more rapidly than for stable processes.

7.2.2 EGLS Estimation of the Cointegration Parameters

For GLS estimation we assume that β is normalized as in (7.1.10),

$$\beta = \begin{bmatrix} I_r \\ \beta_{(K-r)} \end{bmatrix}.$$

Because we are primarily interested in estimating $\beta_{(K-r)}$, we concentrate on the error correction term and replace the short-run parameters $\mathbf{\Gamma}$ by their LS estimators for a given matrix $\mathbf{\Pi}$,

$$\widehat{\mathbf{\Gamma}}(\mathbf{\Pi}) = (\Delta Y - \mathbf{\Pi} Y_{-1})\Delta X'(\Delta X \Delta X')^{-1}.$$

Hence,

$$\Delta Y = \mathbf{\Pi} Y_{-1} + (\Delta Y - \mathbf{\Pi} Y_{-1})\Delta X'(\Delta X \Delta X')^{-1}\Delta X + U^*.$$

Rearranging terms and defining the $(T \times T)$ matrix

$$M := I_T - \Delta X'(\Delta X \Delta X')^{-1}\Delta X,$$

gives

$$R_0 = \mathbf{\Pi} R_1 + U^* = \alpha\beta' R_1 + U^*, \tag{7.2.13}$$

where

$$R_0 := \Delta Y M \quad \text{and} \quad R_1 := Y_{-1} M.$$

Notice that R_0 is just the residual matrix from a (multivariate) regression of Δy_t on ΔX_{t-1} and R_1 is the matrix of residuals from a regression of y_{t-1} on ΔX_{t-1}. Denoting the first r and last $K-r$ rows of R_1 by $R_1^{(1)}$ and $R_1^{(2)}$, respectively, and using the normalization of β, (7.2.13) can be rewritten as

$$R_0 - \alpha R_1^{(1)} = \alpha\beta'_{(K-r)} R_1^{(2)} + U^*. \tag{7.2.14}$$

Based on this "concentrated model" the GLS estimator of $\beta'_{(K-r)}$ is

$$\widehat{\beta}'_{(K-r)} = (\alpha' \Sigma_u^{-1} \alpha)^{-1} \alpha' \Sigma_u^{-1} (R_0 - \alpha R_1^{(1)}) R_1^{(2)\prime} \left(R_1^{(2)} R_1^{(2)\prime} \right)^{-1} \tag{7.2.15}$$

(see Eq. (7.1.12)). Note that the same estimator is obtained if the short-run parameters are not concentrated out first because $\mathbf{\Gamma}$ has been replaced by the optimal matrix for any given matrix $\mathbf{\Pi}$. As in the simple special case model considered in Section 7.1, it is now obvious how to obtain a feasible GLS estimator. In a first estimation round we determine the LS estimator of $[\mathbf{\Pi} : \mathbf{\Gamma}]$ as in (7.2.4) and Σ_u as in (7.2.5). Using the first r columns of $\widehat{\mathbf{\Pi}}$ as an estimator $\widehat{\alpha}$, we get the EGLS estimator

$$\widehat{\widehat{\beta}}'_{(K-r)} = (\widehat{\alpha}' \widehat{\Sigma}_u^{-1} \widehat{\alpha})^{-1} \widehat{\alpha}' \widehat{\Sigma}_u^{-1} (R_0 - \widehat{\alpha} R_1^{(1)}) R_1^{(2)\prime} \left(R_1^{(2)} R_1^{(2)\prime} \right)^{-1}. \tag{7.2.16}$$

This estimator was proposed by Ahn & Reinsel (1990) and Saikkonen (1992) (see also Reinsel (1993, p. 171)). Its asymptotic properties are analogous to those of the EGLS estimator for the simple model considered in Section 7.1. They are summarized in the following proposition which was proven by Ahn & Reinsel (1990).

Proposition 7.2 (*Asymptotic Properties of the EGLS Estimator for the Cointegration Matrix*)
Consider the VECM (7.2.1) with cointegration matrix β normalized as in (7.1.10). Suppose $\widehat{\alpha}$ and $\widehat{\Sigma}_u$ are consistent estimators of α and Σ_u, respectively. Then the EGLS estimator of $\beta'_{(K-r)}$ given in (7.2.16) has the following asymptotic distribution:

$$T(\widehat{\widehat{\beta}}'_{(K-r)} - \beta'_{(K-r)}) \xrightarrow{d} \left(\int_0^1 \mathbf{W}^{\#}_{K-r} d\mathbf{W}^{\#\prime}_r \right)' \left(\int_0^1 \mathbf{W}^{\#}_{K-r} \mathbf{W}^{\#\prime}_{K-r} ds \right)^{-1}, \tag{7.2.17}$$

where $\mathbf{W}^{\#}_{K-r}$ and $\mathbf{W}^{\#}_r$ are suitable independent $(K-r)$- and r-dimensional Wiener processes, respectively, whose parameters depend on those of the VECM. Furthermore,

$$\operatorname{vec}\left[(\hat{\hat{\beta}}'_{(K-r)} - \beta'_{(K-r)})\left(R_1^{(2)}R_1^{(2)\prime}\right)^{1/2}\right] \xrightarrow{d} \mathcal{N}\left(0, I_{K-r} \otimes (\alpha' \Sigma_u^{-1}\alpha)^{-1}\right). \tag{7.2.18}$$

■

Remark 1 The EGLS estimator has the same asymptotic distribution as the GLS estimator. Moreover, it has the same asymptotic distribution one would obtain if all parameters (α, $\mathbf{\Gamma}$, and Σ_u) except $\beta_{(K-r)}$ were known. It converges at rate T. Hence, $\hat{\hat{\beta}}_{(K-r)}$ is a superconsistent estimator of $\beta_{(K-r)}$ and, thus,

$$\hat{\hat{\beta}} = \begin{bmatrix} I_r \\ \hat{\hat{\beta}}_{(K-r)} \end{bmatrix}$$

is a superconsistent estimator of β. The precise form of the Wiener processes $\mathbf{W}^{\#}_{K-r}$ and $\mathbf{W}^{\#}_r$ depends on the short-run dynamics of the process y_t. It is given, for example, in Ahn & Reinsel (1990). ■

Remark 2 The matrix

$$\begin{aligned} T^{-2}R_1R_1' &= T^{-2}Y_{-1}MY_{-1}' \\ &= T^{-2}Y_{-1}Y_{-1}' - T^{-2}Y_{-1}\Delta X'(T^{-1}\Delta X\Delta X')^{-1}T^{-1}\Delta XY_{-1}' \\ &= T^{-2}Y_{-1}Y_{-1}' + o_p(1)O_p(1)O_p(1) \\ &= T^{-2}Y_{-1}Y_{-1}' + o_p(1), \end{aligned}$$

where Lemma 7.2(1) and (7) have been used. This result implies that (7.2.18) could be stated alternatively as

$$\operatorname{vec}\left[(\hat{\hat{\beta}}'_{(K-r)} - \beta'_{(K-r)})\left(Y_{-1}^{(2)}Y_{-1}^{(2)\prime}\right)^{1/2}\right] \xrightarrow{d} \mathcal{N}\left(0, I_K \otimes (\alpha' \Sigma_u^{-1}\alpha)^{-1}\right),$$

where $Y_{-1}^{(2)}$ contains the last $K-r$ rows of Y_{-1}. For practical purposes, the result as stated in (7.2.18) is more useful because it can be used directly for setting up meaningful t-ratios and Wald or F-tests for hypotheses about the coefficients of $\beta_{(K-r)}$. These quantities have the usual asymptotic or approximate distributions. Of course, the same is true if $(R_1^{(2)}R_1^{(2)\prime})^{1/2}$ is replaced by $(Y_{-1}^{(2)}Y_{-1}^{(2)\prime})^{1/2}$. Still, in small samples it is advantageous to take the short-run dynamics into account as in $(R_1^{(2)}R_1^{(2)\prime})^{1/2}$. ■

Remark 3 It is also possible to replace β in $\mathbf{\Pi} = \alpha\beta'$ in (7.2.3) by the EGLS estimator and estimate the other parameters by LS from the model

$$\Delta Y = \alpha\hat{\hat{\beta}}'Y_{-1} + \mathbf{\Gamma}\Delta X + \hat{\hat{U}}^*.$$

The resulting estimator $[\widehat{\widehat{\alpha}} : \widehat{\widehat{\Gamma}}]$ has the same asymptotic properties as $[\widehat{\alpha} : \widehat{\Gamma}]$ in (7.2.7) which is based on a known β. As a consequence, $[\widehat{\widehat{\alpha}}\widehat{\widehat{\beta}}' : \widehat{\widehat{\Gamma}}]$ also has the same asymptotic properties as $[\widehat{\alpha}\beta' : \widehat{\Gamma}]$. ■

Remark 4 The EGLS estimator was actually presented in a slightly different form by Ahn & Reinsel (1990) and Saikkonen (1992). These authors use the representation

$$\widehat{\widehat{\beta}}'_{(K-r)} = (\widehat{\alpha}'\widehat{\Sigma}_u^{-1}\widehat{\alpha})^{-1}\widehat{\alpha}'\widehat{\Sigma}_u^{-1}\widehat{\Pi}_2,$$

where $\widehat{\Pi}_2$ is the $(K \times (K-r))$ matrix of the last $K-r$ columns of the LS estimator $\widehat{\Pi}$ of Π (see Reinsel (1993, p. 171) for a discussion of the equivalence of this estimator and the EGLS estimator (7.2.16)). ■

7.2.3 ML Estimation

If the process y_t is Gaussian or, equivalently, $u_t \sim \mathcal{N}(0, \Sigma_u)$, the VECM (7.2.1) can be estimated by maximum likelihood (ML) taking also the rank restriction for $\Pi = \alpha\beta'$ into account (see Johansen (1988, 1995)). The log-likelihood function for a sample of size T is

$$\begin{aligned} \ln l &= -\frac{KT}{2}\ln 2\pi - \frac{T}{2}\ln|\Sigma_u| \\ &\quad -\frac{1}{2}\mathrm{tr}\left[(\Delta Y - \alpha\beta' Y_{-1} - \Gamma\Delta X)'\Sigma_u^{-1}(\Delta Y - \alpha\beta' Y_{-1} - \Gamma\Delta X)\right]. \end{aligned} \tag{7.2.19}$$

In the following, we will first discuss the computation of the estimators and then consider their asymptotic properties.

The Estimator

For ML estimation we do not assume that β is normalized. We only make the assumption $\mathrm{rk}(\Pi) = r$ which implies that the matrix can be represented as $\Pi = \alpha\beta'$, where α and β are $(K \times r)$ with $\mathrm{rk}(\alpha) = \mathrm{rk}(\beta) = r$. In the next proposition the ML estimators are given. The proposition generalizes the special case estimators given in (7.1.21) and (7.1.22).

Proposition 7.3 (*ML Estimators of a VECM*)
Let $M := I_T - \Delta X'(\Delta X \Delta X')^{-1}\Delta X$, $R_0 := \Delta Y M$ and $R_1 := Y_{-1}M$, as before, and define

$$S_{ij} := R_i R_j'/T, \quad i = 0, 1,$$

$$\lambda_1 \geq \cdots \geq \lambda_K \text{ are the eigenvalues of } S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2},$$

and

$\nu_1, \dots, \nu_K$ are the corresponding orthonormal eigenvectors.

The log-likelihood function in (7.2.19) is maximized for

$$\begin{aligned}
\beta &= \widetilde{\beta} := [\nu_1, \dots, \nu_r]' S_{11}^{-1/2}, \\
\alpha &= \widetilde{\alpha} := \Delta Y M Y_{-p}' \widetilde{\beta} \left(\widetilde{\beta}' Y_{-1} M Y_{-1}' \widetilde{\beta} \right)^{-1} = S_{01} \widetilde{\beta} (\widetilde{\beta}' S_{11} \widetilde{\beta})^{-1}, \\
\mathbf{\Gamma} &= \widetilde{\mathbf{\Gamma}} := (\Delta Y - \widetilde{\alpha} \widetilde{\beta}' Y_{-1}) \Delta X' (\Delta X \Delta X')^{-1}, \\
\Sigma_u &= \widetilde{\Sigma}_u := (\Delta Y - \widetilde{\alpha} \widetilde{\beta}' Y_{-1} - \widetilde{\mathbf{\Gamma}} \Delta X)(\Delta Y - \widetilde{\alpha} \widetilde{\beta}' Y_{-1} - \widetilde{\mathbf{\Gamma}} \Delta X)'/T.
\end{aligned}$$

The maximum is

$$\max \ln l = -\frac{KT}{2} \ln 2\pi - \frac{T}{2} \left[\ln |S_{00}| + \sum_{i=1}^{r} \ln(1 - \lambda_i) \right] - \frac{KT}{2}. \qquad (7.2.20)$$

■

Proof: From Chapter 3, Section 3.4, it is known that for any fixed α and β the maximum of $\ln l$ is attained for

$$\widetilde{\mathbf{\Gamma}}(\alpha\beta') = (\Delta Y - \alpha\beta' Y_{-1}) \Delta X' (\Delta X \Delta X')^{-1}.$$

Thus, we replace $\mathbf{\Gamma}$ in (7.2.19) by $\widetilde{\mathbf{\Gamma}}(\alpha\beta')$ and get the concentrated log-likelihood

$$\begin{aligned}
&-\frac{KT}{2} \ln 2\pi - \frac{T}{2} \ln |\Sigma_u| \\
&\quad -\frac{1}{2} \operatorname{tr} \left[(\Delta Y M - \alpha\beta' Y_{-1} M)' \Sigma_u^{-1} (\Delta Y M - \alpha\beta' Y_{-1} M) \right].
\end{aligned}$$

Hence, we just have to maximize this expression with respect to α, β, and Σ_u. We also know from Chapter 3 that, for given α and β, the maximum is attained if

$$\widetilde{\Sigma}(\alpha\beta') = (\Delta Y M - \alpha\beta' Y_{-1} M)(\Delta Y M - \alpha\beta' Y_{-1} M)'/T$$

is substituted for Σ_u. Consequently, we have to maximize

$$-\frac{T}{2} \ln |(\Delta Y M - \alpha\beta' Y_{-1} M)(\Delta Y M - \alpha\beta' Y_{-1} M)'/T|$$

or, equivalently, minimize the determinant with respect to α and β. Thus, all results of Proposition 7.3 follow from Proposition A.7 of Appendix A.14. ■

The solutions $\widetilde{\beta}$ and $\widetilde{\alpha}$ of the optimization problem given in the proposition are not unique because, for any nonsingular $(r \times r)$ matrix Q, $\widetilde{\alpha} Q^{-1}$

and $\widetilde{\beta}Q'$ represent another set of ML estimators for α and β. However, the proposition shows that explicit expressions for ML estimators are available. If $r = K$, the proposition still remains valid. Also, ML estimators for the levels VAR representation corresponding to the VECM (7.2.1) observing the rank restriction are readily available via the relations in (7.2.10).

The next question concerns the properties of the ML estimators of a cointegrated system. They are discussed in the following.

Asymptotic Properties of the ML Estimator

The following proposition generalizes Result 7 of Section 7.1.

Proposition 7.4 (*Asymptotic Properties of the ML Estimators of a VECM*)
The ML estimators for the VECM (7.2.1) given in Proposition 7.3 have the following asymptotic properties:

$$\sqrt{T}\,\mathrm{vec}([\widetilde{\alpha}\widetilde{\beta}' : \widetilde{\mathbf{\Gamma}}] - [\mathbf{\Pi} : \mathbf{\Gamma}]) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\mathrm{co}}), \tag{7.2.21}$$

where Σ_{co} is as defined in Proposition 7.1, and

$$\sqrt{T}\,\mathrm{vech}(\widetilde{\Sigma}_u - \Sigma_u) \xrightarrow{d} \mathcal{N}(0, 2\mathbf{D}_K^+(\Sigma_u \otimes \Sigma_u)\mathbf{D}_K^{+\prime}). \tag{7.2.22}$$

Furthermore, $\widetilde{\Sigma}_u$ is asymptotically independent of $\widetilde{\alpha}\widetilde{\beta}'$ and $\widetilde{\mathbf{\Gamma}}$. Here, as usual, $\mathbf{D}_K^+ = (\mathbf{D}_K'\mathbf{D}_K)^{-1}\mathbf{D}_K'$ and $\mathbf{D}_K$ is the $(K^2 \times \frac{1}{2}K(K+1))$ duplication matrix. ■

Remark 1 It is clear that the ML estimator of $[\mathbf{\Pi} : \mathbf{\Gamma}]$ must have the same asymptotic distribution as the LS estimator in Proposition 7.1 because the ML estimator with known or given cointegration matrix β also has the same asymptotic distribution. The ML estimator $\widetilde{\alpha}\widetilde{\beta}'$ of $\mathbf{\Pi}$ in Proposition 7.3 may be viewed as a restricted LS estimator which is not as much restricted as the one with known β. Thus, the asymptotic result in (7.2.21) is not surprising. A rigorous proof of the result is given in Johansen (1995). ■

Remark 2 The covariance matrix Σ_{co} is singular, as noted in Remark 1 for Proposition 7.1. The rank of the $(K^2p \times K^2p)$ matrix Σ_{co} cannot be greater than $K(Kp - K + r)$ which is smaller than K^2p if $r < K$. ■

Remark 3 Individually, the matrices α and β cannot be estimated consistently without further constraints. Under the assumptions of Proposition 7.4, these matrices are not identified (not unique). If we make specific identifying assumptions in order to obtain unique parameter values and estimators, consistent estimation is possible. For instance, we may use

$$\beta = \begin{bmatrix} I_r \\ \beta_{(K-r)} \end{bmatrix}.$$

The ML estimator of $\beta_{(K-r)}$ may be obtained from the ML estimator of β given in Proposition 7.3 by denoting the first r rows of $\widetilde{\beta}$ by $\widetilde{\beta}_{(r)}$ and letting $\breve{\beta}_{(K-r)}$ consist of the last $K-r$ rows of $\widetilde{\beta}\widetilde{\beta}_{(r)}^{-1}$. This ML estimator has the same asymptotic properties as the EGLS estimator in Proposition 7.2 (see Ahn & Reinsel (1990)). In other words, inference procedures based on the ML estimator can be derived from the result

$$\operatorname{vec}\left[(\breve{\beta}'_{(K-r)} - \beta'_{(K-r)})\left(R_1^{(2)}R_1^{(2)\prime}\right)^{1/2}\right] \xrightarrow{d} \mathcal{N}\left(0, I_{K-r} \otimes (\alpha'\Sigma_u^{-1}\alpha)^{-1}\right).$$

It was found in a number of studies that the ML estimator $\breve{\beta}_{(K-r)}$ may have some undesirable properties in small samples and, in particular, it may produce occasional outlying estimates which are far away from the true parameter values (e.g., Phillips (1994), Hansen, Kim & Mittnik (1998)). This behavior of the estimator is due to the lack of finite sample moments. Brüggemann & Lütkepohl (2004) compared the EGLS and ML estimators in a small Monte Carlo study and found that the EGLS estimator is more robust in this respect. ■

Remark 4 If β is identified, the corresponding ML estimator of α is asymptotically normal, i.e., $\sqrt{T}\operatorname{vec}(\widetilde{\alpha} - \alpha)$ converges to the same asymptotic distribution as in Remark 2 for Proposition 7.1. ■

Remark 5 The normality of the process is not essential for the asymptotic properties of the estimators $\widetilde{\mathbf{\Gamma}}$ and $\widetilde{\mathbf{\Pi}} = \widetilde{\alpha}\widetilde{\beta}'$. Much of Proposition 7.4 holds under weaker conditions when quasi ML estimators based on the Gaussian likelihood function are considered. We have chosen the normality assumption for convenience. ■

Remark 6 The asymptotic distribution of $\widetilde{\Sigma}_u$ may be different if u_t is not Gaussian. The limiting distribution in (7.2.22) is obtained from the following lemma. ■

Lemma 7.3

$$\operatorname{plim} \sqrt{T}(\widetilde{\Sigma}_u - UU'/T) = 0.$$

■

This lemma not only implies consistency of $\widetilde{\Sigma}_u$ but also shows that the asymptotic distribution of

$$\sqrt{T}\operatorname{vech}(\widetilde{\Sigma}_u - \Sigma_u)$$

is the same as that of

$$\sqrt{T}\operatorname{vech}(T^{-1}UU' - \Sigma_u).$$

In other words, it is independent of the other coefficients of the system and has the form given in (7.2.22) (see also Section 3.4, Proposition 3.4).

Proof of Lemma 7.3:

$$\begin{aligned}
\widetilde{\Sigma}_u &= T^{-1}(\Delta Y - \widetilde{\alpha}\widetilde{\beta}' Y_{-1} - \widetilde{\mathbf{\Gamma}}\Delta X)(\Delta Y - \widetilde{\alpha}\widetilde{\beta}' Y_{-1} - \widetilde{\mathbf{\Gamma}}\Delta X)' \\
&= T^{-1}[U + (\mathbf{\Pi} - \widetilde{\alpha}\widetilde{\beta}')Y_{-1} + (\mathbf{\Gamma} - \widetilde{\mathbf{\Gamma}})\Delta X] \\
&\quad \times [U + (\mathbf{\Pi} - \widetilde{\alpha}\widetilde{\beta}')Y_{-1} + (\mathbf{\Gamma} - \widetilde{\mathbf{\Gamma}})\Delta X]' \\
&= \frac{UU'}{T} + (\mathbf{\Pi} - \widetilde{\alpha}\widetilde{\beta}')\frac{Y_{-1}U'}{T} + \frac{UY'_{-1}}{T}(\mathbf{\Pi} - \widetilde{\alpha}\widetilde{\beta}')' \\
&\quad + (\mathbf{\Pi} - \widetilde{\alpha}\widetilde{\beta}')\frac{Y_{-1}Y'_{-1}}{T}(\mathbf{\Pi} - \widetilde{\alpha}\widetilde{\beta}')' \\
&\quad + (\mathbf{\Pi} - \widetilde{\alpha}\widetilde{\beta}')\frac{Y_{-1}\Delta X'}{T}(\mathbf{\Gamma} - \widetilde{\mathbf{\Gamma}})' + (\mathbf{\Gamma} - \widetilde{\mathbf{\Gamma}})\frac{\Delta X Y'_{-1}}{T}(\mathbf{\Pi} - \widetilde{\alpha}\widetilde{\beta}')' \\
&\quad + (\mathbf{\Gamma} - \widetilde{\mathbf{\Gamma}})\frac{\Delta X U'}{T} + \frac{U\Delta X'}{T}(\mathbf{\Gamma} - \widetilde{\mathbf{\Gamma}})' \\
&\quad + (\mathbf{\Gamma} - \widetilde{\mathbf{\Gamma}})\frac{\Delta X \Delta X'}{T}(\mathbf{\Gamma} - \widetilde{\mathbf{\Gamma}})'.
\end{aligned}$$

Using $\widetilde{\alpha}\widetilde{\beta}' - \mathbf{\Pi} = O_p(T^{-1/2})$, $\widetilde{\mathbf{\Gamma}} - \mathbf{\Gamma} = O_p(T^{-1/2})$ and the results in Lemma 7.2, we get

$$\sqrt{T}(\mathbf{\Gamma} - \widetilde{\mathbf{\Gamma}})\frac{\Delta X U'}{T} = o_p(1),$$

$$\sqrt{T}(\mathbf{\Gamma} - \widetilde{\mathbf{\Gamma}})\frac{\Delta X \Delta X'}{T}(\mathbf{\Gamma} - \widetilde{\mathbf{\Gamma}})' = o_p(1),$$

$$\sqrt{T}(\mathbf{\Pi} - \widetilde{\alpha}\widetilde{\beta}')\frac{Y_{-1}\Delta X'}{T}(\mathbf{\Gamma} - \widetilde{\mathbf{\Gamma}})' = o_p(1),$$

and

$$\sqrt{T}(\mathbf{\Pi} - \widetilde{\alpha}\widetilde{\beta}')\frac{Y_{-1}Y'_{-1}}{T}(\mathbf{\Pi} - \widetilde{\alpha}\widetilde{\beta}')' = o_p(1).$$

Thus, Lemma 7.3 is proven if we can show that

$$\sqrt{T}(\widetilde{\alpha}\widetilde{\beta}' - \mathbf{\Pi})\frac{Y_{-1}U'}{T} = o_p(1). \tag{7.2.23}$$

To prove this result, we define $\widetilde{\alpha}(\beta)$ to be the ML estimator of α given β and note that

$$\begin{aligned}
\sqrt{T}(\widetilde{\alpha}\widetilde{\beta}' - \mathbf{\Pi})\frac{Y_{-1}U'}{T} &= \sqrt{T}[\widetilde{\alpha}\widetilde{\beta}' - \widetilde{\alpha}(\beta)\beta']\frac{Y_{-1}U'}{T} \\
&\quad + \sqrt{T}[\widetilde{\alpha}(\beta) - \alpha]\frac{\beta' Y_{-1}U'}{T}.
\end{aligned}$$

This quantity converges to zero in probability by Lemma 7.2(4), the fact that $\sqrt{T}[\widetilde{\alpha}(\beta) - \alpha] = O_p(1)$ (see (7.2.8)) and because $\sqrt{T}[\widetilde{\alpha}\widetilde{\beta}' - \widetilde{\alpha}(\beta)\beta'] = o_p(1)$. We leave the latter result as an exercise (see Problem 7.3). ■

7.2.4 Including Deterministic Terms

So far we have assumed that there are no deterministic terms in the data generation process, to simplify the exposition. In practice, such terms are typically needed for a proper representation of the data generation process. It turns out, however, that they can be easily accommodated in the estimation procedures for VECMs discussed so far, if the setup of Section 6.4 is used. Suppose the observed process y_t can be represented as

$$y_t = \mu_t + x_t, \tag{7.2.24}$$

where x_t is a zero mean process with VECM representation as in (7.2.1) and μ_t stands for the deterministic term. In general, the latter term may consist of polynomial trends, seasonal and other dummy variables as well as constant means. As in Section 6.4, we can then set up the VECM for the observed y_t variables as

$$\begin{aligned} \Delta y_t &= \alpha[\beta' : \eta'] \begin{bmatrix} y_{t-1} \\ D^{co}_{t-1} \end{bmatrix} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + CD_t + u_t \\ &= \mathbf{\Pi}^+ y^+_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + CD_t + u_t, \end{aligned} \tag{7.2.25}$$

where D^{co}_t contains all the deterministic terms which are present in the cointegration relations, D_t contains all remaining deterministics, and η' and C are the corresponding parameter matrices. Moreover, $\mathbf{\Pi}^+ := \alpha[\beta' : \eta'] = \alpha\beta^{+\prime}$ and

$$y^+_t := \begin{bmatrix} y_t \\ D^{co}_t \end{bmatrix}.$$

Notice that we assume that a specific deterministic term appears only once, either in D^{co}_t or in D_t.

Now we can simply modify the matrices used for representing the estimators in the previous subsections and then use basically the same formulas as before for computing the estimators. For example, defining

$$Y^+_{-1} := [y^+_0, \ldots, y^+_{T-1}],$$

$$\mathbf{\Gamma}^+ := [\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_{p-1}, C],$$

and

$$\Delta X^+ := [\Delta X^+_0, \ldots, \Delta X^+_{T-1}] \quad \text{with} \quad \Delta X^+_{t-1} := \begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \\ D_t \end{bmatrix}$$

gives the LS estimator

$$[\widehat{\mathbf{\Pi}}^{+} : \widehat{\mathbf{\Gamma}}^{+}] = [\Delta Y Y_{-1}^{+\prime} : \Delta Y \Delta X^{+\prime}] \begin{bmatrix} Y_{-1}^{+} Y_{-1}^{+\prime} & Y_{-1}^{+} \Delta X^{+\prime} \\ \Delta X^{+} Y_{-1}^{+\prime} & \Delta X^{+} \Delta X^{+\prime} \end{bmatrix}^{-1}.$$

The EGLS or ML estimators may be obtained analogously.

Hence, the computation of the estimators is equally easy as in the case without deterministic terms. Also, the asymptotic properties of the parameter estimators are essentially unchanged. The asymptotic theory for the deterministic terms requires some care, however, because their convergence rates depend on the specific terms included. For instance, if linear trends are included, the convergence rates of the associated slope parameters are different from $\sqrt{T}$. Generally, if the VECM is specified properly, including the cointegrating rank r, and if EGLS or ML methods are used, the usual inference methods are available. In particular, likelihood ratio tests for parameter restrictions related to the deterministic terms permit standard χ^2 asymptotics (see, e.g., Johansen (1995)).

A question of interest in this context is, for example, whether a particular deterministic term can indeed be constrained to the cointegration relations or needs to be maintained in unrestricted form in the model. The i-th component of D_t can be absorbed in the error correction term if the i-th column of the coefficient matrix C, denoted by C_i, satisfies $C_i = \alpha\eta_i$ for some r-dimensional vector η_i. Thus, the relevant null hypothesis is

$$\alpha_{\perp}' C_i = 0.$$

In other words, there are $K - r$ restrictions for each component that is confined to the cointegration relations. They are easy to test by a likelihood ratio test because the ML estimators and, hence, the likelihood maxima are easy to obtain for both the restricted and unrestricted model by just specifying the terms in D_t^{co} and D_t accordingly. If m deterministic components are restricted to the cointegration relations, the LR statistic has an asymptotic $\chi^2(m(K - r))$-distribution under our usual assumptions.

7.2.5 Other Estimation Methods for Cointegrated Systems

Some other estimation methods for cointegration relations and VECMs have been proposed in the literature. For example, other systems methods for estimating the cointegrating parameters were considered by Phillips (1991) who discussed nonparametric estimation of the short-run parameters. Stock & Watson (1988) proposed an estimator based on principal components and Bossaerts (1988) used canonical correlations. The latter two estimators were shown to be inferior to the ML estimators in a small sample comparison by Gonzalo (1994) and are therefore not considered here.

If there is just a single cointegration relation, it may also be estimated by single equation LS. Suppose that β is normalized as in (7.1.10) such that $\beta = (1, \beta_2, \dots, \beta_K)'$ and $\beta' y_t = y_{1t} + \beta_2 y_{2t} + \cdots + \beta_K y_{Kt}$. Hence,

$$y_{1t} = \gamma_2 y_{2t} + \cdots + \gamma_K y_{Kt} + ec_t,$$

where $\gamma_i := -\beta_i$ and ec_t is a stable, stationary process. Defining

$$y_{(1)} := \begin{bmatrix} y_{11} \\ \vdots \\ y_{1T} \end{bmatrix} \quad \text{and} \quad Y_{(2)} := \begin{bmatrix} y_{21} & \cdots & y_{K1} \\ \vdots & & \vdots \\ y_{2T} & \cdots & y_{KT} \end{bmatrix},$$

the LS estimator for $\gamma' := (\gamma_2, \dots, \gamma_K)$ is

$$\widehat{\gamma}' = y'_{(1)} Y_{(2)} (Y'_{(2)} Y_{(2)})^{-1}.$$

Stock (1987) showed that $\widehat{\gamma}$ is superconsistent and, more precisely, $T(\widehat{\gamma} - \gamma)$ converges in distribution. Thus, $\widehat{\gamma} - \gamma = O_p(T^{-1})$. However, there is some evidence that $\widehat{\gamma}$ is biased in small samples (Phillips & Hansen (1990)). Therefore, using LS estimation of the cointegration parameters without any correction for further dynamics in the model is not recommended.

A large number of single equation estimators for cointegration relations were reviewed and compared by Caporale & Pittis (2004). In addition to the simple LS estimator presented in the foregoing, they also considered estimators which are corrected for short-run dynamics. For example, this may be accomplished by including leads and lags of the differenced regressor variables in the estimation equation (e.g., Stock & Watson (1993)) or by adding also lagged differences of the dependent variable (e.g., Banerjee, Dolado, Galbraith & Hendry (1993), Wickens & Breusch (1988)). Another possible choice in this context is the fully modified estimator of Phillips & Hansen (1990) which takes care of the short-run dynamics nonparametrically and a semiparametric variant of this estimator proposed by Inder (1993). In addition, Caporale & Pittis (2004) presented a large number of modifications. Some of these estimators have rather undesirable small sample properties compared to the systems ML estimator presented in Section 7.2.3. Even those modifications that lead to small sample improvements were only shown to work in a rather limited framework. Also, of course, some of these estimators are only designed for situations where only one cointegration relation exists.

Two-Stage Estimation

Generally, if a superconsistent estimator $\widehat{\beta}$ of the cointegration matrix β is available, this estimator may be substituted for the true β and all the other parameters may be estimated in a second stage from

$$\Delta y_t = \alpha\widehat{\beta}' y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t^*, \tag{7.2.26}$$

where deterministic terms are again ignored for simplicity. If no restrictions are imposed on α and the $\mathbf{\Gamma}_i$'s ($i = 1, \dots, p-1$), LS estimation can be used without loss of asymptotic efficiency. Denoting the two-stage estimators of α and Γ by $\widehat{\alpha}_{2s}$ and $\widehat{\Gamma}_{2s}$, respectively, we have

$$\widehat{\alpha}_{2s} = \Delta Y M Y_{-1}' \widehat{\beta} \left(\widehat{\beta}' Y_{-1} M Y_{-1}' \widehat{\beta} \right)^{-1} \tag{7.2.27}$$

and

$$\widehat{\mathbf{\Gamma}}_{2s} = (\Delta Y - \widehat{\alpha}_{2s} \widehat{\beta}' Y_{-1}) \Delta X' (\Delta X \Delta X')^{-1}, \tag{7.2.28}$$

where the notation from the previous subsections has been used. For these estimators the following proposition holds, which is stated without proof.

Proposition 7.5 (*Asymptotic Properties of the Two-Stage LS Estimator*)
Let y_t be a K-dimensional, cointegrated process with VECM representation (7.2.1). Then the two-stage estimator is consistent and

$$\sqrt{T} \operatorname{vec}([\widehat{\alpha}_{2s} : \widehat{\mathbf{\Gamma}}_{2s}] - [\alpha : \mathbf{\Gamma}]) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\alpha,\mathbf{\Gamma}}), \tag{7.2.29}$$

where $\Sigma_{\alpha,\mathbf{\Gamma}}$ is the same covariance matrix as in (7.2.8). ■

The proposition implies that if a superconsistent estimator of the cointegration matrix β is available, the loading coefficients and short-run parameters of the VECM can be estimated by LS and these estimators have the same asymptotic properties we would obtain by using the true β. Thus, standard inference procedures can be used for the short-run parameters. An analogous result is also available for VECMs with parameter restrictions (see Section 7.3 for the extension).

The second stage in the procedure may be modified. For instance, one may just be interested in the first equation of the system. In this case, the first equation may be estimated separately without taking into account the remaining ones. Thus, the two-stage procedure may be applied in a single equation modelling context.

Results similar to those in Proposition 7.5 were derived by many authors (see, e.g., Stock (1987), Phillips & Durlauf (1986), Park & Phillips (1989), and Johansen (1991)). Generally there has been a considerable amount of research on estimation and hypothesis testing in systems with integrated and cointegrated variables. For instance, Johansen (1991), Johansen & Juselius (1990), and Lütkepohl & Reimers (1992b) considered estimation with restrictions on the cointegration and loading matrices; Park & Phillips (1988, 1989) and Phillips (1988) provided general results on estimating systems with integrated and cointegrated exogenous variables; Stock (1987) considered a so-called nonlinear LS estimator, and Phillips & Hansen (1990) discussed instrumental variables estimation of models containing integrated variables.

7.2.6 An Example

As an example, we use the bivariate system of quarterly, seasonally unadjusted German long-term interest rate ($R_t = y_{1t}$) and inflation rate ($Dp_t = y_{2t}$)

which was also analyzed in Lütkepohl (2004). The sample period is the second quarter of 1972 to the end of 1998. Thus we have $T = 107$ observations. The data are available in File E6 and the two time series are plotted in Figure 7.1. Preliminary tests indicated that both series have a unit root and there are also theoretical reasons for a cointegration relation between them. The so-called Fisher effect implies that the real interest rate is stationary. Because R_t is a nominal yearly interest rate while Dp_t is a quarterly inflation rate, one would therefore expect $R_t - 4Dp_t$ to be stationary, that is, this relation is expected to be a cointegration relation.

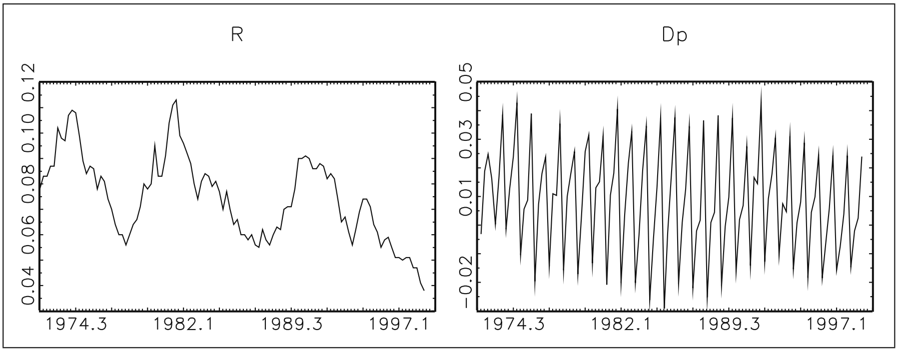

Fig. 7.1. Seasonally unadjusted, quarterly German interest rate (left) and inflation rate (right), 1972.2–1998.4.

We have fitted a VECM with a constant, seasonal dummy variables, and three lagged differences and the pre-specified cointegration relation $R_t - 4Dp_t$ to the data. The results are shown in Table 7.1. Notice that three lagged differences in the VECM imply a model with four lags in the levels. Including at least lags of one year seems plausible because the inflation series has a strong seasonal pattern (see Figure 7.1). Formal statistical procedures for determining the lag length will be discussed in the next chapter. The seasonal movement in Dp_t is also the reason for including seasonal dummy variables in addition to a constant. The deterministic term, $D_t = (1, s_{1t}, s_{2t}, s_{3t})'$, where the s_{it} are seasonal dummy variables, is placed outside the cointegration relation. We have also estimated a VECM with cointegrating rank $r = 1$ using the reduced rank ML procedure and the EGLS method. The estimates are also shown in Table 7.1.

The two estimated cointegration relations are

$$R_t - \underset{(0.63)}{3.96}\ Dp_t = ec_t^{ML} \qquad (7.2.30)$$

and

$$R_t - \underset{(0.61)}{3.63}\ Dp_t = ec_t^{EGLS}, \qquad (7.2.31)$$

Table 7.1. Estimated VECMs for interest rate/inflation example system

	known β	ML estimator	EGLS estimator
α	$\begin{bmatrix} \underset{(-2.3)}{-0.10} \\ \underset{(3.8)}{0.16} \end{bmatrix}$	$\begin{bmatrix} \underset{(-2.3)}{-0.10} \\ \underset{(3.8)}{0.16} \end{bmatrix}$	$\begin{bmatrix} \underset{(-2.8)}{-0.14} \\ \underset{(2.9)}{0.14} \end{bmatrix}$
β'	$[1 : -4]$	$\begin{bmatrix} 1.00 : & \underset{(-6.3)}{-3.96} \end{bmatrix}$	$\begin{bmatrix} 1.00 : & \underset{(-6.0)}{-3.63} \end{bmatrix}$
$\boldsymbol{\Gamma}_1$	$\begin{bmatrix} \underset{(2.7)}{0.27} & \underset{(-1.4)}{-0.21} \\ \underset{(0.7)}{0.07} & \underset{(-2.4)}{-0.34} \end{bmatrix}$	$\begin{bmatrix} \underset{(2.7)}{0.27} & \underset{(-1.4)}{-0.21} \\ \underset{(0.7)}{0.07} & \underset{(-2.4)}{-0.34} \end{bmatrix}$	$\begin{bmatrix} \underset{(2.9)}{0.29} & \underset{(-1.1)}{-0.16} \\ \underset{(0.8)}{0.08} & \underset{(-2.2)}{-0.31} \end{bmatrix}$
$\boldsymbol{\Gamma}_2$	$\begin{bmatrix} \underset{(-0.2)}{-0.02} & \underset{(-1.8)}{-0.22} \\ \underset{(-0.0)}{-0.00} & \underset{(-3.4)}{-0.39} \end{bmatrix}$	$\begin{bmatrix} \underset{(-0.2)}{-0.02} & \underset{(-1.8)}{-0.22} \\ \underset{(-0.0)}{-0.00} & \underset{(-3.4)}{-0.39} \end{bmatrix}$	$\begin{bmatrix} \underset{(0.1)}{0.01} & \underset{(-1.6)}{-0.19} \\ \underset{(0.1)}{0.01} & \underset{(-3.2)}{-0.37} \end{bmatrix}$
$\boldsymbol{\Gamma}_3$	$\begin{bmatrix} \underset{(2.3)}{0.22} & \underset{(-1.3)}{-0.11} \\ \underset{(0.2)}{0.02} & \underset{(-4.5)}{-0.35} \end{bmatrix}$	$\begin{bmatrix} \underset{(2.3)}{0.22} & \underset{(-1.3)}{-0.11} \\ \underset{(0.2)}{0.02} & \underset{(-4.5)}{-0.35} \end{bmatrix}$	$\begin{bmatrix} \underset{(2.6)}{0.26} & \underset{(-1.1)}{-0.09} \\ \underset{(0.4)}{0.04} & \underset{(-4.4)}{-0.34} \end{bmatrix}$
C'	$\begin{bmatrix} \underset{(0.4)}{0.001} & \underset{(3.0)}{0.010} \\ \underset{(0.3)}{0.001} & \underset{(-7.5)}{-0.034} \\ \underset{(1.8)}{0.009} & \underset{(-3.8)}{-0.018} \\ \underset{(-0.1)}{-0.000} & \underset{(-3.6)}{-0.016} \end{bmatrix}$	$\begin{bmatrix} \underset{(0.4)}{0.002} & \underset{(3.0)}{0.010} \\ \underset{(0.3)}{0.001} & \underset{(-7.5)}{-0.034} \\ \underset{(1.8)}{0.009} & \underset{(-3.8)}{-0.018} \\ \underset{(-0.1)}{-0.000} & \underset{(-3.6)}{-0.016} \end{bmatrix}$	$\begin{bmatrix} \underset{(1.2)}{0.005} & \underset{(3.1)}{0.012} \\ \underset{(0.3)}{0.001} & \underset{(-7.5)}{-0.034} \\ \underset{(1.8)}{0.009} & \underset{(-3.8)}{-0.018} \\ \underset{(-0.1)}{-0.000} & \underset{(-3.6)}{-0.016} \end{bmatrix}$

Note: t-values in parentheses underneath parameter estimates; deterministic terms: constant and seasonal dummies ($D_t = (1, s_{1t}, s_{2t}, s_{3t})'$).

where estimated standard errors are given in parentheses. The first coefficient is normalized to be 1. Thereby the t-ratios and the standard errors of the inflation coefficient can be interpreted in the usual way. Clearly, -4 is well within a two-standard error interval around both estimates. Therefore one could argue that restricting the inflation coefficient to 4 is in line with the data. Using the result in Proposition 7.2, a formal test of the null hypothesis $H_0 : \beta_2 = -4$, where β_2 denotes the second component of β, can be based on the t-statistic

$$\frac{-3.96 - (-4)}{0.63} = 0.06$$

for the ML estimator or on

$$\frac{-3.63 - (-4)}{0.61} = 0.61$$

for the EGLS estimator. Both t-values are small compared to critical values from the standard normal distribution corresponding to usual significance levels. Hence, the null hypothesis cannot be rejected for either of the two estimators.

Comparing the other estimates of the three models in Table 7.1, it is obvious that corresponding estimates do not differ much, especially when the sampling uncertainty reflected in the t-ratios is taken into account. In particular, the ML estimates are very close to those of the model with fixed cointegration vector. Thus, imposing the theoretically expected cointegration vector does not appear to be a problematic constraint.

Another observation that can be made in Table 7.1 is that there are some insignificant coefficients in the short-run matrices $\mathbf{\Gamma}_i$ and the estimated deterministic terms (C). Because some of the parameters in $\widehat{\mathbf{\Gamma}}_3$ have rather large t-ratios, it is clear that simply reducing the lag order is not likely to be a good strategy for reducing the number of parameters in the model. It makes sense, however, to consider restricting some of the parameter values to zero. This issue is discussed in the next section.

7.3 Estimating VECMs with Parameter Restrictions

As for other models, restrictions may be imposed on the parameters of VECMs to increase the estimation precision. We will first discuss restrictions for the cointegration relations and then turn to restrictions on the loading coefficients and short-run parameters.

7.3.1 Linear Restrictions for the Cointegration Matrix

In case just-identifying restrictions for the cointegration relations are available, estimation may proceed as described in Section 7.2 and then the identified estimator of β may be obtained by a suitable transformation of the estimator $\widehat{\beta}$. For example, if β is just a single vector and ML estimation is used, a normalization of the first component may be obtained by dividing the vector $\widetilde{\beta}$ by its first component, as discussed earlier.

Sometimes over-identifying restrictions are available for the cointegration matrix. In general, if the restrictions can be expressed in the form

$$\operatorname{vec}(\beta'_{(K-r)}) = \mathbf{R}\gamma + \mathbf{r}, \tag{7.3.1}$$

where $\mathbf{R}$ is a fixed $(r(K-r) \times m)$ matrix of rank m, $\mathbf{r}$ is a fixed $r(K-r)$-dimensional vector, and γ is a vector of free parameters, the EGLS estimator is still available. The GLS estimator may be obtained from the vectorized "concentrated model" (7.2.14),

$$\begin{aligned}\operatorname{vec}(R_0 - \alpha R_1^{(1)}) &= (R_1^{(2)\prime} \otimes \alpha)\operatorname{vec}(\beta'_{(K-r)}) + \operatorname{vec}(U^*)\\ &= (R_1^{(2)\prime} \otimes \alpha)(\mathbf{R}\gamma + \mathbf{r}) + \operatorname{vec}(U^*),\end{aligned}$$

so that

$$\text{vec}(R_0 - \alpha R_1^{(1)}) - (R_1^{(2)\prime} \otimes \alpha)\mathbf{r} = (R_1^{(2)\prime} \otimes \alpha)\mathbf{R}\gamma + \text{vec}(U^*). \tag{7.3.2}$$

Thus, the GLS estimator for γ is

$$\begin{aligned}\widehat{\gamma} &= \left[\mathbf{R}'(R_1^{(2)} R_1^{(2)\prime} \otimes \alpha' \Sigma_u^{-1}\alpha)\mathbf{R}\right]^{-1} \\ &\quad \times \mathbf{R}'(R_1^{(2)} \otimes \alpha' \Sigma_u^{-1}) \left[\text{vec}(R_0 - \alpha R_1^{(1)}) - (R_1^{(2)\prime} \otimes \alpha)\mathbf{r}\right].\end{aligned}$$

Substituting consistent estimators $\widehat{\alpha}$ and $\widehat{\Sigma}_u$ for α and Σ_u, respectively, gives the EGLS estimator

$$\begin{aligned}\widehat{\widehat{\gamma}} &= \left[\mathbf{R}'(R_1^{(2)} R_1^{(2)\prime} \otimes \widehat{\alpha}' \widehat{\Sigma}_u^{-1}\widehat{\alpha})\mathbf{R}\right]^{-1} \\ &\quad \times \mathbf{R}'(R_1^{(2)} \otimes \widehat{\alpha}' \widehat{\Sigma}_u^{-1}) \left[\text{vec}(R_0 - \widehat{\alpha} R_1^{(1)}) - (R_1^{(2)\prime} \otimes \widehat{\alpha})\mathbf{r}\right].\end{aligned} \tag{7.3.3}$$

Extending the arguments used for proving Proposition 7.2, the following asymptotic properties of the EGLS estimator can be shown.

Proposition 7.6 (*Asymptotic Properties of the Restricted EGLS Estimator*)
Suppose y_t is generated by the VECM (7.2.1) and β satisfies the restrictions in (7.3.1). Then

$$\left[\mathbf{R}'(R_1^{(2)} R_1^{(2)\prime} \otimes \widehat{\alpha}' \widehat{\Sigma}_u^{-1}\widehat{\alpha})\mathbf{R}\right]^{1/2} (\widehat{\widehat{\gamma}} - \gamma) \xrightarrow{d} \mathcal{N}(0, I_m). \tag{7.3.4}$$

■

Thus, standard inference procedures can be based on the transformed estimator. It can also be shown that $\widehat{\widehat{\gamma}} - \gamma = O_p(T^{-1})$. In other words, the estimator is superconsistent. Clearly, consistent estimators of α and Σ_u are readily available from unrestricted LS estimation as in Section 7.2.2.

Defining $\widehat{\widehat{\beta}}{}^R_{(K-r)}$ such that $\text{vec}\,\widehat{\widehat{\beta}}{}^R_{(K-r)} = \mathbf{R}\widehat{\widehat{\gamma}} + \mathbf{r}$,

$$\widehat{\widehat{\beta}}{}^R := \begin{bmatrix} I_r \\ \widehat{\widehat{\beta}}{}^R_{(K-r)} \end{bmatrix}$$

is a restricted estimator of the cointegration matrix. It can, for example, be used in the two-stage procedure described in Section 7.2.5.

If the restrictions for the cointegration matrix can be written in the form $\beta = H\varphi$, where H is some known, fixed $(K \times s)$ matrix and φ is $(s \times r)$ with $s \geq r$, ML estimation is also straightforward. For example, in a system with three variables and one cointegration relation, if $\beta_{31} = -\beta_{21}$, we have

$$\beta = \begin{bmatrix} \beta_{11} \\ \beta_{21} \\ -\beta_{21} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{21} \end{bmatrix} = H\varphi,$$

where $\varphi := (\beta_{11}, \beta_{21})'$ and H is defined in the obvious way. If the restrictions can be represented in this form, Y_{-1} is simply replaced by $H'Y_{-1}$ in the quantities entering the eigenvalue problem in Proposition 7.3. Denoting the resulting estimator by $\widetilde{\varphi}$ gives a restricted estimator $\widetilde{\beta} = H\widetilde{\varphi}$ for β and corresponding estimators of α and $\mathbf{\Gamma}$ as in Proposition 7.3. If the restrictions in (7.3.1) can be written in this form, the EGLS and the ML estimators have again identical asymptotic properties.

However, the restrictions in (7.3.1) can in general not be written in the form $\beta = H\varphi$. For instance, if there are three variables ($K = 3$) and two cointegrating relations ($r = 2$), a single zero restriction on the second cointegration vector cannot be expressed in the form $\beta = H\varphi$, whereas it may still be written in the form (7.3.1). Moreover, it may be expressed in the form $\beta = [H_1\varphi_1, H_2\varphi_2]$ with suitable matrices H_1 and H_2 and vectors φ_1 and φ_2. For example, if a zero restriction is placed on the last element of the second cointegrating vector, we get

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & 0 \end{bmatrix} = [H_1\varphi_1, H_2\varphi_2]$$

with $H_1 := I_3$, $\varphi_1 := (\beta_{11}, \beta_{21}, \beta_{31})'$,

$$H_2 := \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

and $\varphi_2 := (\beta_{12}, \beta_{22})'$. In that case, restricted ML estimation is still not difficult but requires an iterative optimization (see Boswijk & Doornik (2002)).

7.3.2 Linear Restrictions for the Short-Run and Loading Parameters

If a superconsistent estimator of the cointegration matrix $\widehat{\beta}$ is available, the two-stage procedure described in Section 7.2.5 can be used for estimating the loading and short-run parameters of a VECM. The method can be readily extended to models with parameter restrictions. Suppose linear restrictions of the form

$$\mathrm{vec}[\alpha : \mathbf{\Gamma}] = \Re\boldsymbol{\varphi}, \tag{7.3.5}$$

where $\Re$ is a fixed $(K(r + K(p-1)) \times n)$ matrix and $\boldsymbol{\varphi}$ is an n-dimensional vector. Then we can write the model in matrix form as

$$\Delta Y = [\alpha : \mathbf{\Gamma}] \begin{bmatrix} \widehat{\beta}' Y_{-1} \\ \Delta X \end{bmatrix} + U^*$$

and in vectorized form we get

$$\begin{aligned}\operatorname{vec}(\Delta Y) &= \left([Y'_{-1}\widehat{\beta} : \Delta X'] \otimes I_K\right) \operatorname{vec}[\alpha : \mathbf{\Gamma}] + \operatorname{vec}(U^*) \\ &= \left([Y'_{-1}\widehat{\beta} : \Delta X'] \otimes I_K\right) \Re\varphi + \operatorname{vec}(U^*).\end{aligned}$$

Hence, the GLS estimator of φ is

$$\begin{aligned}\widehat{\varphi} &= \left[\Re'\left(\begin{bmatrix} \widehat{\beta}'Y_{-1}Y'_{-1}\widehat{\beta} & \widehat{\beta}'Y_{-1}\Delta X' \\ \Delta X Y'_{-1}\widehat{\beta} & \Delta X \Delta X' \end{bmatrix} \otimes \Sigma_u^{-1}\right)\Re\right]^{-1} \\ &\quad \times \Re'\left(\begin{bmatrix} \widehat{\beta}'Y_{-1} \\ \Delta X \end{bmatrix} \otimes \Sigma_u^{-1}\right) \operatorname{vec}(\Delta Y), \qquad (7.3.6)\end{aligned}$$

from which an EGLS estimator $\widehat{\widehat{\varphi}}$ is obtained by replacing the residual covariance matrix Σ_u by a consistent estimator. The latter estimator may, for example, be obtained from an unrestricted estimation of the model. The resulting EGLS estimator has the following asymptotic properties.

Proposition 7.7 (*Asymptotic Properties of the Restricted EGLS Estimator of the Short-Run Parameters*)
Suppose y_t is generated by the VECM (7.2.1), $\widehat{\beta}$ is a superconsistent estimator of β, $\widehat{\Sigma}_u$ is a consistent estimator of Σ_u, and the short-run and loading parameters satisfy (7.3.5). Then

$$\begin{aligned}&\sqrt{T}(\widehat{\widehat{\varphi}} - \varphi) \\ &\xrightarrow{d} \mathcal{N}\left(0, \operatorname{plim} T\left[\Re'\left(\begin{bmatrix} \widehat{\beta}'Y_{-1}Y'_{-1}\widehat{\beta} & \widehat{\beta}'Y_{-1}\Delta X' \\ \Delta X Y'_{-1}\widehat{\beta} & \Delta X \Delta X' \end{bmatrix} \otimes \Sigma_u^{-1}\right)\Re\right]^{-1}\right).\end{aligned} \qquad (7.3.7)$$

■

We do not prove the proposition but just note that it follows from the fact that only stationary variables are involved if $\widehat{\beta}$ is replaced by the true cointegration matrix β and the resulting estimator for φ differs from $\widehat{\widehat{\varphi}}$ by a quantity which is $o_p(T^{-1/2})$. Moreover, the asymptotic normal distribution of $\operatorname{vec}[\widehat{\widehat{\alpha}} : \widehat{\widehat{\mathbf{\Gamma}}}] = \Re\widehat{\widehat{\varphi}}$ follows in the usual way.

It is straightforward to extend these result to the case where the restrictions are of the form

$$\operatorname{vec}[\alpha : \mathbf{\Gamma}] = \Re\varphi + \mathbf{r}, \qquad (7.3.8)$$

where $\mathbf{r}$ is now a fixed $(K(r + K(p-1))) \times 1)$ vector (see Problem 7.6). The more special restrictions in (7.3.5) are considered here for convenience and because they cover most cases of practical importance.

7.3.3 An Example

In Section 7.2.6, we have seen that in the short-run dynamics of the German interest rate/inflation example models a number of coefficients have quite low t-ratios (see Table 7.1). Therefore it makes sense to restrict some of the coefficients to zero. The following model from Lütkepohl (2004, Equation (3.41)) for our data set is an example of a restricted (subset) VECM:

$$\begin{bmatrix} \Delta R_t \\ \Delta Dp_t \end{bmatrix} = \begin{bmatrix} \underset{(-3.1)}{-0.07} \\ \underset{(4.5)}{0.17} \end{bmatrix} (R_{t-1} - 4Dp_{t-1})$$

$$+ \begin{bmatrix} \underset{(2.5)}{0.24} & \underset{(-1.9)}{-0.08} \\ 0 & \underset{(-2.5)}{-0.31} \end{bmatrix} \begin{bmatrix} \Delta R_{t-1} \\ \Delta Dp_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & \underset{(-2.5)}{-0.13} \\ 0 & \underset{(-3.6)}{-0.37} \end{bmatrix} \begin{bmatrix} \Delta R_{t-2} \\ \Delta Dp_{t-2} \end{bmatrix}$$

$$+ \begin{bmatrix} \underset{(2.1)}{0.20} & \underset{(-1.6)}{-0.06} \\ 0 & \underset{(-4.7)}{-0.34} \end{bmatrix} \begin{bmatrix} \Delta R_{t-3} \\ \Delta Dp_{t-3} \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & 0 & \underset{(2.8)}{0.010} & 0 \\ \underset{(3.0)}{0.010} & \underset{(-7.6)}{-0.034} & \underset{(-3.8)}{-0.018} & \underset{(-3.6)}{-0.016} \end{bmatrix} \begin{bmatrix} c \\ s_{1,t} \\ s_{2,t} \\ s_{3,t} \end{bmatrix} + \begin{bmatrix} \widehat{u}_{1,t} \\ \widehat{u}_{2,t} \end{bmatrix}, \tag{7.3.9}$$

$$\widetilde{\Sigma}_u = \begin{bmatrix} 2.61 & -0.15 \\ -0.15 & 2.31 \end{bmatrix} \times 10^{-5}.$$

Here we have used the fixed cointegration vector that was found in Section 7.2.6 and EGLS estimation of the loading coefficients and short-term parameters is used. t-ratios are again given in parentheses underneath the parameter estimates. They are all relatively large. In fact, with two exceptions they are all larger than two. Recall that t-ratios can be interpreted in the usual way as asymptotically standard normally distributed by Proposition 7.7. Comparing the model (7.3.9) to those in Table 7.1, it turns out that the parameters with very small t-ratios in the unrestricted models are just the ones restricted to zero in (7.3.9). The model was actually found by a sequential model selection procedure which will be discussed in the next chapter.

7.4 Bayesian Estimation of Integrated Systems

It is also possible to place Bayesian restrictions on VECMs. A very important constraint in these models is the cointegrating rank, however. In Bayesian analysis, a basic idea is to allow the data to revise the prior restrictions imposed by the analyst. Using this principle also for the unit roots and, hence, for the cointegration relations, setting up the system in VECM form may not be the most plausible approach anymore. Therefore, Bayesian restrictions have often been imposed on the levels VAR form, even if the variables are possibly integrated. A popular prior in this context is the Minnesota or Litterman prior which ignores possible cointegration between the variables altogether. We will present this prior in the following after the general setting has been discussed.

7.4.1 The Model Setup

In Chapter 5, Section 5.4, we have discussed Bayesian estimation of stationary, stable VAR(p) processes. For a Gaussian process with integrated variables and a normal prior, the posterior distribution of the VAR coefficients can be derived in a similar manner. We now consider a levels VAR(p) model of the form

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t.$$

As usual, $\boldsymbol{\beta} := \text{vec}[\nu, A_1, \ldots, A_p]$ is the vector of VAR coefficients including an intercept vector and we assume a prior

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}^*, V_{\boldsymbol{\beta}}). \tag{7.4.1}$$

Then, using the same line of reasoning as in Section 5.4, the posterior mean is

$$\bar{\boldsymbol{\beta}} = [V_{\boldsymbol{\beta}}^{-1} + (ZZ' \otimes \Sigma_u^{-1})]^{-1}[V_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}^* + (Z \otimes \Sigma_u^{-1})\mathbf{y}]$$

and the posterior covariance matrix is

$$\bar{\Sigma}_{\boldsymbol{\beta}} = [V_{\boldsymbol{\beta}}^{-1} + (ZZ' \otimes \Sigma_u^{-1})]^{-1},$$

where

$$\mathbf{y} := \text{vec}[y_1, \ldots, y_T] \quad \text{and} \quad Z := [Z_0, \ldots, Z_{T-1}] \quad \text{with } Z_t := \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}.$$

7.4.2 The Minnesota or Litterman Prior

A possible choice of $\boldsymbol{\beta}^*$ and $V_{\boldsymbol{\beta}}$ for stable processes was discussed in Section 5.4.3. If the variables are believed to be integrated, the following prior discussed by Doan et al. (1984) and Litterman (1986), sometimes known as Minnesota prior, could be used: (1) Set the prior mean of the first lag of each variable equal to one in its own equation and set all other coefficients at zero. In other words, if the prior means were the true parameter values each variable were a random walk. (2) Choose the prior variances of the coefficients as in Section 5.4.3. In other words, the prior variances of the intercept terms are infinite and the prior variance of $\alpha_{ij,l}$, the ij-th element of A_l, is

$$v_{ij,l} = \begin{cases} (\lambda/l)^2 & \text{if } i = j, \\ (\lambda\theta\sigma_i/l\sigma_j)^2 & \text{if } i \neq j, \end{cases}$$

where λ is the prior standard deviation of $\alpha_{ii,1}, 0 < \theta < 1$, and σ_i^2 is the i-th diagonal element of Σ_u. Thus, we get, for instance, for a bivariate VAR(2) system,

$$\begin{array}{rcccccc} y_{1t} = & \underset{(\infty)}{0} & + \underset{(\lambda)}{1 \cdot y_{1,t-1}} & + \underset{(\lambda\theta\sigma_1/\sigma_2)}{0 \cdot y_{2,t-1}} & + \underset{(\lambda/2)}{0 \cdot y_{1,t-2}} & + \underset{(\lambda\theta\sigma_1/2\sigma_2)}{0 \cdot y_{2,t-2}} & + u_{1t}, \\ y_{2t} = & \underset{(\infty)}{0} & + \underset{(\lambda\theta\sigma_2/\sigma_1)}{0 \cdot y_{1,t-1}} & + \underset{(\lambda)}{1 \cdot y_{2,t-1}} & + \underset{(\lambda\theta\sigma_2/2\sigma_1)}{0 \cdot y_{1,t-2}} & + \underset{(\lambda/2)}{0 \cdot y_{2,t-2}} & + u_{2t}, \end{array}$$

where all coefficients are set to their prior means and the numbers in parentheses are their prior standard deviations. Forgetting about the latter numbers for the moment, each of these two equations is seen to specify a random walk for one of the variables. The nonzero prior standard deviations indicate that we are not sure about such a simple model. The standard deviations decline with increasing lag length because more recent lags are assumed to be more likely to belong into the model. The infinite standard deviations for the intercept terms simply reflect that we do not have any prior guess for these coefficients. Also, we do not impose covariance priors and, hence, choose $V_{\boldsymbol{\beta}}$ to be a diagonal matrix. Its inverse is

$$V_{\boldsymbol{\beta}}^{-1} = \begin{bmatrix} 0 & & & & & & & & & \\ & 0 & & & & & & & & \\ & & \frac{1}{\lambda^2} & & & & & 0 & & \\ & & & \frac{\sigma_1^2}{(\lambda\theta\sigma_2)^2} & & & & & & \\ & & & & \frac{\sigma_2^2}{(\lambda\theta\sigma_1)^2} & & & & & \\ & & & & & \frac{1}{\lambda^2} & & & & \\ & & & 0 & & & \frac{2^2}{\lambda^2} & & & \\ & & & & & & & \frac{2^2\sigma_1^2}{(\lambda\theta\sigma_2)^2} & & \\ & & & & & & & & \frac{2^2\sigma_2^2}{(\lambda\theta\sigma_1)^2} & \\ & & & & & & & & & \frac{2^2}{\lambda^2} \end{bmatrix},$$

where 0 is also substituted for the inverse (infinite) standard deviation of the intercept terms.

To compute $\bar{\boldsymbol{\beta}}$ requires the inversion of $V_{\boldsymbol{\beta}}^{-1} + (ZZ' \otimes \Sigma_u^{-1})$. Because this matrix is usually quite large, in the past, Bayesian estimation has often been performed separately for each of the K equations of the system. In that case,

$$\bar{b}_k = [V_k^{-1} + \sigma_k^{-2} ZZ']^{-1}(V_k^{-1} b_k^* + \sigma_k^{-2} Z y_{(k)})$$

is used as an estimator for the parameters b_k of the k-th equation, that is, b_k' is the k-th row of $B := [\nu, A_1, \dots, A_p]$. Here V_k is the prior covariance matrix of b_k, b_k^* is its prior mean, and $y_{(k)} := (y_{k1}, \dots, y_{kT})'$. As in Chapter 5, σ_k^2 is replaced by the k-th diagonal element of the ML estimator

$$\widetilde{\Sigma}_u = Y(I_T - Z'(ZZ')^{-1}Z)Y'/T$$

of the white noise covariance matrix.

Clearly, in this prior, possible cointegration between the variables is not taken into account. Given the growing importance of the concept of cointegration in the recent literature, it is perhaps not surprising that the Minnesota prior has lately lost some of its appeal. Bayesians have responded to the success of the concept of cointegration and of VECMs in classical econometrics. Some recent contributions to Bayesian analysis of VECMs include Kleibergen & van Dijk (1994), Kleibergen & Paap (2002), Strachan (2003), and Strachan & Inder (2004). A survey with many more references was given by Koop, Strachan, van Dijk & Villani (2005).

7.4.3 An Example

As an example illustrating Bayesian estimation based on the Minnesota prior, we consider the following four-dimensional system of U.S. economic variables:

y_1 - logarithm of the real money stock M1 (ln M1),
y_2 - logarithm of GNP in billions of 1982 dollars (ln GNP),
y_3 - discount interest rate on new issues of 91-day Treasury bills (r^s),
y_4 - yield on long-term (20 years) Treasury bonds (r^l).

Quarterly data for the years 1954 to 1987 are used. The data are available in File E3. They are plotted in Figure 7.2. The GNP and Ml data are seasonally adjusted. The variables r^s and r^l are regarded as short- and long-term interest rates, respectively. The plots in Figure 7.2 show that the series are trending. Thus, they may be integrated and, given that this is a small monetary system, there may in fact be cointegration. For example, there may be a long-run money demand relation and perhaps the interest rate spread $r^l - r^s$ may be a stationary variable. Although the system may be cointegrated, we will consider the Minnesota prior in the following.

We have first fitted an unrestricted VAR(2) model to the data and present the results in Table 7.2. It can be seen that at least the last three of the four diagonal elements of A_1 are estimated to be close to 1. The first diagonal

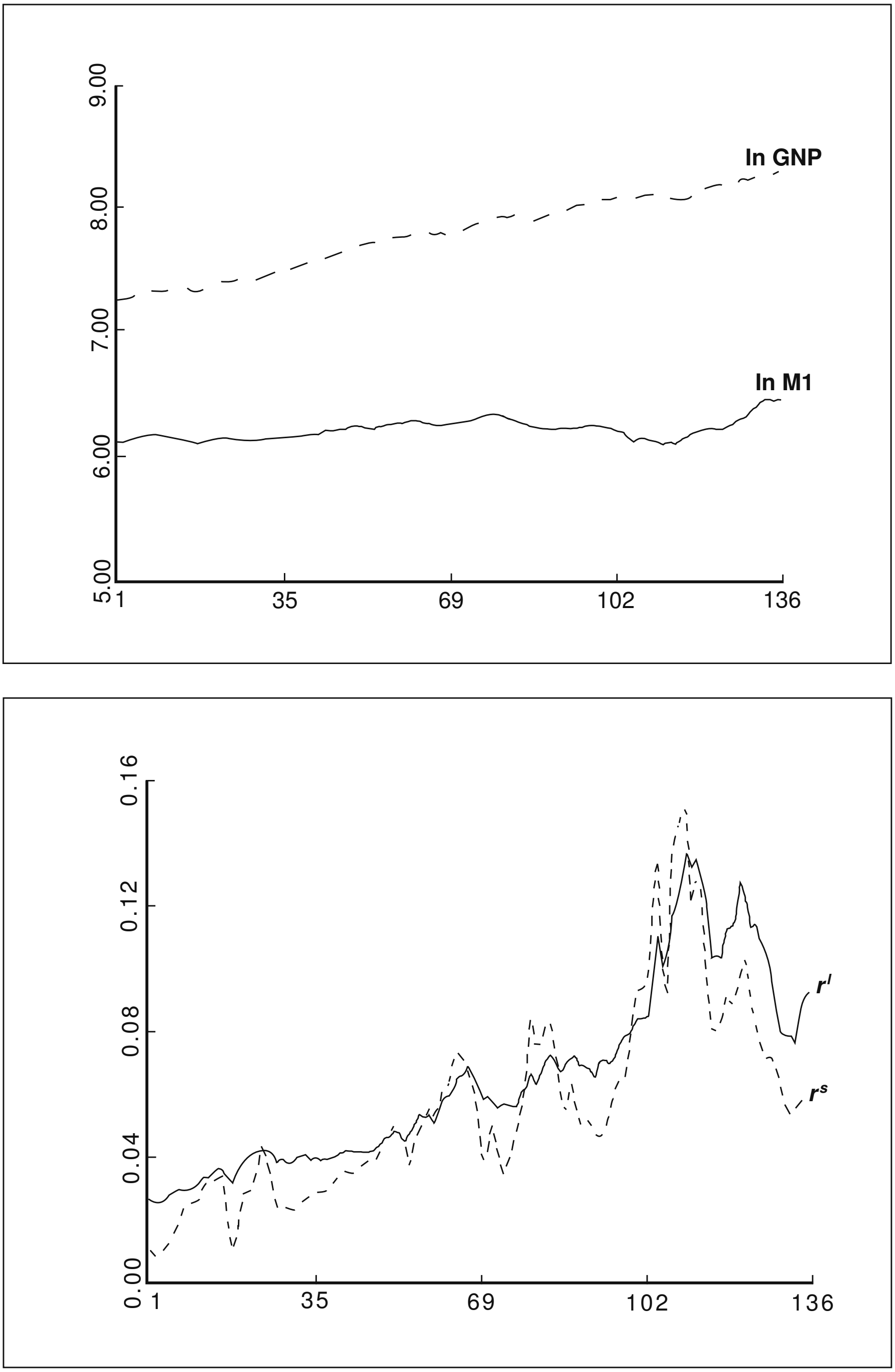

Fig. 7.2. U.S. ln M1, ln GNP, and interest rate time series.

element is also not drastically different from 1, although 1 is not within a two-standard error interval around the estimate. On the basis of the unrestricted estimates, a prior with mean 1 for the diagonal elements of A_1 does not appear to be unreasonable for this example. Of course, in a Bayesian analysis the prior is usually not chosen on the basis of an unrestricted estimation.

Table 7.2. VAR(2) coefficient estimates for the U.S. example system with estimated standard errors in parentheses

estimation method	ν	A_1				A_2			
	.028	1.307	.106	−.554	−.814	−.318	−.101	.318	1.022
		(.070)	(.075)	(.107)	(.224)	(.070)	(.076)	(.115)	(.221)
	.129	.080	1.045	−.177	.473	−.135	−.014	−.197	−.416
unrestricted		(.083)	(.088)	(.126)	(.265)	(.083)	(.090)	(.136)	(.261)
LS	.096	.193	.068	.978	.284	−.248	−.035	.053	−.644
		(.077)	(.081)	(.116)	(.245)	(.077)	(.083)	(.125)	(.240)
	.030	.042	.042	.034	1.065	−.064	−.027	.070	−.308
		(.038)	(.041)	(.058)	(.122)	(.038)	(.041)	(.063)	(.120)
	.041	1.332	.098	−.556	−.838	−.346	−.091	.354	.969
		(.067)	(.073)	(.104)	(.216)	(.064)	(.073)	(.110)	(.207)
	.086	.071	1.052	−.169	.549	−.099	−.039	−.239	−.286
ML		(.079)	(.086)	(.123)	(.256)	(.076)	(.087)	(.131)	(.245)
$(r = 1)$	.005	.179	.080	.991	.425	−.181	−.079	−.022	−.405
		(.076)	(.082)	(.118)	(.245)	(.073)	(.083)	(.125)	(.235)
	−.014	.037	.047	.041	1.138	−.032	−.050	.033	−.186
		(.038)	(.041)	(.059)	(.122)	(.036)	(.042)	(.062)	(.117)

We have estimated the system with the Minnesota prior and different values of λ and θ. Some results for a VAR(2) process are given in Table 7.3 to illustrate the effect of the choice of the prior variance parameters λ and θ. For this particular data set, a combination $\lambda = 1$ and $\theta = .25$ leads to mild changes in the estimates only relative to unrestricted estimates ($\lambda = \infty, \theta = 1$). Decreasing θ has the effect of shrinking the off-diagonal elements towards zero. Thus, a small θ is reasonable if the variables are expected to be unrelated. The effect of a small θ is seen in Table 7.3 in the panel corresponding to $\lambda = 1$ and $\theta = .01$. On the other hand, lowering λ shrinks the diagonal elements of A_1 towards 1 and all other coefficients (except the intercept terms) towards zero. This effect is clearly observed for $\lambda = .01$, $\theta = .25$. Hence, if the analyst has a strong prior in favor of unrelated random walks, a small λ is appropriate.

In practice, one would usually choose a higher VAR order than 2 in a Bayesian analysis because chopping off the process at $p = 2$ implies a very strong prior with mean zero and variances zero for $A_3, A_4, \ldots$, which is a

Table 7.3. Bayesian estimates of the U.S. example system

prior	ν	A_1				A_2			
$\lambda = \infty$	.028	1.307	.106	−.554	−.814	−.318	−.101	.318	1.022
$\theta = 1$	.129	.080	1.045	−.177	.473	−.135	−.014	−.197	−.416
(unrestricted)	.096	.193	.068	.978	.284	−.248	−.035	.053	−.644
	.030	.042	.042	.034	1.065	−.064	−.027	.070	−.308
	.061	1.307	.021	−.514	−.465	−.331	−.009	.212	.679
$\lambda = 1$	.110	.060	1.088	−.173	.283	−.108	−.060	−.162	−.238
$\theta = .25$	.078	.119	.064	1.060	.025	−.167	−.034	−.069	−.316
	.029	.021	.029	.050	1.044	−.043	−.014	.031	−.265
	.083	1.550	.004	−.012	−.007	−.570	−.002	−.000	.004
$\lambda = 1$	−.015	.005	1.270	−.011	−.011	−.001	−.271	−.003	−.002
$\theta = .01$	−.032	−.003	.008	1.095	−.001	−.002	.002	−.216	−.001
	−.016	−.003	.004	.002	1.187	−.001	.001	.000	−.252
	−.045	1.009	.002	−.001	−.000	−.003	.000	−.000	.000
$\lambda = .01$	.018	.001	.999	−.001	−.002	.000	−.002	−.000	−.000
$\theta = .25$	−.004	.001	−.000	.993	−.001	.000	−.000	−.002	−.000
	−.003	.000	.000	.000	.994	.000	.000	−.000	−.002

bit unrealistic. The above analysis is just meant to illustrate the effect of the parameters that determine the prior variances. Also, if the variables are believed to be cointegrated, the Minnesota prior is not a good choice. It is more suited for a process which has a VAR representation in first differences because the basic idea underlying this prior is that the variables are roughly unrelated random walks. Notice, however, that for the present system, if a VECM with cointegration rank $r = 1$ and one lagged difference is fitted by ML and the corresponding levels VAR coefficients are determined via (7.2.10), the estimates in the lower part of Table 7.2 are obtained. If the system is actually cointegrated, the rank restriction should not lead to major distortions in the estimates. Therefore, it should not be surprising that the diagonal elements of the ML estimator of A_1 are again not far from 1. Thus, even if the variables are cointegrated, the Minnesota prior may not lead to substantial distortions. This property may explain why the prior has been used successfully in many applications, in particular, for forecasting (see Litterman (1986)).

7.5 Forecasting Estimated Integrated and Cointegrated Systems

As seen in Chapter 6, Section 6.5, forecasting integrated and cointegrated variables is conveniently discussed in the framework of the levels VAR representation of the data generation process. Therefore we consider a VAR(p) model,

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{7.5.1}$$

with integrated and possibly cointegrated variables. All symbols have their usual meanings (see Section 6.5). Deterministic terms are left out for convenience. Adding them is a straightforward exercise which is left to the reader.

Replacing the coefficients $A_1, \dots, A_p$, and the white noise covariance matrix Σ_u by estimators in the forecasting formulas of Section 6.5 creates similar problems as in the stationary, stable case considered in Chapter 3, Section 3.5. Denoting the h-step forecast based on estimated coefficients by $\widehat{y}_t(h)$ and indicating estimators by hats gives

$$\widehat{y}_t(h) = \widehat{A}_1 \widehat{y}_t(h-1) + \cdots + \widehat{A}_p \widehat{y}_t(h-p), \tag{7.5.2}$$

where $\widehat{y}_t(j) := y_{t+j}$ for $j \leq 0$. For this predictor, the forecast error becomes

$$\begin{aligned} y_{t+h} - \widehat{y}_t(h) &= [y_{t+h} - y_t(h)] + [y_t(h) - \widehat{y}_t(h)] \\ &= \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} + [y_t(h) - \widehat{y}_t(h)], \end{aligned} \tag{7.5.3}$$

where the last equality sign follows from Eq. (6.5.4) in Chapter 6. The last two terms in (7.5.3) are uncorrelated if parameter estimation is based on data up to period t only. In fact, under standard assumptions, the last term has zero probability limit, $y_t(h) - \widehat{y}_t(h) = o_p(1)$, as in the stationary case (see Problem 7.7). Thus, the forecast errors from estimated processes and processes with known coefficients are asymptotically equivalent. However, in the present case, the MSE correction for estimated processes derived in Section 3.5 is difficult to justify (see Problem 7.8 and Basu & Sen Roy (1987)). This problem must be kept in mind when forecast intervals are constructed. One possible MSE estimator is

$$\widehat{\Sigma}_y(h) = \sum_{i=0}^{h-1} \widehat{\Phi}_i \widehat{\Sigma}_u \widehat{\Phi}_i', \tag{7.5.4}$$

where the $\widehat{\Phi}_i$'s are obtained from the estimated A_i's by the recursions in (6.5.5) in Section 6.5. This estimator is likely to underestimate the true forecast uncertainty on average in small samples. Therefore, there is some danger that the confidence level of corresponding forecast intervals is overstated. Reimers (1991) derived a small sample correction especially for models with cointegrated variables and Engle & Yoo (1987) and Reinsel & Ahn (1992) reported on simulation studies in which imposing the cointegration restriction in the estimation gave better long-range forecasts than the use of unrestricted multivariate LS estimators.

7.6 Testing for Granger-Causality

7.6.1 The Noncausality Restrictions

In Section 6.6, we have seen that the restrictions characterizing Granger-noncausality are the same as in the stable case. If the levels VAR(p) representation (7.5.1) of the data generation process is considered again and the vector y_t is partitioned in M- and $(K-M)$-dimensional subvectors z_t and x_t,

$$y_t = \begin{bmatrix} z_t \\ x_t \end{bmatrix} \quad \text{and} \quad A_i = \begin{bmatrix} A_{11,i} & A_{12,i} \\ A_{21,i} & A_{22,i} \end{bmatrix}, \quad i = 1, \dots, p,$$

where the A_i are partitioned in accordance with the partitioning of y_t, then x_t does not Granger-cause z_t if and only if the hypothesis

$$H_0 : A_{12,i} = 0 \quad \text{for} \quad i = 1, \dots, p, \tag{7.6.1}$$

is true. Hence, we just have to test a set of linear restrictions. A Wald test is a standard choice for this purpose. In the present case, it may be problematic, however. We will discuss the potential problem next and then present a modification that has a limiting χ^2-distribution, as usual, and, hence, resolves the problem.

7.6.2 Problems Related to Standard Wald Tests

If the process is estimated by one of the procedures described in Section 7.2 such that the estimator $\widehat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha} := \operatorname{vec}[A_1, \dots, A_p]$ has the asymptotic distribution given in Corollary 7.1.1, then a Wald test can be conducted for the pair of hypotheses

$$H_0 : C\boldsymbol{\alpha} = 0 \quad \text{against} \quad H_1 : C\boldsymbol{\alpha} \neq 0. \tag{7.6.2}$$

Here C is an $(N \times pK^2)$ matrix of rank N. The relevant Wald statistic is

$$\lambda_W = T\widehat{\boldsymbol{\alpha}}'C'(C\widehat{\Sigma}^{\text{co}}_{\boldsymbol{\alpha}}C')^{-1}C\widehat{\boldsymbol{\alpha}}, \tag{7.6.3}$$

where $\widehat{\Sigma}^{\text{co}}_{\boldsymbol{\alpha}}$ is a consistent estimator of $\Sigma^{\text{co}}_{\boldsymbol{\alpha}}$. The statistic λ_W has an asymptotic $\chi^2(N)$-distribution, provided the null hypothesis is true and

$$\operatorname{rk}(C\widehat{\Sigma}^{\text{co}}_{\boldsymbol{\alpha}}C') = \operatorname{rk}(C\Sigma^{\text{co}}_{\boldsymbol{\alpha}}C') = N. \tag{7.6.4}$$

This result follows from standard asymptotic theory (see Appendix C.7). We have chosen to state it here again because the rank condition (7.6.4) now becomes important. It is automatically satisfied for stable, full VAR processes as discussed in Chapter 3, because in that case the asymptotic covariance matrix of the coefficient estimator is nonsingular. Now, however, $\Sigma^{\text{co}}_{\boldsymbol{\alpha}}$ is singular if the cointegration rank r is less than K (see Corollary 7.1.1). Therefore, it is possible in principle that $\operatorname{rk}(C\Sigma^{\text{co}}_{\boldsymbol{\alpha}}C') < N$, even if C has full row rank N.

A limiting χ^2-distribution of λ_W can also be obtained if the inverse of $C\widehat{\Sigma}^{\text{co}}_{\boldsymbol{\alpha}}C'$ in (7.6.3) is replaced by a generalized inverse. In that case, the asymptotic distribution of λ_W is $\chi^2(\operatorname{rk}(C\Sigma^{\text{co}}_{\boldsymbol{\alpha}}C'))$ if

$$\operatorname{rk}(C\widehat{\Sigma}^{\text{co}}_{\boldsymbol{\alpha}}C') = \operatorname{rk}(C\Sigma^{\text{co}}_{\boldsymbol{\alpha}}C') \tag{7.6.5}$$

with probability one (see Andrews (1987)). Unfortunately, the latter condition will not hold in general. In particular, if a cointegrated system is estimated in unconstrained form by multivariate LS and if $\Sigma_{\boldsymbol{\alpha}}^{\text{co}}$ is estimated as in Corollary 7.1.1, $C\widehat{\Sigma}_{\boldsymbol{\alpha}}^{\text{co}}C'$ has rank N with probability 1, while $\text{rk}(C\Sigma_{\boldsymbol{\alpha}}^{\text{co}}C')$ may be less than N. Andrews (1987) showed that in such a case the asymptotic distribution of λ_W may not even be χ^2. A detailed analysis of the problem for the particular case of testing for Granger-causality in cointegrated systems was provided by Toda & Phillips (1993). In this context, it is perhaps worth pointing out that the equality in (7.6.5) may not hold, even if the cointegration rank has been specified correctly and the corresponding restrictions have been imposed in the estimation procedure (see Problem 7.9). For the hypothesis of interest here, a possible solution to the problem was proposed by Dolado & Lütkepohl (1996) and Toda & Yamamoto (1995). It will be presented next. Our discussion follows the former article.

Another possible approach to overcome inference problems in levels VARs with integrated variables was described by Phillips (1995). It is known as *fully modified VAR estimation* and is based on nonparametric corrections. Some of its drawbacks are pointed out by Kauppi (2004).

7.6.3 A Wald Test Based on a Lag Augmented VAR

As discussed in Section 7.2 (see in particular Section 7.2.1), the estimators of coefficients attached to stationary regressors converge at the usual $T^{1/2}$ rate to a nonsingular normal distribution. Therefore, the problem of the previous subsection can be solved if the model can be rewritten in such a way that all parameters under test are attached to stationary regressors. To this end, the following reparameterization is helpful:

$$\begin{aligned} y_t &= \sum_{j=1, j\neq i}^{p} A_j y_{t-j} + A_i y_{t-i} + u_t \\ &= \sum_{j=1, j\neq i}^{p} A_j (y_{t-j} - y_{t-i}) + \left(\sum_{j=1}^{p} A_j \right) y_{t-i} + u_t. \end{aligned}$$

Defining a differencing operator Δ_k such that $\Delta_k y_t = y_t - y_{t-k}$ for $k = \pm 1, \pm 2, \ldots$, the model can be written as

$$\Delta_i y_t = \sum_{j=1, j\neq i}^{p} A_j \Delta_{i-j} y_{t-j} + \mathbf{\Pi} y_{t-i} + u_t, \tag{7.6.6}$$

where $\mathbf{\Pi} = -(I_K - A_1 - \cdots - A_p)$, as usual. For $k > 0$, $\Delta_k y_t = (y_t - y_{t-1}) + (y_{t-1} - y_{t-2}) + \cdots + (y_{t-k+1} - y_{t-k})$ is stationary as the sum of stationary processes and the same is easily seen to hold for $k < 0$. Therefore, it follows from the previously mentioned results in Section 7.2 that the LS estimators of the A_j, $j \neq i$, have a nonsingular joint asymptotic normal distribution.

Notice that these estimators are, of course, identical to those based on the levels VAR model (7.5.1) because we have just reparameterized the model. Hence, the following proposition from Dolado & Lütkepohl (1996, Theorem 1) is obtained.

Proposition 7.8 (*Asymptotic Distribution of the Wald Statistic*)
Let y_t be a K-dimensional $I(1)$ process generated by the VAR(p) process in (7.5.1) and denote the LS estimator of A_i by $\widehat{A}_i$ ($i = 1, \dots, p$). Moreover, let $\boldsymbol{\alpha}_{(-i)}$ be a $K^2(p-1)$-dimensional vector obtained by deleting A_i from $[A_1, \dots, A_p]$ and vectorizing the remaining matrix. Analogously, let $\widehat{\boldsymbol{\alpha}}_{(-i)}$ be a $K^2(p-1)$-dimensional vector obtained by deleting $\widehat{A}_i$ from $[\widehat{A}_1, \dots, \widehat{A}_p]$ and vectorizing the remainder. Then

$$\sqrt{T}(\widehat{\boldsymbol{\alpha}}_{(-i)} - \boldsymbol{\alpha}_{(-i)}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\boldsymbol{\alpha}_{(-i)}}), \tag{7.6.7}$$

where the $(K^2(p-1) \times K^2(p-1))$ covariance matrix $\Sigma_{\boldsymbol{\alpha}_{(-i)}}$ is nonsingular and the Wald statistic λ_W for testing $H_0 : C\boldsymbol{\alpha}_{(-i)} = 0$ has a limiting $\chi^2(N)$-distribution, that is,

$$\lambda_W = T\widehat{\boldsymbol{\alpha}}'_{(-i)}C'(C\widehat{\Sigma}_{\boldsymbol{\alpha}_{(-i)}}C')^{-1}C\widehat{\boldsymbol{\alpha}}_{(-i)} \xrightarrow{d} \chi^2(N)$$

under H_0. Here C is an $(N \times K^2(p-1))$ matrix with $\text{rk}(C) = N$ and $\widehat{\Sigma}_{\boldsymbol{\alpha}_{(-i)}}$ is a consistent estimator of $\Sigma_{\boldsymbol{\alpha}_{(-i)}}$. ■

Note that

$$\Sigma_{\boldsymbol{\alpha}_{(-i)}} = \text{plim}\, T(X_{(-i)}X'_{(-i)})^{11} \otimes \Sigma_u,$$

where $X_{(-i)} = [X_0^{(-i)}, \dots, X_{T-1}^{(-i)}]$ with

$$X_{t-1}^{(-i)} = \begin{bmatrix} \Delta_{i-1}y_{t-1} \\ \vdots \\ \Delta_{i-p}y_{t-p} \\ y_{t-i} \end{bmatrix} \quad (K^2p \times 1)$$

and $(X_{(-i)}X'_{(-i)})^{11}$ denotes the upper left-hand $(K^2(p-1) \times K^2(p-1))$ dimensional submatrix of $(X_{(-i)}X'_{(-i)})^{-1}$. Thereby a consistent estimator of $\Sigma_{\boldsymbol{\alpha}_{(-i)}}$ is obtained as

$$\widehat{\Sigma}_{\boldsymbol{\alpha}_{(-i)}} = T(X_{(-i)}X'_{(-i)})^{11} \otimes \widehat{\Sigma}_u,$$

where $\widehat{\Sigma}_u$ is the residual covariance matrix obtained from the LS residuals.

Proposition 7.8 shows that, whenever the elements in at least one of the complete coefficient matrices A_i are not restricted under H_0, the Wald statistic has its usual asymptotic χ^2-distribution. In other words, if restrictions are placed on all A_i's, $i = 1, \dots, p$, as in the noncausality hypothesis (7.6.1), we can get a χ^2 Wald test by adding an extra lag in estimating the parameters of the process. If the true data generation process is a VAR(p), then a VAR($p+1$) with $A_{p+1} = 0$ is also a correct model. Because we know that $A_{p+1} = 0$, the causality test can be based on the estimator $\widehat{\boldsymbol{\alpha}}_{(-(p+1))}$, that is, an estimator of the first K^2p elements of vec$[\widehat{A}_1, \dots, \widehat{A}_{p+1}]$. Notice that LS estimation may be applied to the levels VAR($p+1$) model. To carry out the causality test, it is not necessary to actually perform the reparameterization of the process in (7.6.6) because the LS estimators of the A_j matrices do not change due to the reparameterization. Also, the covariance matrix of the asymptotic distribution may be estimated as usual from the levels VAR($p+1$).

We do not have to know the cointegration properties of the system to use this lag augmentation test procedure. Of course, there may be a loss of power due to over-specifying the lag length. The loss in power may not be substantial if the true order p is large and the dimension K is small or moderate, because, in this case, the relative reduction in the estimation precision due to one extra VAR coefficient matrix may be small. On the other hand, if the true order is small and K is large, an extra lag of all variables may lead to a sizeable decline in overall estimation precision and, hence, in the power of the modified Wald test. There are in fact cases, where the extra lag is not necessary to obtain the asymptotic χ^2-distribution of the Wald test for Granger-causality. For example, for bivariate processes with cointegrating rank 1, no extra lag is needed, if both variables are $I(1)$ (e.g., Lütkepohl & Reimers (1992a)).

Proposition 7.8 remains valid if deterministic terms are included in the VAR model. This result follows from the discussion in Section 7.2 because including such terms leaves the asymptotic properties of the VAR coefficients unaffected. It may also be of interest that a similar result can be obtained for VAR systems with $I(d)$ variables where $d > 1$. In that case, d coefficient matrices A_i must be unrestricted under H_0 (see Dolado & Lütkepohl (1996)). Alternatively, d lags must be added if all parameter matrices of the original process are restricted. This result can also be obtained from Sims et al. (1990).

7.6.4 An Example

We follow again Lütkepohl (2004) and use the German interest rate/inflation example to illustrate causality testing for cointegrated variables. The data generation process is assumed to be a VAR(4). The model is augmented by one lag and, hence, a VAR(5) is fitted and used in the actual tests for Granger-causality, while a VAR(4) is used for testing instantaneous causality. The results are given in Table 7.4, where F-versions of the Granger-causality test statistics are reported. The asymptotic χ^2-distribution is often a poor approximation to the small sample distribution of the causality test statistics. Therefore, an F-version is preferred which is obtained in the usual way by dividing the χ^2-statistic by its degrees of freedom parameter (see Section 3.6). As in Section 3.6, the test for instantaneous causality is based on the residual covariance matrix. This approach is justified by Lemma 7.3 which shows that the asymptotic distribution of the usual residual covariance matrix estimator is the same as in the stationary case. Hence, the same test for instantaneous causality can be used under normality assumptions.

Table 7.4. Tests for causality between German interest rate and inflation

causality hypothesis	test value	distribution	p-value
R Granger-causal for Dp	2.24	$F(4, 152)$	0.07
Dp Granger-causal for R	0.31	$F(4, 152)$	0.87
R and Dp instantaneously causal	0.61	$\chi^2(1)$	0.44

None of the p-values in Table 7.4 is smaller than 0.05. Therefore, none of the noncausality hypotheses can be rejected at the 5% significance level. Given the subset model (7.3.9), this outcome is somewhat surprising because there are clearly significant estimates in that model. Of course, using the present tests is a different way of looking at the data than considering the individual coefficients in the subset model. The relatively large number of parameters in the presently considered unrestricted model which even includes an extra lag, makes it difficult for the sample information to clearly distinguish the sets of parameters from their values specified in the null hypothesis.

The insignificant value of the test for instantaneous causality is not surprising, however. The correlation matrix corresponding to the covariance matrix in (7.3.9) is

$$\begin{bmatrix} 1 & -0.01 \\ -0.01 & 1 \end{bmatrix}.$$

Thus, the instantaneous correlation between the two residual series is very small. This property is reflected in the test result in Table 7.4.

7.7 Impulse Response Analysis

In Section 6.7, we have seen that, in principle, impulse response analysis in cointegrated systems can be conducted in the same way as for stationary systems. If estimated processes are used, the asymptotic properties of the impulse response coefficients and forecast error variance components follow from Proposition 3.6 in conjunction with Corollary 7.1.1. In other words, the relevant covariance matrices $\Sigma_{\widehat{\boldsymbol{\alpha}}}$ and $\Sigma_{\widehat{\boldsymbol{\sigma}}}$ have to be used in Proposition 3.6. Of course, the remarks on Proposition 3.6 regarding the estimation of standard errors etc. apply for the present case too. In practice, confidence intervals for impulse responses are typically computed with bootstrap methods.

To illustrate the impulse response analysis we use again our German interest rate/inflation example system. We have performed an impulse response analysis on the basis of the subset VECM (7.3.9) and show forecast error impulse responses with bootstrap confidence intervals determined by Hall's percentile method (see Appendix D.3) in Figure 7.3. Using forecast error impulse responses is unproblematic here because no instantaneous causality and no significant instantaneous correlation between the two residual series was diagnosed in Section 7.6.4. The point estimates of the impulse responses look very much like those in Figure 6.4 in Chapter 6. This similarity is not surprising because the model assumed in that chapter is very similar to the present one. Because the variables are integrated of order one, the impulses have permanent effects. This conclusion can be defended even if the estimation uncertainty is taken into account.

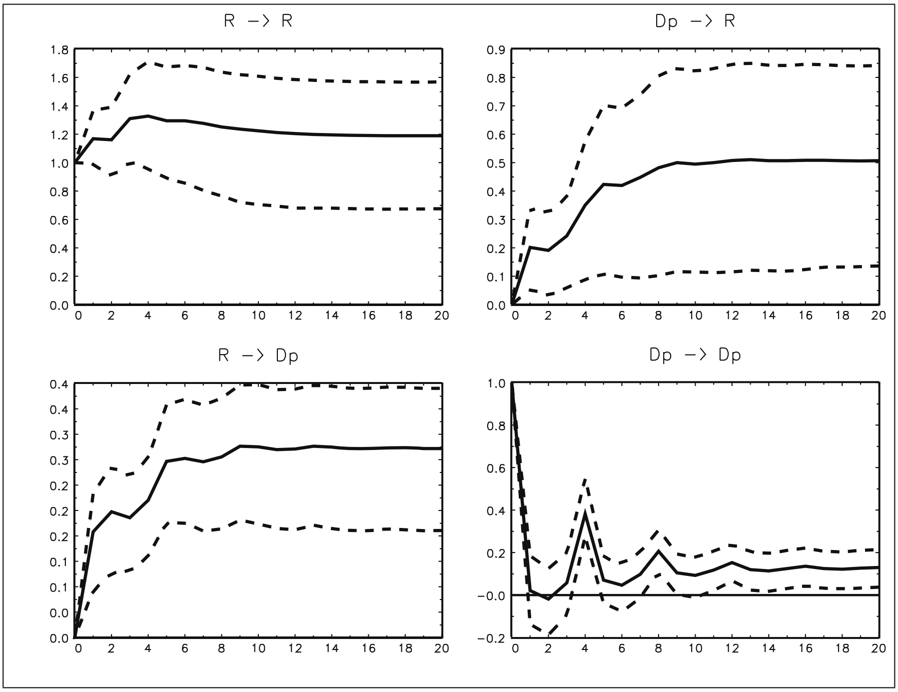

Fig. 7.3. Forecast error impulse responses for model (7.3.9) with 95% Hall percentile bootstrap confidence intervals based on 2000 bootstrap replications.

We emphasize again that an uncritical impulse response analysis is problematic. In particular, different sets of impulse responses exist and it is not clear which one properly reflects the actual reactions of the variables. The caveats of impulse response analysis are discussed in Sections 2.3 and 3.7. They are therefore not repeated here. We will return to impulse response analysis in Chapter 9, when structural restrictions are discussed for identifying meaningful shocks.

7.8 Exercises

7.8.1 Algebraic Exercises

Problem 7.1
Show that, in the proof of Result 6 of Section 7.1,

$$T^{-1}\sum_{t=1}^{T}(u_t^* - u_t)y_{t-1}^{(2)\prime} = o_p(1).$$

$\Big($Hint: Use

$$T^{-1}\sum_{t=1}^{T}(u_t^* - u_t)y_{t-1}^{(2)\prime} = (\widehat{\alpha} - \alpha)T^{-1}\sum_{t=1}^{T}\beta' y_{t-1}y_{t-1}^{(2)\prime}.\Big)$$

Problem 7.2
Prove Proposition 7.1 based on the ideas presented in Section 7.2.1. (Hint: See Ahn & Reinsel (1990).)

Problem 7.3
Prove that $\sqrt{T}[\widetilde{\alpha}\widetilde{\beta}' - \widetilde{\alpha}(\beta)\beta'] = o_p(1)$ holds in the proof of Lemma 7.3. $\big($Hint: note that

$$\widetilde{\alpha}\widetilde{\beta}' - \widetilde{\alpha}(\beta)\beta' = \widetilde{\alpha}[\widetilde{\beta}' - \beta'] + [\widetilde{\alpha} - \widetilde{\alpha}(\beta)]\beta'.\big)$$

Problem 7.4
Determine the ML estimators in a cointegrated VAR(p) process with cointegration rank r, under the assumption that the cointegration matrix satisfies restrictions $\beta = H\varphi$, where H and φ are $(K \times s)$ and $(s \times r)$ matrices, respectively, with $r < s < K$. (Hint: Proceed as in the proof of Proposition 7.3.)

Problem 7.5
Show that the expressions in (7.2.27) and (7.2.28) are the LS estimators of α and $\boldsymbol{\Gamma}$, respectively, conditional on $\beta = \widehat{\beta}$.

Problem 7.6
Derive the EGLS estimator for restrictions of the form $\text{vec}[\alpha : \boldsymbol{\Gamma}] = \Re\boldsymbol{\varphi} + \mathbf{r}$ on the short-run parameters of the VECM (7.2.1) and state its asymptotic distribution (see (7.3.8) for the definition of the notation).

Problem 7.7
Consider a cointegrated VAR(1) process without intercept, $y_t = A_1 y_{t-1} + u_t$, and show that

$$\text{plim}\ [y_T(1) - \widehat{y}_T(1)] = \text{plim}\ (A_1 - \widetilde{A}_1)y_T = 0.$$

Assume that y_t is Gaussian with initial vector $y_0 = 0$ and the ML estimator $\widetilde{A}_1$ is based on $y_1, \dots, y_T$. (Hint: Use Lemma 7.2 and plim $y_T/T = 0$ from Phillips & Durlauf (1986).)

Problem 7.8
Consider the matrix $\Omega(h)$ used in the MSE correction in Section 3.5 and argue why it is problematic for unstable processes. Analyze in particular the derivation in (3.5.12).

Problem 7.9
Consider a three-dimensional VAR(1) process with cointegration rank 1 and suppose the cointegrating matrix has the form $\beta = (\beta_1, \beta_2, 0)'$. Use Corollary 7.1.1 to demonstrate that the elements in the last column of A_1 have zero asymptotic variances. Formulate a linear hypothesis for the coefficients of A_1 for which the rank condition (7.6.4) is likely to be violated if the covariance estimator of Corollary 7.1.1 is used.

7.8.2 Numerical Exercises

The following problems are based on the U.S. data given in File E3 and described in Section 7.4.3. The variables are defined as in that subsection.

Problem 7.10
Apply the ML procedure described in Section 7.2.3 to estimate a VAR(3) process with cointegration rank $r = 1$ and intercept vector. Determine the estimates $\widetilde{\nu}, \widetilde{A}_1, \widetilde{A}_2$, and $\widetilde{A}_3$ and compare them to unrestricted LS estimates of a VAR(3) process.

Problem 7.11
Compute forecasts up to 10 periods ahead using both the unrestricted VAR(3) model and the VAR(3) model with cointegration rank 1. Compare the forecasts.

Problem 7.12
Compare the impulse responses obtained from an unrestricted and restricted VAR(3) model with cointegration rank 1.