Cointegration

16

As already mentioned in Chap. 7, many raw economic time series are nonstationary and become stationary only after some transformation. The most common of these transformations is the formation of differences, perhaps after having taken logs. In most cases first differences are sufficient to achieve stationarity. The stationarized series can then be analyzed in the context of VAR models as explained in the previous chapters. However, many economic theories are formalized in terms of the original series so that we may want to use the VAR methodology to infer also the behavior of the untransformed series. Yet, by taking first differences we loose probably important information on the levels. Thus, it seems worthwhile to develop an approach which allows us to take the information on the levels into account and at the same time take care of the nonstationary character of the variables. The concept of *cointegration* tries to achieve this double requirement.

In the following we will focus our analysis on variables which are integrated of order one, i.e. on time series which become stationary after having taken first differences. However, as we have already mentioned in Sect. 7.5.1, a regression between integrated variables may lead to spurious correlations which make statistical inferences and interpretations of the estimated coefficients a delicate issue (see Sect. 7.5.3). A way out of this dilemma is presented by the theory of *cointegrated processes*. Loosely speaking, a multivariate process is cointegrated if there exists a linear combination of the processes which is stationary although each process taken individually may be integrated. In many cases, this linear combination can be directly related to economic theory which has made the analysis of cointegrated processes an important research topic. In the bivariate case, already been dealt with in Sect. 7.5.2, the cointegrating relation can be immediately read off from the cointegrating regression and the cointegration test boils down to a unit root test for the residuals of the cointegrating regression. However, if more than two variables are involved, the single equation residual based test is, as explained in Sect. 7.5.2, no longer satisfactory. Thus, a genuine multivariate is desirable.

K. Neusser, *Time Series Econometrics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-319-32862-1_16

The concept of cointegration goes back to the work of Engle and Granger (1987) which is itself based on the precursor study of Davidson et al. (1978). In the meantime the literature has grown tremendously. Good introductions can be found in Banerjee et al. (1993), Watson (1994) or Lütkepohl (2006). For the more statistically inclined reader Johansen (1995) is a good reference.

16.1 A Theoretical Example

Before we present the general theory of cointegration within the VAR context, it is instructive to introduce the concept in the well-known class of present discounted value models. These models relate some variable X_t to present discounted value of another variable Y_t:

$$X_t = \gamma(1-\beta)\sum_{j=0}^{\infty}\beta^j \mathbb{P}_t Y_{t+j} + u_t, \qquad 0 < \beta < 1,$$

where $u_t \sim \text{WN}(0, \sigma_u^2)$ designates a preference shock. Thereby, β denotes the subjective discount factor and γ is some unspecified parameter. The present discounted value model states that the variable X_t is proportional to the sum of future Y_{t+j}, $j = 0, 1, 2, \ldots$, discounted by the factor β. We can interpret X_t and Y_t as the price and the dividend of a share, as the interest rate on long- and short-term bonds, or as consumption and income. In order to operationalize this model, we will assume that forecasts are computed as linear mean-squared error forecasts. The corresponding forecast operator is denoted by $\mathbb{P}_t$. Furthermore, we will assume that the forecaster observes Y_t and its past $Y_{t-1}, Y_{t-2}, \ldots$ The goal of the analysis is the investigation of the properties of the bivariate process $\{(X_t, Y_t)'\}$. The analysis of this important class models presented below is based on Campbell and Shiller (1987).[1]

The model is closed by assuming some specific time series model for $\{Y_t\}$. In this example, we will assume that $\{Y_t\}$ is an integrated process of order one (see Definition 7.1 in Sect. 7.1) such that $\{\Delta Y_t\}$ follows an AR(1) process:

$$\Delta Y_t = \mu(1-\phi) + \phi \Delta Y_{t-1} + v_t, \qquad |\phi| < 1 \text{ and } v_t \sim \text{WN}(0, \sigma_v^2).$$

This specification of the $\{Y_t\}$ process implies that $\mathbb{P}_t \Delta Y_{t+h} = \mu(1-\phi^h) + \phi^h \Delta Y_t$. Because $\mathbb{P}_t Y_{t+h} = \mathbb{P}_t \Delta Y_{t+h} + \ldots + \mathbb{P}_t \Delta Y_{t+1} + Y_t$, $h = 0, 1, 2, \ldots$, the present discounted value model can be manipulated to give:

[1] A more recent interesting application of this model is given by the work of Beaudry and Portier (2006).

$$
\begin{aligned}
X_t &= \gamma(1-\beta)\left[Y_t + \beta\mathbb{P}_t Y_{t+1} + \beta^2\mathbb{P}_t Y_{t+2} + \ldots\right] + u_t \\
&= \gamma(1-\beta)[\ Y_t \\
&\quad + \beta\ Y_t + \beta\ \mathbb{P}_t\Delta Y_{t+1} \\
&\quad + \beta^2 Y_t + \beta^2\mathbb{P}_t\Delta Y_{t+1} + \beta^2\mathbb{P}_t\Delta Y_{t+2} \\
&\quad + \beta^3 Y_t + \beta^3\mathbb{P}_t\Delta Y_{t+1} + \beta^3\mathbb{P}_t\Delta Y_{t+2} + \beta^3\mathbb{P}_t\Delta Y_{t+3} \\
&\quad + \ldots] + u_t \\
&= \gamma(1-\beta)\left[\frac{1}{1-\beta}Y_t + \frac{\beta}{1-\beta}\mathbb{P}_t\Delta Y_{t+1} + \frac{\beta^2}{1-\beta}\mathbb{P}_t\Delta Y_{t+2} + \ldots\right] + u_t
\end{aligned}
$$

This expression shows that the integratedness of $\{Y_t\}$ is transferred to $\{X_t\}$. Bringing Y_t to the left we get the following expression:

$$S_t = X_t - \gamma Y_t = \gamma\sum_{j=1}^{\infty}\beta^j\mathbb{P}_t\Delta Y_{t+j} + u_t.$$

The variable S_t is occasionally referred to as the spread. If γ is greater than zero, expected increases in ΔY_{t+j}, $j \geq 1$, have a positive impact on the spread today. For $\gamma = 1$, S_t can denote the log of the price-dividend ratio of a share, or minus the logged savings ratio as in the permanent income model of Campbell (1987). If investors expect positive (negative) change in the dividends tomorrow, they want to buy (sell) the share thereby increasing (decreasing) its price already today. In the context of the permanent income hypothesis expected positive income changes lead to a reduction in today's saving rate. If, on the contrary, households expect negative income changes to occur in the future, they will save already today ("saving for the rainy days").

Inserting for $\mathbb{P}_t\Delta Y_{t+j}$, $j = 0, 1, \ldots$, the corresponding forecast equation $\mathbb{P}_t\Delta Y_{t+h} = \mu(1-\phi^h) + \phi^h\Delta Y_t$, we get:

$$S_t = \frac{\beta\gamma\mu(1-\phi)}{(1-\beta)(1-\beta\phi)} + \frac{\beta\gamma\phi}{1-\beta\phi}\Delta Y_t + u_t.$$

The remarkable feature about this relation is that $\{S_t\}$ is a stationary process because both $\{\Delta Y_t\}$ and $\{u_t\}$ are stationary, despite the fact that $\{Y_t\}$ and $\{X_t\}$ are both integrated processes of order one. The mean of S_t is:

$$\mathbb{E}S_t = \frac{\beta\gamma\mu}{1-\beta}.$$

From the relation between S_t and ΔY_t and the AR(1) representation of $\{\Delta Y_t\}$ we can deduce a VAR representation of the joint process $\{(S_t, \Delta Y_t)'\}$:

$$\begin{pmatrix} S_t \\ \Delta Y_t \end{pmatrix} = \mu(1-\phi)\begin{pmatrix} \frac{\beta\gamma}{(1-\beta)(1-\beta\phi)} + \frac{\beta\gamma\phi}{1-\beta\phi} \\ 1 \end{pmatrix} + \begin{pmatrix} 0 & \frac{\beta\gamma\phi^2}{1-\beta\phi} \\ 0 & \phi \end{pmatrix}\begin{pmatrix} S_{t-1} \\ \Delta Y_{t-1} \end{pmatrix} + \begin{pmatrix} u_t + \frac{\beta\gamma\phi}{1-\beta\phi}v_t \\ v_t \end{pmatrix}.$$

Further algebraic transformation lead to a VAR representation of order two for the level variables $\{(X_t, Y_t)'\}$:

$$\begin{aligned} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} &= c + \Phi_1\begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \Phi_2\begin{pmatrix} X_{t-2} \\ Y_{t-2} \end{pmatrix} + Z_t \\ &= \mu(1-\phi)\begin{pmatrix} \frac{\gamma}{(1-\beta)(1-\beta\phi)} \\ 1 \end{pmatrix} + \begin{pmatrix} 0 & \gamma + \frac{\gamma\phi}{1-\beta\phi} \\ 0 & 1+\phi \end{pmatrix}\begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} \\ &\quad + \begin{pmatrix} 0 & \frac{-\gamma\phi}{1-\beta\phi} \\ 0 & -\phi \end{pmatrix}\begin{pmatrix} X_{t-2} \\ Y_{t-2} \end{pmatrix} + \begin{pmatrix} 1 & \frac{\gamma}{1-\beta\phi} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} u_t \\ v_t \end{pmatrix} \end{aligned}$$

Next we want to check whether this stochastic difference equation possesses a stationary solution. For this purpose, we must locate the roots of the equation $\det \Phi(z) = \det(I_2 - \Phi_1 z - \Phi_2 z^2) = 0$ (see Theorem 12.1). As

$$\det \Phi(z) = \det\begin{pmatrix} 1 & \left(-\gamma - \frac{\gamma\phi}{1-\beta\phi}\right)z + \frac{\gamma\phi}{1-\beta\phi}z^2 \\ 0 & 1-(1+\phi)z+\phi z^2 \end{pmatrix} = 1-(1+\phi)z+\phi z^2,$$

the roots are $z_1 = 1/\phi$ and $z_2 = 1$. Thus, only the root z_1 lies outside the unit circle whereas the root z_2 lies on the unit circle. The existence of a *unit root* precludes the existence of a stationary solution. Note that we have just one unit root, although each of the two processes taken by themselves are integrated of order one.

The above VAR representation can be further transformed to yield a representation of process in first differences $\{(\Delta X_t, \Delta Y_t)'\}$:

$$\begin{aligned} \begin{pmatrix} \Delta X_t \\ \Delta Y_t \end{pmatrix} &= \mu(1-\phi)\begin{pmatrix} \frac{\gamma}{(1-\beta)(1-\beta\phi)} \\ 1 \end{pmatrix} - \begin{pmatrix} 1 & -\gamma \\ 0 & 0 \end{pmatrix}\begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} \\ &\quad + \begin{pmatrix} 0 & \frac{\gamma\phi}{1-\beta\phi} \\ 0 & \phi \end{pmatrix}\begin{pmatrix} \Delta X_{t-1} \\ \Delta Y_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & \frac{\gamma}{1-\beta\phi} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} u_t \\ v_t \end{pmatrix}. \end{aligned}$$

This representation can be considered as a generalization of the Dickey-Fuller regression in first difference form (see Eq. (7.1)). In the multivariate case, it is known as the vector error correction model (VECM) or vector error correction representation. In this representation the matrix

$$\Pi = -\Phi(1) = \begin{pmatrix} -1 & \gamma \\ 0 & 0 \end{pmatrix}$$

is of special importance. This matrix is singular and of rank one. This is not an implication which is special to this specification of the present discounted value model, but arises generally as shown in Campbell (1987) and Campbell and Shiller (1987). In the VECM representation all variables except $(X_{t-1}, Y_{t-1})'$ are stationary by construction. This implies that $-\Pi(X_{t-1}, Y_{t-1})'$ must be stationary too, despite the fact that $\{(X_t, Y_t)'\}$ is not stationary as shown above. Multiplying $-\Pi(X_{t-1}, Y_{t-1})'$ out, one obtains two linear combinations which define stationary processes. However, as Π has only rank one, there is just one linearly independent combination. The first one is $X_{t-1} - \gamma Y_{t-1}$ and equals S_{t-1} which was already shown to be stationary. The second one is degenerate because it yields zero. The phenomenon is called *cointegration*.

Because Π has rank one, it can be written as the product of two vectors α and β:

$$\Pi = \alpha\beta' = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ -\gamma \end{pmatrix}'.$$

Clearly, this decomposition of Π is not unique because $\tilde{\alpha} = a\alpha$ and $\tilde{\beta} = a^{-1}\beta$, $a \neq 0$, would also qualify for such a decomposition as $\Pi = \tilde{\alpha}\tilde{\beta}'$. The vector β is called a *cointegration vector*. It has the property that $\{\beta'(X_t, Y_t)'\}$ defines a stationary process despite the fact that $\{(X_t, Y_t)'\}$ is non-stationary. The cointegration vector thus defines a linear combination of X_t and Y_t which is stationary. The matrix α, here only a vector, is called the *loading matrix* and its elements the *loading coefficients*.

The VAR and the VECM representations are both well suited for estimation. However, if we want to compute the impulse responses, we need a causal representation. Such a causal representation does not exist due to the unit root in the VAR process for $\{(X_t, Y_t)'\}$ (see Theorem 12.1). To circumvent this problem we split the matrix $\Phi(z)$ into the product of two matrices $M(z)$ and $V(z)$. $M(z)$ is a diagonal matrix which encompasses all unit roots on its diagonal. $V(z)$ has all its roots outside the unit circle so that $V^{-1}(z)$ exists for $|z| < 1$. In our example, we get:

$$\begin{aligned} \Phi(z) &= M(z)V(z) \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1-z \end{pmatrix} \begin{pmatrix} 1 & \left(-\gamma - \frac{\gamma\phi}{1-\beta\phi}\right) z + \frac{\gamma\phi}{1-\beta\phi} z^2 \\ 0 & 1-\phi z \end{pmatrix}. \end{aligned}$$

Multiplying $\Phi(z)$ with $\widetilde{M}(z) = \begin{pmatrix} 1-z & 0 \\ 0 & 1 \end{pmatrix}$ from the left, we find:

$$\widetilde{M}(z)\Phi(z) = \widetilde{M}(z)M(z)V(z) = (1-z)I_2V(z) = (1-z)V(z).$$

The application of this result to the VAR representation of $\{(X_t, Y_t)\}$ leads to a causal representation of $\{(\Delta X_t, \Delta Y_t)\}$:

$$\begin{aligned}
\Phi(\mathrm{L})\begin{pmatrix} X_t \\ Y_t \end{pmatrix} &= M(\mathrm{L})V(\mathrm{L})\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = c + Z_t \\
\widetilde{M}(\mathrm{L})\Phi(\mathrm{L})\begin{pmatrix} X_t \\ Y_t \end{pmatrix} &= (1-\mathrm{L})V(\mathrm{L})\begin{pmatrix} X_t \\ Y_t \end{pmatrix} \\
&= \mu(1-\phi)\begin{pmatrix} 1-\mathrm{L} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \frac{\gamma}{(1-\beta)(1-\beta\phi)} \\ 1 \end{pmatrix} + \begin{pmatrix} 1-\mathrm{L} & 0 \\ 0 & 1 \end{pmatrix} Z_t \\
V(\mathrm{L})\begin{pmatrix} \Delta X_t \\ \Delta Y_t \end{pmatrix} &= \begin{pmatrix} 0 \\ \mu(1-\phi) \end{pmatrix} + \begin{pmatrix} 1-\mathrm{L} & 0 \\ 0 & 1 \end{pmatrix} Z_t \\
\begin{pmatrix} \Delta X_t \\ \Delta Y_t \end{pmatrix} &= \mu\begin{pmatrix} \gamma \\ 1 \end{pmatrix} + V^{-1}(\mathrm{L})\begin{pmatrix} 1-\mathrm{L} & 0 \\ 0 & 1 \end{pmatrix} Z_t \\
\begin{pmatrix} \Delta X_t \\ \Delta Y_t \end{pmatrix} &= \mu\begin{pmatrix} \gamma \\ 1 \end{pmatrix} + \Psi(\mathrm{L})Z_t.
\end{aligned}$$

The polynomial matrix $\Psi(\mathrm{L})$ can be recovered by the method of undetermined coefficients from the relation between $V(\mathrm{L})$ and $\Psi(\mathrm{L})$:

$$V(\mathrm{L})\Psi(\mathrm{L}) = \begin{pmatrix} 1-\mathrm{L} & 0 \\ 0 & 1 \end{pmatrix}$$

In this exposition, we abstain from the explicit computation of $V^{-1}(\mathrm{L})$ and $\Psi(\mathrm{L})$. However, the following relation holds:

$$V(1) = \begin{pmatrix} 1 & -\gamma \\ 0 & 1-\phi \end{pmatrix} \Longrightarrow V^{-1}(1) = \begin{pmatrix} 1 & \frac{\gamma}{1-\phi} \\ 0 & \frac{1}{1-\phi} \end{pmatrix}.$$

Implying that

$$\Psi(1) = V^{-1}(1)\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = (1-\phi)^{-1}\begin{pmatrix} 0 & \gamma \\ 0 & 1 \end{pmatrix}.$$

The cointegration vector $\beta = (1, -\gamma)'$ and loading matrix $\alpha = (-1, 0)'$ therefore have the following properties:

$$\beta'\Psi(1) = \begin{pmatrix} 0 & 0 \end{pmatrix} \quad \text{and} \quad \Psi(1)\alpha = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Like in the univariate case (see Theorem 7.1 in Sect. 7.1.4), we can also construct the Beveridge-Nelson decomposition in the multivariate case. For this purpose, we decompose $\Psi(\mathrm{L})$ as follows:

$$\Psi(\mathrm{L}) = \Psi(1) + (\mathrm{L} - 1)\widetilde{\Psi}(\mathrm{L})$$

with $\widetilde{\Psi}_j = \sum_{i=j+1}^{\infty} \Psi_i$. This result can be used to derive the multivariate Beveridge-Nelson decomposition (see Theorem 16.1 in Sect. 16.2.3):

$$\begin{aligned}
\begin{pmatrix} X_t \\ Y_t \end{pmatrix} &= \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} + \mu \begin{pmatrix} \gamma \\ 1 \end{pmatrix} t + \Psi(1) \sum_{j=1}^{t} Z_j + \text{stationary process} \\
&= \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} + \mu \begin{pmatrix} \gamma \\ 1 \end{pmatrix} t + \frac{1}{1-\phi} \begin{pmatrix} 0 & \gamma \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{\gamma}{1-\beta\phi} \\ 0 & 1 \end{pmatrix} \sum_{j=1}^{t} \begin{pmatrix} u_j \\ v_j \end{pmatrix} \\
&\quad + \text{stationary process} \\
&= \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} + \mu \begin{pmatrix} \gamma \\ 1 \end{pmatrix} t + \frac{1}{1-\phi} \begin{pmatrix} 0 & \gamma \\ 0 & 1 \end{pmatrix} \sum_{j=1}^{t} \begin{pmatrix} u_j \\ v_j \end{pmatrix} \\
&\quad + \text{stationary process.}
\end{aligned} \tag{16.1}$$

The Beveridge-Nelson decomposition represents the bivariate integrated process $\{(X_t, Y_t)'\}$ as a sum of three components: a linear trend, a multivariate random walk and a stationary process. Multiplying the Beveridge-Nelson decomposition from the left by the cointegration vector $\beta = (1, -\gamma)'$, we see that both the trend and the random walk component are eliminated and that only the stationary component remains.

Because the first column of $\Psi(1)$ consists of zeros, only the second structural shock, namely $\{v_t\}$, will have a long-run (permanent) effect. The long-run effect is $\gamma/(1-\phi)$ for the first variable, X_t, and $1/(1-\phi)$ for the second variable, Y_t. The first structural shock (preference shock) $\{u_t\}$ has non long-run effect, its impact is of a transitory nature only. This decomposition into permanent and transitory shocks is not typical for this model, but can be done in general as part of the so-called common trend representation (see Sect. 16.2.4).

Finally, we will simulate the reaction of the system to a unit valued shock in v_t. Although this shock only has a temporary influence on ΔY_t, it will have a permanent effect on the level Y_t. Taking $\phi = 0.8$, we get long-run effect (persistence) of $1/(1-\phi) = 5$ as explained in Sect. 7.1.3. The present discounted value model then implies that this shock will also have a permanent effect on X_t too. Setting $\gamma = 1$, this long-run effect is given by $\gamma(1-\beta)\sum_{j=0}^{\infty} \beta^j (1-\phi)^{-1} = \gamma/(1-\phi) = 5$. Because this long-run effect is anticipated in period t, the period of the occurrence of the shock, X_t will increase by more than one. The spread turns, therefore, into positive. The error correction mechanism will then dampen the effect on future changes of X_t

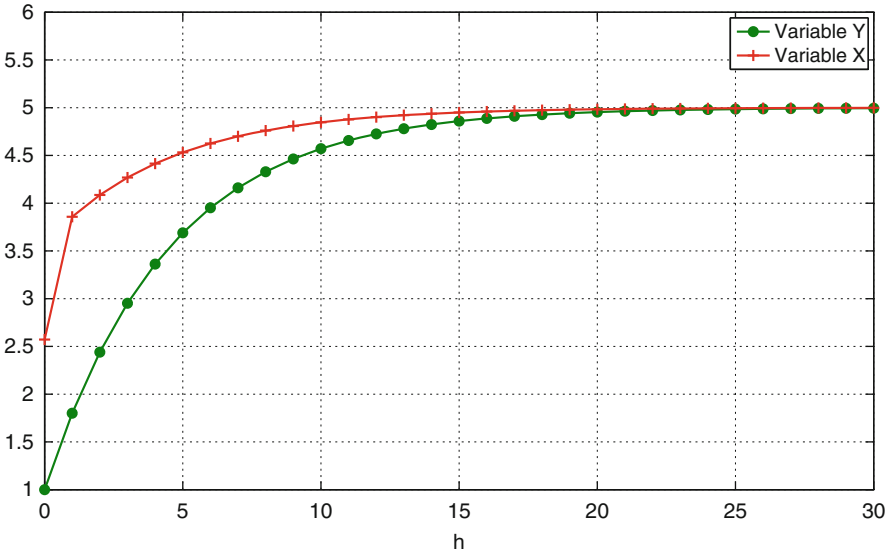

Fig. 16.1 Impulse response functions of the present discounted value model after a unit shock to Y_t ($\gamma = 1, \beta = 0.9, \phi = 0.8$)

so that the spread will return steadily to zero. The corresponding impulse responses of both variables are displayed in Fig. 16.1.

Figure 16.2 displays the trajectories of both variables after a stochastic simulation where both shocks $\{u_t\}$ and $\{v_t\}$ are drawn from a standard normal distribution. One can clearly discern the non-stationary character of both series. However, as it is typically for cointegrated series, they move more or less in parallel to each other. This parallel movement is ensured by the error correction mechanism. The difference between both series which is equal to the spread under this parameter constellation is mean reverting around zero.

16.2 Definition and Representation of Cointegrated Processes

16.2.1 Definition

We now want to make the concepts introduced earlier more precise and give a general definition of cointegrated processes and derive the different representations we have seen in the previous section. Given an arbitrary regular (purely non-deterministic) stationary process $\{U_t\}_{t\in\mathbb{Z}}$ of dimension n, $n \geq 1$, with mean zero and some distribution for the starting random variable X_0, we can define recursively a process $\{X_t\}$, $t = 0, 1, 2, \ldots$ as follows:

$$X_t = \mu + X_{t-1} + U_t, \qquad t = 1, 2, \ldots$$

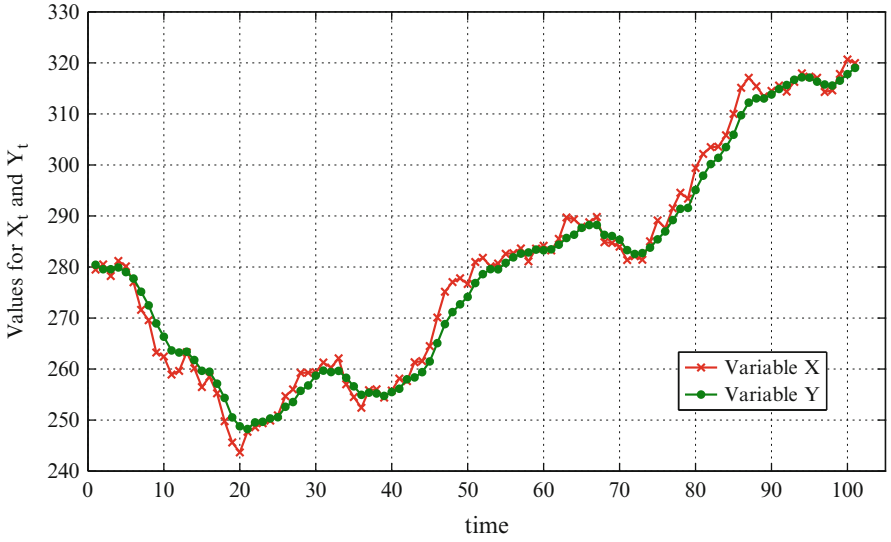

Fig. 16.2 Stochastic simulation of the present discounted value model under standard normally distributed shocks ($\gamma = 1, \beta = 0.9, \phi = 0.8$)

Thereby, μ denotes an arbitrary constant vector of dimension n. If $U_t \sim \text{WN}(0, \Sigma)$, then $\{X_t\}$ is a multivariate random walk with drift μ. In general, however, $\{U_t\}$ is autocorrelated and possesses a Wold representation $U_t = \Psi(L)Z_t$ (see Sect. 14.1.1) such that

$$\Delta X_t = \mu + U_t = \mu + \Psi(\text{L})Z_t = \mu + Z_t + \Psi_1 Z_{t-1} + \Psi_2 Z_{t-2} + \ldots, \tag{16.2}$$

where $Z_t \sim \text{WN}(0, \Sigma)$ and $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$ with $\Psi_0 = I_n$. We now introduce the following definitions.

Definition 16.1. *A regular stationary process $\{U_t\}$ with mean zero is integrated of order zero, $I(0)$, if and only if it can be represented as*

$$U_t = \Psi(\text{L})Z_t = Z_t + \Psi_1 Z_{t-1} + \Psi_2 Z_{t-2} + \ldots$$

such that $Z_t \sim \text{WN}(0, \Sigma)$, $\sum_{j=0}^{\infty} j\|\Psi_j\| < \infty$, and $\Psi(1) = \sum_{j=0}^{\infty} \Psi_j \neq 0$.

Definition 16.2. *A stochastic process $\{X_t\}$ is integrated of order d, $I(d)$, $d = 0, 1, 2, \ldots$, if and only if $\Delta^d(X_t - \mathbb{E}(X_t))$ is integrated of order zero.*

In the following we concentrate on I(1) processes. The definition of an I(1) process implies that $\{X_t\}$ equals $X_t = X_0 + \mu t + \sum_{j=1}^{t} U_j$ and is thus non-stationary even if $\mu = 0$. The condition $\Psi(1) \neq 0$ corresponds to the one in the univariate case (compare Definition 7.1 in Sect. 7.1). On the one hand, it precludes the case that a trend-stationary process is classified as an integrated process. On the

other hand, it implies that $\{X_t\}$ is in fact non-stationary. Indeed, if the condition is violated so that $\Psi(1) = 0$, we could express $\Psi(\mathrm{L})$ as $(1 - \mathrm{L})\widetilde{\Psi}(\mathrm{L})$. Thus we could cancel $1 - \mathrm{L}$ on both sides of Eq. (16.2) to obtain a stationary representation of $\{X_t\}$, given some initial distribution for X_0. This would then contradict our primal assumption that $\{X_t\}$ is non-stationary. The condition $\sum_{j=0}^{\infty} j\|\Psi_j\| < \infty$ is stronger than $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$ which follows from the Wold's Theorem. It guarantees the existence of the Beveridge-Nelson decomposition (see Theorem 16.1 below).[2] In particular, the condition is fulfilled if $\{U_t\}$ is a causal ARMA process which is the prototypical case.

Like in the univariate case, we can decompose an I(1) process additively into several components.

Theorem 16.1 (Beveridge-Nelson Decomposition)**.** *If $\{X_t\}$ is an integrated process of order one, it can be decomposed as*

$$X_t = X_0 + \mu t + \Psi(1) \sum_{j=1}^{t} Z_j + V_t,$$

where $V_t = \widetilde{\Psi}(\mathrm{L})Z_0 - \widetilde{\Psi}(\mathrm{L})Z_t$ with $\widetilde{\Psi}_j = \sum_{i=j+1}^{\infty} \Psi_i$, $j = 0, 1, 2, \ldots$ and $\{V_t\}$ stationary.

Proof. Following the proof of the univariate case (see Sect. 7.1.4):

$$\Psi(\mathrm{L}) = \Psi(1) + (\mathrm{L} - 1)\widetilde{\Psi}(\mathrm{L})$$

with $\widetilde{\Psi}_j = \sum_{i=j+1}^{\infty} \Psi_i$. Thus,

$$\begin{aligned} X_t &= X_0 + \mu t + \sum_{j=1}^{t} U_j = X_0 + \mu t + \sum_{j=1}^{t} \Psi(\mathrm{L})Z_j \\ &= X_0 + \mu t + \sum_{j=1}^{t} \left(\Psi(1) + (\mathrm{L} - 1)\widetilde{\Psi}(\mathrm{L})\right) Z_j \\ &= X_0 + \mu t + \Psi(1) \sum_{j=1}^{t} Z_j + \sum_{j=1}^{t} (\mathrm{L} - 1)\widetilde{\Psi}(\mathrm{L})Z_j \\ &= X_0 + \mu t + \Psi(1) \sum_{j=1}^{t} Z_j + \widetilde{\Psi}(\mathrm{L})Z_0 - \widetilde{\Psi}(\mathrm{L})Z_t. \end{aligned}$$

The only point left is to show that $\widetilde{\Psi}(\mathrm{L})Z_0 - \widetilde{\Psi}(\mathrm{L})Z_t$ is stationary. Based on Theorem 10.2, it is sufficient to show that the coefficient matrices are

[2] This condition could be relaxed and replaced by the condition $\sum_{j=0}^{\infty} j^2\|\Psi_j\|^2 < \infty$. In addition, this condition is an important assumption for the application of the law of large numbers and for the derivation of the asymptotic distribution (Phillips and Solo 1992).

absolutely summable. This can be derived by applying the triangular inequality and the condition for integrated processes:

$$\sum_{j=0}^{\infty} \|\widetilde{\Psi}_j\| = \sum_{j=0}^{\infty} \left\| \sum_{i=j+1}^{\infty} \Psi_i \right\| \leq \sum_{j=0}^{\infty} \sum_{i=j+1}^{\infty} \|\Psi_i\| = \sum_{j=1}^{\infty} j \|\Psi_j\| < \infty.$$

□

The process $\{X_t\}$ can therefore be viewed as the sum of a linear trend, $X_0 + \mu t$, with stochastic intercept, a multivariate random walk, $\Psi(1) \sum_{j=0}^{t} Z_t$, and a stationary process $\{V_t\}$. Based on this representation, we can then define the notion of cointegration (Engle and Granger 1987).

Definition 16.3 (Cointegration)**.** *A multivariate stochastic process $\{X_t\}$ is called* cointegrated *if $\{X_t\}$ is integrated of order one and if there exists a vector $\beta \in \mathbb{R}^n$, $\beta \neq 0$, such that $\{\beta' X_t\}$, is integrated of order zero, given a corresponding distribution for the initial random variable X_0. β is called the* cointegrating *or* cointegration vector. *The* cointegrating rank *is the maximal number, r, of linearly independent cointegrating vectors $\beta_1, \ldots, \beta_r$. These vectors span a linear space called the* cointegration space.

The Beveridge-Nelson decomposition implies that β is a cointegrating vector if and only if $\beta' \Psi(1) = 0$. In this case the random walk component $\sum_{j=1}^{t} Z_j$ is annihilated and only the deterministic and the stationary component remain.[3] For some issues it is of interest whether the cointegration vector β also eliminates the trend component. This would be the case if $\beta' \mu = 0$. See Sect. 16.3 for details.

The cointegration vectors are determined only up to some basis transformations. If $\beta_1, \ldots, \beta_r$ is a basis for the cointegration space then $(\beta_1, \ldots, \beta_r)R$ is also a basis for the cointegration space for any nonsingular $r \times r$ matrix R because $((\beta_1, \ldots, \beta_r)R)' \Psi(1) = 0$.

16.2.2 Vector Autoregressive (VAR) and Vector Error Correction Models (VECM)

Although the Beveridge-Nelson decomposition is very useful from a theoretical point of view, in practice it is often more convenient to work with alternative representations. Most empirical investigations of integrated processes start from a VAR(p) model which has the big advantage that it can be easily estimated:

$$X_t = c + \Phi_1 X_{t-1} + \ldots + \Phi_p X_{t-p} + Z_t, \qquad Z_t \sim \mathrm{WN}(0, \Sigma) \tag{16.3}$$

where $\Phi(\mathrm{L}) = I_n - \Phi_1 \mathrm{L} - \ldots - \Phi_p \mathrm{L}^p$ and c is an arbitrary constant. Subtracting X_{t-1} on both sides of the difference equation, the VAR model can be rewritten as:

[3] The distribution of X_0 is thereby chosen such that $\beta' X_0 = \beta' \widetilde{\Psi}(\mathrm{L}) Z_0$.

$$\Delta X_t = c + \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \ldots + \Gamma_{p-1} \Delta X_{t-p+1} + Z_t \tag{16.4}$$

where $\Pi = -\Phi(1) = -I_n + \Phi_1 + \ldots \Phi_p$ and $\Gamma_i = -\sum_{j=i+1}^{p} \Phi_j$. We will make the following assumptions:

(i) All roots of the polynomial $\det \Phi(z)$ are outside the unit circle or equal to one, i.e.

$$\det \Phi(z) = 0 \Longrightarrow \begin{cases} |z| > 1 \text{ or} \\ z = 1, \end{cases}$$

(ii) The matrix Π is singular with rank r, $1 \leq r < n$.
(iii) $\text{Rank}(\Pi) = \text{Rank}(\Pi^2)$.

Assumption (i) makes sure that $\{X_t\}$ is an integrated process with order of integration $d \geq 1$. Moreover, it precludes other roots on the unit circles than one. The case of seasonal unit roots is treated in Hylleberg et al. (1990) and Johansen and Schaumburg (1998).[4] Assumption (ii) implies that there exists at least $n - r$ unit roots and two $n \times r$ matrices α and β with full column rank r such that

$$\Pi = \alpha\beta'.$$

The columns of β thereby represent the cointegration vectors whereas α denotes the so-called loading matrix. The decomposition of Π in the product of α and β' is not unique. For every non-singular $r \times r$ matrix R we can generate an alternative decomposition $\Pi = \alpha\beta' = (\alpha R'^{-1})(\beta R)'$. Finally, assumption (iii) implies that the order of integration is exactly one and not greater. The number of unit roots is therefore exactly $n - r$.[5] This has the implication that $\Phi(z)$ can be written as

$$\Phi(z) = U(z)M(z)V(z)$$

where the roots of the matrix polynomials $U(z)$ and $V(z)$ are all outside the unit circle and where $M(z)$ equals

$$M(z) = \begin{pmatrix} (1-z)I_{n-r} & 0 \\ 0 & I_r \end{pmatrix}.$$

This representation of $\Phi(z)$ is a special form of the Smith-McMillan factorization of polynomial matrices (see Kailath (1980) and Yoo (1987)). This factorization isolates the unit roots in one simple matrix so that the system can be analyzed more easily.

[4] The seasonal unit roots are the roots of $z^s - 1 = 0$ where s denotes the number of seasons. These roots can be expressed as $\cos(2k\pi/s) + \iota \sin(2k\pi/s)$, $k = 0, 1, \ldots, s-1$.

[5] For details see Johansen (1995), Neusser (2000) and Bauer and Wagner (2003).

These assumptions will allow us to derive from the VAR(p) model several representations where each of them brings with it a particular interpretation. Replacing Π by $\alpha\beta'$ in Eq. (16.4), we obtain the *vector error correction representation* or *vector error correction model* (VECM):

$$\Delta X_t = c + \alpha\beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + \ldots + \Gamma_{p-1} \Delta X_{t-p+1} + Z_t. \tag{16.5}$$

Multiplying both sides of the equation by $(\alpha'\alpha)^{-1}\alpha'$ and solving for $\beta' X_{t-1}$, we get:

$$\beta' X_{t-1} = (\alpha'\alpha)^{-1}\alpha' \left(\Delta X_t - c - \sum_{j=1}^{p-1} \Gamma_j \Delta X_{t-j} - Z_t \right).$$

α has full column rank r so that $\alpha'\alpha$ is a non-singular $r \times r$ matrix. As the right hand side of the equation represents a stationary process, also the left hand side must be stationary. This means that the r-dimensional process $\{\beta' X_{t-1}\}$ is stationary despite the fact that $\{X_t\}$ is integrated and has potentially a unit root with multiplicity n.

The term error correction was coined by Davidson et al. (1978). They interpret the mean of $\beta' X_t$, $\mu^* = \mathbb{E}\beta' X_t$, as the long-run equilibrium or steady state around which the system fluctuates. The deviation from equilibrium (error) is therefore given by $\beta' X_{t-1} - \mu^*$. The coefficients of the loading matrix α should then guarantee that deviations from the equilibrium are corrected over time by appropriate changes (corrections) in X_t.

An Illustration

To illustrate the concept of the error correction model, we consider the following simple system with $\alpha = (\alpha_1, \alpha_2)'$, $\alpha_1 \neq \alpha_2$, and $\beta = (1, -1)'$. For simplicity, we assume that the long-run equilibrium μ^* is zero. Ignoring higher order lags, we consider the system:

$$\begin{aligned} \Delta X_{1t} &= \alpha_1 (X_{1,t-1} - X_{2,t-1}) + Z_{1t} \\ \Delta X_{2t} &= \alpha_2 (X_{1,t-1} - X_{2,t-1}) + Z_{2t}. \end{aligned}$$

The autoregressive polynomial of this system is:

$$\Phi(z) = \begin{pmatrix} 1 - (1 + \alpha_1)z & \alpha_1 z \\ -\alpha_2 z & 1 - (1 - \alpha_2)z \end{pmatrix}.$$

The determinant of this polynomial is $\det \Phi(z) = 1 - (2 + \alpha_1 - \alpha_2)z + (1 + \alpha_1 - \alpha_2)z^2$ with roots equal to $z = 1$ and $z = 1/(1 + \alpha_1 - \alpha_2)$. This shows that assumption (i) is fulfilled. As $\Pi = \begin{pmatrix} \alpha_1 & -\alpha_1 \\ \alpha_2 & -\alpha_2 \end{pmatrix}$, the rank of Π equals one which implies assumption (ii). Finally,

$$\Pi^2 = \begin{pmatrix} \alpha_1^2 - \alpha_1\alpha_2 & -\alpha_1^2 + \alpha_1\alpha_2 \\ -\alpha_2^2 + \alpha_1\alpha_2 & \alpha_2^2 - \alpha_1\alpha_2 \end{pmatrix}.$$

Thus, the rank of Π^2 is also one because $\alpha_1 \neq \alpha_2$. Hence, assumption (iii) is also fulfilled.

We can gain an additional insight into the system by subtracting the second equation from the first one to obtain:

$$X_{1t} - X_{2t} = (1 + \alpha_1 - \alpha_2)(X_{1,t-1} - X_{2,t-1}) + Z_{1t} - Z_{2t}.$$

The process $\beta' X_t = X_{1t} - X_{2t}$ is stationary and causal with respect to $Z_{1t} - Z_{2t}$ if and only if $|1 + \alpha_1 - \alpha_2| < 1$, or equivalently if and only if $-2 < \alpha_1 - \alpha_2 < 0$. Note the importance of the assumption that $\alpha_1 \neq \alpha_2$. It prevents that $X_{1t} - X_{2t}$ becomes a random walk and thus a non-stationary (integrated) process. A sufficient condition is that $-1 < \alpha_1 < 0$ and $0 < \alpha_2 < 1$ which imply that a positive (negative) error, i.e. $X_{1,t-1} - X_{2,t-1} > 0 (< 0)$, is corrected by a negative (positive) change in X_{1t} and a positive (negative) change in X_{2t}. Although the shocks Z_{1t} and Z_{2t} push $X_{1t} - X_{2t}$ time and again away from its long-run equilibrium, the error correction mechanism ensures that the variables are adjusted in such a way that the system moves back to its long-run equilibrium.

16.2.3 The Beveridge-Nelson Decomposition

We next want to derive from the VAR representation a causal representation or MA(∞) representation for $\{\Delta X_t\}$. In contrast to a normal causal VAR model, the presence of unit roots precludes the simple application of the method of undetermined coefficients, but requires an additional effort. Multiplying the VAR representation in Eq. (16.3), $\Phi(\mathrm{L})X_t = U(\mathrm{L})M(\mathrm{L})V(\mathrm{L})X_t = c + Z_t$, from the left by $U^{-1}(\mathrm{L})$ we obtain:

$$M(\mathrm{L})V(\mathrm{L})X_t = U^{-1}(1)c + U^{-1}(\mathrm{L})Z_t.$$

Multiplying this equation by $\widetilde{M}(\mathrm{L}) = \begin{pmatrix} I_{n-r} & 0 \\ 0 & (1-\mathrm{L})I_r \end{pmatrix}$ leads to:

$$V(\mathrm{L})\Delta X_t = \widetilde{M}(1)U^{-1}(1)c + \widetilde{M}(\mathrm{L})U^{-1}(\mathrm{L})Z_t$$

which finally leads to

$$\begin{aligned} \Delta X_t &= V^{-1}(1)\widetilde{M}(1)U^{-1}(1)c + V^{-1}(\mathrm{L})\widetilde{M}(\mathrm{L})U^{-1}(\mathrm{L})Z_t \\ &= \mu + \Psi(\mathrm{L})Z_t. \end{aligned}$$

This is the MA(∞) representation of $\{\Delta X_t\}$ and corresponds to Eq. (16.2).

Because $\Pi = -\Phi(1) = -U(1)M(1)V(1)$, the following relation holds for the partitioned matrices:

$$\Phi(1) = \begin{pmatrix} U_{11}(1) & U_{12}(1) \\ U_{21}(1) & U_{22}(1) \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & I_r \end{pmatrix} \begin{pmatrix} V_{11}(1) & V_{12}(1) \\ V_{21}(1) & V_{22}(1) \end{pmatrix} = \begin{pmatrix} U_{12}(1) \\ U_{22}(1) \end{pmatrix} \begin{pmatrix} V_{21}(1) & V_{22}(1) \end{pmatrix}.$$

This implies that we can define α and β as

$$\alpha = -\begin{pmatrix} U_{12}(1) \\ U_{22}(1) \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} V_{21}(1)' \\ V_{22}(1)' \end{pmatrix}.$$

$U(1)$ and $V(1)$ are non-singular so that α and β have full column rank r. Based on this derivation we can formulate the following lemma.

Lemma 16.1. *The columns of the so-defined matrix β are the cointegration vectors for the process $\{X_t\}$. The corresponding matrix of loading coefficients is α which fulfills $\Psi(1)\alpha = 0$.*

Proof. We must show that $\beta'\Psi(1) = 0$ which is the defining property of cointegration vectors. Denoting by $(V^{(ij)}(1))_{i,j=1,2}$ the appropriately partitioned matrix of $V(1)^{-1}$, we obtain:

$$\begin{aligned} \beta'\Psi(1) &= \beta' \begin{pmatrix} V^{(11)}(1) & V^{(12)}(1) \\ V^{(21)}(1) & V^{(22)}(1) \end{pmatrix} \begin{pmatrix} I_{n-r} & 0 \\ 0 & 0 \end{pmatrix} U^{-1}(1) \\ &= \begin{pmatrix} V_{21}(1) & V_{22}(1) \end{pmatrix} \begin{pmatrix} V^{(11)}(1) & 0 \\ V^{(21)}(1) & 0 \end{pmatrix} U^{-1}(1) \\ &= \left(V_{21}(1)V^{(11)}(1) + V_{22}(1)V^{(21)}(1) \vdots 0 \right) U^{-1}(1) = 0_n \end{aligned}$$

where the last equality is a consequence of the property of the inverse matrix.

With the same arguments, we can show that $\Psi(1)\alpha = 0$. □

The equivalence between the VEC and the MA representation is known as Granger's representation theorem in the literature. Granger's representation theorem immediately implies the Beveridge-Nelson decomposition:

$$X_t = X_0 + \Psi(1)c\,t + \Psi(1)\sum_{j=1}^{t} Z_j + V_t \tag{16.6}$$

$$= X_0 + V^{-1}(1)\widetilde{M}(1)U^{-1}(1)c\,t + V^{-1}(1)\widetilde{M}(1)U^{-1}(1)\sum_{j=1}^{t} Z_j + V_t \tag{16.7}$$

where the stochastic process $\{V_t\}$ is stationary and defined as $V_t = \widetilde{\Psi}(\mathrm{L})Z_0 - \widetilde{\Psi}(\mathrm{L})Z_t$ with $\widetilde{\Psi}_j = \sum_{i=j+1}^{\infty} \Psi_i$ and $\Psi(\mathrm{L}) = V^{-1}(\mathrm{L})\widetilde{M}(\mathrm{L})U^{-1}(\mathrm{L})$. As $\beta'\Psi(1) = \beta'V^{-1}(1)\widetilde{M}(1)U^{-1}(1) = 0$, β eliminates the stochastic trend (random walk), $\sum_{j=1}^{t} Z_t$, as well as the deterministic linear trend $\mu t = V^{-1}(1)\widetilde{M}(1)U^{-1}(1)c\,t$.

An interesting special case is obtained when the constant c is a linear combination of the columns of α, i.e. if there exists a vector g such that $c = \alpha g$. Under this circumstance, $\Psi(1)c = \Psi(1)\alpha g = 0$ and the linear trend vanishes and we have $\mathbb{E}\Delta X_t = 0$. In this case, the data will exhibit no trend although the VAR model contains a constant. A similar consideration can be made if the VAR model is specified to contain a constant and a linear time trend $d\,t$. The Beveridge-Nelson decomposition would then imply that the data should follow a quadratic trend. However, in the special case that d is a linear combination of the columns of α, the quadratic trend disappears and only the linear remains because of the constant.

16.2.4 Common Trend and Triangular Representation

The $\Psi(1)$ in the Beveridge-Nelson decomposition is singular. This implies that the multivariate random walk $\Psi(1)\sum_{j=1}^{\infty} Z_j$ does not consist of n independent univariate random walks. Instead only $n-r$ independent random walks make up the stochastic trend so that $\{X_t\}$ is driven by $n-r$ stochastic trends. In order to emphasize this fact, we derive from the Beveridge-Nelson decomposition the so-called *common trend representation* (Stock and Watson 1988a).

As $\Psi(1)$ has rank $n-r$, there exists a $n\times r$ matrix γ such that $\Psi(1)\gamma = 0$. Denote by $\gamma^{\perp}$ the $n\times(n-r)$ matrix whose columns are orthogonal to γ, i.e. $\gamma'\gamma^{\perp} = 0$. The Beveridge-Nelson decomposition can then be rewritten as:

$$
\begin{aligned}
X_t &= X_0 + \Psi(1)\left(\gamma^{\perp} \vdots \gamma\right)\left(\gamma^{\perp} \vdots \gamma\right)^{-1} c\,t \\
&\quad + \Psi(1)\left(\gamma^{\perp} \vdots \gamma\right)\left(\gamma^{\perp} \vdots \gamma\right)^{-1} \sum_{j=1}^{t} Z_j + V_t \\
&= X_0 + \left(\Psi(1)\gamma^{\perp} \vdots 0\right)\left(\gamma^{\perp} \vdots \gamma\right)^{-1} c\,t \\
&\quad + \left(\Psi(1)\gamma^{\perp} \vdots 0\right)\left(\gamma^{\perp} \vdots \gamma\right)^{-1} \sum_{j=1}^{t} Z_j + V_t \\
&= X_0 + \left(\Psi(1)\gamma^{\perp} \vdots 0\right)\tilde{c}\,t + \left(\Psi(1)\gamma^{\perp} \vdots 0\right)\sum_{j=1}^{t} \widetilde{Z}_j + V_t
\end{aligned}
$$

where $\tilde{c} = \left(\gamma^{\perp} \vdots \gamma\right)^{-1} c$ and $\widetilde{Z}_j = \left(\gamma^{\perp} \vdots \gamma\right)^{-1} Z_j$. Therefore, only the first $n-r$ elements of the vector $\tilde{c}$ are relevant for the deterministic linear trend. The remaining elements are multiplied by zero and are thus irrelevant. Similarly, for the multivariate random walk only the first $n-r$ elements of the process $\{\widetilde{Z}_t\}$ are responsible for the stochastic trend. The remaining elements of $\widetilde{Z}_t$ are multiplied

by zero and thus have no permanent, but only a transitory influence. The above representation decomposes the shocks orthogonally into *permanent* and *transitory* ones (Gonzalo and Ng 2001). The previous lemma shows that one can choose for γ the matrix of loading coefficients α.

Summarizing the first $n - r$ elements of $\tilde{c}$ and $\widetilde{Z}_t$ to $\tilde{c}_1$ and $\widetilde{Z}_{1t}$, respectively, we arrive at the *common trend representation*:

$$X_t = X_0 + B\tilde{c}_1\, t + B \sum_{j=1}^{t} \widetilde{Z}_{1j} + V_t$$

where the $n \times (n - r)$ matrix B is equal to $\Psi(1)\gamma^{\perp}$.

Applying these results to our introductory example, we arrive at

$$B = \frac{1}{1-\phi}\begin{pmatrix}\gamma\\ 1\end{pmatrix}, \quad \tilde{c} = \mu(1-\phi), \quad \text{and} \quad \widetilde{Z}_{1t} = u_t.$$

This again demonstrates that the trend, the linear as well as the stochastic trend, are exclusively stemming from the nonstationary variables $\{Y_t\}$ (compare with Eq. (16.1)).

Finally, we want to present a triangular representation which is well suited to deal with the nonparametric estimation approach advocated by Phillips (1991) and Phillips and Hansen (1990) (see Sect. 16.4). In this representation we normalize the cointegration vector such $\beta = (I_r, -b')'$. In addition, we partition the vector X_t into X_{1t} and X_{2t} such that X_{1t} contains the first r and X_{2t} the last $n - r$ elements. $X_t = (X'_{1t}, X_{2t})'$ can then be expressed as:

$$X_{1t} = b'X_{2t} + \pi_1 D_t + u_{1t} \tag{16.8a}$$

$$\Delta X_{2t} = \pi_2 \Delta D_t + u_{2t} \tag{16.8b}$$

where D_t summarizes the deterministic components such as constant and linear time trend. $\{u_{1t}\}$ and $\{u_{2t}\}$ denote potentially autocorrelated and cross-correlated stationary time series.

16.3 Johansen's Test for Cointegration

In Sect. 7.5.2 we have already discussed a regression based test for cointegration among two variables. It was based on a unit root of the residuals from a bivariate regression of one variable against the other. In this regression, it turned out to be irrelevant which of the two variables was chosen as the regressor and which one as the regressand. This method can, in principle, be extended to more than two variables. However, with more than two variables, the choice of the regressand

becomes more crucial as not all variables may be part of the cointegrating relation. Moreover, more than one independent cointegrating relation may exist. For these reasons, it is advantageous to use a method which treats all variables symmetrically. The *cointegration test* developed by Johansen fulfills this criterion because it is based on a VAR which does not single out a particular variable. This test has received wide recognition and is most often used in practice. The test serves two purposes. First, we want to determine the number r of cointegrating relationships. Second, we want to test properties of the cointegration vector β and the loading matrix α.

The exposition of the Johansen test follows closely the work of Johansen where the derivations and additional details can be found (Johansen 1988, 1991, 1995). We start with a VAR(p) model with constant c in VEC form (see Eq. (16.4)):

$$\begin{aligned} \Delta X_t &= c + \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \ldots + \Gamma_{p-1} \Delta X_{t-p+1} + Z_t, \\ t &= 1, 2, \ldots, T \end{aligned} \tag{16.9}$$

where $Z_t \sim \text{IIDN}(0, \Sigma)$ and given starting values $X_0 = x_0, \ldots, X_{-p+1} = x_{-p+1}$. The problem can be simplified by regressing ΔX_t as well as X_{t-1} against $c, \Delta X_{t-1}, \ldots, \Delta X_{t-p+1}$ and working with the residuals from these regressions.[6] This simplification results in a VAR model of order one. We therefore start our analysis without loss of generality with a VAR(1) model without constant term:

$$\Delta X_t = \Pi X_{t-1} + Z_t$$

where $Z_t \sim \text{IIDN}(0, \Sigma)$.[7]

The phenomenon of cointegration manifests itself in the singularity of the matrix Π. In particular, we want to determine the rank of Π which gives the number of linearly independent cointegrating relationships. Denoting by r, the rank of Π, $0 \leq r \leq n$, , we can formulate a sequence of hypotheses:

$$\text{H}(r) : \text{rank}(\Pi) \leq r, \qquad r = 0, 1, \ldots, n.$$

Hypothesis H(r), thus, implies that there exists *at most* r linearly independent cointegrating vectors. The sequence of hypotheses is nested in the sense that H(r) implies H($r + 1$):

$$\text{H}(0) \subseteq \text{H}(1) \subseteq \ldots \subseteq \text{H}(n).$$

The hypothesis H(0) means that $\text{rank}(\Pi) = 0$. In this case, $\Pi = 0$ and there are no cointegration vectors. $\{X_t\}$ is thus driven by n independent random walks and

[6] If the VAR model (16.9) contains further deterministic components besides the constant, these components have to be accounted for in these regressions.

[7] This two-stage least-squares procedure is also known as partial regression and is part of the Frisch-Waugh-Lowell Theorem (Davidson and MacKinnon 1993, 19–24).

the VAR can be transformed into a VAR model for $\{\Delta X_t\}$ which in our simplified version just means that $\Delta X_t = Z_t \sim \text{IIDN}(0, \Sigma)$. The hypothesis H($n$) places no restriction on Π and includes in this way the case that the level of $\{X_t\}$ is already stationary. Of particular interest are the hypotheses between these two extreme ones where non-degenerate cointegrating vectors are possible. In the following, we not only want to test for the number of linearly independent cointegrating vectors, r, but we also want to test hypotheses about the structure of the cointegrating vectors summarized in β.

Johansen's test is conceived as a likelihood-ratio test. This means that we must determine the likelihood function for a sample $X_1, X_2, \ldots, X_T$ where T denotes the sample size. For this purpose, we assume that $\{Z_t\} \sim \text{IIDN}(0, \Sigma)$ so that logged likelihood function of the parameters α, β, and Σ conditional on the starting values is given by :

$$\ell(\alpha, \beta, \Sigma) = -\frac{Tn}{2}\ln(2\pi) + \frac{T}{2}\ln\det(\Sigma^{-1}) - \frac{1}{2}\sum_{t=1}^{T}(\Delta X_t - \alpha\beta' X_{t-1})'\Sigma^{-1}(\Delta X_t - \alpha\beta' X_{t-1})$$

where $\Pi = \alpha\beta'$. For a fixed given β, α can be estimated by a regression of ΔX_t on $\beta' X_{t-1}$:

$$\hat{\alpha} = \hat{\alpha}(\beta) = S_{01}\beta(\beta' S_{11}\beta)^{-1}$$

where the moment matrices S_{00}, S_{11}, S_{01} and S_{10} are defined as:

$$S_{00} = \frac{1}{T}\sum_{t=1}^{T}(\Delta X_t)(\Delta X_t)'$$

$$S_{11} = \frac{1}{T}\sum_{t=1}^{T} X_{t-1}X_{t-1}'$$

$$S_{01} = \frac{1}{T}\sum_{t=1}^{T}(\Delta X_t)X_{t-1}'$$

$$S_{10} = S_{01}'.$$

The covariance matrix of the residuals then becomes:

$$\widehat{\Sigma} = \widehat{\Sigma}(\beta) = S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}.$$

Using these results, we can concentrate the log-likelihood function to obtain:

$$\ell(\beta) = \ell(\hat{\alpha}(\beta), \beta, \widehat{\Sigma}(\beta)) = -\frac{Tn}{2}\ln(2\pi) - \frac{T}{2}\ln\det(\widehat{\Sigma}(\beta)) - \frac{Tn}{2}$$
$$= -\frac{Tn}{2}\ln(2\pi) - \frac{Tn}{2} - \frac{T}{2}\ln\det\left(S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}\right). \qquad (16.10)$$

The expression $\frac{Tn}{2}$ in the above equation is derived as follows:

$$\frac{1}{2}\sum_{t=1}^{T}(\Delta X_t - \hat{\alpha}\beta' X_{t-1})'\widehat{\Sigma}^{-1}(\Delta X_t - \hat{\alpha}\beta' X_{t-1})$$
$$= \frac{1}{2}\,\mathrm{tr}\left(\sum_{t=1}^{T}(\Delta X_t - \hat{\alpha}\beta' X_{t-1})(\Delta X_t - \hat{\alpha}\beta' X_{t-1})'\widehat{\Sigma}^{-1}\right)$$
$$= \frac{1}{2}\,\mathrm{tr}\left((TS_{00} - T\hat{\alpha}\beta' S_{10} - TS_{01}\beta\hat{\alpha}' + T\hat{\alpha}\beta' S_{11}\beta\hat{\alpha}')\widehat{\Sigma}^{-1}\right)$$
$$= \frac{T}{2}\,\mathrm{tr}\left(\underbrace{(S_{00} - \hat{\alpha}\beta' S_{10})}_{=\widehat{\Sigma}}\,\widehat{\Sigma}^{-1}\right) = \frac{Tn}{2}.$$

The log-likelihood function is thus maximized if

$$\det(\widehat{\Sigma}(\beta)) = \det\left(S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}\right)$$
$$= \det S_{00}\,\frac{\det\left(\beta'(S_{11} - S_{10}S_{00}^{-1}S_{01})\beta\right)}{\det(\beta' S_{11}\beta)}$$

is minimized over β.[8] The minimum is obtained by solving the following generalized eigenvalue problem (Johansen 1995):

$$\det\left(\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}\right) = 0.$$

This eigenvalue problem delivers n eigenvalues

$$1 \geq \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_n \geq 0$$

[8] Thereby we make use of the following equality for partitioned matrices:

$$\det\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \det A_{11}\det(A_{22} - A_{21}A_{11}^{-1}A_{12}) = \det A_{22}\det(A_{11} - A_{12}A_{22}^{-1}A_{21})$$

where A_{11} and A_{22} are invertible matrices (see for example Meyer 2000, p. 475).

with corresponding n eigenvectors $\hat{\beta}_1, \ldots, \hat{\beta}_n$. These eigenvectors are normalized such that $\hat{\beta}' S_{11} \hat{\beta} = I_n$. Therefore we have that

$$\operatorname{argmin}_\beta \det(\widehat{\Sigma}(\beta)) = (\det S_{00}) \prod_{i=1}^{n} \hat{\lambda}_i.$$

Remark 16.1. In the case of cointegration, Π is singular with $\operatorname{rank}\widehat{\Pi} = r$. To estimate r, it seems natural to investigate the number of nonzero eigenvalues of $\widehat{\Pi} = S_{01} S_{11}^{-1}$. However, because eigenvalues may be complex, it is advantageous not to investigate the eigenvalues of $\widehat{\Pi}$ but those of $\widehat{\Pi}'\widehat{\Pi}$ which are all real and positive due to the symmetry of $\widehat{\Pi}'\widehat{\Pi}$. These eigenvalues are called the *singular values* of $\widehat{\Pi}$.[9] Noting that

$$\begin{aligned} 0 &= \det\left(\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}\right) = \det S_{11} \det\left(\lambda I_n - S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2}\right) \\ &= \det S_{11} \det\left(\lambda I_n - (S_{00}^{-1/2} S_{01} S_{11}^{-1/2})'(S_{00}^{-1/2} S_{01} S_{11}^{-1/2})\right), \end{aligned}$$

the generalized eigenvalue problem above therefore just determines the singular values of $S_{00}^{-1/2} S_{01} S_{11}^{-1/2} = S_{00}^{-1/2} \widehat{\Pi} S_{11}^{1/2}$.

Remark 16.2. Based on the observation that, for $n = 1$, $\lambda = \frac{S_{01} S_{10}}{S_{11} S_{00}}$ equals the squared empirical correlation coefficient between ΔX_t and X_{t-1}, we find that the eigenvalues λ_j, $j = 1, \ldots, n$, are nothing but the squared canonical correlation coefficients (see Johansen 1995; Reinsel 1993). Thereby, the largest eigenvalue, $\hat{\lambda}_1$, corresponds to the largest squared correlation coefficient that can be achieved between linear combinations of ΔX_1 and X_{t-1}. Thus β_1 gives the linear combination of the integrated variable X_{t-1} which comes closest in the sense of correlation to the stationary variable $\{\Delta X_t\}$. The second eigenvalue λ_2 corresponds to the maximal squared correlation coefficient between linear combinations of ΔX_t and X_{t-1} which are orthogonal to the linear combination corresponding to λ_1. The remaining squared canonical correlation coefficients are obtained by iterating this procedure n times.

If the dimension of the cointegrating space is r then $\hat{\beta}$ consists of those eigenvectors which correspond to the r largest eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_r$. The remaining eigenvalues $\lambda_{r+1}, \ldots, \lambda_n$ should be zero. Under the null hypothesis H(r), the log-likelihood function (16.10) can be finally expressed as:

[9] An appraisal of the singular values of a matrix can be found in Strang (1988) or Meyer (2000).

$$\ell(\hat{\beta}_r) = -\frac{Tn}{2}\ln\pi - \frac{Tn}{2} - \frac{T}{2}\ln\det S_{00} - \frac{T}{2}\sum_{i=1}^{r}\ln(1-\lambda_i).$$

The expression for the optimized likelihood function can now be used to construct the Johansen likelihood-ratio test. There are two versions of the test depending on the alternative hypothesis:

$$\begin{aligned} \text{trace test:} \quad & \mathrm{H}_0 : \mathrm{H}(r) \text{ against } \mathrm{H}(n), \\ \text{max test:} \quad & \mathrm{H}_0 : \mathrm{H}(r) \text{ against } \mathrm{H}(r+1). \end{aligned}$$

The corresponding likelihood ratio test statistics are therefore:

$$\begin{aligned} \text{trace test:} \quad & 2(\ell(\hat{\beta}_n) - \ell(\hat{\beta}_r)) = -T\sum_{j=r+1}^{n}\ln(1-\hat{\lambda}_j) \approx T\sum_{j=r+1}^{n}\hat{\lambda}_j, \\ \text{max test:} \quad & 2(\ell(\hat{\beta}_{r+1}) - \ell(\hat{\beta}_r)) = -T\ln(1-\hat{\lambda}_{r+1}) \approx T\hat{\lambda}_{r+1}. \end{aligned}$$

In practice it is useful to adopt a sequential test strategy based on the trace test. Given some significance level, we test in a first step the null hypothesis H(0) against H(n). If, on the one hand, the test does not reject the null hypothesis, we conclude that $r = 0$ and that there is no cointegrating relation. If, on the other hand, the test rejects the null hypothesis, we conclude that there is at least one cointegrating relation. We then test in a second step the null hypothesis H(1) against H(n). If the test does not reject the null hypothesis, we conclude that there exists one cointegrating relation, i.e. that $r = 1$. If the test rejects the null hypothesis, we examine the next hypothesis H(2), and so on. In this way we obtain a test sequence. If in this sequence, the null hypothesis H(r) is not rejected, but H($r+1$) is, we conclude that exist r linearly independent cointegrating relations as explained in the diagram below.

$$\begin{array}{ccccccc} \mathrm{H}(0) \text{ against } \mathrm{H}(n) & \xrightarrow{\text{rejection}} & \mathrm{H}(1) \text{ against } \mathrm{H}(n) & \xrightarrow{\text{rejection}} & \mathrm{H}(2) \text{ against } \mathrm{H}(n) \ldots \\ \downarrow & & \downarrow & & \downarrow \\ r = 0 & & r = 1 & & r = 2 \end{array}$$

If in this sequence we do not reject H(r) for some r, it is useful to perform the max test H(r) against H($r+1$) as a robustness check. The asymptotic distributions of the test statistics are, like in the Dickey-Fuller unit root test, nonstandard and depend on the specification of the deterministic components.

16.3.1 Specification of the Deterministic Components

As mentioned previously, the asymptotic distribution of Johansen's test depends on the specification of the deterministic components. Thus, some care must be devoted to this issue. We illustrate this point by decomposing the model additively into a linear deterministic and a stochastic component in vector error correction form (see Johansen (1995, 80–84) and Lütkepohl (2006, section 6.4)):

$$X_t = \mu_0 + \mu_1 t + Y_t \tag{16.11}$$

$$\Delta Y_t = \Pi Y_{t-1} + Z_t = \alpha\beta' Y_{t-1} + Z_t. \tag{16.12}$$

For the ease of exposition, we have omit the autoregressive corrections. Eliminating Y_t using $Y_t = X_t - \mu_0 - \mu_1 t$ and $\Delta Y_t = \Delta X_t - \mu_1$ leads to

$$\Delta X_t - \mu_1 = \alpha\beta' \left(X_{t-1} - \mu_0 - \mu_1(t-1)\right) + Z_t.$$

This equation can be rewritten as

$$\Delta X_t = c_0 + c_1(t-1) + \alpha\beta' X_{t-1} + Z_t \tag{16.13}$$

with $c_0 = \mu_1 - \alpha\beta'\mu_0$ and $c_1 = -\alpha\beta'\mu_1$

$$= c_0 + \alpha(\beta', -\beta'\mu_1)X_{t-1}^0 + Z_t \tag{16.14}$$

where $X_t^0 = (X_t', t)'$. Equation (16.13) is just the vector error correction model (16.4) augmented by the linear trend term $c_1 t$. If the term c_1 would be left unrestricted arguments similar to those in Sect. 16.2.3 would show that X_t exhibits a deterministic quadratic trend with coefficient vector $\Psi(1)c_1$. This, however, contradicts the specification in Eq. (16.11). However, if we recognize that c_1 in Eq. (16.13) is actually restricted to lie in the span of α, i.e. that $c_1 = \alpha\gamma_1$ with $\gamma_1 = -\beta'\mu_1$, no quadratic trend would emerge in the levels because $\Psi(1)\alpha = 0$ by Granger's representation Theorem 16.1. Alternatively, one may view the time trend as showing up in the error correction term, respectively being part of the cointegrating relation, as in Eq. (16.14).

Similarly, one may consider the case that X_t has a constant mean μ_0, i.e. that $\mu_1 = 0$ in Eq. (16.11). This leads to the same error correction specification (16.13), but without the term $c_1 t$. Leaving the constant c_0 unrestricted, this will generate a linear trend $\Psi(1)c_0 t$ as shown in Sect. 16.2.3. In order to reconcile this with the assumption of a constant mean, we must recognize that $c_0 = \alpha\gamma_0$ with $\gamma_0 = -\beta'\mu_0$.

Table 16.1 Trend specifications in vector error correction models

Case	Deterministic term in VECM or VAR	Restriction	Trend in X_t	$\mathbb{E}\Delta X_t$	$\mathbb{E}(\beta' X_t)$
I	None	–	Zero	Zero	Zero
II	c_0	$c_0 = \alpha\gamma_0$	Constant	Zero	Constant
III	c_0	None	Linear	Constant	Constant
IV	$c_0 + c_1 t$	$c_1 = \alpha\gamma_1$	Linear	Constant	Linear
V	$c_0 + c_1 t$	None	Quadratic	Linear	Linear

Table inspired by Johansen (2007)

Based on these arguments, we can summarize the discussion by distinguishing five different cases displayed in Table 16.1.[10] This table also shows the implications for $\mathbb{E}\Delta X_t$ and $\mathbb{E}(\beta' X_t)$. These can read off from Eqs. (16.13) and (16.14).

The corresponding asymptotic distributions of the trace as well as the max test statistic in these five cases are tabulated in Johansen (1995), MacKinnon et al. (1999), and Osterwald-Lenum (1992).[11] The finite sample properties of theses tests can be quite poor. Thus, more recently, bootstrap methods have been proven to provide a successful alternative in practice (Cavaliere et al. 2012).

16.3.2 Testing Hypotheses on Cointegrating Vectors

As mentioned previously, the cointegrating vectors are not unique, only the cointegrating space is. This makes the cointegrating vectors often difficult to interpret economically, despite some basis transformation. It is therefore of interest to see whether the space spanned by the cointegrating vectors summarized in the columns of $\hat{\beta}$ can be viewed as a subspace spanned by some hypothetical vectors $H = (h_1, \ldots, h_s)$, $r \leq s < n$. If this hypothesis is true, the cointegrating vectors should be linear combinations of the columns of H so that the null hypothesis can be formulated as

$$\mathrm{H}_0 : \beta = H\varphi \tag{16.15}$$

for some $s \times r$ matrix φ. Under this null hypothesis, this amounts to solve an analogous general eigenvalue problem:

$$\det\left(\varphi H' S_{11} H - H' S_{10} S_{00}^{-1} S_{01} H\right) = 0.$$

The solution of this problem is given by the eigenvalues $1 > \tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots \tilde{\lambda}_s > 0$ with corresponding normalized eigenvectors. The likelihood ratio test statistic for this hypothesis is then

[10] It is instructive to compare theses cases to those of the unit root test (see Sect. 7.3.1).

[11] The tables by MacKinnon et al. (1999) allow for the possibility of exogenous integrated variables.

$$T \sum_{j=1}^{r} \ln \frac{1 - \tilde{\lambda}_j}{1 - \hat{\lambda}_j}.$$

This test statistic is asymptotically distributed as a χ^2 distribution with $r(n - s)$ degrees of freedom.

With similar arguments it is possible to construct a test of the null hypothesis that the cointegrating space spanned by the columns of $\hat{\beta}$ contains some hypothetical vectors $K = (h_1, \ldots, h_s)$, $1 \leq s \leq r$. The null hypothesis can then be formulated as

$$\mathrm{H}_0 : K\varphi = \beta \tag{16.16}$$

for some $s \times r$ matrix φ. Like in the previous case, this hypothesis can also be tested by the corresponding likelihood ratio test statistic which is asymptotically distributed as a χ^2 distribution with $s(n - r)$ degrees of freedom. Similarly, it is possible to test hypotheses on α and joint hypotheses on α and β (see Johansen 1995; Kunst and Neusser 1990; Lütkepohl 2006).

16.4 Estimation and Testing of Cointegrating Relationships

Johansen's approach has become very popular because it presents an integrated framework for testing and estimating cointegrating relationships based on the maximum likelihood method. However, it requires the specification of a concrete VAR model. This proves sometimes difficult it practice, especially when the true data generating process is not purely autoregressive. Similar to the Phillips-Perron test discussed in Sect. 7.3.2, Phillips and Hansen (1990) propose a nonparametric approach for the estimation and hypothesis testing of cointegrating relationships. This approach is especially appropriate if the long-run relationships are the prime objective of the investigation as f.e. in Neusser and Kugler (1998).

The Phillips and Hansen approach is based on the triangular representation of cointegrated processes given in the equation system (16.8). Thereby the r cointegration vectors are normalized such that $\beta = (I_r, -b')'$ where b is the regression coefficient matrix from a regression of X_{1t} on X_{2t} controlling for deterministic components D_t (see Eq. (16.8a)).[12] The least-squares estimate of b is (super) consistent as already noted in Sect. 7.5.2. However, the estimator is not directly suitable for hypothesis testing because the conventional test statistics do not have the usual asymptotic distributions. The idea of Phillips and Hansen (1990) is to correct the conventional least-squares estimates to account for serial correlation and for the endogeneity arising from the cointegrating relationship. This leads to the fully-modified ordinary least-squares estimator (FMOLS estimator).

[12] The choice of the variables used for normalization turns out to be important in practice. See the application in Sect. 16.5.

As the endogeneity shows up in the long-run correlation between the variables, the proposed modification uses of the long-run variance J of $u_t = (u'_{1t}, u'_{2t})'$. According to Sect. 11.1 this entity defined as:

$$J = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} = \sum_{h=-\infty}^{\infty} \Gamma(h) = \Lambda + \Lambda' - \Sigma$$

where

$$\Lambda = \sum_{h=0}^{\infty} \Gamma(h) = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix},$$

$$\Sigma = \mathbb{E}(u_t u'_t) = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

The fully-modified ordinary least-squares estimator of (b, π_1) is then constructed as follows. Estimate the Eqs. (16.8a) and (16.8b) by ordinary least-squares to compute the residuals $\hat{u}_t = (\hat{u}'_{1t}, \hat{u}'_{2t})'$. From these residuals estimate Σ as $\widehat{\Sigma} = \sum_{t=1}^{T} u_t u'_t$ and the long-run variance J and its one-sided counterpart Λ. Estimates of J and Λ, denoted by $\hat{J}$ and $\hat{\Lambda}$, can be obtained by applying a kernel estimator as explained in Sect. 4.4, respectively Sect. 11.1. The estimators $\widehat{\Sigma}$, $\hat{J}$ and $\hat{\Lambda}$ are consistent because ordinary least-squares is. The corresponding estimates are then used to correct the data for X_{1t} and to construct the bias correction term $\hat{\Lambda}_{21}^{(+)}$:

$$X_{1t}^{(+)} = X_{1t} - \hat{J}_{12}\hat{J}_{22}^{-1}\hat{u}_{2t},$$
$$\hat{\Lambda}_{21}^{(+)} = \hat{\Lambda}_{21} - \hat{J}_{12}\hat{J}_{22}^{-1}\hat{\Lambda}_{22}.$$

The fully-modified ordinary least-squares estimator (FMOLS estimator) is then given by

$$\left(\hat{b}, \hat{\pi}_1\right) = \left(\left(\sum_{t=1}^{T} X_{1t}^{(+)}(X'_{2t}, D'_t)\right) - T(\hat{\Lambda}_{21}^{(+)'}, 0)\right) \left(\sum_{t=1}^{T}(X'_{2t}, D'_t)'(X'_{2t}, D'_t)\right)^{-1}.$$

It turns out that this estimator is asymptotically equivalent to full maximum likelihood with limiting distributions free of nuisance parameters.

The main advantage of the FMOLS estimator is that conventional Wald test statistics, appropriately modified, have limiting χ^2 distributions. This brings statistical inference back to the realm of traditional econometric analysis. Consider testing the null hypothesis

$$\mathrm{H}_0 : R \operatorname{vec} b = q.$$

where q is a vector of dimension g and R selects the appropriate elements of $\operatorname{vec} b$. Thus, in effect we are considering hypothesis of the form $\mathrm{H}_0 : b = b_0$. The hypothesis $b = 0$ is thereby of particular interest. The Wald test statistic is then defined as

$$W = (R \operatorname{vec} \hat{b} - q)' \left[R \left(\hat{J}_{11.2} \otimes \left(\sum_{t=1}^{T} (X'_{2t}, D'_t)'(X'_{2t}, D'_t) \right)^{-1} \right) R' \right]^{-1} (R \operatorname{vec} \hat{b} - q)$$

where $\hat{J}_{11.2} = \hat{J}_{11} - \hat{J}_{12}\hat{J}_{22}^{-1}\hat{J}_{21}$. It can be shown that the so defined modified Wald test statistic is asymptotically distributed as χ^2 with g degrees of freedom (see Phillips and Hansen 1990; Hansen 1992).

16.5 An Example

This example reproduces the study by Neusser (1991) with actualized data for the United States over the period first quarter 1950 to fourth quarter 2005. The starting point is a VAR model which consists of four variables: real gross domestic product (Y), real private consumption (C), real gross investment (I), and the ex-post real interest rate (R). All variables, except the real interest rate, are in logs. First, we identify a VAR model for these variables where the order is determined by Akaike's (AIC), Schwarz' (BIC) or Hannan-Quinn' (HQ) information criteria. The AIC suggests seven lags whereas the other criteria propose a VAR of order two. As the VAR(7) consists of many statistically insignificant coefficients, we prefer the more parsimonious VAR(2) model which produces the following estimates:

$$X_t = \begin{pmatrix} Y_t \\ C_t \\ I_t \\ R_t \end{pmatrix} = \begin{pmatrix} 0.185 \\ (0.047) \\ 0.069 \\ (0.043) \\ 0.041 \\ (0.117) \\ -0.329 \\ (0.097) \end{pmatrix} + \begin{pmatrix} 0.951 & 0.254 & 0.088 & 0.042 \\ (0.086) & (0.091) & (0.033) & (0.032) \\ 0.157 & 0.746 & 0.065 & -0.013 \\ (0.079) & (0.084) & (0.031) & (0.030) \\ 0.283 & 0.250 & 1.304 & 0.026 \\ (0.216) & (0.229) & (0.084) & (0.081) \\ 0.324 & -0.536 & -0.024 & 0.551 \\ (0.178) & (0.189) & (0.069) & (0.067) \end{pmatrix} X_{t-1}$$

$$+\begin{pmatrix} \begin{matrix}-0.132\\(0.085)\end{matrix} & \begin{matrix}-0.085\\(0.093)\end{matrix} & \begin{matrix}-0.089\\(0.033)\end{matrix} & \begin{matrix}-0.016\\(0.031)\end{matrix} \\ \begin{matrix}-0.213\\(0.078)\end{matrix} & \begin{matrix}0.305\\(0.085)\end{matrix} & \begin{matrix}-0.066\\(0.031)\end{matrix} & \begin{matrix}0.112\\(0.029)\end{matrix} \\ \begin{matrix}-0.517\\(0.214)\end{matrix} & \begin{matrix}0.040\\(0.233)\end{matrix} & \begin{matrix}-0.364\\(0.084)\end{matrix} & \begin{matrix}0.098\\(0.079)\end{matrix} \\ \begin{matrix}-0.042\\(0.176)\end{matrix} & \begin{matrix}0.296\\(0.192)\end{matrix} & \begin{matrix}0.005\\(0.069)\end{matrix} & \begin{matrix}0.163\\(0.065)\end{matrix} \end{pmatrix} X_{t-2} + Z_t$$

where the estimated standard errors of the corresponding coefficients are reported in parenthesis. The estimate covariance matrix Σ, $\widehat{\Sigma}$, is

$$\widehat{\Sigma} = 10^{-4} \begin{pmatrix} 0.722 & 0.428 & 1.140 & 0.002 \\ 0.428 & 0.610 & 1.026 & -0.092 \\ 1.140 & 1.026 & 4.473 & -0.328 \\ 0.002 & -0.092 & -0.328 & 3.098 \end{pmatrix}.$$

The sequence of the hypotheses starts with H(0) which states that there exists no cointegrating relation. The alternative hypothesis is always H(n) which says that there are n cointegrating relations. According to Table 16.2 the value of the trace test statistic is 111.772 which is clearly larger than the 5 % critical value of 47.856. Thus, the null hypothesis H(0) is rejected and we consider next the hypothesis **H**(1). This hypothesis is again clearly rejected so that we move on to the hypothesis H(2). Because H(3) is not rejected, we conclude that there exists 3 cointegrating relations. To check this result, we test the hypothesis H(2) against H(3) using the max test. As this test also rejects H(2), we can be pretty confident that there are three cointegrating relations given as:

$$\hat{\beta} = \begin{pmatrix} 1.000 & 0.000 & 0.000 \\ 0.000 & 1.000 & 0.000 \\ 0.000 & 0.000 & 1.000 \\ -258.948 & -277.869 & -337.481 \end{pmatrix}.$$

Table 16.2 Evaluation of the results of Johansen's cointegration test

		Trace statistic		Max statistic	
Null hypothesis	Eigenvalue	Test statistic	Critical value	Test statistic	Critical value
H(0) : $r = 0$	0.190	111.772	47.856	47.194	27.584
H(1) : $r \leq 1$	0.179	64.578	29.797	44.075	21.132
H(2) : $r \leq 2$	0.081	20.503	15.495	18.983	14.265
H(3) : $r \leq 3$	0.007	1.520	3.841	1.520	3.841

Critical 5 % values are taken from MacKinnon et al. (1999)

This matrix is actually the outcome from the EVIEWS econometrics software package. It should be noted that EVIEWS, like other packages, chooses the normalization mechanically. This can become a problem if the variable on which the cointegration vectors are normalized is not part of the cointegrating relation.

In this form, the cointegrating vectors are economically difficult to interpret. We therefore ask whether they are compatible with the following hypotheses:

$$\beta_C = \begin{pmatrix} 1.0 \\ -1.0 \\ 0.0 \\ 0.0 \end{pmatrix}, \qquad \beta_I = \begin{pmatrix} 1.0 \\ 0.0 \\ -1.0 \\ 0.0 \end{pmatrix}, \qquad \beta_R = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}.$$

These hypotheses state that the log-difference (ratio) between consumption and GDP, the log-difference (ratio) between investment and GDP, and the real interest rate are stationary. They can be rationalized in the context of the neoclassical growth model (see King et al. 1991; Neusser 1991). Each of them can be brought into the form of Eq. (16.16) where β is replaced by its estimate $\hat{\beta}$. The corresponding test statistics for each of the three cointegrating relations is distributed as a χ^2 distribution with one degree of freedom,[13] which gives a critical value of 3.84 at the 5 % significance level. The corresponding values for the test statistic are 12.69, 15.05 and 0.45, respectively. This implies that we must reject the first two hypotheses β_C and β_I. However, the conjecture that the real interest is stationary, cannot be rejected. Finally, we can investigate the joint hypothesis $\beta_0 = (\beta_C, \beta_I, \beta_R)$ which can be represented in the form (16.15). In this case the value of the test statistic is 41.20 which is clearly above the critical value of 7.81 inferred from the χ^2_3 distribution.[14] Thus, we must reject this joint hypothesis.

As a matter of comparison, we perform a similar investigation using the fully-modified approach of Phillips and Hansen (1990). For this purpose we restrict the analysis to Y_t, C_t, and I_t because the real interest rate cannot be classified unambiguously as being stationary, respectively integrated of order one. The long-run variance J and its one-sided counterpart Λ are estimated using the quadratic spectral kernel with VAR(1) prewhitening as advocated by Andrews and Monahan (1992) (see Sect. 4.4). Assuming two cointegrating relations and taking Y_t and C_t as the left hand side variables in the cointegrating regression (Eq. (16.8a)), the following results are obtained:

$$\begin{pmatrix} Y_t \\ C_t \end{pmatrix} = \begin{pmatrix} 0.234 \\ (0.166) \\ 0.215 \\ (0.171) \end{pmatrix} I_t + \begin{pmatrix} 6.282 \\ (0.867) \\ 5.899 \\ (0.892) \end{pmatrix} + \begin{pmatrix} 0.006 \\ (0.002) \\ 0.007 \\ (0.002) \end{pmatrix} t + \hat{u}_{1t}$$

[13] The degrees of freedom are computed according to the formula: $s(n-r) = 1(4-3) = 1$.

[14] The degrees of freedom are computed according to the formula: $r(n-s) = 3(4-3) = 3$.

where the estimated standard deviations are reported in parenthesis. The specification allows for a constant and a deterministic trend as well as a drift in the equation for ΔI_t (Eq. (16.8b), not shown).

Given these results we can test a number of hypotheses to get a better understanding of the cointegrating relations. First we test the hypothesis of no cointegration of Y_t, respectively C_t with I_t. Thus, we test $\mathrm{H}_0 : b(1) = b(2) = 0$. The value of the corresponding Wald test statistic is equal to 2.386 which is considerably less than the 5 % critical value of 5.992. Therefore we can not reject the null hypothesis no cointegration Another interesting hypothesis is $\mathrm{H}_0 : b(1) = b(2)$ which would mean that Y_t and C_t are cointegrated with cointegration vector $(1, -1)$. As the corresponding Wald statistic is equal to 0.315, this hypothesis can not be rejected at the 5 % critical value of 3.842. This suggests a long-run relation between Y_t and C_t.

Repeating the analysis with C_t and I_t as the left hand side variables leads to the following results:

$$\begin{pmatrix} C_t \\ I_t \end{pmatrix} = \begin{pmatrix} 0.834 \\ (0.075) \\ 2.192 \\ (0.680) \end{pmatrix} Y_t + \begin{pmatrix} 0.767 \\ (0.561) \\ -11.27 \\ (5.102) \end{pmatrix} + \begin{pmatrix} 0.002 \\ (0.001) \\ -0.008 \\ (0.006) \end{pmatrix} t + \hat{u}_{1t}$$

As before first the hypothesis $H_0 : b(1) = b(2) = 0$ is tested. The corresponding value of the test statistic is 137.984 which is clearly above the 5 % critical value. Thus, the null hypothesis of no cointegration is rejected. Next, the hypothesis $H_0 : b(1) = b(2) = 1$ is tested. This hypothesis is rejected as the value 7.717 of the test statistic is above the critical value. If these hypotheses are not tested jointly, but individually, the null hypothesis $b(1) = 1$ can be rejected, but $b(2) = 1$ can not. These findings conform reasonably well with those based on the Johansen approach.

The diverse result between the two specifications demonstrates that the sensitivity of cointegration analysis with respect to the normalization. It is important that the variable on which the cointegrating vector is normalized is indeed in the cointegrating space. Otherwise, insensible results may be obtained.