12
Cointegration

12.1 Introduction

The regression theory of Chapter 6 and the VAR models discussed in the previous chapter are appropriate for modeling $I(0)$ data, like asset returns or growth rates of macroeconomic time series. Economic theory often implies equilibrium relationships between the levels of time series variables that are best described as being $I(1)$. Similarly, arbitrage arguments imply that the $I(1)$ prices of certain financial time series are linked. This chapter introduces the statistical concept of cointegration that is required to make sense of regression models and VAR models with $I(1)$ data.

The chapter is organized as follows. Section 12.2 gives an overview of the concepts of spurious regression and cointegration, and introduces the error correction model as a practical tool for utilizing cointegration with financial time series. Section 12.3 discusses residual-based tests for cointegration. Section 12.4 covers regression-based estimation of cointegrating vectors and error correction models. In Section 12.5, the connection between VAR models and cointegration is made, and Johansen's maximum likelihood methodology for cointegration modeling is outlined. Some technical details of the Johansen methodology are provided in the appendix to this chapter.

Excellent textbook treatments of the statistical theory of cointegration are given in Hamilton (1994), Johansen (1995) and Hayashi (2000). Applications of cointegration to finance may be found in Campbell, Lo and

MacKinlay (1997), Mills (1999), Alexander (2001), Cochrane (2001) and Tsay (2001).

12.2 Spurious Regression and Cointegration

12.2.1 Spurious Regression

The time series regression model discussed in Chapter 6 required all variables to be $I(0)$. In this case, the usual statistical results for the linear regression model hold. If some or all of the variables in the regression are $I(1)$ then the usual statistical results may or may not hold[1]. One important case in which the usual statistical results do not hold is *spurious regression* when all the regressors are $I(1)$ and not cointegrated. The following example illustrates.

Example 71 *An illustration of spurious regression using simulated data*

Consider two independent and not cointegrated $I(1)$ processes y_{1t} and y_{2t} such that

$$y_{it} = y_{it-1} + \varepsilon_{it}, \text{ where } \varepsilon_{it} \sim GWN(0,1), \ i = 1,2$$

Following Granger and Newbold (1974), 250 observations for each series are simulated and plotted in Figure 12.1 using

```
> set.seed(458)
> e1 = rnorm(250)
> e2 = rnorm(250)
> y1 = cumsum(e1)
> y2 = cumsum(e2)
> tsplot(y1, y2, lty=c(1,3))
> legend(0, 15, c("y1","y2"), lty=c(1,3))
```

The data in the graph resemble stock prices or exchange rates. A visual inspection of the data suggests that the levels of the two series are positively related. Regressing y_{1t} on y_{2t} reinforces this observation:

```
> summary(OLS(y1~y2))

Call:
OLS(formula = y1 ~y2)
```

---

[1] A systematic technical analysis of the linear regression model with $I(1)$ and $I(0)$ variables is given in Sims, Stock and Watson (1990). Hamilton (1994) gives a nice summary of these results and Stock and Watson (1989) provides useful intuition and examples.
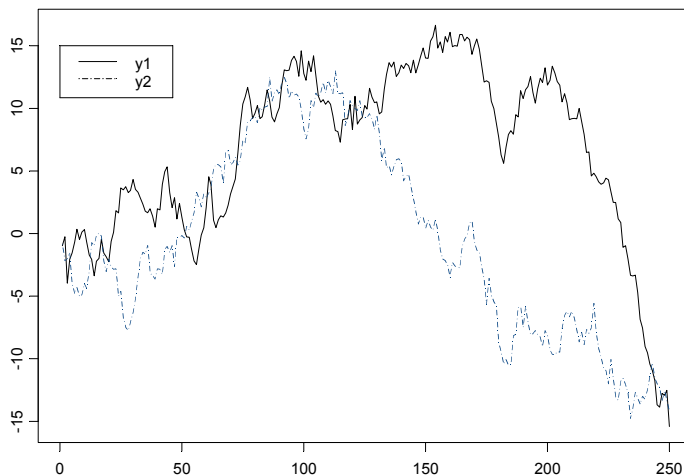
FIGURE 12.1. Two simulated independent $I(1)$ processes.

```
Residuals:
     Min       1Q  Median       3Q      Max
 -16.360   -4.352  -0.128    4.979   10.763


Coefficients:
             Value Std. Error t value Pr(>|t|)
(Intercept)  6.7445  0.3943    17.1033  0.0000
        y2  0.4083  0.0508     8.0352  0.0000


Regression Diagnostics:

        R-Squared 0.2066
Adjusted R-Squared 0.2034
Durbin-Watson Stat 0.0328


Residual standard error: 6.217 on 248 degrees of freedom
F-statistic: 64.56 on 1 and 248 degrees of freedom, the
p-value is 3.797e-014
```

The estimated slope coefficient is 0.408 with a large t-statistic of 8.035 and the regression R^2 is moderate at 0.201. The only suspicious statistic is the very low Durbin-Watson statistic suggesting strong residual auto-correlation. These statistics are representative of the spurious regression

phenomenon with $I(1)$ that are not cointegrated. If Δy_{1t} is regressed on Δy_{2t} the correct relationship between the two series is revealed

```
> summary(OLS(diff(y1)~diff(y2)))

Call:
OLS(formula = diff(y1) ~diff(y2))

Residuals:
     Min      1Q  Median      3Q     Max
 -3.6632 -0.7706 -0.0074  0.6983  2.7184

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -0.0565  0.0669     -0.8447  0.3991
   diff(y2)  0.0275  0.0642      0.4290  0.6683

Regression Diagnostics:

         R-Squared  0.0007
Adjusted R-Squared -0.0033
Durbin-Watson Stat  1.9356

Residual standard error: 1.055 on 247 degrees of freedom
F-statistic: 0.184 on 1 and 247 degrees of freedom, the
p-value is 0.6683
```

Similar results to those above occur if $cov(\varepsilon_{1t}, \varepsilon_{2t}) \neq 0$. The levels regression remains spurious (no real long-run common movement in levels), but the differences regression will reflect the non-zero contemporaneous correlation between Δy_{1t} and Δy_{2t}.

Statistical Implications of Spurious Regression

Let $\mathbf{Y}_t = (y_{1t}, \ldots, y_{nt})'$ denote an $(n \times 1)$ vector of $I(1)$ time series that are not cointegrated. Using the partition $\mathbf{Y}_t = (y_{1t}, \mathbf{Y}_{2t}')'$, consider the least squares regression of y_{1t} on $\mathbf{Y}_{2t}$ giving the fitted model

$$y_{1t} = \hat{\boldsymbol{\beta}}_2' \mathbf{Y}_{2t} + \hat{u}_t \tag{12.1}$$

Since y_{1t} is not cointegrated with $\mathbf{Y}_{2t}$ (12.1) is a *spurious regression* and the true value of $\boldsymbol{\beta}_2$ is zero. The following results about the behavior of $\hat{\boldsymbol{\beta}}_2$ in the spurious regression (12.1) are due to Phillips (1986):

- $\hat{\boldsymbol{\beta}}_2$ does not converge in probability to zero but instead converges in distribution to a non-normal random variable not necessarily centered at zero. This is the spurious regression phenomenon.

- The usual OLS t-statistics for testing that the elements of $\boldsymbol{\beta}_2$ are zero diverge to $\pm\infty$ as $T \to \infty$. Hence, with a large enough sample it will appear that $\mathbf{Y}_t$ is cointegrated when it is not if the usual asymptotic normal inference is used.

- The usual R^2 from the regression converges to unity as $T \to \infty$ so that the model will appear to fit well even though it is misspecified.

- Regression with $I(1)$ data only makes sense when the data are cointegrated.

12.2.2 Cointegration

Let $\mathbf{Y}_t = (y_{1t}, \ldots, y_{nt})'$ denote an $(n \times 1)$ vector of $I(1)$ time series. $\mathbf{Y}_t$ is *cointegrated* if there exists an $(n \times 1)$ vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)'$ such that

$$\boldsymbol{\beta}'\mathbf{Y}_t = \beta_1 y_{1t} + \cdots + \beta_n y_{nt} \sim I(0) \tag{12.2}$$

In words, the nonstationary time series in $\mathbf{Y}_t$ are cointegrated if there is a linear combination of them that is stationary or $I(0)$. If some elements of $\boldsymbol{\beta}$ are equal to zero then only the subset of the time series in $\mathbf{Y}_t$ with non-zero coefficients is cointegrated. The linear combination $\boldsymbol{\beta}'\mathbf{Y}_t$ is often motivated by economic theory and referred to as a *long-run equilibrium* relationship. The intuition is that $I(1)$ time series with a long-run equilibrium relationship cannot drift too far apart from the equilibrium because economic forces will act to restore the equilibrium relationship.

Normalization

The cointegration vector $\boldsymbol{\beta}$ in (12.2) is not unique since for any scalar c the linear combination $c\boldsymbol{\beta}'\mathbf{Y}_t = \boldsymbol{\beta}^{*\prime}\mathbf{Y}_t \sim I(0)$. Hence, some *normalization* assumption is required to uniquely identify $\boldsymbol{\beta}$. A typical normalization is

$$\boldsymbol{\beta} = (1, -\beta_2, \ldots, -\beta_n)'$$

so that the cointegration relationship may be expressed as

$$\boldsymbol{\beta}'\mathbf{Y}_t = y_{1t} - \beta_2 y_{2t} - \cdots - \beta_n y_{nt} \sim I(0)$$

or

$$y_{1t} = \beta_2 y_{2t} + \cdots + \beta_n y_{nt} + u_t \tag{12.3}$$

where $u_t \sim I(0)$. In (12.3), the error term u_t is often referred to as the *disequilibrium error* or the *cointegrating residual*. In long-run equilibrium, the disequilibrium error u_t is zero and the long-run equilibrium relationship is

$$y_{1t} = \beta_2 y_{2t} + \cdots + \beta_n y_{nt}$$

Multiple Cointegrating Relationships

If the $(n \times 1)$ vector $\mathbf{Y}_t$ is cointegrated there may be $0 < r < n$ linearly independent cointegrating vectors. For example, let $n = 3$ and suppose there are $r = 2$ cointegrating vectors $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}, \beta_{13})'$ and $\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22}, \beta_{23})'$. Then $\boldsymbol{\beta}_1' \mathbf{Y}_t = \beta_{11} y_{1t} + \beta_{12} y_{2t} + \beta_{13} y_{3t} \sim I(0)$, $\boldsymbol{\beta}_2' \mathbf{Y}_t = \beta_{21} y_{1t} + \beta_{22} y_{2t} + \beta_{23} y_{3t} \sim I(0)$ and the (3×2) matrix

$$\mathbf{B}' = \left(\begin{array}{c} \boldsymbol{\beta}_1' \\ \boldsymbol{\beta}_2' \end{array} \right) = \left(\begin{array}{ccc} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{33} \end{array} \right)$$

forms a *basis* for the space of cointegrating vectors. The linearly independent vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in the cointegrating basis $\mathbf{B}$ are not unique unless some normalization assumptions are made. Furthermore, any linear combination of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, e.g. $\boldsymbol{\beta}_3 = c_1 \boldsymbol{\beta}_1 + c_2 \boldsymbol{\beta}_2$ where c_1 and c_2 are constants, is also a cointegrating vector.

Examples of Cointegration and Common Trends in Economics and Finance

Cointegration naturally arises in economics and finance. In economics, cointegration is most often associated with economic theories that imply equilibrium relationships between time series variables. The permanent income model implies cointegration between consumption and income, with consumption being the common trend. Money demand models imply cointegration between money, income, prices and interest rates. Growth theory models imply cointegration between income, consumption and investment, with productivity being the common trend. Purchasing power parity implies cointegration between the nominal exchange rate and foreign and domestic prices. Covered interest rate parity implies cointegration between forward and spot exchange rates. The Fisher equation implies cointegration between nominal interest rates and inflation. The expectations hypothesis of the term structure implies cointegration between nominal interest rates at different maturities. The equilibrium relationships implied by these economic theories are referred to as *long-run equilibrium* relationships, because the economic forces that act in response to deviations from equilibriium may take a long time to restore equilibrium. As a result, cointegration is modeled using long spans of low frequency time series data measured monthly, quarterly or annually.

In finance, cointegration may be a high frequency relationship or a low frequency relationship. Cointegration at a high frequency is motivated by arbitrage arguments. The *Law of One Price* implies that identical assets must sell for the same price to avoid arbitrage opportunities. This implies cointegration between the prices of the same asset trading on different markets, for example. Similar arbitrage arguments imply cointegration between spot and futures prices, and spot and forward prices, and bid and

ask prices. Here the terminology long-run equilibrium relationship is somewhat misleading because the economic forces acting to eliminate arbitrage opportunities work very quickly. Cointegration is appropriately modeled using short spans of high frequency data in seconds, minutes, hours or days. Cointegration at a low frequency is motivated by economic equilibrium theories linking assets prices or expected returns to fundamentals. For example, the present value model of stock prices states that a stock's price is an expected discounted present value of its expected future dividends or earnings. This links the behavior of stock prices at low frequencies to the behavior of dividends or earnings. In this case, cointegration is modeled using low frequency data and is used to explain the long-run behavior of stock prices or expected returns.

12.2.3   Cointegration and Common Trends

If the $(n \times 1)$ vector time series $\mathbf{Y}_t$ is cointegrated with $0 < r < n$ cointegrating vectors then there are $n - r$ *common $I(1)$ stochastic trends*. To illustrate the duality between cointegration and common trends, let $\mathbf{Y}_t = (y_{1t}, y_{2t})' \sim I(1)$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})' \sim I(0)$ and suppose that $\mathbf{Y}_t$ is cointegrated with cointegrating vector $\boldsymbol{\beta} = (1, -\beta_2)'$. This cointegration relationship may be represented as

$$
y_{1t} = \beta_2 \sum_{s=1}^{t} \varepsilon_{1s} + \varepsilon_{3t}
$$

$$
y_{2t} = \sum_{s=1}^{t} \varepsilon_{1s} + \varepsilon_{2t}
$$

The common stochastic trend is $\sum_{s=1}^{t} \varepsilon_{1s}$. Notice that the cointegrating relationship annihilates the common stochastic trend:

$$
\boldsymbol{\beta}' \mathbf{Y}_t = \beta_2 \sum_{s=1}^{t} \varepsilon_{1s} + \varepsilon_{3t} - \beta_2 \left(\sum_{s=1}^{t} \varepsilon_{1s} + \varepsilon_{2t} \right) = \varepsilon_{3t} - \beta_2 \varepsilon_{2t} \sim I(0).
$$

12.2.4   Simulating Cointegrated Systems

Cointegrated systems may be conveniently simulated using Phillips' (1991) *triangular representation*. For example, consider a bivariate cointegrated system for $\mathbf{Y}_t = (y_{1t}, y_{2t})'$ with cointegrating vector $\boldsymbol{\beta} = (1, -\beta_2)'$. A triangular representation has the form

$$
y_{1t} = \beta_2 y_{2t} + u_t, \text{ where } u_t \sim I(0) \tag{12.4}
$$

$$
y_{2t} = y_{2t-1} + v_t, \text{ where } v_t \sim I(0) \tag{12.5}
$$

The first equation describes the long-run equilibrium relationship with an $I(0)$ disequilibrium error u_t. The second equation specifies y_{2t} as the common stochastic trend with innovation v_t:

$$y_{2t} = y_{20} + \sum_{j=1}^{t} v_j.$$

In general, the innovations u_t and v_t may be contemporaneously and serially correlated. The time series structure of these innovations characterizes the short-run dynamics of the cointegrated system. The system (12.4)-(12.5) with $\beta_2 = 1$, for example, might be used to model the behavior of the logarithm of spot and forward prices, spot and futures prices or stock prices and dividends.

Example 72 *Simulated bivariate cointegrated system*

Consider simulating $T = 250$ observations from the system (12.4)-(12.5) using $\boldsymbol{\beta} = (1, -1)'$, $u_t = 0.75u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iid\, N(0, 0.5^2)$ and $v_t \sim iid\, N(0, 0.5^2)$. The S-PLUS code is

```
> set.seed(432)
> e = rmvnorm(250, mean=rep(0,2), sd=c(0.5,0.5))
> u.ar1 = arima.sim(model=list(ar=0.75), innov=e[,1])
> y2 = cumsum(e[,2])
> y1 = y2 + u.ar1
> par(mfrow=c(2,1))
> tsplot(y1, y2, lty=c(1,3),
+ main="Simulated bivariate cointegrated system",
+ sub="1 cointegrating vector, 1 common trend")
> legend(0, 7, legend=c("y1","y2"), lty=c(1,3))
> tsplot(u.ar1, main="Cointegrating residual")
```

Figure 12.2 shows the simulated data for y_{1t} and y_{2t} along with the cointegrating residual $u_t = y_{1t} - y_{2t}$. Since y_{1t} and y_{2t} share a common stochastic trend they follow each other closely. The impulse response function for u_t may be used to determine the speed of adjustment to long-run equilibrium. Since u_t is an AR(1) with $\phi = 0.75$ the half life of a shock is $\ln(0.5)/\ln(0.75) = 2.4$ time periods.

Next, consider a trivariate cointegrated system for $\mathbf{Y}_t = (y_{1t}, y_{2t}, y_{3t})'$. With a trivariate system there may be one or two cointegrating vectors. With one cointegrating vector there are two common stochastic trends and with two cointegrating vectors there is one common trend. A triangular representation with one cointegrating vector $\boldsymbol{\beta} = (1, -\beta_2, -\beta_3)'$ and two
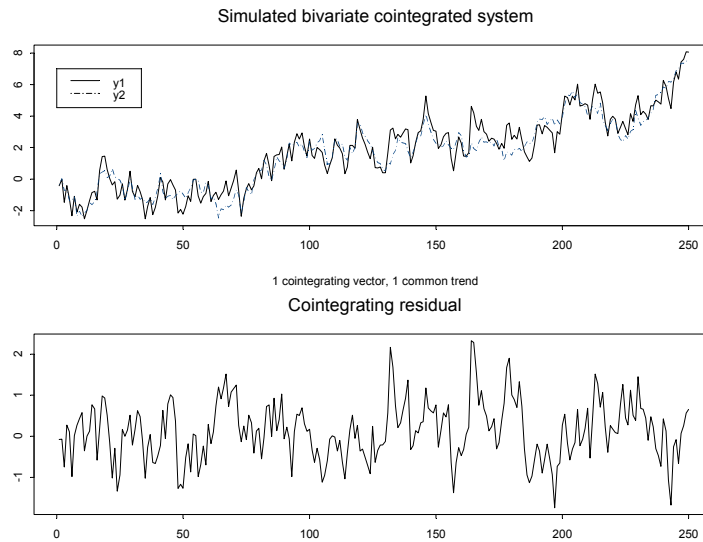
Simulated bivariate cointegrated system

1 cointegrating vector, 1 common trend

Cointegrating residual

FIGURE 12.2. Simulated bivariate cointegrated system with $\boldsymbol{\beta} = (1, -1)'$.

stochastic trends is

$$
\begin{align}
y_{1t} &= \beta_2 y_{2t} + \beta_3 y_{3t} + u_t, \text{ where } u_t \sim I(0) & (12.6) \\
y_{2t} &= y_{2t-1} + v_t, \text{ where } v_t \sim I(0) & (12.7) \\
y_{3t} &= y_{3t-1} + w_t, \text{ where } w_t \sim I(0) & (12.8)
\end{align}
$$

The first equation describes the long-run equilibrium and the second and third equations specify the common stochastic trends. An example of a trivariate cointegrated system with one cointegrating vector is a system of nominal exchange rates, home country price indices and foreign country price indices. A cointegrating vector $\boldsymbol{\beta} = (1, -1, -1)'$ implies that the real exchange rate is stationary.

Example 73 *Simulated trivariate cointegrated system with 1 cointegrating vector*

The S-PLUS code for simulating $T = 250$ observation from (12.6)-(12.8) with $\boldsymbol{\beta} = (1, -0.5, -0.5)'$, $u_t = 0.75u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iid\,N(0, 0.5^2)$, $v_t \sim iid\,N(0, 0.5^2)$ and $w_t \sim iid\,N(0, 0.5^2)$ is

```
> set.seed(573)
> e = rmvnorm(250, mean=rep(0,3), sd=c(0.5,0.5,0.5))
> u1.ar1 = arima.sim(model=list(ar=0.75), innov=e[,1])
> y2 = cumsum(e[,2])
```
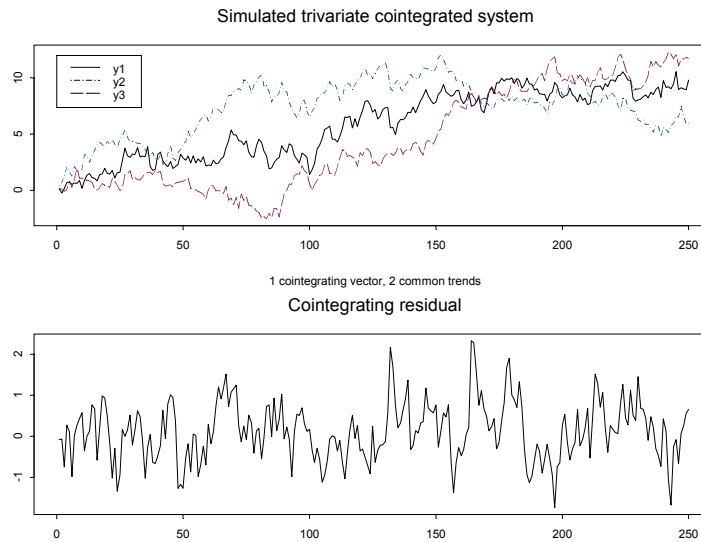
FIGURE 12.3. Simulated trivariate cointegrated system with one cointegrating vector $\boldsymbol{\beta} = (1, -0.5, -0.5)'$ and two stochastic trends.

```
> y3 = cumsum(e[,3])
> y1 = 0.5*y2 + 0.5*y3 + u1.ar1
> par(mfrow=c(2,1))
> tsplot(y1, y2, y3, lty=c(1,3,4),
+ main="Simulated trivariate cointegrated system",
+ sub="1 cointegrating vector, 2 common trends")
> legend(0, 12, legend=c("y1","y2","y3"), lty=c(1,3,4))
> tsplot(u.ar1, main="Cointegrating residual")
```

Figure 12.3 illustrates the simulated data. Here, y_{2t} and y_{3t} are the two independent common trends and $y_{1t} = 0.5y_{2t} + 0.5y_{3t} + u_t$ is the average of the two trends plus an AR(1) residual.

Finally, consider a trivariate cointegrated system with two cointegrating vectors and one common stochastic trend. A triangular representation for this system with cointegrating vectors $\boldsymbol{\beta}_1 = (1, 0, -\beta_{13})'$ and $\boldsymbol{\beta}_2 = (0, 1, -\beta_{23})'$ is

$$y_{1t} = \beta_{13}y_{3t} + u_t, \text{ where } u_t \sim I(0) \tag{12.9}$$

$$y_{2t} = \beta_{23}y_{3t} + v_t, \text{ where } v_t \sim I(0) \tag{12.10}$$

$$y_{3t} = y_{3t-1} + w_t, \text{ where } w_t \sim I(0) \tag{12.11}$$

Here the first two equations describe two long-run equilibrium relations and the third equation gives the common stochastic trend. An example in

finance of such a system is the term structure of interest rates where y_3 represents the short rate and y_1 and y_2 represent two different long rates. The cointegrating relationships would indicate that the spreads between the long and short rates are stationary.

Example 74 *Simulated trivariate cointegrated system with 2 cointegrating vectors*

The S-PLUS code for simulating $T = 250$ observation from (12.9)-(12.11) with $\boldsymbol{\beta}_1 = (1, 0, -1)'$, $\beta_2 = (0, 1, -1)'$, $u_t = 0.75 u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iid\, N(0, 0.5^2)$, $v_t = 0.75 v_{t-1} + \eta_t$, $\eta_t \sim iid\, N(0, 0.5^2)$ and $w_t \sim iid\, N(0, 0.5^2)$ is

```
> set.seed(573)
> e = rmvnorm(250,mean=rep(0,3), sd=c(0.5,0.5,0.5))
> u.ar1 = arima.sim(model=list(ar=0.75), innov=e[,1])
> v.ar1 = arima.sim(model=list(ar=0.75), innov=e[,2])
> y3 = cumsum(e[,3])
> y1 = y3 + u.ar1
> y2 = y3 + v.ar1
> par(mfrow=c(2,1))
> tsplot(y1, y2, y3, lty=c(1,3,4),
+ main="Simulated trivariate cointegrated system",
+ sub="2 cointegrating vectors, 1 common trend")
> legend(0, 10, legend=c("y1","y2","y3"), lty=c(1,3,4))
> tsplot(u.ar1, v.ar1, lty=c(1,3),
+ main="Cointegrated residuals")
> legend(0, -1, legend=c("u","v"), lty=c(1,3))
```

12.2.5 Cointegration and Error Correction Models

Consider a bivariate $I(1)$ vector $\mathbf{Y}_t = (y_{1t}, y_{2t})'$ and assume that $\mathbf{Y}_t$ is cointegrated with cointegrating vector $\boldsymbol{\beta} = (1, -\beta_2)'$ so that $\boldsymbol{\beta}' \mathbf{Y}_t = y_{1t} - \beta_2 y_{2t}$ is $I(0)$. In an extremely influential and important paper, Engle and Granger (1987) showed that cointegration implies the existence of an *error correction model* (ECM) of the form

$$\Delta y_{1t} \;\; = \;\; c_1 + \alpha_1(y_{1t-1} - \beta_2 y_{2t-1}) \qquad\qquad (12.12)$$
$$+ \sum_j \psi_{11}^j \Delta y_{1t-j} + \sum_j \psi_{12}^j \Delta y_{2t-j} + \varepsilon_{1t}$$
$$\Delta y_{2t} \;\; = \;\; c_2 + \alpha_2(y_{1t-1} - \beta_2 y_{2t-1}) \qquad\qquad (12.13)$$
$$+ \sum_j \psi_{21}^j \Delta y_{1t-j} + \sum_j \psi_{22}^2 \Delta y_{2t-j} + \varepsilon_{2t}$$

that describes the dynamic behavior of y_{1t} and y_{2t}. The ECM links the long-run equilibrium relationship implied by cointegration with the short-run dynamic adjustment mechanism that describes how the variables react
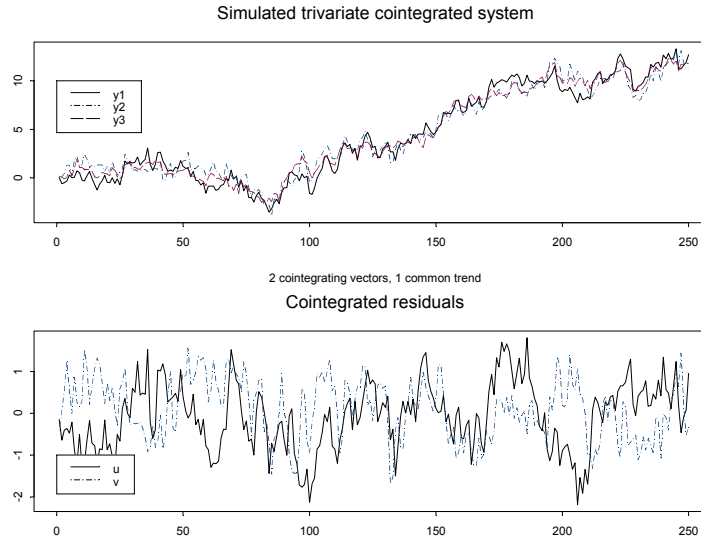
FIGURE 12.4. Simulated trivatiate cointegrated system with two cointegrating vectors $\boldsymbol{\beta}_1 = (1, 0, -1)'$, $\boldsymbol{\beta}_2 = (0, 1, -1)'$ and one common trend.

when they move out of long-run equilibrium. This ECM makes the concept of cointegration useful for modeling financial time series.

Example 75 *Bivariate ECM for stock prices and dividends*

As an example of an ECM, let s_t denote the log of stock prices and d_t denote the log of dividends and assume that $\mathbf{Y}_t = (s_t, d_t)'$ is $I(1)$. If the log dividend-price ratio is $I(0)$ then the logs of stock prices and dividends are cointegrated with $\boldsymbol{\beta} = (1, -1)'$. That is, the long-run equilibrium is

$$d_t = s_t + \mu + u_t$$

where μ is the mean of the log dividend-price ratio, and u_t is an $I(0)$ random variable representing the dynamic behavior of the log dividend-price ratio (disequilibrium error). Suppose the ECM has the form

$$\begin{aligned} \Delta s_t &= c_s + \alpha_s(d_{t-1} - s_{t-1} - \mu) + \varepsilon_{st} \\ \Delta d_t &= c_d + \alpha_d(d_{t-1} - s_{t-1} - \mu) + \varepsilon_{dt} \end{aligned}$$

where $c_s > 0$ and $c_d > 0$. The first equation relates the growth rate of dividends to the lagged disequilibrium error $d_{t-1} - s_{t-1} - \mu$, and the second equation relates the growth rate of stock prices to the lagged disequilibrium as well. The reactions of s_t and d_t to the disequilibrium error are captured by the *adjustment coefficients* α_s and α_d.

Consider the special case of (12.12)-(12.13) where $\alpha_d = 0$ and $\alpha_s = 0.5$. The ECM equations become

$$\begin{aligned}\Delta s_t &= c_s + 0.5(d_{t-1} - s_{t-1} - \mu) + \varepsilon_{st}, \\ \Delta d_t &= c_d + \varepsilon_{dt}.\end{aligned}$$

so that only s_t responds to the lagged disequilibrium error. Notice that $E[\Delta s_t|\mathbf{Y}_{t-1}] = c_s + 0.5(d_{t-1} - s_{t-1} - \mu)$ and $E[\Delta d_t|\mathbf{Y}_{t-1}] = c_d$. Consider three situations:

1. $d_{t-1} - s_{t-1} - \mu = 0$. Then $E[\Delta s_t|\mathbf{Y}_{t-1}] = c_s$ and $E[\Delta d_t|Y_{t-1}] = c_d$, so that c_s and c_d represent the growth rates of stock prices and dividends in long-run equilibrium.

2. $d_{t-1} - s_{t-1} - \mu > 0$. Then $E[\Delta s_t|\mathbf{Y}_{t-1}] = c_s + 0.5(d_{t-1} - s_{t-1} - \mu) > c_s$. Here the dividend yield has increased above its long-run mean (positive disequilibrium error) and the ECM predicts that s_t will grow faster than its long-run rate to restore the dividend yield to its long-run mean. Notice that the magnitude of the adjustment coefficient $\alpha_s = 0.5$ controls the speed at which s_t responds to the disequilibrium error.

3. $d_{t-1} - s_{t-1} - \mu < 0$. Then $E[\Delta s_t|\mathbf{Y}_{t-1}] = c_s + 0.5(d_{t-1} - s_{t-1} - \mu) < c_s$. Here the dividend yield has decreased below its long-run mean (negative disequilibrium error) and the ECM predicts that s_t will grow more slowly than its long-run rate to restore the dividend yield to its long-run mean.

In Case 1, there is no expected adjustment since the model was in long-run equilibrium in the previous period. In Case 2, the model was above long-run equilibrium last period so the expected adjustment in s_t is downward toward equilibrium. In Case 3, the model was below long-run equilibrium last period and so the expected adjustment is upward toward the equilibrium. This discussion illustrates why the model is called an error correction model. When the variables are out of long-run equilibrium, there are economic forces, captured by the adjustment coefficients, that push the model back to long-run equilibrium. The speed of adjustment toward equilibrium is determined by the magnitude of α_s. In the present example, $\alpha_s = 0.5$ which implies that roughly one half of the disequilibrium error is corrected in one time period. If $\alpha_s = 1$ then the entire disequilibrium is corrected in one period. If $\alpha_s = 1.5$ then the correction overshoots the long-run equilibrium.

12.3   Residual-Based Tests for Cointegration

Let the $(n \times 1)$ vector $\mathbf{Y}_t$ be $I(1)$. Recall, $\mathbf{Y}_t$ is cointegrated with $0 < r < n$ cointegrating vectors if there exists an $(r \times n)$ matrix $\mathbf{B}'$ such that

$$\mathbf{B}'\mathbf{Y}_t = \begin{pmatrix} \boldsymbol{\beta}_1'\mathbf{Y}_t \\ \vdots \\ \boldsymbol{\beta}_r'\mathbf{Y}_t \end{pmatrix} = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{rt} \end{pmatrix} \sim I(0)$$

Testing for cointegration may be thought of as testing for the existence of long-run equilibria among the elements of $\mathbf{Y}_t$. Cointegration tests cover two situations:

- There is at most one cointegrating vector

- There are possibly $0 \leq r < n$ cointegrating vectors.

The first case was originally considered by Engle and Granger (1986) and they developed a simple two-step residual-based testing procedure based on regression techniques. The second case was originally considered by Johansen (1988) who developed a sophisticated sequential procedure for determining the existence of cointegration and for determining the number of cointegrating relationships based on maximum likelihood techniques. This section explains Engle and Granger's two-step procedure. Johansen's more general procedure will be discussed later on.

Engle and Granger's two-step procedure for determining if the $(n \times 1)$ vector $\boldsymbol{\beta}$ is a cointegrating vector is as follows:

- Form the cointegrating residual $\boldsymbol{\beta}'\mathbf{Y}_t = u_t$

- Perform a unit root test on u_t to determine if it is $I(0)$.

The null hypothesis in the Engle-Granger procedure is no-cointegration and the alternative is cointegration. There are two cases to consider. In the first case, the proposed cointegrating vector $\boldsymbol{\beta}$ is pre-specified (not estimated). For example, economic theory may imply specific values for the elements in $\boldsymbol{\beta}$ such as $\boldsymbol{\beta} = (1, -1)'$. The cointegrating residual is then readily constructed using the prespecified cointegrating vector. In the second case, the proposed cointegrating vector is estimated from the data and an estimate of the cointegrating residual $\hat{\boldsymbol{\beta}}'\mathbf{Y}_t = \hat{u}_t$ is formed. Tests for cointegration using a pre-specified cointegrating vector are generally much more powerful than tests employing an estimated vector.

12.3.1   Testing for Cointegration When the Cointegrating Vector Is Pre-specified

Let $\mathbf{Y}_t$ denote an $(n \times 1)$ vector of $I(1)$ time series, let $\boldsymbol{\beta}$ denote an $(n \times 1)$ prespecified cointegrating vector and let $u_t = \boldsymbol{\beta}'\mathbf{Y}_t$ denote the prespecified
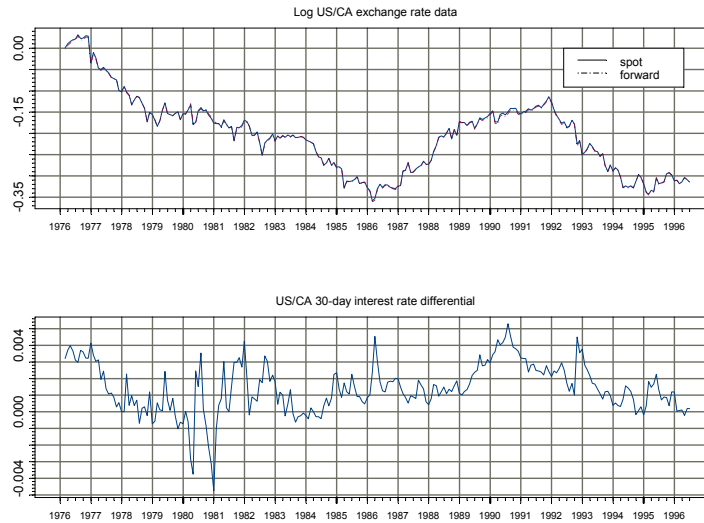
FIGURE 12.5. Log of US/CA spot and 30-day exchange rates and 30-day interest rate differential.

cointegrating residual. The hypotheses to be tested are

$$H_0 \quad : \quad u_t = \boldsymbol{\beta}' \mathbf{Y}_t \sim I(1) \text{ (no cointegration)} \qquad (12.14)$$
$$H_1 \quad : \quad u_t = \boldsymbol{\beta}' \mathbf{Y}_t \sim I(0) \text{ (cointegration)}$$

Any unit root test statistic may be used to evaluate the above hypotheses. The most popular choices are the ADF and PP statistics. Cointegration is found if the unit root test rejects the no-cointegration null. It should be kept in mind, however, that the cointegrating residual may include deterministic terms (constant or trend) and the unit root tests should account for these terms accordingly. See Chapter 4 for details about the application of unit root tests.

Example 76 *Testing for cointegration between spot and forward exchange rates using a known cointegrating vector*

In international finance, the covered interest rate parity arbitrage relationship states that the difference between the logarithm of spot and forward exchange rates is equal to the difference between nominal domestic and foreign interest rates. It seems reasonable to believe that interest rate spreads are $I(0)$ which implies that spot and forward rates are cointegrated with cointegrating vector $\boldsymbol{\beta} = (1, -1)'$. To illustrate, consider the log monthly spot, s_t, and 30 day forward, f_t, exchange rates between the

US and Canada over the period February 1976 through June 1996 taken
from the S+FinMetrics "timeSeries" object lexrates.dat

```
> uscn.s = lexrates.dat[,"USCNS"]
> uscn.s@title = "Log of US/CA spot exchange rate"
> uscn.f = lexrates.dat[,"USCNF"]
> uscn.f@title = "Log of US/CA 30-day forward exchange rate"
> u = uscn.s - uscn.f
> colIds(u) = "USCNID"
> u@title = "US/CA 30-day interest rate differential"
```

The interest rate differential is constructed using the pre-specified cointe-
grating vector $\boldsymbol{\beta} = (1, -1)'$ as $u_t = s_t - f_t$. The spot and forward exchange
rates and interest rate differential are illustrated in Figure 12.5. Visually,
the spot and forward exchange rates clearly share a common trend and the
interest rate differential appears to be $I(0)$. In addition, there is no clear de-
terministic trend behavior in the exchange rates. The S+FinMetrics func-
tion unitroot may be used to test the null hypothesis that s_t and f_t are
not cointegrated ($u_t \sim I(1)$). The ADF t-test based on 11 lags and a con-
stant in the test regression leads to the rejection at the 5% level of the
hypothesis that s_t and f_t are not cointegrated with cointegrating vector
$\boldsymbol{\beta} = (1, -1)'$:

```
> unitroot(u, trend="c", method="adf", lags=11)

Test for Unit Root: Augmented DF Test

Null Hypothesis: there is a unit root
   Type of Test: t-test
 Test Statistic: -2.881
        P-value: 0.04914

Coefficients:
    lag1    lag2    lag3    lag4    lag5    lag6    lag7
 -0.1464 -0.1171 -0.0702 -0.1008 -0.1234 -0.1940  0.0128

    lag8    lag9   lag10   lag11 constant
 -0.1235  0.0550  0.2106 -0.1382  0.0002

Degrees of freedom: 234 total; 222 residual
Time period: from Jan 1977 to Jun 1996
Residual standard error: 8.595e-4
```

12.3.2 Testing for Cointegration When the Cointegrating Vector Is Estimated

Let $\mathbf{Y}_t$ denote an $(n \times 1)$ vector of $I(1)$ time series and let $\boldsymbol{\beta}$ denote an $(n \times 1)$ unknown cointegrating vector. The hypotheses to be tested are given in (12.14). Since $\boldsymbol{\beta}$ is unknown, to use the Engle-Granger procedure it must be first estimated from the data. Before $\boldsymbol{\beta}$ can be estimated some normalization assumption must be made to uniquely identify it. A common normalization is to specify the first element in $\mathbf{Y}_t$ as the dependent variable and the rest as the explanatory variables. Then $\mathbf{Y}_t = (y_{1t}, \mathbf{Y}'_{2t})'$ where $\mathbf{Y}_{2t} = (y_{2t}, \ldots, y_{nt})'$ is an $((n-1) \times 1)$ vector and the cointegrating vector is normalized as $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}'_2)'$. Engle and Granger propose estimating the normalized cointegrating vector $\boldsymbol{\beta}_2$ by least squares from the regression

$$y_{1t} = c + \boldsymbol{\beta}'_2 \mathbf{Y}_{2t} + u_t \tag{12.15}$$

and testing the no-cointegration hypothesis (12.14) with a unit root test using the estimated cointegrating residual

$$\hat{u}_t = y_{1t} - \hat{c} - \hat{\boldsymbol{\beta}}_2 \mathbf{Y}_{2t} \tag{12.16}$$

where $\hat{c}$ and $\hat{\boldsymbol{\beta}}_2$ are the least squares estimates of c and $\boldsymbol{\beta}_2$. The unit root test regression in this case is without deterministic terms (constant or constant and trend). Phillips and Ouliaris (1990) show that ADF and PP unit root tests (t-tests and normalized bias) applied to the estimated cointegrating residual (12.16) *do not* have the usual Dickey-Fuller distributions under the null hypothesis (12.14) of no-cointegration. Instead, due to the spurious regression phenomenon under the null hypothesis (12.14), the distribution of the ADF and PP unit root tests have asymptotic distributions that are functions of Wiener processes that depend on the deterministic terms in the regression (12.15) used to estimate $\boldsymbol{\beta}_2$ *and* the number of variables, $n-1$, in $\mathbf{Y}_{2t}$. These distributions are known as the *Phillips-Ouliaris* (PO) distributions, and are described in Phillips and Ouliaris (1990). To further complicate matters, Hansen (1992) showed the appropriate PO distributions of the ADF and PP unit root tests applied to the residuals (12.16) also depend on the trend behavior of y_{1t} and $\mathbf{Y}_{2t}$ as follows:

Case I: $\mathbf{Y}_{2t}$ and y_{1t} are both $I(1)$ without drift. The ADF and PP unit root test statistics follow the PO distributions, adjusted for a constant, with dimension parameter $n-1$.

Case II: $\mathbf{Y}_{2t}$ is $I(1)$ with drift and y_{1t} may or may not be $I(1)$ with drift. The ADF and PP unit root test statistics follow the PO distributions, adjusted for a constant and trend, with dimension parameter $n-2$. If $n-2=0$ then the ADF and PP unit root test statistics follow the DF distributions adjusted for a constant and trend.

Case III: $\mathbf{Y}_{2t}$ is $I(1)$ without drift and y_{1t} is $I(1)$ with drift. In this case, $\boldsymbol{\beta}_2$ should be estimated from the regression

$$y_{1t} = c + \delta t + \boldsymbol{\beta}_2' \mathbf{Y}_{2t} + u_t \qquad (12.17)$$

The resulting ADF and PP unit root test statistics on the residuals from (12.17) follow the PO distributions, adjusted for a constant and trend, with dimension parameter $n - 1$.

Computing Quantiles and P-values from the Phillips-Ouliaris Distributions Using the S+FinMetrics Functions `pcoint` and `qcoint`

The S+FinMetrics functions `qcoint` and `pcoint`, based on the response surface methodology of MacKinnon (1996), may be used to compute quantiles and p-values from the PO distributions. For example, to compute the 10%, 5% and 1% quantiles from the PO distribution for the ADF t-statistic, adjusted for a constant, with $n - 1 = 3$ and a sample size $T = 100$ use

```
> qcoint(c(0.1,0.05,0.01), n.sample=100, n.series=4,
+ trend="c", statistic="t")
[1] -3.8945 -4.2095 -4.8274
```

Notice that the argument `n.series` represents the total number of variables n. To adjust the PO distributions for a constant and trend set `trend="ct"`. To compute the PO distribution for the ADF normalized bias statistic set `statistic="n"`. The quantiles from the PO distributions can be very different from the quantiles from the DF distributions, especially if $n - 1$ is large. To illustrate, the 10%, 5% and 1% quantiles from the DF distribution for the ADF t-statistic with a sample size $T = 100$ are

```
> qunitroot(c(0.1,0.05,0.01), n.sample=100,
+ trend="c", statistic="t")
[1] -2.5824 -2.8906 -3.4970
```

The following examples illustrate testing for cointegration using an estimated cointegrating vector.

Example 77 *Testing for cointegration between spot and forward exchange rates using an estimated cointegrating vector*

Consider testing for cointegration between spot and forward exchange rates assuming the cointegrating vector is not known using the same data as in the previous example. Let $\mathbf{Y}_t = (s_t, f_t)'$ and normalize the cointegrating vector on s_t so that $\boldsymbol{\beta} = (1, -\beta_2)'$. The normalized cointegrating coefficient β_2 is estimated by least squares from the regression

$$s_t = c + \beta_2 f_t + u_t$$

giving the estimated cointegrating residual $\hat{u}_t = s_t - \hat{c} - \hat{\beta}_2 f_t$. The `OLS` function in S+FinMetrics is used to estimate the above regression:

```
> uscn.ts = seriesMerge(uscn.s,uscn.f)
> ols.fit = OLS(USCNS~USCNF,data=uscn.ts)
> ols.fit

Call:
OLS(formula = USCNS ~USCNF, data = uscn.ts)

Coefficients:
 (Intercept)  USCNF
 0.0023       1.0041

Degrees of freedom: 245 total; 243 residual
Time period: from Feb 1976 to Jun 1996
Residual standard error: 0.001444
```

The estimated value of β_2 is 1.004 and is almost identical to the value $\beta_2 = 1$ implied by covered interest parity. The estimated cointegrating residual $\hat{u}_t$ is extracted from the least squres fit using

```
> u.hat = residuals(ols.fit)
```

Next, the no-cointegration hypothesis (12.14) is tested using the ADF and PP t-tests. Because the mean of $\hat{u}_t$ is zero, the unit root test regressions are estimated without a constant or trend. The ADF t-statistic is computed using 11 lags, as in the previous example, and the PP t-statistic is computed using an automatic lag truncation parameter:

```
> adf.fit = unitroot(u.hat,trend="nc",method="adf",lags=11)
> adf.tstat = adf.fit$sval
> adf.tstat
   lag1
 -2.721

> pp.fit = unitroot(u.hat,trend="nc",method="pp")
> pp.tstat = pp.fit$sval
> pp.tstat
   lag1
 -5.416
```

The ADF t-statistic is -2.721 whereas the PP t-statistic is -5.416. Since s_t and f_t are both $I(1)$ without drift, the 10%, 5% and 1% quantiles from the approrpiate Phillips-Ouliaris distribution for the ADF t-statistic is

```
> qcoint(c(0.10,0.05,0.01),n.sample=nrow(uscn.s),n.series=2,
+ trend="c",statistic="t")
[1] -3.062 -3.361 -3.942
```

The no-cointegration null hypothesis is not rejected at the 10% level using the ADF t-statistic but is rejected at the 1% level using the PP t-statistic. The p-values for the ADF and PP t-statistics are

```
> pcoint(adf.tstat, n.sample=nrow(uscn.s), n.series=2,
+ trend="c", statistic="t")
[1] 0.1957
```

```
> pcoint(pp.tstat, n.sample=nrow(uscn.s), n.series=2,
+ trend="c", statistic="t")
[1] 0.00003925
```

12.4   Regression-Based Estimates of Cointegrating Vectors and Error Correction Models

12.4.1   Least Square Estimator

Least squares may be used to consistently estimate a normalized cointegrating vector. However, the asymptotic behavior of the least squares estimator is non-standard. The following results about the behavior of $\hat{\boldsymbol{\beta}}_2$ if $\mathbf{Y}_t$ is cointegrated are due to Stock (1987) and Phillips (1991):

- $T(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)$ converges in distribution to a non-normal random variable not necessarily centered at zero.

- The least squares estimate $\hat{\boldsymbol{\beta}}_2$ is consistent for $\boldsymbol{\beta}_2$ and converges to $\boldsymbol{\beta}_2$ at rate T instead of the usual rate $T^{1/2}$. That is, $\hat{\boldsymbol{\beta}}_2$ is *super consistent*.

- $\hat{\boldsymbol{\beta}}_2$ is consistent even if Y_{2t} is correlated with u_t so that there is no asymptotic simultaneity bias.

- In general, the asymptotic distribution of $T(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)$ is asymptotically biased and non-normal. The usual OLS formula for computing $\widehat{avar}(\hat{\boldsymbol{\beta}}_2)$ is incorrect and so the usual OLS standard errors are not correct.

- Even though the asymptotic bias goes to zero as T gets large $\hat{\boldsymbol{\beta}}_2$ may be substantially biased in small samples. The least squres estimator is also not efficient.

The above results indicate that the least squares estimator of the cointegrating vector $\boldsymbol{\beta}_2$ could be improved upon. A simple improvement is suggested by Stock and Watson (1993).

12.4.2 Stock and Watson's Efficient Lead/Lag Estimator

Stock and Watson (1993) provide a very simple method for obtaining an asymptotically efficient (equivalent to maximum likelihood) estimator for the normalized cointegrating vector $\boldsymbol{\beta}_2$ as well as a valid formula for computing its asymptotic variance[2].

Let $\mathbf{Y}_t = (y_{1t}, \mathbf{Y}_{2t}')'$ where $\mathbf{Y}_{2t} = (y_{2t}, \ldots, y_{nt})'$ is an $((n-1) \times 1)$ vector and let the cointegrating vector be normalized as $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2')'$. Stock and Watson's efficient estimation procedure is:

- Augment the cointegrating regression of y_{1t} on $\mathbf{Y}_{2t}$ with appropriate deterministic terms $\mathbf{D}_t$ with p leads and lags of $\Delta\mathbf{Y}_{2t}$

$$
\begin{aligned}
y_{1t} &= \boldsymbol{\gamma}'\mathbf{D}_t + \boldsymbol{\beta}_2'\mathbf{Y}_{2t} + \sum_{j=-p}^{p} \boldsymbol{\psi}_j'\Delta\mathbf{Y}_{2t-j} + u_t \qquad (12.18) \\
&= \boldsymbol{\gamma}'\mathbf{D}_t + \boldsymbol{\beta}_2'\mathbf{Y}_{2t} + \boldsymbol{\psi}_p'\Delta\mathbf{Y}_{2t+p} + \cdots + \boldsymbol{\psi}_1'\Delta\mathbf{Y}_{2t+1} \\
&\quad + \boldsymbol{\psi}_0'\Delta\mathbf{Y}_{2t} + \boldsymbol{\psi}_{-1}'\Delta\mathbf{Y}_{2t-1} + \cdots + \boldsymbol{\psi}_{-p}'\Delta\mathbf{Y}_{2t-p} + u_t
\end{aligned}
$$

- Estimate the augmented regression by least squares. The resulting estimator of $\boldsymbol{\beta}_2$ is called the *dynamic OLS* estimator and is denoted $\hat{\boldsymbol{\beta}}_{2,DOLS}$. It will be consistent, asymptotically normally distributed and efficient (equivalent to MLE) under certain assumptions (see Stock and Watson, 1993).

- Asymptotically valid standard errors for the individual elements of $\hat{\boldsymbol{\beta}}_{2,DOLS}$ are given by the OLS standard errors from (12.18) multiplied by the ratio

$$
\left(\frac{\hat{\sigma}_u^2}{\widehat{lrv(u_t)}} \right)^{1/2}
$$

where $\hat{\sigma}_u^2$ is the OLS estimate of $var(u_t)$ and $\widehat{lrv(u_t)}$ is any consistent estimate of the long-run variance of u_t using the residuals $\hat{u}_t$ from (12.18). Alternatively, the Newey-West HAC standard errors may also be used.

Example 78 *DOLS estimation of cointegrating vector using exchange rate data*[3]

Let s_t denote the log of the monthly spot exchange rate between two currencies at time t and let f_t^k denote the log of the forward exchange rate at time t for delivery of foreign currency at time $t + k$. Under rational

---

[2]Hamilton (1994) chapter 19, and Hayashi (2000) chapter 10, give nice discussions of the Stock and Watson procedure.

[3]This example is based on Zivot (2000).

expectations and risk neutrality f_t^k is an unbiased predictor of s_{t+k}, the spot exchange rate at time $t + k$. That is

$$s_{t+k} = f_t^k + \varepsilon_{t+k}$$

where ε_{t+k} is a white noise error term. This is known as the *forward rate unbiasedness hypothesis* (FRUH). Assuming that s_t and f_t^k are $I(1)$ the FRUH implies that s_{t+k} and f_t^k are cointegrated with cointegrating vector $\boldsymbol{\beta} = (1, -1)'$. To illustrate, consider again the log monthly spot, s_t, and one month forward, f_t^1, exchange rates between the US and Canada over the period February 1976 through June 1996 taken from the S+FinMetrics "timeSeries" object lexrates.dat. The cointegrating vector between s_{t+1} and f_t^1 is estimated using least squares and Stock and Watson's dynamic OLS estimator computed from (12.18) with $y_{1t} = s_{t+1}$, $\mathbf{D}_t = 1$, $\mathbf{Y}_{2t} = f_t^1$ and $p = 3$. The data for the DOLS regression equation (12.18) are constucted as

```
> uscn.df = diff(uscn.f)
> colIds(uscn.df) = "D.USCNF"
> uscn.df.lags = tslag(uscn.df,-3:3,trim=T)
> uscn.ts = seriesMerge(uscn.s,uscn.f,uscn.df.lags)
> colIds(uscn.ts)
[1] "USCNS"          "USCNF"          "D.USCNF.lead3"
[4] "D.USCNF.lead2"  "D.USCNF.lead1"  "D.USCNF.lag0"
[7] "D.USCNF.lag1"   "D.USCNF.lag2"   "D.USCNF.lag3"
```

The least squares estimator of the normalized cointegrating coefficient β_2 computed using the S+FinMetrics function OLS is

```
> summary(OLS(tslag(USCNS,-1)~USCNF,data=uscn.ts,na.rm=T))

Call:
OLS(formula = tslag(USCNS, -1) ~USCNF, data = uscn.ts,
na.rm = T)

Residuals:
    Min      1Q  Median      3Q     Max
 -0.0541 -0.0072  0.0006  0.0097  0.0343

Coefficients:
             Value Std. Error  t value Pr(>|t|)
(Intercept) -0.0048    0.0025  -1.9614   0.0510
     USCNF   0.9767    0.0110  88.6166   0.0000

Regression Diagnostics:

        R-Squared 0.9709
```

```
Adjusted R-Squared 0.9708
Durbin-Watson Stat 2.1610

Residual standard error: 0.01425 on 235 degrees of freedom
Time period: from Jun 1976 to Feb 1996
F-statistic: 7853 on 1 and 235 degrees of freedom,
the p-value is 0
```

Notice that in the regression formula, `tslag(USCN,-1)` computes s_{t+1}. The least squares estimate of β_2 is 0.977 with an estimated standard error of 0.011 indicating that f_t^1 underpredicts s_{t+1}. However, the usual formula for computing the estimated standard error is incorrect and should not be trusted.

The DOLS estimator of β_2 based on (12.18) is computed using

```
> dols.fit = OLS(tslag(USCNS,-1)~USCNF +
+ D.USCNF.lead3+D.USCNF.lead2+D.USCNF.lead1 +
+ D.USCNF.lag0+D.USCNF.lag1+D.USCNF.lag2+D.USCNF.lag3,
+ data=uscn.ts,na.rm=T)
```

The Newey-West HAC standard errors for the estimated coefficients are computed using `summary` with `correction="nw"`:

```
> summary(dols.fit,correction="nw")

Call:
OLS(formula = tslag(USCNS, -1) ~USCNF + D.USCNF.lead3 +
D.USCNF.lead2 + D.USCNF.lead1 + D.USCNF.lag0 +
D.USCNF.lag1 + D.USCNF.lag2 + D.USCNF.lag3, data =
uscn.ts, na.rm = T)

Residuals:
     Min     1Q  Median     3Q     Max
 -0.0061 -0.0008  0.0000  0.0009  0.0039

Coefficients:
                 Value Std. Error   t value  Pr(>|t|)
   (Intercept)  0.0023     0.0005    4.3948    0.0000
         USCNF  1.0040     0.0019  531.8862    0.0000
 D.USCNF.lead3  0.0114     0.0063    1.8043    0.0725
 D.USCNF.lead2  0.0227     0.0068    3.3226    0.0010
 D.USCNF.lead1  1.0145     0.0090  112.4060    0.0000
  D.USCNF.lag0  0.0005     0.0073    0.0719    0.9427
  D.USCNF.lag1 -0.0042     0.0061   -0.6856    0.4937
  D.USCNF.lag2 -0.0056     0.0061   -0.9269    0.3549
  D.USCNF.lag3 -0.0014     0.0045   -0.3091    0.7575
```

Regression Diagnostics:

```
        R-Squared 0.9997
Adjusted R-Squared 0.9997
Durbin-Watson Stat 0.4461
```

```
Residual standard error: 0.001425 on 228 degrees of freedom
Time period: from Jun 1976 to Feb 1996
F-statistic: 101000 on 8 and 228 degrees of freedom,
the p-value is 0
```

The DOLS estimator of β_2 is 1.004 with a very small estimated standard error of 0.0019 and indicates that f_t^1 is essentially an unbiased predictor of the future spot rate s_{t+1}.

12.4.3 Estimating Error Correction Models by Least Squares

Consider a bivariate $I(1)$ vector $\mathbf{Y}_t = (y_{1t}, y_{2t})'$ and assume that $\mathbf{Y}_t$ is cointegrated with cointegrating vector $\boldsymbol{\beta} = (1, -\beta_2)'$ so that $\boldsymbol{\beta}'\mathbf{Y}_t = y_{1t} - \beta_2 y_{2t}$ is $I(0)$. Suppose one has a consistent estimate $\hat{\beta}_2$ (by OLS or DOLS) of the cointegrating coefficient and is interested in estimating the corresponding error correction model (12.12)-(12.13) for Δy_{1t} and Δy_{2t}. Because $\hat{\beta}_2$ is super consistent it may be treated as known in the ECM, so that the estimated disequilibrium error $y_{1t} - \hat{\beta}_2 y_{2t}$ may be treated like the known disequilibrium error $y_{1t} - \beta_2 y_{2t}$. Since all variables in the ECM are $I(0)$, the two regression equations may be consistently estimated using ordinary least squares (OLS). Alternatively, the ECM system may be estimated by seemingly unrelated regressions (SUR) to increase efficiency if the number of lags in the two equations are different.

Example 79 *Estimation of error correction model for exchange rate data*

Consider again the monthly log spot rate, s_t, and log forward rate, f_t, data between the U.S. and Canada. Earlier it was shown that s_t and f_t are cointegrated with an estimated cointegrating coefficient $\hat{\beta}_2 = 1.004$. Now consider estimating an ECM of the form (12.12)-(12.13) by least squares using the estimated disequilibrium error $s_t - 1.004 \cdot f_t$. In order to estimate the ECM, the number of lags of Δs_t and Δf_t needs to be determined. This may be done using test statistics for the significance of the lagged terms or model selection criteria like AIC or BIC. An initial estimation using one lag of Δs_t and Δf_t may be performed using

```
> u.hat = uscn.s - 1.004*uscn.f
> colIds(u.hat) = "U.HAT"
> uscn.ds = diff(uscn.s)
> colIds(uscn.ds) = "D.USCNS"
```

```
> uscn.df = diff(uscn.f)
> colIds(uscn.df) = "D.USCNF"
> uscn.ts = seriesMerge(uscn.s,uscn.f,uscn.ds,uscn.df,u.hat)
> ecm.s.fit = OLS(D.USCNS~tslag(U.HAT)+tslag(D.USCNS)
+ +tslag(D.USCNF),data=uscn.ts,na.rm=T)
> ecm.f.fit = OLS(D.USCNF~tslag(U.HAT)+tslag(D.USCNS)+
+ tslag(D.USCNF),data=uscn.ts,na.rm=T)
```

The estimated coefficients from the fitted ECM are

```
> ecm.s.fit

Call:
OLS(formula = D.USCNS ~tslag(U.HAT) + tslag(D.USCNS) + tslag(
D.USCNF), data = uscn.ts, na.rm = T)

Coefficients:
 (Intercept) tslag(U.HAT) tslag(D.USCNS) tslag(D.USCNF)
 -0.0050       1.5621         1.2683          -1.3877

Degrees of freedom: 243 total; 239 residual
Time period: from Apr 1976 to Jun 1996
Residual standard error: 0.013605

> ecm.f.fit

Call:
OLS(formula = D.USCNF ~tslag(U.HAT) + tslag(D.USCNS) + tslag(
D.USCNF), data = uscn.ts, na.rm = T)

Coefficients:
 (Intercept) tslag(U.HAT) tslag(D.USCNS) tslag(D.USCNF)
 -0.0054       1.7547         1.3595          -1.4702

Degrees of freedom: 243 total; 239 residual
Time period: from Apr 1976 to Jun 1996
Residual standard error: 0.013646
```

12.5 VAR Models and Cointegration

The Granger representation theorem links cointegration to error correction models. In a series of important papers and in a marvelous textbook, Soren Johansen firmly roots cointegration and error correction models in a vector autoregression framework. This section outlines Johansen's approach to cointegration modeling.

12.5.1   The Cointegrated VAR

Consider the levels VAR(p) model for the ($n \times 1$) vector $\mathbf{Y}_t$

$$\mathbf{Y}_t = \boldsymbol{\Phi}\mathbf{D}_t + \boldsymbol{\Pi}_1\mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Pi}_p\mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t, \ t = 1, \ldots, T, \qquad (12.19)$$

where $\mathbf{D}_t$ contains deterministic terms (constant, trend, seasonal dummies etc.). Recall, the VAR(p) model is stable if

$$\det(\mathbf{I}_n - \boldsymbol{\Pi}_1 z - \cdots - \boldsymbol{\Pi}_p z^p) = 0 \qquad (12.20)$$

has all roots outside the complex unit circle. If (12.20) has a root on the unit circle then some or all of the variables in $\mathbf{Y}_t$ are $I(1)$ and they may also be cointegrated. Recall, $\mathbf{Y}_t$ is cointegrated if there exists some linear combination of the variables in $\mathbf{Y}_t$ that is $I(0)$. Suppose $\mathbf{Y}_t$ is $I(1)$ and possibly cointegrated. Then the VAR representation (22.11) is not the most suitable representation for analysis because the cointegrating relations are not explicitly apparent. The cointegrating relations become apparent if the levels VAR (22.11) is transformed to the *vector error correction model* (VECM)

$$\Delta\mathbf{Y}_t = \boldsymbol{\Phi}\mathbf{D}_t + \boldsymbol{\Pi}\mathbf{Y}_{t-1} + \boldsymbol{\Gamma}_1\Delta\mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t \qquad (12.21)$$

where $\boldsymbol{\Pi} = \boldsymbol{\Pi}_1 + \cdots + \boldsymbol{\Pi}_p - \mathbf{I}_n$ and $\boldsymbol{\Gamma}_k = -\sum_{j=k+1}^{p} \boldsymbol{\Pi}_j, \ k = 1, \ldots, p-1$. The matrix $\boldsymbol{\Pi}$ is called the *long-run impact matrix* and $\boldsymbol{\Gamma}_k$ are the *short-run impact matrices*. Notice that the VAR parameters $\boldsymbol{\Pi}_i$ may be recovered from the VECM parameters $\boldsymbol{\Pi}$ and $\boldsymbol{\Gamma}_k$ via

$$\begin{aligned} \boldsymbol{\Pi}_1 &= \boldsymbol{\Gamma}_1 + \boldsymbol{\Pi} + \mathbf{I}_n, & (12.22) \\ \boldsymbol{\Pi}_k &= \boldsymbol{\Gamma}_k - \boldsymbol{\Gamma}_{k-1}, \ k = 2, \ldots, p. \end{aligned}$$

In the VECM (12.21), $\Delta\mathbf{Y}_t$ and its lags are $I(0)$. The term $\boldsymbol{\Pi}\mathbf{Y}_{t-1}$ is the only one which includes potential $I(1)$ variables and for $\Delta\mathbf{Y}_t$ to be $I(0)$ it must be the case that $\boldsymbol{\Pi}\mathbf{Y}_{t-1}$ is also $I(0)$. Therefore, $\boldsymbol{\Pi}\mathbf{Y}_{t-1}$ must contain the cointegrating relations if they exist. If the VAR(p) process has unit roots then from (12.20) it is clear that $\boldsymbol{\Pi}$ is a singular matrix. If $\boldsymbol{\Pi}$ is singular then it has *reduced rank*; that is $rank(\boldsymbol{\Pi}) = r < n$. There are two cases to consider:

1. $rank(\boldsymbol{\Pi}) = 0$. This implies that $\boldsymbol{\Pi} = \mathbf{0}$ and $\mathbf{Y}_t$ is $I(1)$ and not cointegrated. The VECM (12.21) reduces to a VAR($p-1$) in first differences

$$\Delta\mathbf{Y}_t = \boldsymbol{\Phi}\mathbf{D}_t + \boldsymbol{\Gamma}_1\Delta\mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t.$$

2. $0 < rank(\boldsymbol{\Pi}) = r < n$. This implies that $\mathbf{Y}_t$ is $I(1)$ with r linearly independent cointegrating vectors and $n-r$ common stochastic trends

(unit roots)[4]. Since $\mathbf{\Pi}$ has rank r it can be written as the product

$$\underset{(n \times n)}{\mathbf{\Pi}} = \underset{(n \times r)}{\boldsymbol{\alpha}} \underset{(r \times n)}{\boldsymbol{\beta}}{}'$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $(n \times r)$ matrices with $rank(\boldsymbol{\alpha}) = rank(\boldsymbol{\beta}) = r$. The rows of $\boldsymbol{\beta}'$ form a basis for the r cointegrating vectors and the elements of $\boldsymbol{\alpha}$ distribute the impact of the cointegrating vectors to the evolution of $\Delta \mathbf{Y}_t$. The VECM (12.21) becomes

$$\Delta \mathbf{Y}_t = \boldsymbol{\Phi} \mathbf{D}_t + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{Y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta \mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t,$$
$$(12.23)$$

where $\boldsymbol{\beta}' \mathbf{Y}_{t-1} \sim I(0)$ since $\boldsymbol{\beta}'$ is a matrix of cointegrating vectors.

It is important to recognize that the factorization $\mathbf{\Pi} = \boldsymbol{\alpha} \boldsymbol{\beta}'$ is not unique since for any $r \times r$ nonsingular matrix $\mathbf{H}$ we have

$$\boldsymbol{\alpha} \boldsymbol{\beta}' = \boldsymbol{\alpha} \mathbf{H} \mathbf{H}^{-1} \boldsymbol{\beta}' = (\mathbf{aH})(\boldsymbol{\beta} \mathbf{H}^{-1\prime})' = \mathbf{a}^* \boldsymbol{\beta}^{*\prime}.$$

Hence the factorization $\mathbf{\Pi} = \boldsymbol{\alpha} \boldsymbol{\beta}'$ only identifies the space spanned by the cointegrating relations. To obtain unique values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}'$ requires further restrictions on the model.

Example 80 *A bivariate cointegrated VAR(1) model*

Consider the bivariate VAR(1) model for $\mathbf{Y}_t = (y_{1t}, y_{2t})'$

$$\mathbf{Y}_t = \mathbf{\Pi}_1 \mathbf{Y}_{t-1} + \boldsymbol{\epsilon}_t.$$

The VECM is

$$\Delta \mathbf{Y}_t = \mathbf{\Pi} \mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t$$

where $\mathbf{\Pi} = \mathbf{\Pi}_1 - \mathbf{I}_2$. Assuming $\mathbf{Y}_t$ is cointegrated there exists a 2×1 vector $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ such that $\boldsymbol{\beta}' \mathbf{Y}_t = \beta_1 y_{1t} + \beta_2 y_{2t}$ is $I(0)$. Using the normalization $\beta_1 = 1$ and $\beta_2 = -\beta$ the cointegrating relation becomes $\boldsymbol{\beta}' \mathbf{Y}_t = y_{1t} - \beta y_{2t}$. This normalization suggests the stochastic long-run equilibrium relation

$$y_{1t} = \beta y_{2t} + u_t$$

where u_t is $I(0)$ and represents the stochastic deviations from the long-run equilibrium $y_{1t} = \beta y_{2t}$.

Since $\mathbf{Y}_t$ is cointegrated with one cointegrating vector, $rank(\mathbf{\Pi}) = 1$ and can be decomposed as

$$\mathbf{\Pi} = \boldsymbol{\alpha} \boldsymbol{\beta}' = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \begin{pmatrix} 1 & -\beta \end{pmatrix} = \begin{pmatrix} \alpha_1 & -\alpha_1 \beta \\ \alpha_2 & -\alpha_2 \beta \end{pmatrix}.$$

---

[4]To see that $\mathbf{Y}_t$ has $n - r$ common stochastic trends we have to look at the Beveridge-Nelson decomposition of the moving average representation of $\Delta \mathbf{Y}_t$.

The elements in the vector $\boldsymbol{\alpha}$ are interpreted as *speed of adjustment* coefficients. The cointegrated VECM for $\Delta \mathbf{Y}_t$ may be rewritten as

$$\Delta \mathbf{Y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{12.24}$$

Writing the VECM equation by equation gives

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

The first equation relates the change in y_{1t} to the lagged disequilibrium error $\boldsymbol{\beta}'\mathbf{Y}_{t-1} = (y_{1t-1} - \beta y_{2t-1})$ and the second equation relates the change in Δy_{2t} to the lagged disequilibrium error as well. Notice that the reactions of y_1 and y_2 to the disequilibrium errors are captured by the adjustment coefficients α_1 and α_2.

The stability conditions for the bivariate VECM are related to the stability conditions for the disequilibrium error $\boldsymbol{\beta}'\mathbf{Y}_t$. By pre-multiplying (12.24) by $\boldsymbol{\beta}'$, it is straightforward to show that $\boldsymbol{\beta}'\mathbf{Y}_t$ follows an AR(1) process

$$\boldsymbol{\beta}'\mathbf{Y}_t = (1 + \boldsymbol{\beta}'\boldsymbol{\alpha})\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t$$

or

$$u_t = \phi u_{t-1} + v_t$$

where $u_t = \boldsymbol{\beta}'\mathbf{Y}_t$, $\phi = 1 + \boldsymbol{\beta}'\boldsymbol{\alpha} = 1 + (\alpha_1 - \beta\alpha_2)$ and $v_t = \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t = u_{1t} - \beta u_{2t}$. The AR(1) model for u_t is stable as long as $|\phi| = |1 + (\alpha_1 - \beta\alpha_2)| < 1$. For example, suppose $\beta = 1$. Then the stability condition is $|\phi| = |1 + (\alpha_1 - \alpha_2)| < 1$ which is satisfied if $\alpha_1 - \alpha_2 < 0$ and $\alpha_1 - \alpha_2 > -2$. If $\alpha_2 = 0$ then $-2 < \alpha_1 < 0$ is the required stability condition.

12.5.2   Johansen's Methodology for Modeling Cointegration

The basic steps in Johansen's methodology are:

- Specify and estimate a VAR(p) model for $\mathbf{Y}_t$.

- Construct likelihood ratio tests for the rank of $\boldsymbol{\Pi}$ to determine the number of cointegrating vectors.

- If necessary, impose normalization and identifying restrictions on the cointegrating vectors.

- Given the normalized cointegrating vectors estimate the resulting cointegrated VECM by maximum likelihood.

12.5.3 Specification of Deterministic Terms

Following Johansen (1995), the deterministic terms in (12.23) are restricted to the form

$$\mathbf{\Phi D}_t = \boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$$

If the deterministic terms are unrestricted then the time series in $\mathbf{Y}_t$ may exhibit quadratic trends and there may be a linear trend term in the cointegrating relationships. Restricted versions of the trend parameters $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ limit the trending nature of the series in $\mathbf{Y}_t$. The trend behavior of $\mathbf{Y}_t$ can be classified into five cases:

1. Model $H_2(r)$: $\boldsymbol{\mu}_t = 0$ (no constant). The restricted VECM is

$$\Delta \mathbf{Y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\Gamma}_1\Delta\mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t,$$

   and all the series in $\mathbf{Y}_t$ are $I(1)$ without drift and the cointegrating relations $\boldsymbol{\beta}'\mathbf{Y}_t$ have mean zero.

2. Model $H_1^*(r)$: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 = \boldsymbol{\alpha}\boldsymbol{\rho}_0$ (restricted constant). The restricted VECM is

$$\Delta \mathbf{Y}_t = \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\rho}_0) + \boldsymbol{\Gamma}_1\Delta\mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t,$$

   the series in $\mathbf{Y}_t$ are $I(1)$ without drift and the cointegrating relations $\boldsymbol{\beta}'\mathbf{Y}_t$ have non-zero means $\boldsymbol{\rho}_0$.

3. Model $H_1(r)$: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$ (unrestricted constant). The restricted VECM is

$$\Delta \mathbf{Y}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\Gamma}_1\Delta\mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t$$

   the series in $\mathbf{Y}_t$ are $I(1)$ with drift vector $\boldsymbol{\mu}_0$ and the cointegrating relations $\boldsymbol{\beta}'\mathbf{Y}_t$ may have a non-zero mean.

4. Model $H^*(r)$: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}\boldsymbol{\rho}_1 t$ (restricted trend). The restricted VECM is

$$\begin{aligned}\Delta \mathbf{Y}_t \quad &= \quad \boldsymbol{\mu}_0 + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\rho}_1 t) \\ &\quad + \boldsymbol{\Gamma}_1\Delta\mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t\end{aligned}$$

   the series in $\mathbf{Y}_t$ are $I(1)$ with drift vector $\boldsymbol{\mu}_0$ and the cointegrating relations $\boldsymbol{\beta}'\mathbf{Y}_t$ have a linear trend term $\boldsymbol{\rho}_1 t$.

5. Model $H(r)$: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$ (unrestricted constant and trend). The unrestricted VECM is

$$\Delta \mathbf{Y}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\Gamma}_1\Delta\mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t,$$

   the series in $\mathbf{Y}_t$ are $I(1)$ with a linear trend (quadratic trend in levels) and the cointegrating relations $\boldsymbol{\beta}'\mathbf{Y}_t$ have a linear trend.
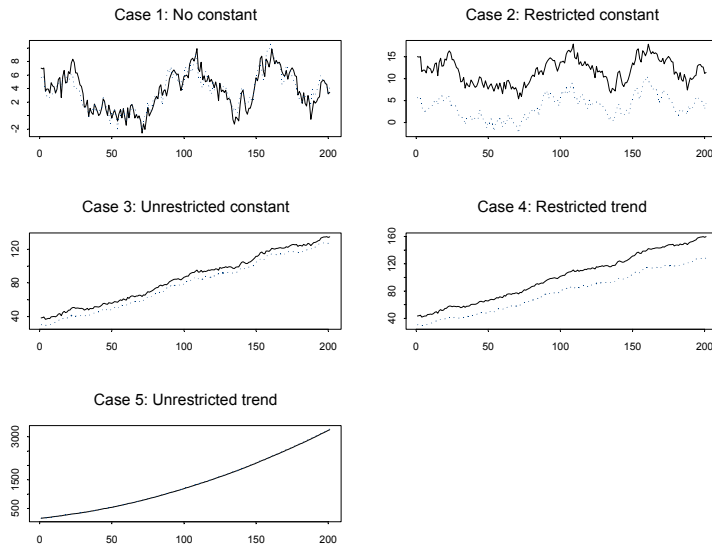
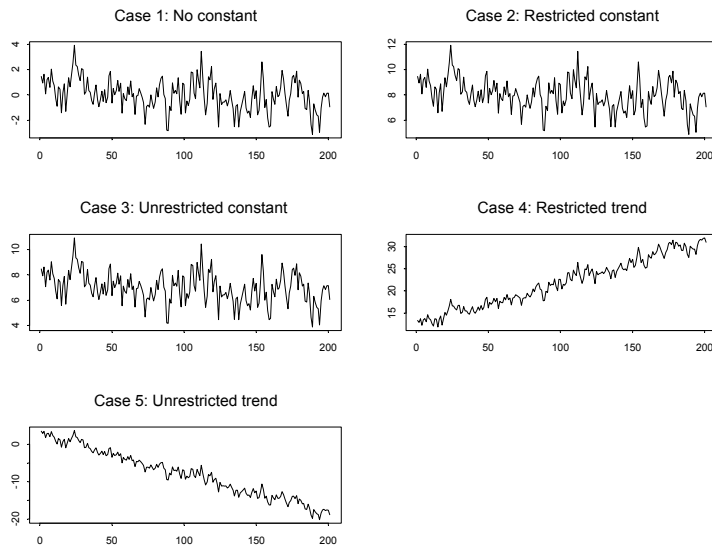FIGURE 12.6. Simulated $\mathbf{Y}_t$ from bivariate cointegrated VECM for five trend cases.

FIGURE 12.7. Simulated $\boldsymbol{\beta}'\mathbf{Y}_t$ from bivariate cointegrated VECM for five trend cases.

Simulated data from the five trend cases for a bivariate cointegrated VAR(1) model are illustrated in Figures 12.6 and 12.7. Case I is not really relevant for empirical work. The restricted contstant Case II is appropriate for non-trending $I(1)$ data like interest rates and exchange rates. The unrestriced constant Case III is appropriate for trending $I(1)$ data like asset prices, macroeconomic aggregates (real GDP, consumption, employment etc). The restricted trend case IV is also appropriate for trending $I(1)$ as in Case III. However, notice the deterministic trend in the cointegrating residual in Case IV as opposed to the stationary residual in case III. Finally, the unrestricted trend Case V is appropriate for $I(1)$ data with a quadratic trend. An example might be nominal price data during times of extreme inflation.

12.5.4 Likelihood Ratio Tests for the Number of Cointegrating Vectors

The unrestricted cointegrated VECM (12.23) is denoted $H(r)$. The $I(1)$ model $H(r)$ can be formulated as the condition that the rank of $\mathbf{\Pi}$ is less than or equal to r. This creates a nested set of models

$$H(0) \subset \cdots \subset H(r) \subset \cdots \subset H(n)$$

where $H(0)$ represents the non-cointegrated VAR model with $\mathbf{\Pi} = \mathbf{0}$ and $H(n)$ represents an unrestricted stationary VAR(p) model. This nested formulation is convenient for developing a sequential procedure to test for the number r of cointegrating relationships.

Since the rank of the long-run impact matrix $\mathbf{\Pi}$ gives the number of cointegrating relationships in $\mathbf{Y}_t$, Johansen formulates likelihood ratio (LR) statistics for the number of cointegrating relationships as LR statistics for determining the rank of $\mathbf{\Pi}$. These tests are based on the estimated eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_n$ of the matrix $\mathbf{\Pi}$[5]. These eigenvalues also happen to equal the squared *canonical correlations* between $\Delta\mathbf{Y}_t$ and $\mathbf{Y}_{t-1}$ corrected for lagged $\Delta\mathbf{Y}_t$ and $\mathbf{D}_t$ and so lie between 0 and 1. Recall, the rank of $\mathbf{\Pi}$ is equal to the number of non-zero eigenvalues of $\mathbf{\Pi}$.

Johansen's Trace Statistic

Johansen's LR statistic tests the nested hypotheses

$$H_0(r) : r = r_0 \text{ vs. } H_1(r_0) : r > r_0$$

The LR statistic, called the *trace statistic*, is given by

$$LR_{trace}(r_0) = -T \sum_{i=r_0+1}^{n} \ln(1 - \hat{\lambda}_i)$$

---

[5]The calculation of the eigenvalues $\hat{\lambda}_i$ $(i = 1, \dots, n)$ is described in the appendix.

If $rank(\Pi) = r_0$ then $\hat{\lambda}_{r_0+1}, \ldots, \hat{\lambda}_n$ should all be close to zero and $LR_{trace}(r_0)$ should be small. In contrast, if $rank(\mathbf{\Pi}) > r_0$ then some of $\hat{\lambda}_{r_0+1}, \ldots, \hat{\lambda}_n$ will be nonzero (but less than 1) and $LR_{trace}(r_0)$ should be large. The asymptotic null distribution of $LR_{trace}(r_0)$ is not chi-square but instead is a multivariate version of the Dickey-Fuller unit root distribution which depends on the dimension $n - r_0$ and the specification of the deterministic terms. Critical values for this distribution are tabulated in Osterwald-Lenum (1992) for the five trend cases discussed in the previous section for $n - r_0 = 1, \ldots, 10$.

Sequential Procedure for Determining the Number of Cointegrating Vectors

Johansen proposes a sequential testing procedure that consistently determines the number of cointegrating vectors. First test $H_0(r_0 = 0)$ against $H_1(r_0 > 0)$. If this null is not rejected then it is concluded that there are no cointegrating vectors among the n variables in Y_t. If $H_0(r_0 = 0)$ is rejected then it is concluded that there is at least one cointegrating vector and proceed to test $H_0(r_0 = 1)$ against $H_1(r_0 > 1)$. If this null is not rejected then it is concluded that there is only one cointegrating vector. If the null is rejected then it is concluded that there is at least two cointegrating vectors. The sequential procedure is continued until the null is not rejected.

Johansen's Maximum Eigenvalue Statistic

Johansen also derives a LR statistic for the hypotheses

$$H_0(r_0) : r = r_0 \text{ vs. } H_1(r_0) : r_0 = r_0 + 1$$

The LR statistic, called the maximum eigenvalue statistic, is given by

$$LR_{\max}(r_0) = -T \ln(1 - \hat{\lambda}_{r_0+1})$$

As with the trace statistic, the asymptotic null distribution of $LR_{\max}(r_0)$ is not chi-square but instead is a complicated function of Brownian motion, which depends on the dimension $n - r_0$ and the specification of the deterministic terms. Critical values for this distribution are tabulated in Osterwald-Lenum (1992) for the five trend cases discussed in the previous section for $n - r_0 = 1, \ldots, 10$.

Finite Sample Correction to LR Tests

Reinsel and Ahn (1992) and Reimars (1992) have suggested that the LR tests perform better in finite samples if the factor $T - np$ is used instead of T in the construction of the LR tests.

12.5.5   Testing Hypothesis on Cointegrating Vectors Using the *S+FinMetrics* Function `coint`

This section describes how to test for the number of cointegrating vectors, and how to perform certain tests for linear restrictions on the long-run $\boldsymbol{\beta}$ coefficients.

Testing for the Number of Cointegrating Vectors

The Johansen LR tests for determining the number of cointegrating vectors in multivariate time series may be computed using the S+FinMetrics function `coint`. The function `coint` has arguments

```
> args(coint)
function(Y, X = NULL, lags = 1, trend = "c", H = NULL,
b = NULL, save.VECM = T)
```

where Y is a matrix, data frame or "`timeSeries`" containing the $I(1)$ variables in $\mathbf{Y}_t$, X is a numeric object representing exogenous variables to be added to the VECM, `lags` denotes the number of lags in the VECM (one less than the number of lags in the VAR representation), `trend` determines the trend case specification, and `save.VECM` determines if the VECM information is to be saved. The arguments H and b will be explained later. The result of `coint` is an object of class "`coint`" for which there are `print` and `summary` methods. The use of `coint` is illustrated with the following examples.

Example 81 *Exchange rate data*

Consider testing for the number of cointegrating relations among the logarithms of the monthly spot and forward exchange rates in the "`timeSeries`" `uscn.ts` examined earlier. The first step is to determine the number of lags to use in the VECM. Using the S+FinMetrics function VAR, the lag length that minimizes the AIC with a maximum lag of 6 is $p = 2$:

```
> uscn.ts = seriesMerge(uscn.s, uscn.f)
> var.fit = VAR(uscn.ts,max.ar=6,criterion="AIC")
> var.fit$ar.order
[1] 2
```

The lag length for the VECM is then $p-1 = 1$. Since the monthly exchange rate data are not trending, the Johansen LR tests are computed assuming the restricted constant case II:

```
> coint.rc = coint(uscn.ts,trend="rc",lags=1)
> class(coint.rc)
[1] "coint"
> coint.rc
```

```
Call:
coint(Y = uscn.ts, lags = 1, trend = "rc")

Trend Specification:
H1*(r): Restricted constant

Trace tests significant at the 5% level are flagged by ' +'.
Trace tests significant at the 1% level are flagged by '++'.
Max Eigenvalue tests significant at the 5% level are flagged
by ' *'.
Max Eigenvalue tests significant at the 1% level are flagged
by '**'.

Tests for Cointegration Rank:
        Eigenvalue Trace Stat  95% CV  99% CV Max Stat
H(0)++** 0.0970     32.4687    19.9600 24.6000 24.8012
H(1)     0.0311      7.6675     9.2400 12.9700  7.6675


          95% CV   99% CV
H(0)++** 15.6700  20.2000
H(1)      9.2400  12.9700
```

Recall, the number of cointegrating vectors is equal to the number of non-zero eigenvalues of $\mathbf{\Pi}$. The two estimated eigenvalues are 0.0970 and 0.0311. The first row in the table gives $LR_{trace}(0)$ and $LR_{\max}(0)$ for testing the null of $r_0 = 0$ cointegrating vectors as well as the 95% and 99% quantiles of the appropriate asymptotic distributions taken from the tables in Osterwald-Lenum (1992). Both the trace and maximum eigenvalue statistics reject the $r_0 = 0$ null at the 1% level. The second row in the table gives $LR_{trace}(1)$ and $LR_{\max}(1)$ for testing the null of $r_0 = 1$. Neither statistic rejects the null that $r_0 = 1$.

The `summary` method gives the same output as `print` as well as the un-normalized cointegrating vectors, adjustment coefficients and the estimate of $\mathbf{\Pi}$.

Testing Linear Restrictions on Cointegrating Vectors

The `coint` function can also be used to test linear restrictions on the cointegrating vectors $\boldsymbol{\beta}$. Two types of restrictions are currently supported: the same linear restrictions on all cointegrating vectors in $\boldsymbol{\beta}$; some cointegrating vectors in $\boldsymbol{\beta}$ are assumed known. Following Johansen (1995), two examples are given illustrating how to use the `coint` function to test linear restrictions on $\boldsymbol{\beta}$.

Example 82 *Johansen's Danish data*

The `"timeSeries"` data set `johansen.danish` in S+FinMetrics contains the monthly Danish data used in Johansen (1995), with the columns `LRM`, `LRY`, `LPY`, `IBO`, `IBE` representing the log real money supply (m_t), log real income (y_t), log prices, bond rate (i_t^b) and deposit rate (i_t^d), respectively. Johansen (1995) considered testing the cointegrating relationship among m_t, y_t, i_t^b and i_t^d. A natural hypothesis is that the velocity of money is a function of the interest rates, or the cointegrating relation contains m_t and y_t only through the term $m_t - y_t$. For $\mathbf{R}' = (1, 1, 0, 0)$, this hypothesis can be represented as a linear restriction on $\boldsymbol{\beta}$:

$$H_0 : \mathbf{R}'\boldsymbol{\beta} = 0 \text{ or } \boldsymbol{\beta} = \mathbf{H}\boldsymbol{\Psi} \tag{12.25}$$

where $\boldsymbol{\Psi}$ are the unknown parameters in the cointegrating vectors $\boldsymbol{\beta}$, $\mathbf{H} = \mathbf{R}_\perp$ and $\mathbf{R}_\perp$ is the orthogonal complement of $\mathbf{R}$ such that $\mathbf{R}'\mathbf{R}_\perp = \mathbf{0}$. Johansen (1995) showed that the null hypothesis (12.25) against the alternative of unrestricted r cointegrating relations $H(r)$ can be tested using a likelihood ratio statistic, which is asymptotically distributed as a χ^2 with $r(n - s)$ degree of freedom where s is the number of columns in $\mathbf{H}$.

To test the hypothesis that the coefficients of m_t and y_t add up to zero in the cointegrating relations, Johansen (1995) considered a restricted constant model. In this case, $\mathbf{R}' = (1, 1, 0, 0, 0)$ since the restricted constant also enters the cointegrating space. Given the restriction matrix $\mathbf{R}$, the matrix $\mathbf{H}$ can be computed using the `perpMat` function in S+FinMetrics:

```
> R = c(1, 1, 0, 0, 0)
> H = perpMat(R)
> H
     [,1] [,2] [,3] [,4]
[1,]   -1    0    0    0
[2,]    1    0    0    0
[3,]    0    1    0    0
[4,]    0    0    1    0
[5,]    0    0    0    1
```

Now the test can be simply performed by passing the matrix $\mathbf{H}$ to the `coint` function:

```
> restr.mod1 = coint(johansen.danish[,c(1,2,4,5)],
+ trend="rc", H=H)
```

The result of the test can be shown by calling the generic `print` method on `restr.mod1` with the optional argument `restrictions=T`:

```
> print(restr.mod1, restrictions=T)

Call:
coint(Y = johansen.danish[, c(1, 2, 4, 5)], trend = "rc",
     H = H)
```

```
Trend Specification:
H1*(r): Restricted constant

Tests for Linear Restriction on Coint Vectors:
Null hypothesis: the restriction is true
      Stat        Dist df P-value
H(1) 0.0346 chi-square  1  0.8523
H(2) 0.2607 chi-square  2  0.8778
H(3) 4.6000 chi-square  3  0.2035
H(4) 6.0500 chi-square  4  0.1954
```

For unrestricted sequential cointegration testing, the statistics in the output are labeled according to the null hypothesis, such as $H(0)$, $H(1)$, etc. However, when restrictions are imposed, the statistics in the output printed with `restrictions=T` are labeled according to the alternative hypothesis, such as $H(1)$, $H(2)$, etc. In the above output, the null hypothesis cannot be rejected against the alternatives of $H(1)$, $H(2)$, $H(3)$ and $H(4)$ at conventional levels of significance.

After confirming that the cointegrating coefficients on m_t and y_t add up to zero, it is interesting to see if $(1, -1, 0, 0, 0)$ actually is a cointegrating vector. In general, to test the null hypothesis that some cointegrating vectors in $\boldsymbol{\beta}$ are equal to $\mathbf{b}$:

$$H_0 : \boldsymbol{\beta} = (\mathbf{b}, \boldsymbol{\Psi})$$

where $\mathbf{b}$ is the $n \times s$ matrix of known cointegrating vectors and $\boldsymbol{\Psi}$ is the $n \times (r-s)$ matrix of unknown cointegrating vectors, Johansen (1995) showed that a likelihood ratio statistic can be used, which is asymptotically distributed as a χ^2 with $s(n-r)$ degrees of freedom. This test can be simply performed by setting the optional argument $\mathbf{b}$ to the known cointegrating vectors. For example,

```
> b = as.matrix(c(1,-1,0,0,0))
> restr.mod2 = coint(johansen.danish[,c(1,2,4,5)],
+ trend="rc", b=b)
> print(restr.mod2, restrictions=T)

Trend Specification:
H1*(r): Restricted constant

Tests for Known Coint Vectors:
Null hypothesis: the restriction is true
       Stat        Dist df P-value
H(1) 29.7167 chi-square  4  0.0000
H(2)  8.3615 chi-square  3  0.0391
```

```
H(3)  4.8759 chi-square  2  0.0873
H(4)  0.5805 chi-square  1  0.4461
```

Again, the statistics in the above output are labeled according to the alternative hypothesis $H(1)$, $H(2)$, etc. Although the hypothesis that the cointegrating coefficients on m_t and y_t sum up to zero cannot be rejected, $m_t - y_t$ does not seem to be stationary because the hypothesis of the known value of **b** is rejected at conventional levels of significance against $H(1)$ and $H(2)$.

12.5.6   Maximum Likelihood Estimation of the Cointegrated VECM

If it is found that $rank(\mathbf{\Pi}) = r$, $0 < r < n$, then the cointegrated VECM (12.23) becomes a reduced rank multivariate regression. The details of the maximum likelihood estimation of (12.23) under the reduced rank restriction $rank(\Pi) = r$ is briefly outlined in the Appendix to this chapter. There it is shown that

$$\hat{\boldsymbol{\beta}}_{mle} = (\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_r), \tag{12.26}$$

where $\hat{\mathbf{v}}_i$ are the eigenvectors associated with the eigenvalues $\hat{\lambda}_i$, and that the MLEs of the remaining parameters are obtained by multivariate least squares estimation of (12.23) with $\boldsymbol{\beta}$ replaced by $\hat{\boldsymbol{\beta}}_{mle}$.

Normalized Estimates

Recall, the factorization $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ is not unique and so the columns of $\hat{\boldsymbol{\beta}}_{mle}$ in (12.26) may be interpreted as linear combinations of the underlying cointegrating relations. For interpretations, it is often convenient to normalize or identify the cointegrating vectors by choosing a specific coordinate system in which to express the variables. One arbitrary way to do this, suggested by Johansen, is to solve for the triangular representation of the cointegrated system. The details of this normalization process is given in the appendix, and the S+FinMetrics function VECM utilizes this normalization scheme by default. The resulting normalized cointegrating vector is denoted $\hat{\boldsymbol{\beta}}_{c,mle}$. The normalization of the MLE for $\boldsymbol{\beta}$ to $\hat{\boldsymbol{\beta}}_{c,mle}$ will affect the MLE of $\boldsymbol{\alpha}$ but not the MLEs of the other parameters in the VECM.

It must be emphasized that it is not possible to estimate the individual elements of $\boldsymbol{\beta}$ without a specific normalization or identification scheme and that the normalization based on Phillips' triangular representation is arbitrary and the resulting normalized cointegrating vectors (12.29) may not have any economic meaning. Only in the case $r = 1$ can a unique cointegrating vector be found after normalization.

Example 83 *Unnormalzed MLEs for exchange rate data*

The unnormalized cointegrating vector assuming $r_0 = 1$ may also be extracted directly from the "`coint`" object:

```
> coint.rc$coint.vectors[1,]
     USCNS   USCNF Intercept*
 -739.0541 743.314   2.023532
```

Notice in the case of a restricted constant, the last coefficient in $\hat{\boldsymbol{\beta}}_{mle}$ is an estimate of the restricted constant. Normalizing on `USCNS` by dividing each element in $\hat{\boldsymbol{\beta}}_{mle}$ by -739.0541 gives

```
> coint.rc$coint.vectors[1,]/
+ as.numeric(-coint.rc$coint.vectors[1,1])
 USCNS    USCNF  Intercept*
    -1 1.005764 0.002738003
```

The normalized MLEs, $\hat{\boldsymbol{\beta}}_{c,mle} = (-1, 1.006)'$ and $\hat{\mu}_c = 0.0027$ are almost identical to the least squares estimates $\hat{\boldsymbol{\beta}} = (1, -1.004)'$ and $\hat{\mu} = 0.0023$ found earlier.

Asymptotic Distributions

Let $\hat{\boldsymbol{\beta}}_{c,mle}$ denote the MLE of the normalized cointegrating matrix $\boldsymbol{\beta}_c$. Johansen (1995) shows that $T(vec(\hat{\boldsymbol{\beta}}_{c,mle}) - vec(\boldsymbol{\beta}_c))$ is asymptotically (mixed) normally distributed and that a consistent estimate of the asymptotic covariance of $\hat{\boldsymbol{\beta}}_{c,mle}$ is given by

$$\widehat{avar}(vec(\hat{\boldsymbol{\beta}}_{c,mle})) =$$
$$T^{-1}(\mathbf{I}_n - \hat{\boldsymbol{\beta}}_{c,mle}\mathbf{c}')\mathbf{S}_{11}^{-1}(\mathbf{I}_n - \hat{\boldsymbol{\beta}}_{c,mle}\mathbf{c}')' \otimes \left(\hat{\boldsymbol{\alpha}}'_{c,mle}\hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\alpha}}_{c,mle}\right)^{-1} \quad (12.27)$$

Notice that this result implies that $\hat{\boldsymbol{\beta}}_{c,mle} \overset{p}{\to} \boldsymbol{\beta}_c$ at rate T instead of the usual rate $T^{1/2}$. Hence, like the least squares estimator, $\hat{\boldsymbol{\beta}}_{c,mle}$ is *super consistent*. However, unlike the least squares estimator, asymptotically valid standard errors may be compute using the square root of the diagonal elements of (12.27).

12.5.7 Maximum Likelihood Estimation of the Cointegrated VECM Using the S+FinMetrics Function VECM

Once the number of cointegrating vectors is determined from the `coint` function, the maximum likelihood estimates of the full VECM may be obtained using the S+FinMetrics function `VECM`. The arguments expected by `VECM` are

```
> args(VECM)
function(object, coint.rank = 1, coint.vec = NULL, X = NULL,
         unbiased = T, lags = 1, trend = "c", levels = F)
```

where `object` is either a "`coint`" object, usually produced by a call to the function `coint`, or a rectangular data object. If `object` is a "`coint`" object then `coint.rank` specifies the rank of $\mathbf{\Pi}$ to determine the number of cointegrating vectors to be used in the fitted VECM. The cointegrating vectors are then normalized using the Phillips' triangular representation described in the appendix. The lag length and trend specification for the VECM is obtained from the information in the "`coint`" object. The lag length in the fitted VECM, however, is one less than the lag length specified in the "`coint`" object. If `object` is a rectangular data object, then `coint.vec` must be assigned a matrix whose columns represent pre-specified cointegrating vectors. The argument `lags` is then used to specify the lag length of the VECM, and `trend` is used to set the trend specification. The optional argument `X` is used to specify any exogenous variables (e.g. dummy variables for events) other than a constant or trend. The optional argument `levels` determines if the VECM is to be fit to the levels $\mathbf{Y}_t$ or to the first differences $\Delta\mathbf{Y}_t$, and determines if forecasts are to be computed for the levels or the first differences. The result of `VECM` is an object of class "`VECM`", which inherits from "`VAR`" for which there are `print`, `summary`, `plot`, `cpredict` and `predict` methods and extractor functions `coef`, `fitted`, `residuals` and `vcov`. Since "`VECM`" objects inherit from "`VAR`" objects, most of the method and extractor functions for "`VECM`" objects work similarly to those for "`VAR`" objects. The use of `VECM` is illustrated with the following examples.

Example 84 *Maximum likelihood estimation of the VECM for exchange rate data*

Using the "`coint`" object `coint.rc` computed from the VAR(2) model with a restricted constant, the VECM(1) with a restricted constant for the exchange rate data is computed using

```
> vecm.fit = VECM(coint.rc)
> class(vecm.fit)
[1] "VECM"
> inherits(vecm.fit,"VAR")
[1] T
```

The `print` method gives the basic output

```
> vecm.fit

Call:
VECM(test = coint.rc)

Cointegrating Vectors:
          coint.1
    USCNS  1.0000
```

```
      USCNF -1.0058
Intercept* -0.0027

VECM Coefficients:
            USCNS    USCNF
   coint.1  1.7771   1.9610
USCNS.lag1  1.1696   1.2627
USCNF.lag1 -1.2832  -1.3679

Std. Errors of Residuals:
  USCNS   USCNF
 0.0135 0.0136

Information Criteria:
    logL     AIC      BIC       HQ
  2060.2 -4114.4 -4103.9 -4110.1


                  total residual
Degree of freedom:    243        240
Time period: from Apr 1976 to Jun 1996
```

The `print` method output is similar to that created by the `VAR` function. The output labeled `Cointegrating Vectors:` gives the estimated cointegrating vector coefficients normalized on the first variable in the specification of the VECM. To see standard errors for the estimated coefficients use the `summary` method

```
> summary(vecm.fit)

Call:
VECM(test = coint.rc)

Cointegrating Vectors:
            coint.1
             1.0000

     USCNF   -1.0058
 (std.err)    0.0031
  (t.stat) -326.6389

Intercept*   -0.0027
 (std.err)    0.0007
  (t.stat)   -3.9758


VECM Coefficients:
```

```
              USCNS    USCNF
  coint.1  1.7771   1.9610
 (std.err)  0.6448   0.6464
 (t.stat)  2.7561   3.0335


USCNS.lag1  1.1696   1.2627
 (std.err)  0.9812   0.9836
 (t.stat)  1.1921   1.2837


USCNF.lag1 -1.2832 -1.3679
 (std.err)  0.9725   0.9749
 (t.stat) -1.3194 -1.4030


Regression Diagnostics:
              USCNS   USCNF
    R-squared 0.0617 0.0689
Adj. R-squared 0.0538 0.0612
  Resid. Scale 0.0135 0.0136


Information Criteria:
   logL    AIC     BIC      HQ
 2060.2 -4114.4 -4103.9 -4110.1


                total residual
Degree of freedom:   243      240
Time period: from Apr 1976 to Jun 1996
```

The VECM fit may be inspected graphically using the generic **plot** method

```
> plot(vecm.fit)


Make a plot selection (or 0 to exit):

1: plot: All
2: plot: Response and Fitted Values
3: plot: Residuals
4: plot: Normal QQplot of Residuals
5: plot: ACF of Residuals
6: plot: PACF of Residuals
7: plot: ACF of Squared Residuals
8: plot: PACF of Squared Residuals
9: plot: Cointegrating Residuals
10: plot: ACF of Cointegrating Residuals
11: plot: PACF of Cointegrating Residuals
12: plot: ACF of Squared Cointegrating Residuals
```
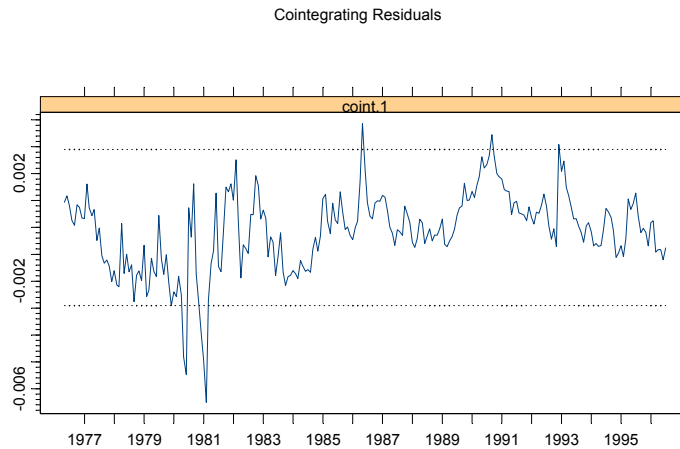
Cointegrating Residuals

FIGURE 12.8. Cointegrating residual from maximum likelihood estimation of VECM(1) for exchange rate data.

```
13: plot: PACF of Squared Cointegrating Residuals
Selection:
```

The first eight plot options are the same as those created for a `"VAR"` object. The remaining plot options allow a graphical inspection of the cointegrating residual. For example, plot option 9 is illustrated in Figure 12.8. The estimated cointegrating residual appears to be $I(0)$.

Example 85 *Estimate VECM with pre-specified cointegrating vector*

For the exchange rate data, the MLE of the normalized cointegrating vector is close to $(1, -1)'$, and the estimate of the restricted constant is close to zero. These are the values implied by the FRUH. To estimate a VECM(1) imposing $\boldsymbol{\beta} = (1, -1)'$ and $\mu_t = 0$ use

```
> beta.true = as.matrix(c(1,-1, 0))
> dimnames(beta.true) = list(c("USCNS","USCNF","Intercept"),
+                            "coint.1")
> vecm.fruh.fit = VECM(uscn.ts, coint.vec = beta.true,
+                 lags = 1, trend = "rc")
```

Since the restricted constant lies in the cointegrating space, the last element of the pre-specified cointegrating vector is the value of the restricted constant. A summary of the restricted VECM fit is:

```
> summary(vecm.fruh.fit)

Call:
VECM(data = uscn.ts, coint.vec = beta.true, lags = 1, trend =
    "rc")

Cointegrating Vectors:
          coint.1
    USCNS       1
    USCNF      -1
Intercept       0

VECM Coefficients:
             USCNS    USCNF
   coint.1  0.0337   0.1354
 (std.err)  0.4442   0.4466
  (t.stat)  0.0758   0.3032


USCNS.lag1   2.1330   2.2781
 (std.err)   0.9535   0.9588
  (t.stat)   2.2371   2.3760


USCNF.lag1  -2.2226  -2.3573
 (std.err)   0.9472   0.9525
  (t.stat)  -2.3465  -2.4748


Regression Diagnostics:
                 USCNS   USCNF
     R-squared  0.0354  0.0393
Adj. R-squared  0.0273  0.0312
  Resid. Scale  0.0137  0.0138


Information Criteria:
      logL        AIC        BIC         HQ
  2053.662  -4101.324  -4090.845  -4097.103


                    total  residual
Degree of freedom:    243       240
Time period: from Apr 1976 to Jun 1996
```

Notice that the VECM with the pre-specified cointegrating vector does not
fit as well as the VECM using the estimated cointegrating vector.

12.5.8   Forecasting from the VECM

Forecasts from a VECM are computed by first transforming the VECM to a VAR using (12.22), and then using the forecasting algorithms for VAR models described in the previous chapter. For VECM models, one may forecast the changes in the variables, $\Delta\mathbf{Y}_t$, or the levels of the variables $\mathbf{Y}_t$. The generic S+FinMetrics functions `predict` and `cpredict` are used to compute unconditional and conditional forecasts from a "VECM" object. The following example illustrates the use of the `predict` method to compute forecasts for the differences and levels of the exchange rate data.

Example 86 *Forecasts from VECM fit to exchange rate data*

The "VECM" object `vecm.fit` was produced with the optional argument `levels=F`. Consequently, the `predict` method will produce forecasts for the changes in s_t and f_t. To compute h-step forecasts for Δs_t and Δf_t for $h = 1, \ldots, 12$ use

```
> vecm.fcst = predict(vecm.fit,n.predict=12)
> class(vecm.fcst)
[1] "forecast"
```

To see the forecast and forecast standard errors use

```
> summary(vecm.fcst)

Predicted Values with Standard Errors:

                 USCNS    USCNF
 1-step-ahead -0.0105  -0.0110
    (std.err)  0.0136   0.0136
 2-step-ahead -0.0130  -0.0139
    (std.err)  0.0183   0.0183
...
12-step-ahead -0.0237  -0.0260
    (std.err)  0.0435   0.0432
```

By default, the forecasts are computed using the chain-rule of forecasting. To compute simulation-based forecasts use `method = "mc"` or `method = "bootstrap"` in the call to `predict`.

To see the forecasts with standard error bands along the original data use

```
> plot(vecm.fcst, xold=diff(uscn.ts), n.old=12)
```

Since the forecasts are of the first differenced data, the data passed to `xold` must be first differenced. The resulting plot is shown in Figure 12.9.

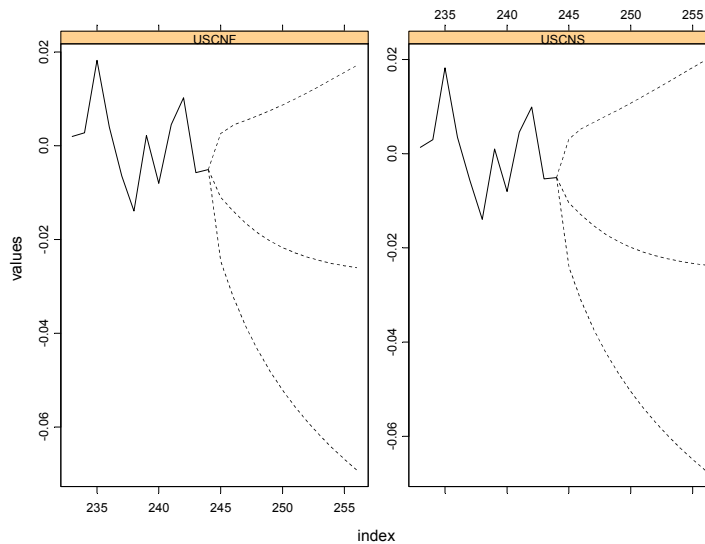To compute forecasts for the levels s_t and f_t, re-fit the VECM with the optional argument `levels=T`

FIGURE 12.9. VECM forecasts of first differences of exchange rate data.

```
> vecm.fit.level = VECM(coint.rc, levels=T)
```

and then call the **predict** method as before

```
> vecm.fcst.level = predict(vecm.fit.level, n.predict=12)
> summary(vecm.fcst.level)

Predicted Values with Standard Errors:

                 USCNS    USCNF
 1-step-ahead -0.3150 -0.3154
    (std.err)  0.0136  0.0136
 2-step-ahead -0.3157 -0.3161
    (std.err)  0.0183  0.0183
...
12-step-ahead -0.3185 -0.3193
    (std.err)  0.0435  0.0432
```

To plot the forecasts use

```
> plot(vecm.fcst.level, xold=uscn.ts, n.old=12)
```

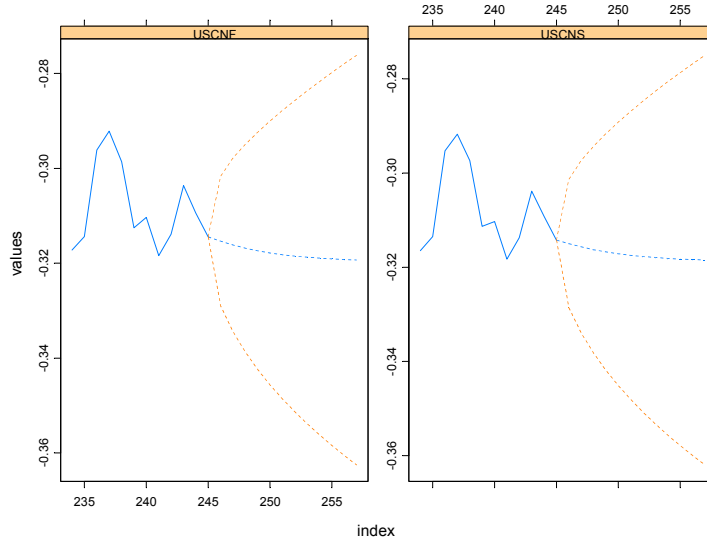The resulting plot is shown in Figure 12.10.

FIGURE 12.10. VECM forecasts of levels of exchange rate data.

12.6 Appendix: Maximum Likelihood Estimation of a Cointegrated VECM

The following brief discussion of maximum likelihood estimation of the cointegrated VECM (12.23) follows Hamilton (1994) and Johansen (1995). For simplicity, assume the absence of deterministic terms.

- Concentrate the likelihood function with respect to the error covariance matrix and short-run dynamics by estimating the regressions

$$\Delta \mathbf{Y}_t \;=\; \hat{\boldsymbol{\Gamma}}_1 \Delta \mathbf{Y}_{t-1} + \cdots \hat{\boldsymbol{\Gamma}}_{p-1} \Delta \mathbf{Y}_{t-p+1} + \hat{\mathbf{U}}_t$$
$$\mathbf{Y}_t \;=\; \hat{\boldsymbol{\Phi}}_1 \Delta \mathbf{Y}_{t-1} + \cdots \hat{\boldsymbol{\Phi}}_{p-1} \Delta \mathbf{Y}_{t-p+1} + \hat{\mathbf{V}}_t$$

- Form the sample covariance matrices

$$\mathbf{S}_{00} = \frac{1}{T} \sum_{t=1}^{T} \hat{\mathbf{U}}_t \hat{\mathbf{U}}_t', \;\; \mathbf{S}_{01} = \frac{1}{T} \sum_{t=1}^{T} \hat{\mathbf{U}}_t \hat{\mathbf{V}}_t', \;\; \mathbf{S}_{11} = \frac{1}{T} \sum_{t=1}^{T} \hat{\mathbf{V}}_t \hat{\mathbf{V}}_t'$$

- Solve the eigenvalue problem

$$|\lambda \mathbf{S}_{11} - \mathbf{S}_{10} \mathbf{S}_{00}^{-1} \mathbf{S}_{01}| = 0$$

giving ordered eigenvalues[6] $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_n$ and associated eigenvectors $\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_n$ that satisfy

$$
\begin{aligned}
\hat{\lambda}_i \mathbf{S}_{11} \hat{\mathbf{v}}_i &= \mathbf{S}_{10} \mathbf{S}_{00}^{-1} \mathbf{S}_{01}, \ i = 1, \ldots, n \\
\hat{\mathbf{V}} \mathbf{S}_{11} \hat{\mathbf{V}} &= \mathbf{I}_n
\end{aligned}
$$

where $\hat{\mathbf{V}} = [\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_n]$

- The unnormalized MLE for the $(n \times r)$ matrix $\boldsymbol{\beta}$ based on $0 < r < n$ cointegrating vectors is given by the first r eigenvectors

$$
\hat{\boldsymbol{\beta}}_{mle} = (\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_r)
$$

- Form the normalized estimator $\hat{\boldsymbol{\beta}}_{c,mle}$ by imposing the appropriate normalizing and identifying restrictions. The MLE for the normalized estimator of α may be computed as

$$
\hat{\boldsymbol{\alpha}}_{c,mle} = \mathbf{S}_{01} \hat{\boldsymbol{\beta}}_{c,mle}
$$

- The maximum likelihood estimators for the remaining parameters may be obtained by multivariate least squares of the VECM with $\boldsymbol{\beta}$ replaced by $\hat{\boldsymbol{\beta}}_{c,mle}$

$$
\Delta \mathbf{Y}_t = \boldsymbol{\alpha}_c \hat{\boldsymbol{\beta}}'_{c,mle} \mathbf{Y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta \mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t
$$

- The maximized value of the likelihood function based on r cointegrating vectors used in the construction of LR tests for the number of cointegrating vectors is

$$
L_{\max}^{-2/T} \propto |\mathbf{S}_{00}| \prod_{i=1}^{r} (1 - \hat{\lambda}_i)
$$

- Estimates of the orthogonal complements of $\boldsymbol{\alpha}_c$ and $\boldsymbol{\beta}_c$ are given by

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}}_{c,\perp} &= \mathbf{S}_{00}^{-1} \mathbf{S}_{11} (\hat{\mathbf{v}}_{r+1}, \ldots, \hat{\mathbf{v}}_n) \\
\hat{\boldsymbol{\beta}}_{c,\perp} &= \mathbf{S}_{11} (\hat{\mathbf{v}}_{r+1}, \ldots, \hat{\mathbf{v}}_n)
\end{aligned}
$$

Let $\mathbf{c}$ be any $(n \times r)$ matrix such that $\boldsymbol{\beta}' \mathbf{c}$ has full rank. Then $\boldsymbol{\beta}$ may be normalized as

$$
\boldsymbol{\beta}_c = \boldsymbol{\beta} (\mathbf{c}' \boldsymbol{\beta})^{-1}
$$

---

[6]These eigenvalues are the squared canonical correlations between $\mathbf{Y}_t$ and $\Delta \mathbf{Y}_t$ corrected for $\Delta \mathbf{Y}_{t-1}, \ldots, \Delta \mathbf{Y}_{t-p+1}$. Johansen (1995) describes how to solve for the eigenvalues.

satisfying $\mathbf{c}'\boldsymbol{\beta}_c = \mathbf{I}_r$ provided $|\mathbf{c}'\boldsymbol{\beta}| \neq \mathbf{0}$. Johansen suggests setting

$$\mathbf{c} = (\mathbf{I}_r \vdots \mathbf{0})' \qquad (12.28)$$

This choice of $\mathbf{c}$ corresponds to solving the cointegrating relations $\boldsymbol{\beta}'\mathbf{Y}_t$ for the first r variables. To see this, let $\mathbf{Y}_t = (\mathbf{Y}'_{1t}, \mathbf{Y}'_{2t})'$, where $\mathbf{Y}_{1t}$ contains the first r variables and $\mathbf{Y}_{2t}$ contains the remaining $n-r$ variables, and let $\boldsymbol{\beta}' = (-\boldsymbol{\beta}_1 \vdots \boldsymbol{\beta}_2)$, where $\boldsymbol{\beta}_1$ is $(r \times r)$ and $\boldsymbol{\beta}_2$ is $(r \times (n-r))$. Then $\boldsymbol{\beta}'\mathbf{c} = -\boldsymbol{\beta}_1$ and

$$\boldsymbol{\beta}_c = \begin{pmatrix} \mathbf{I}_r \\ -\boldsymbol{\beta}_1^{-1}\boldsymbol{\beta}_2 \end{pmatrix} \qquad (12.29)$$

provided $\boldsymbol{\beta}_1$ has full rank r.

   Some examples will help clarify the normalization scheme described above. First, suppose there is only one cointegrating vector so that $r = 1$. Let the $(n \times 1)$ vector $\boldsymbol{\beta} = (-\beta_1, \beta_2, \ldots, \beta_n)'$ and define $\mathbf{c} = (1, 0, \ldots, 0)'$ so that $\boldsymbol{\beta}'\mathbf{c} = -\beta_1$ and $\boldsymbol{\beta}_c = (1, -\beta_2/\beta_1, \ldots, -\beta_n/\beta_1)'$ is the normalized cointegrating vector. Notice that this normalization requires $\beta_1 \neq 0$. Next, suppose there are two cointegrating vectors, $r = 2$, and let

$$\boldsymbol{\beta}' = \begin{pmatrix} -\beta_{11} & -\beta_{12} & \beta_{13} & \cdots & \beta_{1n} \\ -\beta_{21} & -\beta_{22} & \beta_{23} & \cdots & \beta_{2n} \end{pmatrix} = \left(-\boldsymbol{\beta}_1 \vdots \boldsymbol{\beta}_2\right)$$

$$\mathbf{c}' = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \end{pmatrix} = (\mathbf{I}_2 \vdots \mathbf{0})$$

such that $\boldsymbol{\beta}_1$ has full rank. Then $\boldsymbol{\beta}'\mathbf{c} = -\boldsymbol{\beta}_1$ and

$$\boldsymbol{\beta}'_c = \begin{pmatrix} 1 & 0 & \beta^*_{13} & \cdots & \beta^*_{1n} \\ 0 & 1 & \beta^*_{23} & \cdots & \beta^*_{2n} \end{pmatrix} = (\mathbf{I}_2 \vdots \boldsymbol{\beta}^*)$$

where $\boldsymbol{\beta}^* = -\boldsymbol{\beta}_1^{-1}\boldsymbol{\beta}_2$. The rows of $\boldsymbol{\beta}'_c$ are the normalized cointegrating vectors.

12.7   References

[1] ALEXANDER, C. (2001). *Market Models: A Guide to Financial Data Analysis*, John Wiley and Sons.

[2] COCHRANE, J. (2001). *Asset Pricing.* Princeton University Press, New Jersey.

[3] ENGLE, R.F. AND C.W.J. GRANGER (1987). "Co-Integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251-276.

[4] GRANGER, C.W.J. AND P.E. NEWBOLD (1974). "Spurious Regression in Econometrics," *Journal of Econometrics*, 2, 111-120.

[5] HAMILTON, J.D. (1994). *Time Series Analysis*, Princeton Unversity Press, New Jersey.

[6] HANSEN, B.E. (1992). "Efficient Estimation and Testing of Cointegrating Vectors in the Presence of Deterministic Trends," *Journal of Econometrics*, 53, 87-121.

[7] HAYASHI, F. (2000). *Econometrics*, Princeton University Press, New Jersey.

[8] JOHANSEN, S. (1988). "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

[9] JOHANSEN, S. (1995). *Likelihood Based Inference in Cointegrated Vector Error Correction Models*, Oxford University Press, Oxford.

[10] MacKINNON, J. (1996). "Numerical Distribution Functions for Unit Root and Cointegration Tests," *Journal of Applied Econometrics*, 11, 601-618.

[11] MILLS, T. (1999). *The Econometric Analysis of Financial Time Series*, Cambridge University Press, Cambridge.

[12] OSTERWALD-LENUM, M. (1992). "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Statistics," *Oxford Bulletin of Economics and Statistics*,54, 461-472.

[13] PHILLIPS, P.C.B. (1986). "Understanding Spurious Regression in Econometrics," *Journal of Econometrics*, 33, 311-340.

[14] PHILLIPS, P.C.B. (1991). "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283-306.

[15] PHILLIPS, P.C.B. AND S. OULIARIS (1990). "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, 73-93.

[16] REIMARS, H.-E. (1992). "Comparisons of Tests for Multivariate Cointegration," *Statistical Papers*, 33, 335-359.

[17] REINSEL, G.C. AND S.K. AHN (1992). "Vector Autoregression Models with Unit Roots and Reduced Rank Structure: Estimation, Likelihood Ratio Test, and Forecasting," *Journal of Time Series Analysis*, 13, 353-375.

[18] SIMS, C.A., J.H. STOCK AND M.W. WATSON (1990). "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113-144.

[19] STOCK, J.H. (1987). "Asymptotic Properties of Least Squares Estimation of Cointegrating Vectors," *Econometrica*, 55, 1035-1056.

[20] STOCK, J.H. AND M.W. WATSON (1989). "Variable Trends in Economic Time Series," *Journal of Economic Perspectives*, Vol. 2(3), 147-174.

[21] STOCK, J.H. AND M.W. WATSON (1993). "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783-820.

[22] TSAY, R. (2001). *The Analysis of Financial Time Series,* John Wiley & Sons, New York.

[23] ZIVOT, E. (2000). "Cointegration and Forward and Spot Exchange Rate Regressions," *Journal of International Money and Finance*, 19, 785-812.