6
Further Topics in the Analysis of Non-Stationary Time Series

6.1 Introduction

In this chapter three further topics are considered in some detail: estimation of models with I(2) variables; forecasting; and structural models with short-run behaviour driven by expectations. Though mathematically the notions of order of integration and cointegration are exact, in practice they are valid to the best approximation or resolution that the data may permit. To define an order of integration as a specific integer quantity is to assume that the series is approximated by a single well-defined time series process across the sample. Time series data for developed economies have exhibited many features, from behaviour that might be viewed as purely stationary through to series that require first or second differencing to render them stationary. Some nominal series in first differences may require further differencing, which suggests that the original nominal series are of order I(2) or higher when further differencing is required. In this chapter, discussion is limited to processes up until I(2).

The condition required for a series to be considered to be I(1), as compared with one exhibiting further features only consistent with I(2) behaviour, is necessary and sufficient for cointegration amongst I(1) series, but beyond testing this condition, there is a well defined procedure for inference and estimation of I(2) processes (Johansen 1992, 1995). It might often be difficult to distinguish between an I(1) and an I(2) series, which suggests that series, which appear to be I(2), are being approximated to some order of accuracy by second differences. Alternatively, these series may be better modelled using non-integer orders of differencing (Granger and Joyeux 1980; Hosking 1981). To this end, the question of fractional processes and long-memory will be discussed briefly after the section on I(2) behaviour. A further reason why it might be difficult to detect the order of integration of a series may be due to the existence of structural breaks. This opens up a plethora of potential difficulties for any form of structural modelling. Breaks in structure have a

number of forms when conventional (I(0)) linear econometrics is considered, but beyond slope and intercept shifts, there are other types of intercept correction used in macro modelling (see Clements and Hendry 1998, 2001). The break may also apply to the cointegrating relations (co-breaks) or in the order of integration and cointegration. Testing was limited in chapter 4 to recursive break tests and tests with a known break in structure that could be corrected by the use of dummy variables. In this chapter forecast performance is compared by considering the difference between forecasts made with and without the imposition of cointegration. Specifically, the simulation results of Hendry and Clements and Lin and Tsay are evaluated.

Once the notion of forecast failure is considered, then issues associated with our ability to detect short-run structure arise. In this context, there can be no difference between estimating a structural relationship as compared with a reduced form, except for the added efficiency that might derive from the imposition of further restrictions on the long-run and short-run parameters. There are a number of approaches to defining structural models under cointegration of which the best defined follows from the work of Pesaran et al. (2000). The elegance of the Johansen approach is lost once the long-and the short-run coefficients are interrelated, as testing for a unit root in multivariate processes cannot be readily disentangled from the estimation of the long-run and short-run parameters. In particular, when the long-run parameters are embedded within the short run, as occurs with models with future expectations, then testing for cointegration is less straightforward. Here, the impact of forward-looking behaviour is considered in terms of exogenous processes that are weakly and cointegrating exogenous and then processes that have unit roots in the exogenous variables. The simple method suggested by Dolado et al. (1991) is considered along with an extension of this method to the multivariate context by Engsted and Haldrup (1997). An alternative maximum likelihood approach is discussed here, though the inference is contaminated by both the unit root and generated regressor problem.

6.2 Inference and estimation when series are not I(1)

In this section the I(2) approach advanced in Johansen (1992a), is considered along with some discussion of multi-cointegrated and fractional processes. Whether a series is I(1), close to I(1) in levels or differences, is a matter of debate. To some extent cointegration operates beyond the framework of this debate, because long memory processes may also interact, as has been observed recently by Abadir and Talmain (2002). From the original definition of cointegration due to Engle and Granger (1987) series of order I(j) cointegrate and I(1) and I(0) series may also combine in the manner described by Flôres and Szafarz (1996). One estimator, which combines I(0), I(1) and I(2)

processes is that given in Johansen (1992a). This assumes that differenced series are of integer order, which rules out the possibility that series such as inflation rates are fractional processes. The distinction between long memory and non-stationarity might be viewed as semantic for the data sets readily available, but one cannot dismiss the possibility that series may move across orders of integration from non-stationarity through long-memory to stationarity. In this light the series might never be purely stationary or non-stationary. Where this would appear to accord with sound economic principle then one might have to look for the best approximation.[1]

6.2.1 Cointegration when series are I(2)

Consider the cointegration case developed by Engle and Granger (1987), where all the series are I(2). It follows from our discussion of cointegration in chapter 4 that second differences have the following Wold decomposition:

$$\Delta^2 x_t = C(L)\epsilon_t$$

and Δx_t cointegrate when $\beta'_{I(2)}\, C\,(1) = 0$ and $\beta'_{I(2)}\, \Delta x_t = \eta_{I(2)t} \sim I(0)$. If a left-hand factor can be extracted in the manner described in section 4.5, then:

$$\Delta^2 x_t = C_0(L)C_1(L)\epsilon_t. \tag{6.1}$$

It is possible to transform the Wold form into an error-correcting VARMA when $FC(1) = 0$, and F is an idempotent matrix. Therefore:

$$(\Delta I - FL)\Delta x_t = C_1(L)\epsilon_t. \tag{6.2}$$

When $C_1\,(L)$ has no more unit roots, then an I(2) cointegrating VAR exists in second differences:

$$\Gamma(L)\Delta^2 x_t = \prod_{I(2)} \Delta x_{t-1} + \epsilon_t$$

where $\Pi_{I(2)} = \alpha_{I(2)}\beta'_{I(2)} = F$. This has been called balanced I(2) behaviour by Juselius (1995). Now consider the case where $C(1)$ has further unit roots, then it might be possible to undertake a further factorization when a left-hand term $C_{01}(L) = (I - GL)$ can be extracted and $GC_1(1) = 0$. Therefore:

$$(\Delta I - FL)\Delta x_t = (I - GL)C_{11}(L)\epsilon_t \tag{6.3}$$

$$(\Delta I - GL)(\Delta I - FL)x_t = C_{11}(L)\epsilon_t \tag{6.4}$$

The following I(2) representation can be readily derived from multiplying through the two left-hand divisors above. Therefore:

$$\Delta^2 x_t - F\Delta x_{t-1} - G(\Delta x_{t-1} - Fx_{t-2}) = C_{11}(L)\epsilon_t,$$

transforming to the VAR by inverting $C_{11}(L)$ and applying the reparameterization $(A(1)L + (1 - L)A^{*}(L))$ to produce terms in first differences and $(A(0) + (1 - L)A^{+}(L) + (1 - L)^2 A^{++}(L))$ terms in levels,

$$\Psi(L)\Delta x_t - A(1)F\Delta x_{t-2} + A(0)GFx_{t-2} - A(1)G\Delta x_{t-2} + A^{+}(0)GF\Delta x_{t-2} = \epsilon_t$$

or

$$\Psi(L)\Delta x_t = A(1)F\Delta x_{t-2} - A(0)(GF)x_{t-2} + (A^{x}(1)G - A^{x}(0)GF)\Delta x_{t-2}) + \epsilon_t \quad (6.5)$$

where $\Psi(L) = (A(L) - A^{*}(L)(F + G)L + A^{++}(L)GFL^2)$, $A^{x}(1) = A(0)^{-1}A(1)$ and $A^{x}(0) = A(0)^{-1}A^{+}(0)$. Assuming a VAR(2) system with $A(1)F = \Sigma\alpha_{\perp}(\alpha'_{\perp}\Sigma\alpha_{\perp})^{-1}\kappa'\tau'$, $F = H^{-1}M_{I(2)}H$, $M_{I(2)} = diag(1 \ldots 1, 0 \ldots 0)$, $A(0)G = \alpha\rho'\tau'$ and $(A^{x}(1)G - A^{x}(0)GF) = \alpha\psi'$, then (6.5) is a restricted version of the I(2) representation in Hansen and Johansen (1998):

$$\Delta^2 x_t = \sum \alpha_{\perp}(\alpha'_{\perp} \sum \alpha_{\perp})^{-1}\kappa'\tau'\Delta x_{t-1} + \alpha(\rho'\tau' x_{t-1} - \psi'\Delta x_{t-1}) + \epsilon_t. \quad (6.6)$$

In the notation of Hansen and Johansen, α is $n \times r$, ρ is $(r + s) \times r$, τ is $n \times (r + s)$, ψ is $n \times r$, κ is $(r + s) \times (n - r)$ and Σ is $n \times n$.

Next the approach due to Johansen (1992) is considered for testing for cointegration in I(2) systems, then an example is discussed along with identification and estimation.

6.2.1.1 The Johansen procedure for testing cointegrating rank with I(2) variables

Prior to any discussion of the appropriate method of estimation the more conventional VECM for the I(2) case is presented (Johansen 1995a):

$$\Delta^2 x_t = \alpha\beta' x_{t-1} - \Gamma\Delta x_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta^2 x_{t-i} + N_0 D_t + \epsilon_t. \quad (6.7)$$

Where $\Gamma = \Sigma\alpha_{\perp}(\alpha'_{\perp}\Sigma\alpha_{\perp})^{-1}\kappa'\tau' + \alpha\psi'$, α and $\beta' = \rho'\tau'$ are the conventional loadings and cointegrating vectors for the case in which series of any order may collapse to a stationary linear combination. If $\Gamma = 0$, then this is the cointegration case considered by Engle and Granger (1987) where all the series are I(2) and:

$$\Delta^2 x_t = \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta^2 x_{t-i} + N_0 D_t + \epsilon_t. \quad (6.8)$$

Alternatively, when $\alpha\beta' = 0$ and the differenced I(1) series have linear combinations that are stationary:

$$\Delta^2 x_t = -\Gamma\Delta x_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta^2 x_{t-i} + N_0 D_t + \epsilon_t \quad (6.9)$$

where $-\Gamma = (\alpha'_{\perp})^{-1}\kappa'\tau' = \alpha_{I(2)}\beta'_{I(2)}$ as $\alpha'_{\perp}$ has full rank, because $\alpha\beta' = 0$ implies $\alpha = 0$ and $\rho = 0$. The full I(2) case allows for the possibility of cointegration amongst I(2) series that become I(0) in combination, and cointegration amongst I(1) series that become I(0).

Clearly, (6.8) can be estimated using the Johansen procedure, except the regression that is purged of short-run behaviour in, for example the VAR(1) case is:

$$R_{0,t} = \alpha\beta' R_{1,t}$$

or

$$\Delta^2 x_t = \prod x_{t-1},$$

and decomposition and testing follows in the usual way (see sections 4.3–4.4).

Alternatively, for the VAR(1) case associated with (6.9) the estimation procedure is in every respect the same as that derived by Johansen (1991), except the data are first and second differenced. For the VAR(1) case this involves estimating the following model:

$$R_{0,t} = \alpha_{I(2)}\beta'_{I(2)}R_{1,t} = (\alpha'_{\perp})^{-1}\kappa'\tau'R_{1,t}$$

or

$$\Delta^2 x_t = -\Gamma\Delta x_{t-1}$$

This becomes more complicated when the two types of cointegration are combined, then (6.7) needs to be estimated, but this requires two blocks of reduced rank tests to be undertaken. One procedure for undertaking this analysis would be to consider the unit roots associated with cointegration amongst I(2) series whose first differences cointegrate. However: when $\alpha\beta' \neq 0$, then the model to be estimated will either require very long lags as the moving average terms $\beta' x_{t-1} = J(L)\varepsilon_{t-1}$ have been omitted or the Johansen approach might be applied to a VARMA(1,q) model. To see this re-write (6.2) as:

$$\Delta^2 x_t - F\Delta x_{t-1} = C_1(L)\epsilon_t. \tag{6.10}$$

If (6.10) were to be estimated, then the method must account for roots on the unit circle as when the level terms cointegrate, $C_1(L)$ contains further unit roots. Otherwise, the conventional VAR associated with this problem is of infinite order and not conventionally invertible. There is no unique way of deriving the estimator and in general the existence of the time series representation cannot be proven.

In general, the case with both I(2) and I(1) interdependencies can be handled by considering the solution to two reduced rank problems:

$$\prod = \alpha\beta'$$
$$\alpha'_{\perp}\Gamma\beta_{\perp} = \xi\gamma'$$

where ξ and γ are $(n - r) \times s$ dimensioned matrices. To simplify the exposition quadratic trends are not considered here. Johansen (1995) suggests the problem is made tractable by correcting the short-run behaviour firstly for the usual cointegration case as the I(2) series collapse to linear combinations that are stationary. When the Frisch–Waugh theorem is applied to purge the short-run relationship of the nuisance terms, then $\Delta^2 x_t$ and x_{t-1} are both regressed on Δx_{t-1} and Δx^2_{t-i} $i = 1, 2, \ldots, n - 1$ by ordinary least squares. The residuals from these regressions will not be correlated with the lagged second differences and the influence of the first form of cointegration will be removed. Again $R_{0,t}$ and $R_{1,t}$ are, in essence, the $n \times 1$ residual vectors from regressions with Δx_t and x_{t-1} as the dependent variables. The following regressions, yield estimates of the first long-run parameter matrix:

$$R_{0,t} = \alpha\beta' R_{1,t} = \prod R_{1,t}. \tag{6.11}$$

Now β is calculated by solving the conventional eigenvalue problem for the I(1) case and the usual I(1) analysis is undertaken to determine cointegrating

rank (section 4.4). To confirm that the I(1) analysis is valid the test for I(2) components discussed previously in 4.4.5 needs to be undertaken, this relates to the solution to the second reduced rank problem, that is rank($\alpha'_\perp\Gamma\beta_\perp$) = $n - r$. Should this matrix not have full rank, then there are I(2) components not accounted for. Next an analysis of the I(2) components of the model is undertaken, controlling for the I(1) variables. Subject to knowledge of (α, β, r) the I(1) terms are eliminated by pre-multiplying (6.7) by $\alpha'_\perp$:

$$\begin{aligned}\alpha'_\perp\Delta^2 x_t &= \alpha'_\perp\alpha\beta' x_{t-1} - \alpha'_\perp\Gamma\Delta x_{t-1} + \sum_{i=1}^{p-1}\alpha'_\perp\Psi_i\Delta^2 x_{t-i} + \alpha'_\perp N_0 D_t + \alpha'_\perp\epsilon_t \\ &= -\alpha'_\perp\Gamma\Delta x_{t-1} + \sum_{i=1}^{p-1}\alpha'_\perp\Psi_i\Delta^2 x_{t-i} + \alpha'_\perp N_0 D_t + \alpha'_\perp\epsilon_t. \end{aligned} \tag{6.12}$$

This is an $n - r$ dimensioned system and in the pure I(1) case $rank(\alpha'_\perp\Gamma) = n - r$. The test for further I(2) trends is undertaken by regressing $\alpha'_\perp\Delta^2 x_t$ and $\alpha'_\perp\Delta x_{t-1}$ on $\alpha'_\perp\Delta x^2_{t-i}$ $i = 1, 2, \ldots, n-1$. The residuals from the regressions of $R_{0,t}$ and $R_{1,t}$ for this case yield an eigenvalue problem that can be solved in the usual way. The Johansen test for this case determines the *rank* ($\alpha'_\perp\Gamma\beta_\perp$) = s, where $0 \le s \le n - r$ and associated with s significant eigenvalues is the $s \times n - r$ matrix of eigenvectors γ' that define common trends. If all the variables are I(1), then the system separates into r stationary variables ($\beta' x_{t-1}$) and $n - r$ common trends $\gamma'\Delta x_{t-1}$. Otherwise there are s common trends and $n - r - s$, $I(2)$ trends.

To complete the I(2) analysis, (6.7) is now multiplied by the $r \times n$ matrix $\overline{\alpha}'$:

$$\begin{aligned}\overline{\alpha}'\Delta^2 x_t &= \overline{\alpha}'\alpha\beta' x_{t-1} - \overline{\alpha}'\Gamma\Delta x_{t-1} + \sum_{i=1}^{p-1}\overline{\alpha}'\Psi_i\Delta^2 x_{t-i} + \overline{\alpha}' N_0 D_t + \overline{\alpha}'\epsilon_t \\ &= \beta' x_{t-1} - \overline{\alpha}'\Gamma\Delta x_{t-1} + \sum_{i=1}^{p-1}\overline{\alpha}'\Psi_i\Delta^2 x_{t-i} + \overline{\alpha}' F D_t + \overline{\alpha}'\epsilon_t \end{aligned} \tag{6.13}$$

where $\overline{\alpha}'\alpha = I_r$. Subtracting (6.13) from $\omega \times$ (6.12):

$$\begin{aligned}\overline{\alpha}'\Delta^2 x_t - \omega\alpha'_\perp\Delta^2 x_t = \beta' x_{t-1} - \overline{\alpha}'\Gamma\Delta x_{t-1} + \sum_{i=1}^{p-1}\overline{\alpha}'\Psi_i\Delta^2 x_{t-i} + \overline{\alpha}' N_0 D_t + \overline{\alpha}'\epsilon_t + \\ \omega(\alpha'_\perp\Gamma\Delta x_{t-1} - \sum_{i=1}^{p-1}\alpha'_\perp\Psi_i\Delta^2 x_{t-i} - \alpha'_\perp N_0 D_t - \alpha'_\perp\epsilon_t) \\ \overline{\alpha}'\Delta^2 x_t - \beta' x_{t-1} = \omega\alpha'_\perp\Delta^2 x_t - (\overline{\alpha}' - \omega\alpha'_\perp)(\Gamma\Delta x_{t-1} - \\ \sum_{i=1}^{p-1}\Psi_i\Delta^2 x_{t-i} - N_0 D_t) + (\overline{\alpha}' - \omega\alpha'_\perp)\epsilon_t \end{aligned} \tag{6.14}$$

where $\omega = \Sigma_{\alpha\alpha_\perp}\Sigma^{-1}_{\alpha_\perp\alpha_\perp}$, $\Sigma_{\alpha\alpha_\perp} = \overline{\alpha}'\Sigma\alpha_\perp$ and $\Sigma_{\alpha_\perp\alpha_\perp} = \alpha'_\perp\Sigma\alpha_\perp$ The errors of (6.12) and (6.14) are independent by construction. While the parameters of (6.12), ($\alpha'_\perp\Gamma$, $\alpha'_\perp\Psi_i\Sigma_{\alpha_\perp\alpha_\perp}$) and (6.14), ($\omega$,($\overline{\alpha}' - \omega\alpha'_\perp$)$\Gamma$,($\overline{\alpha}' - \omega\alpha'_\perp$)$\Psi_i$, ($\overline{\alpha}' - \omega\alpha'_\perp$)$N_0$) are variation free. It follows that the parameters (Γ, Ψ_i, N_0, Σ) can be disentangled from the

above reparameterization. If there are no further cross-equation restrictions on the higher-order dynamics and cointegration, then (6.12) and (6.14) can be analyzed separately, while the dependence that operates on the common trends applies to (6.12) alone.

The second reduced rank hypothesis is:

$$H_{r,s}: rank(\alpha_\perp'\Gamma\beta_\perp) = rank(\xi\gamma') = s$$

where $0 \leq s \leq n - r$. Using the identity $I = \bar{\beta}\beta' + \beta_\perp\bar{\beta}'_\perp$, the variables $\beta'\Delta x_{t-1}$ and $\bar{\beta}'\Delta x_{t-1}$ may be introduced into (6.12):

$$\begin{aligned}\alpha_\perp'\Delta^2 x_t &= -\alpha_\perp'\Gamma(\bar{\beta}\beta' + \beta_\perp\bar{\beta}_\perp')\Delta x_{t-1} + \sum_{i=1}^{p-1}\alpha_\perp'\Psi_i\Delta^2 x_{t-i} + \alpha_\perp' N_0 D_t + \alpha_\perp'\epsilon_t \\ &= -\alpha_\perp'\Gamma\bar{\beta}\beta'\Delta x_{t-1} + \alpha_\perp'\Gamma\beta_\perp\bar{\beta}_\perp'\Delta x_{t-1} + \sum_{i=1}^{p-1}\alpha_\perp'\Psi_i\Delta^2 x_{t-i} + \alpha_\perp' N_0 D_t + \alpha_\perp'\epsilon_t \qquad (6.15) \\ &= -\alpha_\perp'\Gamma\bar{\beta}\beta'\Delta x_{t-1} + \xi\gamma'\bar{\beta}_\perp'\Delta x_{t-1} + \sum_{i=1}^{p-1}\alpha_\perp'\Psi_i\Delta^2 x_{t-i} + \alpha_\perp' N_0 D_t + \alpha_\perp'\epsilon_t. \qquad (6.16)\end{aligned}$$

The parameters (Γ, Ψ_i for $i = 1, \ldots n - 1$ N_0) can be estimated by regressing $\bar{\alpha}'\Delta^2 x_t - \beta' x_{t-1}$ on $\alpha_\perp'\Delta^2 x_t$, Δx_{t-1}, $\Delta^2 x_{t-i}$ and D_t. The dependence amongst the s common trends can be determined from the regression:

$$R_{0,t} = -\xi\gamma' R_{1,t} \qquad (6.17)$$

where $R_{0,t}$ and $R_{1,t}$ are residuals based on regressing $\alpha_\perp'\Delta^2 x_t$ and $\bar{\beta}_\perp'\Delta x_t$ respectively on $\beta'\Delta x_{t-1}$, $\Delta^2 x_{t-i}$ for $i = 1, \ldots p - 1$ and D_t. The likelihood ratio test statistic is based on the solution to the eigenvalue problem $|\lambda S_{1,1} - S_{1,0}\, S^{-1}{}_{0,0} S_{0,1}| = 0$, calculated from sample product moments derived for the I(2) case using:

$$S_{i,j} = T^{-1}\sum_{i=1}^{T} R_{i,t}R_{j,t}' \text{ for } i = 0,1$$

It follows that s is selected by calculating the maximal eigenvalue test:

$$LR(s_0, s_1) = -T\left[\sum_{i=s_0+1}^{s_1}\log(1-\lambda_i)\right]$$

and for an appropriate choice of s the matrix γ' is the matrix whose columns are the eigenvectors associated with the first s significant eigenvalues.

An alternative approach is derived in Johansen (1997) and Hansen and Johansen (1998) using (6.6) where the parameters to be estimated that are variation free are (α, ρ, τ, Σ, κ, ψ).

6.2.1.2 An example of I(2)

Identification and model selection in the I(2) case is more complicated than in the I(1) case and partial consideration of the null of cointegration conditioned

on the notion that the series are all I(1) may not be valid (Paruolo 1996). When the series are I(2) they become stationary by virtue of a combination of I(1) and I(2) processes and from (6.6) the cointegrating relations have the following form:

$$\begin{aligned}\alpha(\rho'\tau'x_{t-1} - \psi'\Delta x_{t-1}) &= \alpha\rho'\tau'x_{t-1} - \alpha\psi'\Delta x_{t-1} \\ &= \alpha\beta'x_{t-1} - \alpha\psi'\Delta x_{t-1} = \alpha(\beta'x_{t-1} - \psi'\Delta x_{t-1}).\end{aligned}$$

Engle and Yoo (1989) defined cointegrating relationships of the form $\beta'x_{t-1} - \psi'\Delta x_{t-1}$ as polynomial cointegration. To observe this re-write the cointegrating vectors as a lag polynomial $\beta(L)$ in x:

$$\beta(L)'x_t = \beta'x_{t-1} - \psi'x_{t-1} + \psi'x_{t-2} = ((\beta' - \psi')I + \psi'L)x_{t-1}.$$

The cointegrating vectors reduce to linear combinations ($\beta'x_{t'}$) of x_{t-1} (Engle and Granger 1987, when either $\psi' = 0$ or $\psi' = \alpha_\perp$ and $\Gamma = \kappa'\tau'$. In general, (6.7) has r linear combination of I(2) variables that are I(0), s independent linear combinations of I(1) variables that are I(0) and $n - r - s$ variables that follow I(2) trends. If, in addition, $\Gamma = \kappa'\tau' = 0$, then $s = 0$ and there are $n - r$, I(2) trends rendered stationary by the second difference operator; the case considered by Engle and Granger (1987).

It was suggested in Hunter (1992a) that some of the series analyzed by Johansen and Juselius (1992) were I(2). In response to this suggestion Hunter and Simpson (1995) analyzed a system in which the UK inflation series enters the model in first difference form, but they based their analysis on a longer data set. Here, the extended VAR(2) model estimated by Hunter (1992a) is tested for I(2) behaviour. For this example, $n = 6$, $x_t' = [p_{0t}p_{1t}p_{2t}e_{12t}r_{1t}r_{2t}]$, the variables are described in section 4.3.1.2 and the statistics are calculated for the period 1973Q2–1987Q3. When the first reduced rank regression (6.11) is undertaken to calculate $\alpha\beta'$, the intercept is unrestricted and a trend is introduced into the model. At the second stage the trend is restricted to exclude quadratic trends. The problem is addressed firstly using the approach adopted by Paruolo (1996) and this is then compared with that described in Johansen (1995).

Paruolo (1996) derives critical values for the test of the joint hypothesis:[2]

$$H_{r,s} : rank(\Gamma) + rank(\Pi) = s + r.$$

The test statistic ($_1Q_{r,s}$) is compared with associated points on the null distribution, the comparison is made either with [*p.value*] calculated by PCGIVE 10.1 (Doornik and Hendry 2001) or 5% critical values ($c_{r,n-r-s}$ (5%)) taken from Paruolo (1996). It is suggested in Doornik and Hendry (2001) that testing is applied from the top left of the table, while Paruolo (1996) suggests progressing from the top to the bottom of each column to a point at which the null can no longer be rejected. Paruolo (1996) advises that tests are applied to the specific case, moving to the general or from the most restricted to less

Table 6.1 I(2) Cointegration tests

r	$1Q_{r,s}(Q^{??}_{r,s})$ [5% c.v.] [p.value]						Q^{*}_{r} c_{n-r}	$c_{r,n-r-s}$
n-r-s	6	5	4	3	2	1		
	314.01	254.23	199.22	163.69	141.7	126.62		
0	[194.32]	[134.54]	[79.53]	[44.0	[22.01]	[6.93]		
	240.35	203.12	174.83	148.54	126.69	109.21	119.69	93.92
	[0.0000]	[0.0000]	[0.0031]	[0.0105]	[0.0073]	[0.0028]		
		203.82	148.4	114.58	90.026	74.347		
1		[134.96]	[79.539]	[45.719]	[21.165]	[5.486]		
		171.89	142.57	117.63	97.97	81.93	68.861	68.68
		[0.0009]	[0.0429]	[0.1335]	[0.2082]	[0.1840]		
			124.56	88.233	65.029	49.417		
2			[80.184]	[43.857]	[20.653]	[5.041]		
			116.31	91.41	72.99	57.95	44.376	47.21
			[0.0226]	[0.1234]	[0.2247]	[0.2537]		
3				83.798	56.535	35.023		
				[59.868]	[32.605	[11.093]		
				70.87	51.35	38.82	23.938	29.38
				[0.0039]	[0.0176]	[0.1215]		
					48.922	27.513		
4					[35.512]	[14.103]		
					36.12	22.6	13.413	15.34
					[0.0016]	[0.0084]		
5						13.576		
						[8.392]	5.184	3.84
						12.93		
						[0.0601]		
c^{*}_{n-r-s}		75.33	53.35	35.07	20.17	9.09		

restricted cases. Following this approach, the first diagonal element implies $r = 0$, $n - r - s = 6$ and the test statistic for the case with unrestricted constant ($\mu_1 \neq 0$) is ${}_1Q_{0,0} = 314.01 > c_{0,6}$ (5%) = 240.35. Based on the calculated statistic the null hypothesis ($rank(\Gamma) = s = rank(\Pi) = r = 0$) cannot be accepted. Progressing to the next column, where $r = 0$ and $n - r - s = 5$, ${}_1Q_{0,1} = 254.23 > c_{0,5}(5\%) = 203.12$, the null is rejected, that $rank(\Gamma) = s = 1$ and $rank(\Pi) = r = 0$. At this point using Paruolo's (1996) suggestion to move down the column, $r = 1$, $n - r - s = 4$, $s = 1$, the joint test statistic ${}_1Q_{0,1} = 203.82 > c_{0,5}(5\%) = 177.89$ and the [p-value]=.0009 confirms that the null hypothesis cannot be accepted at either the 5% or the 1% level. Now the next column is considered, $r = 0$, $n - r - s = 4$, $s = 2$ and the [p-value]=0.0031 implies the null ($rank(\Gamma) = s = 2$, $rank(\Pi) = r = 0$) cannot be accepted.

Following this approach, testing stops and the correct decomposition of the long-run is detected once a null in the above table is accepted. Looking at the

[*p.values*] in the column headed $n - r - s = 4$, there is no case where the null hypothesis can be accepted. The final rejection of the null implies that there are at least $r = 2$ cointegrating vectors and $6 - r - s \leq 3$, I(1) trends. Now progression is from the top of the next column ($n - r - s = 3$) and again to a point at which the null cannot be rejected. From the size of the [*p.value*] = 0.1335, this occurs when $r = 1$, $n - r - s = 3$ and $s = 2$. The Paruolo approach implies that there are $r = 1$ stationary linear combinations (cointegrating vectors), $n - r - s = 6 - 1 - s = 3$, I(1) trends and $s = 2$, I(2) trends. Were one to follow the direction in Doornik and Hendry (2001), to progress down and to the right, then this suggests shifting to the next column at the point at which $r = 2$ and then progressing down that column.[3] The direction of Doornik and Hendry is consistent with the proposition that the first step of the Johansen I(2) estimator correctly determines the number, but not necessarily the exact nature of the cointegrating vectors.

In comparison, Johansen (1995a) suggests that the cointegrating rank calculated from the first step estimation is still reliable, which suggests testing the hypothesis associated with I(2) trends conditional on selecting a particular value for *r*. The null hypothesis that Johansen (1995a) tests is:

$$H_{r,s} \big| H(r) : rank(\xi\gamma') = s.$$

Based on the first rank test it is suggested that $r = 2$ is selected and then s is determined by moving along that row to the point at which the null cannot be rejected. The Johansen test along each row considers the specific case and moves towards the more general, but this now occurs for different values of $n - r - s$, which for fixed r imply different values of s. Given $r = 2$, the test statistic $Q_{2,s}$ is considered for $s = 0, 1, 2, 3$. Starting from the left $n - r - s = 6 - 2 - 0 = 4$, the Johansen tests statistic is $Q_{2,0} = 80.184$, which exceeds the 5% critical value ($c^{*}_{6-2-0} = 53.35$) taken from Johansen (1995a), implying that the null ($r = 2$, $s = 0$) cannot be accepted. Continuing along the row where $r = 2$, the null eventually cannot be rejected when $n - r - s = 6 - 2 - 2$ and $s = 2$ ($Q_{2,2} = 20.653 < c^{*}_{6-2-2} = 20.17$). In line with Doornik and Hendry, the Johansen testing procedure implies that there are $r = 2$ stationary linear combinations (cointegrating vectors), $n - r - s = 6 - 2 - s = 2$, I(1) trends and $s = 2$, I(2) trends.

The two test procedures advanced by Johansen (1995a) and Paruolo (1996) imply that $s = 2$, but they disagree about the number of cointegrating vectors and I(1) trends. Johansen (1995a) shows that by progressing from $s = 0, 1, 2, 3$, the $Q_{2,2}$ test has the same optimal properties in the limit as the Johansen test statistic for cointegration. Furthermore, looking at the Johansen I(2) tests presented in the table above ($Q_{r,s}$), when $r = 0, 1, 2$ the tests are not materially different whatever value $n - r - s$ is selected. Partial confirmation of the optimality of the test may be observed by comparing values of $Q_{r,s}$. For the

column headed $n - r - s = 3$, $Q_{0,3} = 44 \approx Q_{1,2} = 45.719 \approx Q_{2,1} = 43.857$ and all these values exceed the critical value ($c^*_{6-2-2} = 35.07$) at the 5% level.

Inspection of the roots of the companion matrix of the VAR is often viewed as a useful tool in determining the number of unit roots and as a result some idea of the likely number of non-stationary processes driving x_t (Johansen 1995a). The VAR(2) written as a first order model in state space from is:

$$x_t^* = \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix} = A_c x_{t-1}^* + \varepsilon_t^* = \begin{bmatrix} A_1 & A_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix} = \begin{bmatrix} A_1 x_{t-1} + A_2 x_{t-2} + \varepsilon_t \\ x_{t-1} \end{bmatrix}$$

or

$$\begin{bmatrix} A(L)x_t \\ x_{t-1} - x_{t-1} \end{bmatrix} = \begin{bmatrix} x_t - A_1 x_{t-1} - A_2 x_{t-2} \\ x_{t-1} - x_{t-1} \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}.$$

Dhrymes (1984) shows that the characteristic roots of the dynamic process described by the polynomial $A(L)$ can be calculated from the eigenroots of the companion matrix A_c. The eigenvalues (roots) for the VAR(2) model estimated above and for comparison a similar VAR(1) are given in Table 6.2.

The Australian exchange rate example in Johansen (1991a), summarized in Johansen (1995a), yields the clear-cut conclusion that there are three unit roots when $n - r = 5 - 2 = 3$. By contrast, the VAR(2) case considered here appears to reveal three roots close to the unit circle, a real root (.9719) and a complex conjugate pair of roots with modulus (.9001), but, according to the I(2) test produced by Johansen, $n - r = 4$. This suggests that detecting the

Table 6.2 Eigenvalues of companion matrix

VAR(2)			*VAR(1)*		
real	*imag*	*modulus*	*real*	*imag*	*modulus*
–0.01897	0.3874	0.3879			
–0.01897	–0.3874	0.3879			
0.1327	0.0000	0.1327			
0.4550	0.3193	0.5559			
0.4550	–0.3193	0.5559			
0.9719	0.0000	0.9719	0.9574	0.0000	0.9574
0.8877	0.1486	0.9001	0.9222	0.1115	0.9289
0.8877	–0.1486	0.9001	0.9222	–0.1115	0.9289
0.6553	0.2302	0.6946	0.6587	0.2145	0.6927
0.6553	–0.2302	0.6946	0.6587	–0.2145	0.6927
0.4910	0.0000	0.4910	0.9252	0.0000	0.9252
0.7729	0.0000	0.7729			

number of unit roots from the companion matrix is not always straightforward. Firstly, a VAR(2) system can be decomposed into two stationary processes ($r = 2$), two non-stationary processes (either $n - 2 - s = 2$ or $s = 2$) and a pair of common I(2) or I(1) trends driven by a single unit root. Secondly, should the roots of the VAR(1) be considered for comparison, then the estimates are quite consistent with the proposition that there are $n - r = 4$ unit roots. Analysis associated with both sets of eigenvalues for the two companion matrices does not appear to support the approach due to Paruolo (1996), which suggests $r = 1$ and $n - r = 4$.

Having found that some of the series are I(2), the usual cointegrating vectors may not be valid as the stationary linear combinations may require combinations of I(2) processes that are I(1) to make them stationary or polynomial cointegration. Consider these following suggestions for the long-run relationships associated with the VAR(2) system developed above. Based on the findings in Hunter (1992a) and Johansen and Juselius (1992), there are two cointegrating vectors that accept PPP and UIRP restrictions. The conclusion of the I(2) analysis for PPP is that the series may only be rendered stationary when the cointegrating vector is augmented by differences in I(2) variables. For example, relative movements in the cross-country inflation rates may be what is required. With $s = 2$ common I(2) trends driving the price series ($p_0 p_1 p_2$) then the cointegrating vectors could take the following form:

$$\beta' x_{t-1} - \psi' \Delta x_{t-1} = \left(\begin{bmatrix} 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ e_{12} \\ r_1 \\ r_2 \end{bmatrix}_{t-1} - \begin{bmatrix} 0 & 0 & \psi_{31} & -\psi_{31} & 0 & 0 \\ \psi_{12} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Delta \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ e_{12} \\ r_1 \\ r_2 \end{bmatrix}_{t-1} \right).$$

A similar type of long run occurs with polynomial cointegration (Engle and Yoo 1991; Gregoir and Laroque 1993):

$$\beta' x_{t-1} - \psi' \Delta x_{t-1} = \begin{bmatrix} 0 & 1 & \psi_{31} - 1 - \psi_{31} L & -1 - \psi_{31} + \psi_{31} L & \beta_{51} & \beta_{61} \\ (\beta_{12} - \psi_{12} + \psi_{12} L) & 0 & 0 & 0 & 1 & -1 \end{bmatrix} x_{t-1}$$

where $x'_t = [p_{0t}\, p_{1t}\, p_{2t}\, e_{12t}\, r_{1t}\, r_{2t}]$. The two forms of I(2) cointegration are equivalent when $\beta_{51} = 0$, $\beta_{61} = 0$ and $\beta_{12} = 0$. Unfortunately, prior to any evaluation

of the long run, the system needs to be identified, but identification of the type discussed in chapter 4 is considerably more complicated in the I(2) case as three sets of matrices lack identification:[121]

$$\begin{aligned}\alpha\rho'\tau' &= \alpha\varsigma\varsigma^{-1}\rho'\theta'\theta'^{-1}\tau' = \alpha^{*}\rho^{+'}\tau^{*'}\\ \alpha\psi' &= \alpha\varsigma\varsigma^{-1}\psi' = \alpha^{*}\psi^{*'}\\ \alpha_{\perp}(\alpha_{\perp}'\Sigma\alpha_{\perp})^{-1}\kappa'\tau' &= \alpha_{\perp}\upsilon^{-1}(\alpha_{\perp}'\Sigma\alpha_{\perp})^{-1}\upsilon\kappa'\theta'\theta'^{-1}\tau'\\ &= \alpha_{\perp}^{*}(\alpha_{\perp}'\Sigma\alpha_{\perp})^{-1}\kappa^{+'}\tau^{*'}\end{aligned}$$

Hence, the same likelihood can be defined for (6.6) using parameters [α, ρ', τ', κ', ψ', $\alpha_{\perp}'$] and [α^{*}, $\rho^{+'}$, $\tau^{*'}$, $\kappa^{+'}$, $\psi^{*'}$, $\alpha_{\perp}^{*'}$]. The two sets of parameterizations are observationally equivalent and observational equivalence leads to a fundamental loss of identification.

Although inflation seemed to be I(1) in the late 1980s and early 1990s the argument appears less compelling in a world where inflation is predominantly under control, which suggests that economic and financial time series might be better described as long-memory.

6.2.2 Fractional cointegration

The notion of fractional differenced series was introduced in chapter 2. When such processes are considered then the possibility of fractional cointegration ought to be entertained. Robinson and Yajima (2002) explain that this notion of fractional cointegration is quite consistent with the original definition of cointegration due to Engle and Granger (1987). Consider a pair of series x_{1t} and x_{2t} that require fractional differencing for them to be rendered stationary, then:

$$\Delta^{d}x_{it} = (1-L)^{d}x_{it} \sim I(0) \text{ for } i = 1,\ 2$$

where $(1-L)^{d} = \sum_{j=0}^{\infty}\frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)}$. For $a > 0$, $\Gamma(a) = \int_{0}^{\infty} z^{l-1}e^{-z}dz$ and $a = -l$, $l = 0$, 1, ..., $\Gamma(a)$ has simple poles with residues $\frac{-1^{l}}{l}$, otherwise $\Gamma(a) = \Gamma(a+1)/a$. It follows that x_t is cointegrated when:

$$\beta'x_t = J(L)\varepsilon_t \sim C(i,d)$$

$$\text{when } x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}.$$

Proofs exist for the analysis of stationary fractional series with $-.5 < d < .5$ (Robinson and Yajima 2002). The conventional question arises over the rank of the matrix of cointegrating vectors, $rank(\beta) = r$. Do there exist r linear combinations of variables x_t that require the fractional difference operator $(1-L)^{d}$ to be applied for the series to be I(0). Robinson (1994) explains how to use non-parametric estimates of the dynamic process to calculate the cointegrating relationships when series have the same order of integration. Robinson

and Marinucci (1998) apply this approach to stationary fractionally integrated series to estimate the long-run parameters from the equation:

$$\left[1-\beta_1\right]\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = J(L)\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}.$$

The estimator is similar to that used by Phillips and Hansen (1990) to estimate long-run parameters when the series are I(1). The unknown moving average parameters in $J(L)$ are captured by a frequency domain estimator, which also appears to compare well with Phillips and Hansen (1990) when the series are I(1) (Marinucci and Robinson 2001). Although there is evidence that this type of approach is able to estimate long-run parameters when r is known or not large, the method, though efficient in calculating well-known long-run relationships, does not provide a formal test of the proposition that either fractional or integer integrated series are cointegrated. The method can determine the extent to which the variables in the regression are related by determining whether β_1 is significant or not. Clearly, any such conclusion is conditional on the appropriateness of this normalization.

Robinson and Yajima have attempted to determine the order of integration and cointegration by two different methods. They consider three different crude oil prices (WTI, Dubai and Brent). Based on an Augmented Dickey–Fuller test with an intercept, the three series are found to be stationary at the 5% level of significance. But when the order of difference is assumed to be fractional, the estimates of d for the three series are [.5336, .4367, .4538].[5]

Robinson and Yajima (2002) suggest two approaches to the problem of selecting the cointegrating rank, but they use one of them in their example. Consider the Vector Auto-Regressive Fractionally Integrated Moving Average (VARFIMA) model:

$$E(L)x_t = C(L)\varepsilon_t$$

where $E(L) = diag[(1-L)^{d_1}, (1-L)^{d_2} \ldots (1-L)^{d_n}]$.[6] The series are ordered on the basis of the prior estimate of the difference order. The test is based as is usually the case on the rank of the matrix $C(1)$, which, under conventional cointegration, has rank $n - r$ associated with the extent to which there is any over-differencing. The test, as is the case with integer cointegration, progresses from the most restricted model, where $C(1)$ has full rank, $n - r = n$ and $r = 0$, there is no cointegration to the cointegration cases, $r = 1, 2, 3$. The test for fractional cointegration is:

$$H_i : rank(G) = rank(C(1)) = n - r$$
$$where\ G = \frac{1}{2\pi}C(1)C(1)'.$$

To make the test operational, Robinson and Yajima use the following non-parametric estimator of G:

$$\hat{G} = \frac{1}{m_1}\sum_{j=1}^{m_1} \mathrm{Rc}\left\{\hat{\Lambda}(\lambda_j)^{-1} I_j \hat{\Lambda}(\lambda_j)^{-1}\right\}.$$

Where $I_j = \omega(\lambda_j)\omega(\lambda_j)'$, $\omega(\lambda_j) = (\omega_1(\lambda_j)\omega_2(\lambda_j) \ldots \omega_n(\lambda_j))'$, $Re\{\cdot\}$ is the real component, $\hat{\Lambda}(\lambda_j) = diag(e^{\frac{i\pi d_\cdot}{2}}\lambda_j^{-\bar{d}_\cdot}, \ldots e^{\frac{i\pi d_\cdot}{2}}\lambda_j^{-\bar{d}_\cdot}), \lambda_j = \frac{2\pi j}{T}$, and $m < \frac{T}{2}$. It has been assumed that d_a is replaced by a pooled estimate $\bar{d}_\cdot = (\hat{d}_1 + \hat{d}_2 + \hat{d}_3)/3$ and $\omega_a(\lambda_j) = \frac{1}{\sqrt{2\pi T}}\Sigma_{t=1}^{T} x_{at}e^{it\lambda_j}$ is the discrete Fourier transform of the original data. The effective bandwidth m_1 is set to increase at a faster rate than m to counteract the effect of using an estimate of d_a Robinson and Yajima (2002) provide estimates of G evaluated with $m = 13$ and $m_1 = 15$:

$$\hat{G} = \begin{bmatrix} .00493 & .00542 & .00575 \\ .00542 & .00625 & .00653 \\ .00575 & .00653 & .0073 \end{bmatrix},$$

where $\hat{G}$ has the following eigenvalues [.01807, .000275, .000124]. The most important eigenvector is associated with the largest root, which given that the other two roots are small suggests that $n - r = 1$ or with $n = 3$ variables then there are $r = 2$ cointegrating relationships. Robinson and Yajima (2002) proceed to analyze the case where the three series have two distinct orders of differencing. This suggests that the WTI oil price series is handled differently than that for Brent and Dubai. Once Brent and Dubai crude prices are considered together with two types of difference, the reduced rank calculation is applied to a 2×2 sub-matrix, which from the obvious rank deficiency in $\hat{G}$ above implies $r = 1$.

6.3 Forecasting in cointegrated systems

6.3.1 VMA analysis

Cointegration describes how, in the long run, the levels of a set of variables should move together. A similar property should therefore be expected of forecasts from such a system. That is, the forecasts of a set of variables from a cointegrated system should be related to one another such that, although individually subject to the implications of non-stationarity, there remain linear combinations of the forecasts that are zero, or constant (depending on the deterministic terms in the model). If valid long-run relationships are imposed on an empirical model of the data, this ought to improve the quality of long-run forecasts, as additional information is being exploited. But is the value of the long-run restrictions, in terms of forecast improvement, greater than for other types of restriction, or restrictions on stationary systems? Engle and Yoo (1987) provide an analysis of this problem in the CI(1, 1) case.

Consider the usual VMA representation of an $n \times 1$, CI(1, 1) system considered in section 4.2:

$$\Delta x_t = C(L)\epsilon_t, \tag{6.18}$$

where $C(L) = \sum_{i=0}^{\infty} C_i L^i$, *rank* $(C(1)) = n - r$, and $C_0 = I_n$. In order to obtain an expression for x_t, which is to be the object of the forecast, sum both sides of (6.18) from $i = 1, \ldots, t$ to give

$$x_t - x_0 = \sum_{i=1}^{t} C(L)\epsilon_i.$$

In addition, assume initial values x_0 and ϵ_q, $q = 0$ are zero. Then,

$$C(L)\epsilon_i = \sum_{j=0}^{\infty} C_j \epsilon_{i-j} = \sum_{j=0}^{i-1} C_j \epsilon_{i-j}$$

and so

$$x_t = \sum_{i=1}^{t} \sum_{j=0}^{i-1} C_j \epsilon_{i-j}. \tag{6.19}$$

Equation (6.19) can be rewritten in terms of ϵ_s, $S = 1, \ldots, t$

$$x_t = \sum_{s=1}^{t} \sum_{r=0}^{t-s} C_r \epsilon_s.$$

Moving forward another h periods,

$$x_{t+h} = \sum_{s=1}^{t+h} \sum_{r=0}^{t+h-s} C_r \epsilon_s = \sum_{s=1}^{t} \sum_{r=0}^{t+h-s} C_r \epsilon_s + \sum_{s=t+1}^{t+h} \sum_{r=0}^{t+h-s} C_r \epsilon_s$$

and redefining the index on the last summation to emphasize that it contains terms in the disturbances beyond t only, gives

$$x_{t+h} = \sum_{s=1}^{t} \sum_{r=0}^{t+h-s} C_r \epsilon_s + \sum_{q=1}^{h} \sum_{r=0}^{h-q} C_r \epsilon_{t+q}. \tag{6.20}$$

Equation (6.20) expresses x_{t+h} as the sum of two terms that partition the disturbances between those occurring up to and including time t, and later values.

The forecast of x_{t+h} based on information available at time t is the expected value of x_{t+h} given the information, and is denoted $x_{t+h|t}$. In this context, h is known as the forecast horizon and t is called the forecast origin. Using the fact that the conditional expectation of a future disturbance term is zero, and the conditional expectation of any current or past value is the expectation of a realized value, from (6.20),

$$x_{t+h|t} = \sum_{s=1}^{t} \sum_{r=0}^{t+h-s} C_r \epsilon_s. \tag{6.21}$$

This does not yet establish that the forecasts are linearly related. The requirement for this is for there to exist a linear combination of the forecasts that is zero (in the absence of deterministic terms). That is, there must exist an $n \times 1$ vector γ such that $\gamma' x_{t+h|t} = 0$. From (6.21), a sufficient condition for this is that

$$\gamma' \sum_{r=0}^{t+h-s} C_r = 0.$$

But this does not follow from the properties of the VMA, as it requires each of $\sum_{r=0}^{t+h-s} C_r$, $s = 1, \ldots, t$ to be of reduced rank and to have the same null space. However, cointegration is a long-run property and its implications can only be expected to follow in the long run. In a forecasting context, this means that any special properties of the forecast arising from cointegration can only be expected to become apparent as the forecast horizon, h, becomes large. So consider the limit of $\sum_{r=0}^{t+h-s} C_r$, as $h \to \infty$:

$$Lim_{h\to\infty} \sum_{r=0}^{t+h-s} C_r = \sum_{r=0}^{\infty} C_r = C(1), \tag{6.22}$$

and define what can be called the long-run forecast, $x_{\infty|t}$, as:

$$x_{\infty|t} = Lim_{h\to\infty}\left[x_{t+h|t}\right]. \tag{6.23}$$

Then, from (6.21) and (6.22), $x_{\infty|t}$, is given by

$$\begin{aligned} x_{\infty|t} = Lim_{h\to\infty}\left[x_{t+h|t}\right] &= Lim_{h\to\infty}\left[\sum_{s=1}^{t}\sum_{r=0}^{t+h-s} C_r \epsilon_s\right] \\ &= \sum_{s=1}^{t}\left[Lim_{h\to\infty}\sum_{r=0}^{t+h-s} C_r\right]\epsilon_s = C(1)\sum_{s=1}^{t}\epsilon_s. \end{aligned}$$

The long-run forecast therefore follows a linear combination of the realized value of a vector stochastic trend. But *rank* $(C(1)) = n - r$, and so there exist r linearly independent vectors, that is the cointegrating vectors, β, such that $\beta' C(1) = 0$. Therefore:

$$\beta' x_{\infty|t} = \beta' C(1)\sum_{s=1}^{t}\epsilon_s = 0. \tag{6.24}$$

The extent to which fixed horizon forecasts approximate to (6.24) depends how quickly the matrix coefficients $C_i, i = 0, 1 \ldots$, decay. From (6.21)

$$x_{t+h|t} = \sum_{s=1}^{t}\left[C(1) - \sum_{r=t+h-s+1}^{\infty} C_r\right]\epsilon_s = x_{\infty|t} - \sum_{s=1}^{t}\sum_{r=t+h-s+1}^{\infty} C_r \epsilon_s$$

and so

$$\beta' x_{t+h|t} = -\beta' \sum_{s=1}^{t} \sum_{r=t+h-s+1}^{\infty} C_r \epsilon_s.$$

Thus the smallest index on the C_r is $r = h + 1$, indicating that, assuming the C_r do decay with r, the greater is the forecast horizon, the smaller will be the deviation of the forecasts from their long-run relationship. Thus, empirically, the evidence for cointegration restrictions improving forecasts should be weaker for short horizons, than longer ones. The more rapidly the coefficients decay, the fewer steps ahead the forecasts need to be before they display a functional relationship similar to the cointegrating relations.

Turning to the h-step ahead forecast error, denoted $e_{t+h|t}$, and its variance, from (6.20) and (6.21), this error is

$$e_{t+h|t} = \sum_{s=1}^{t} \sum_{r=0}^{t+h-s} C_r \epsilon_s + \sum_{q=1}^{h} \sum_{r=0}^{h-q} C_r \epsilon_{t+q} - \sum_{s=1}^{t} \sum_{r=0}^{t+h-s} C_r \epsilon_s = \sum_{q=1}^{h} \sum_{r=0}^{h-q} C_r \epsilon_{t+q} \tag{6.25}$$

and, since the disturbances are not autocorrelated

$$\text{var } (e_{t+h|t}) = \sum_{q=1}^{h} \left[\left(\sum_{r=0}^{h-q} C_r \right) \Omega \left(\sum_{r=0}^{h-q} C_r' \right) \right],$$

where $\Omega = E(\epsilon_t \epsilon_t')$, for all t. That is, the forecast error variance grows with h. Interestingly, it is also the case that the forecast errors are cointegrated, with precisely the same time series structure as the original process, x_t, under the condition that all forecasts are made using the same information, that available at time t. To see this use (6.25) to construct the forecast error difference process

$$\begin{aligned}
\Delta e_{t+h|t} &= e_{t+h|t} - e_{t+h-1|t} \\
&= \sum_{q=1}^{h} \sum_{r=0}^{h-q} C_r \epsilon_{t+q} - \sum_{q=1}^{h-1} \sum_{r=0}^{h-1-q} C_r \epsilon_{t+q} \\
&= C_0 \epsilon_{t+h} + \sum_{q=1}^{h-1} \left(\sum_{r=0}^{h-q} C_r - \sum_{r=0}^{h-q-1} C_r \right) \epsilon_{t+q} \\
&= C_0 \epsilon_{t+h} + \sum_{q=1}^{h-1} C_{h-q} \epsilon_{t+q} = \sum_{q=1}^{h} C_{h-q} \epsilon_{t+q} = \sum_{q=1}^{h} C_{h-q} \epsilon_{t+h-(h-q)} \\
&= \sum_{k=0}^{h-1} C_k \epsilon_{t+h-k} = C(L) \epsilon_{t+h}, \quad \epsilon_q = 0, \ q \le t,
\end{aligned}$$

where the initial values are now relative to the forecast origin, and consistent with the original VMA, have been set to zero. Thus

$$\Delta e_{t+h|t} = C(L) \epsilon_{t+h}$$

and hence, from the original VMA, all h-step ahead forecast errors are cointegrated of order (1,1). That is, the difference between the h-step ahead and the $h-1$-step ahead forecast errors, both made conditional on information available at time t, is stationary, but the sequence of h-step ahead forecast errors, for $h = 1,2, \ldots$, is I(1).

An intuition for the non-stationarity of the forecast error can be provided by expressing a future value of the process as a sum of the forecast and the forecast error,

$$x_{t+h} = x_{t+h|t} + e_{t+h|t}. \tag{6.26}$$

Since, $x_{t+h|t}$ depends only on realized values (the disturbance values at time t and before), it is non-stochastic. Thus the stochastic non-stationarity properties of x_{t+h} and $e_{t+h|t}$ must be the same, so they must both be integrated of order 1. Applying the initial value condition $\epsilon_q = 0$, $q \leq t$, equation (6.26) gives $x_{t+h|t} = x_{\infty|t}$ and hence:

$$x_{t+h} = x_{\infty|t} + e_{t+h|t},$$

from which, pre-multiplication by the cointegrating vector gives

$$\beta' x_{t+h} = \beta' x_{\infty|t} + \beta' e_{t+h|t} = \beta' e_{t+h|t}. \tag{6.27}$$

The left-hand side of (6.27) is I(0) from the VMA, and therefore so is $\beta' e_{t+h|t}$, hence $e_{t+h|t}$ is CI(1,1).

6.3.2 Forecasting from the VAR

The property that the long-run forecasts should be linearly constrained can also be obtained from a VAR. Again, let x_t be an $n \times 1$ CI(1,1) vector, this time having the VAR(p) structure

$$x_t = \sum_{i=1}^{p} A_i x_{t-i} + \epsilon_t. \tag{6.28}$$

Reparameterize this in the usual way as the VECM

$$\Delta x_t = \Pi x_{t-1} - \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + \epsilon_t \tag{6.29}$$

where, again $\Pi = \alpha\beta'$ with α and β dimensioned $n \times r$. Following Lin and Tsay (1996), in order to understand how the forecasts from (6.28) have the same long-run properties as the series themselves, note that Δx_t is I(0), and that forecasts of a stationary series converge to the expected value of the process as the forecast horizon tends to infinity. That is

$$Lim_{h\to\infty} \Delta x_{t+h|t} = \mu_{\Delta x} \tag{6.30}$$

where $\mu_{\Delta x} = E(\Delta x_t)$. The properties of the forecasts of the difference process are used to obtain those of the levels via the VECM. Using (6.29), the h-step ahead forecast equation for the difference process is

$$\Delta x_{t+h|t} = \Pi x_{t+h-1|t} - \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t+h-i|t}. \tag{6.31}$$

In order to derive the properties of the long-run forecasts, take the limit of (6.31) as $h \to \infty$, and substitute from (6.30) to give

$$\mu_{\Delta x} = \Pi\left[Lim_{h\to\infty} x_{t+h-1|t}\right] - \sum_{i=1}^{p-1} \Gamma_i \mu_{\Delta x}.$$

Rearranging, and using the notation of (6.23) for the long-run forecast of the level,

$$\Pi x_{\infty|t} = \left[I_n + \sum_{i=1}^{p-1} \Gamma_i\right] \mu_{\Delta x}. \tag{6.32}$$

The right-hand side of (6.32) is a constant matrix, and so shows that the long-run forecasts, $x_{\infty|t}$, are tied together. The analysis can be taken further to complete the analogy with equation (6.24) for the VMA case. Pre-multiplying (6.32) by α' and replacing Π by $\alpha\beta'$ gives

$$(\alpha'\alpha)\beta' x_{\infty|t} = \alpha'\left[I_n + \sum_{i=1}^{p-1} \Gamma_i\right] \mu_{\Delta x}$$

where $(\alpha'\alpha)$ is non-singular, so that

$$\beta' x_{\infty|t} = (\alpha'\alpha)^{-1}\left[I_n + \sum_{i=1}^{p-1} \Gamma_i\right] \mu_{\Delta x}.$$

This is directly comparable with (6.24) (except that in 6.24 initial values have been set to zero), and shows that each cointegrating vector constitutes a constraint on the long run forecasts.

6.3.3 The mechanics of forecasting from a VECM

In order to benefit from any perceived advantages to forecasting from cointegrated models, it is necessary to impose the cointegrating relationships. In the VAR setting, this may be undertaken as follows.

For given β and by implication, known cointegrating rank, r, construct cointegrating combinations $\eta_t = \beta' x_t$, and estimate the VECM, conditional on r, as

$$\Delta x_t = \alpha \eta_{t-1} - \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + \epsilon_t.$$

Estimation may be performed by OLS, to give

$$\Delta x_t = \hat{\alpha}\eta_{t-1} + \sum_{i=1}^{p-1} \hat{\Gamma}_i \Delta x_{t-i} + e_t = \hat{\Pi} x_{t-1} + \sum_{i=1}^{p-1} \hat{\Gamma}_i \Delta x_{t-i} + e_t,$$

where $\hat{\Pi} = \hat{\alpha}\beta'$. Now, rearrange the VECM as the VAR

$$x_t = \sum_{i=1}^{p} \hat{A}_i x_{t-i} + e_t,$$

$$\hat{A}_1 = I_n + \hat{\Pi} - \hat{\Gamma}_1, \quad \hat{A}_i = \hat{\Gamma}_i - \hat{\Gamma}_{i-1} \text{ and } \hat{A}_p = \hat{\Gamma}_{p-1}.$$

The *h*-step ahead forecasts can then be produced recursively using

$$x_{t+h|t} = \sum_{i=1}^{p} \hat{A}_i \Delta x_{t+h-i|h}, \tag{6.33}$$

where $x_{t+h-i|t} = x_{t+h-i}$ for $h \leq i$. If r and β are unknown, they may be replaced by values $\hat{r}$ and $\hat{\beta}$ estimated using the Johansen procedure. This is the approach used by Lin and Tsay (1996).

The order of the forecasting VAR in (6.33), and that used for the Johansen pre-whitening, should be the same, determined, for example, using an information criterion, such as the Schwarz (SIC) (see Reimers 1992; Lütkepohl 1991). Otherwise, as was explained in section 4.3.3, programs such as PCGIVE provide systems and single equation diagnostic test for each equation in the VAR (Doornik and Hendry 2001).

The details of information criteria vary according to the weight put on additional parameters, but they are generally of the form

$$IC = \ln\left|\frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t' \hat{\epsilon}_t\right| + mf(T), \tag{6.34}$$

where $f(T)$ is an increasing function of T, $m = pn^2$, the number of estimated coefficients in an unrestricted VAR, and $\hat{\epsilon}_t$ the vector of VAR residuals. A criterion which often preferred is the SIC, for which $f(T) = \frac{\ln(T)}{T}$. Amongst the criteria most commonly used, this penalizes additional parameters (increasing VAR order) the most heavily, leading to relatively parsimonious models. The favoured model is that for which the information criterion value is minimized. When used in this way, the SIC provides consistent model selection in the sense that, as the sample size tends to infinity, it will select the correct model order with probability tending to one.

6.3.4 Forecast performance

The imposition of cointegrating restrictions on a model of I(1) series should lead to forecast improvements for two reasons. Firstly, valid long-run relationships should improve the accuracy of long-run forecasts by exploiting information about the interrelatedness of the series. Secondly, fewer parameters are estimated. In the unrestricted VECM, Π has n^2 elements, whereas when restricted, it has $2nr$. However, a number of practical issues arise:

(i) How useful is the long-run information in providing long but finite time horizon forecasts?

(ii) How are short-run forecasts affected?
(iii) What are the costs of mistakenly identifying series as I(1) when they are really I(0)?
(iv) What is the cost of incorrectly estimating r?
(v) What is the cost of imposing invalid long-run restrictions (getting the cointegrating vectors wrong)?

These issues are discussed by Clements and Hendry (1995, 1998), Lin and Tsay (1996) and Engle and Yoo (1987), among others. The three studies report Monte Carlo results; their findings are summarized below.

6.3.4.1 Engle and Yoo

These authors consider a bivariate model (representable as a first-order VAR) and discuss two types of forecast that can be made from it, one ignoring any long-run restrictions, and one imposing them. These forecasts are based on an unrestricted VAR (UVAR) and the Engle and Granger two-step methodology (EG) respectively. In the latter case, at each replication, a preliminary static regression is used to estimate the cointegrating relations and the lagged residuals from this model being included as the lagged levels term in a dynamic ECM.[7] The putative long relations are not subject to prior testing for cointegration.

The sample size is 100 and the forecast horizon from 1 to 20, so that in this case, a long-run forecast is being defined as one with a horizon 20 per cent beyond the sample, if not less. The finding is that, in terms of the mean square forecast error as measured by the trace of the sample covariance matrix of the forecast errors (see section 6.3.4.4 for more detail on forecast evaluation), the unrestricted VAR provides superior forecasts up to and including the 5-step ahead forecast (5 per cent of sample size), thereafter, the imposition of estimated long run restrictions improve the forecast monotonically, to an advantage of 40 per cent over the unrestricted forecast at 20 steps ahead. This is, of course, against a background of worsening forecast performance as forecast horizon increases.

6.3.4.2 Clements and Hendry

In their book and earlier paper, Clements and Hendry (1998, 1995) generalize the study of Engle and Yoo. They present the results of a bivariate VAR(1) system estimated on 100 observations, but for a wider range of parameter values and models. In addition to UVAR and EG, they consider the Johansen maximum likelihood estimator (ML) and a misspecified model in differences alone (DV), the lagged levels term being excluded. The DV model can be used to forecast the level of the process by adding successive forecasts of the differences to the known value of the level at the forecast origin. They also

introduce another issue, which is the form of the process used to compare forecasts: the levels, the differences, or the stationary combinations. The last of these representations is obtained by transforming the model to one in terms of the cointegrating combinations and the differenced common trends. Thus, the number of processes is unaltered, and their integration and cointegration properties preserved. Their notation for the I(0) variables is w_t where $w'_t = (x'_t\beta \quad \beta'_\perp \Delta x_t)$. Consider the partition $\alpha' = (\alpha'_a \quad \alpha'_b)$ with α_a dimensioned $r \times r$ and α_b dimensioned $r \times (n - r)$ and defining

$$J' = (0 \quad I_{n-r}) \text{and} Q = (\beta \quad J)'$$

the representation is

$$w_t = Gw_{t-1} + \nu_t \tag{6.35}$$

$$\text{where } G = \begin{pmatrix} (I_r + \beta'\alpha) & 0 \\ \alpha_b & 0 \end{pmatrix} \text{and } \nu_t = \begin{pmatrix} \beta' \\ J' \end{pmatrix} \epsilon_t.$$

Clements and Hendry produce forecasts of x_t and Δx_t using each of the four estimation methods, UVAR, ML, EG, and DV. These primary forecasts are transformed to produce forecasts of each of x_t, Δx_t and w_t. That is, each forecast is one of x_t or Δx_t, initially, but all are transformed (as necessary) into x_t, Δx_t and w_t. The purpose of the exercise is to emphasize that the superiority of one forecast method over another depends not only on what model is used to produce the forecast, but also on what properties of the forecast are being compared.

In particular, in comparing EG and UVAR to forecast x_t, the level of the process, the importance of the imposition of a valid long-run restriction is examined. But the question then arises as to whether it matters that the restriction is specifically a long-run restriction. In other words, are the advantages available from the imposition of correct restrictions markedly different in a non-stationary cointegrated environment compared to a stationary one? The way to get at this issue is to transform the forecasts to stationarity before comparing them, effectively filtering out long-run variation. The appropriate transformation is that of equation (6.35), applied to the forecasts. This procedure is only available in the context of simulations (using parameter values from the DGP), since the UVAR, by its very nature, brings with it no estimation of the cointegrating combinations. It is still the case that the forecasts differ in the method of their production, but are now being compared on a more appropriately matched basis – that is, in stationary terms. If relative forecasting performance is different in stationary space, then it suggests that the long-run nature of the restrictions is relevant in determining forecast behaviour.

If it is the long run nature of the restrictions that improve the long-run forecasts, then direct comparisons of the forecasts of the level of the process

where the restrictions are, and are not imposed, should favour the forecasts made subject to the restrictions. However, if the long-run components are removed prior to comparison, these transformed forecasts should not differ significantly. Equation (6.35) is a very useful device for decomposing the causes of relative forecast behaviour.

In their simplest case (among 13 parameterizations), Clements and Hendry generate data according to a bivariate VECM model with a single lag,

$$\Delta x_t = \Pi x_{t-1} + \epsilon_t.$$

Forecast comparisons are made in a number of ways, the simplest of which is based on the trace of the estimated variance–covariance matrix of the forecast errors (see section 6.3.4.4 for more detail on forecast evaluation). One parameterization is very similar to that used by Engle and Yoo, and therefore comparable with the earlier results. It is shown that, at longer forecast horizons, material improvement in the levels forecast are available by imposing cointegrating restrictions. That is, EG and ML are superior in levels forecasting to UVAR when the forecast horizon is relatively long. In addition, the superiority is more marked with smaller sample sizes due to the enhanced role of the degrees of freedom saved by imposing the restrictions.

When the forecasts are transformed to stationarity (using equation (6.35)) and compared again, UVAR is no longer inferior. This suggests that the gains in forecast performance from the imposition of the restrictions are due to their long-run characteristics, as no further restrictions have been imposed. In contrast to these findings, the misspecified DV model performs only slightly worse than EG and ML (and therefore better than UVAR) in levels forecasts at longer forecast horizons, but notably under-performs the other three when the forecasts are compared in stationary space.

These findings must be interpreted with care because, in practice, VAR order and cointegration rank are decided from the data. In addition, systems will normally consist of more than two variables. Clements and Hendry summarize the results of their more widely parameterized study using response surfaces, presenting their conclusions with a number of warnings about the additional complexities that enter in the practical forecasting setting. The results represent a benchmark case only.

6.3.4.3 Lin and Tsay

Lin and Tsay (1996) generalize the model for forecast performance comparisons to one involving four variables. Their Monte Carlo study is necessarily restricted in terms of the parameter values used, but the DGPs used are chosen to mimic observed data characteristics, so in this sense are calibrated so as to apply to a relevant parameter space. The structures used have the following characteristics.

(i) All systems are second order (VAR(2)).
(ii) Five DGPs are considered in all, being respectively, from model 1 to model 5, strongly stationary, but with two roots close to the unit circle, stationary with two roots very close to the unit circle, non-stationary system with cointegrating rank 2, non-stationary and non-cointegrating.[8] Of these, the stationary and unit root non-cointegrated cases are diagonal.
(iii) The in-sample period consists of 400 observations, with 100 additional out-of-sample data points generated for forecasting comparison. Forecast horizons of 1 to 60 are used. Each replication gives rise to a set of forecasts at each forecast horizon.
(iv) All models are estimated as ECMs with cointegrating rank $r = 0, 1, 2, 3, 4$ using Johansen's (1988, 1991, 1995) approach, and then recast as VECMs for the purpose of forecasting the levels.
(v) The forecasting metric, $E(L)$, where L is the forecast horizon (see equation 6.36), is based on the trace of the estimated variance–covariance matrix of the forecast errors. Each replication gives rise to an estimated variance–covariance matrix of forecast errors, and these are then averaged across replications. The larger is the statistic, the poorer the forecast.

The results of these exercises are presented in Figure 6.1.

Lin and Tsay gather their conclusions on these results into the following principal points:

(i) When the system is stationary the long-run forecasts approach a constant quickly as the forecast horizon increases. (The size of the forecast errors, in terms of their variance is also relatively small.)
(ii) If the system is stationary, then under-specifying the rank of the long-run matrix leads to under-performance. That is, imposing long-run restrictions that do not exist in practice (which are not valid) damages long-run forecast performance. The more of these there are, the worse the performance of the forecasts.
(iii) Unless the system is very close to non-stationarity (the near non-stationary DGP is model 3), correct specification of the cointegrating rank is best.
(iv) Under specification of the cointegrating rank is not serious if the processes concerned are non-stationary. This should be contrasted with the stationary case, where, although cointegration is not defined, the rank of the long-run matrix still is, and where this is under-specified, there is a deterioration in forecast performance.

Clearly, non-stationary and near non-stationary systems are harder to forecast than stationary ones. As a matter of design, it should be noted that while

Figure 6.1a

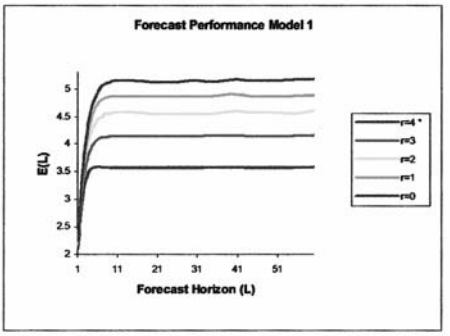

Figure 6.1b

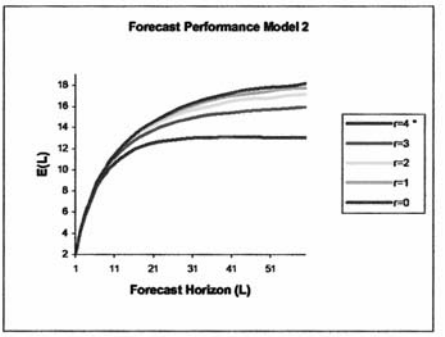

Figure 6.1c

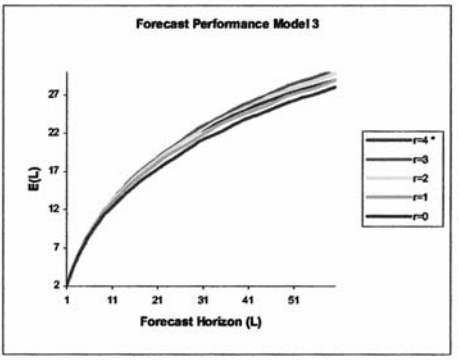

Figure 6.1d

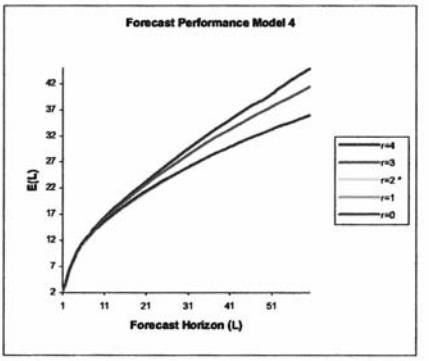

Figure 6.1e

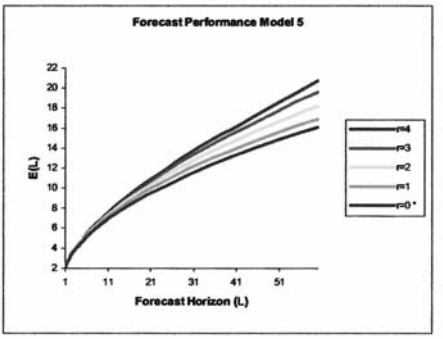

Figure 6.1 Forecasting performance from Lin and Tsay study, by model

Lin and Tsay control carefully for the roots of the processes involved, only their cointegrated structure displays common features, in this case of the unit root. All the other models are diagonal, meaning that, in the case of model 3 for example, although there are roots very close to being unit roots, they do

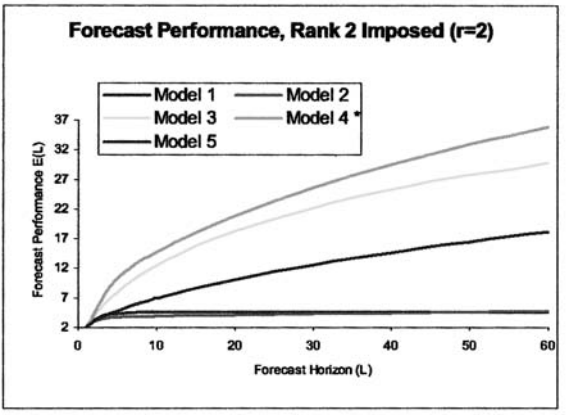

Figure 6.2 Lin and Tsay results, all models, rank 2 system

not constitute a common feature. For this to be so, the determinant of the VAR lag operator evaluated at that root would have to be less than full rank, but not zero. Diagonality results in its being zero (Burke 1996).[9] Model 3 also has the interesting property that the quality of forecasts is least affected by the choice of (cointegrating) rank.

By grouping these results differently, a further conclusion can be made. Instead of looking at the results by model and varying the cointegrating rank imposed, it is possible to fix the imposed cointegrating rank, and see which model is easiest or hardest to forecast for that restriction. Figure 6.2 demonstrates the case for the imposition of rank 2, which is correct for model 4. It is immediately obvious that, using the trace measure (see Forecast Evaluation below), the cointegrated system is the hardest to forecast at medium and long horizons. It is even harder to forecast than the non-stationary non-cointegrated case.[10] In fact, no matter what cointegrating rank is imposed (0 to 4), the cointegrated system is the most difficult to forecast, in the sense that it has the largest trace statistic. However, it remains the case that, if the system is cointegrated, it is best to impose the appropriate cointegrating rank (figure 6.1d).[11]

These forecast comparisons are more limited since they are compared in levels terms only. Clements and Hendry demonstrate that once transformed to stationarity, there is much less difference between forecasts based on different procedures. It is not clear from Lin and Tsay if the same transformation would result in less obvious distinctions between the forecasts based on the imposition of different cointegrating ranks at the estimation stage. Broadly speaking, the extension to the multivariate case is not found to undermine the findings on Clements and Hendry for the bivariate case. However, the four-variable setting makes it even more difficult to generalize the findings,

and the multiplicity of possible cases should lead to reticence when interpreting the results in a wider setting.

In order to reduce the impact of such criticisms, Lin and Tsay present two real data examples, one financial and one macroeconomic. They observe that the problem of roots close to the unit circle, but not actually being unit roots, is observable in data (that is, similarity to model 2, or, more extremely, model 3). In such circumstances, the under-specification of the rank (imposing unit roots that are not present) can be expected to result in poor long term forecasts.[12] Secondly, they observe that forecast error variances from a stationary system converge fairly rapidly as a function of forecast horizon. This is used to explore the stationarity of a system of bond yields. In this case, the unit root and cointegration tests performed suggest cointegration. This could be a case where the process is near non-stationary, and with a common feature, but the common feature is a root close to, but not on, the unit circle. It is clear from their investigations that, at a practical level, cointegrating restrictions cannot be assumed to improve long term forecasts, even where there is within-sample statistical evidence to support them.

6.3.4.4 Forecast evaluation

In both the Lin and Tsay and Clements and Hendry studies, the basic measure of forecast accuracy is the trace of the Monte Carlo estimate of the variance–covariance matrix of the forecasts. It has the following form. Let $e_{k,t}(j)$ be the j–step ahead vector forecast error made at time t arising from the k^{th} replication. Let the total number of replications be K. Then let

$$\hat{\Omega}_{\kappa,\tau}(j) = e_{k,t}(j)e_{k,t}(j)'.$$

One of the measures used by Clements and Hendry, and the one relevant to most of the results reported above, is

$$T(j) = trace\left[\frac{\sum_{k=1}^{K}\hat{\Omega}_{k,t}(j)}{K}\right],$$

which is referred to as the trace mean-square forecast error (TMSFE). Lin and Tsay use a modified version of this criterion since each replication gives rise to a set of j–step ahead forecasts, as a result of rolling forward the forecast origin within the same replication. They construct a within replication estimate of the forecast error variance–covariance matrix as

$$\hat{\Omega}_{k}(j) = \frac{\sum_{t=300}^{400-j}\hat{\Omega}_{k,t}(j)}{100-j+1}$$

This is then averaged across replications, the final measure being

$$E(j) = \sqrt{trace\left(\frac{\sum_{k=1}^{K} \hat{\Omega}_k(j)}{K} \right)} \tag{6.36}$$

Clements and Hendry (1998) discuss the choice of criterion, and use others in addition to TMSFE. An important aspect of these is their sensitivity to linear transformations of the data, although extensive use continues to be made of it.

6.3.4.5 Other issues relevant to forecasting performance in practice

In practice, forecasting will be subject to a number of other possible sources of error (Clements and Hendry, 1998, chapter 7, for a taxonomy). In the context of forecasting in cointegrated systems, these include the uncertainties associated with the selection of VAR order, the reliability of unit root and cointegration tests, and the estimation of the cointegrating vectors. This analysis has dealt exclusively with CI(1, 1) systems, elsewhere in this book, the case of cointegration in I(2) systems has been considered. This raises the question not just of how forecasting might be affected by choice of cointegration rank, but also types of (linear) cointegration, especially where there exists the possibility of variables being integrated of order up to 2.

All forecasting is predicated on at least two assumptions regarding model stability. That is, that the model structure has remained constant during the in-sample period, and that this same structure will remain into the forecasting period. Clements and Hendry (2001) have considered the implications for forecasting of some types of model instability in depth. Other procedures allow model switching (usually in a univariate setting, however), or non-linear adjustment to equilibrium. Any or all of these methods may be appropriate where a simple linear approximation fails to provide adequate forecasting performance.

Typically, the order of underlying VAR model is chosen by the optimization of some form of parsimonious information criterion, such as the SIC. These do not all have the same model selection properties, however (Reimers 1992). A potentially important variant of these criteria is to jointly select over VAR order and cointegrating rank. The criteria given by equation (6.34) are easily modified for this purpose. The VAR(p) can be estimated as a VECM as this does not alter the value of the $\left| \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}'_t \right|$, but cointegrating restrictions can be placed on the long-run matrix, via the Johansen procedure for example, such that

$$\underset{n\times n}{\Pi} = \underset{n\times r}{\alpha}\ \underset{r\times n}{\beta'}$$

such that there are only $2nr$ parameters of Π to be freely estimated. The information criterion is therefore of the form of (6.34) with $m = (p-1)n^2 + 2nr$, the selected model being that for which the criterion is minimized over a grid of values of p and $r = 0, 1 \ldots, n$ (the upper limit on the range of r allowing for stationarity). The evidence on the appropriate form of the penalty term, $f(T)$, is mixed (Reimers 1992), and while SIC can dominate, relative performance depends on simulation design. In practice, it is best to compute a range of criteria and search for corroborative evidence amongst them as to model order and cointegrating rank, and, if there is significant deviation in the findings, to check that subsequent inferences are not sensitive across the set of models suggested.[13]

Lin and Tsay (1996) point out that a model should be selected (and estimated) according to its purpose. In their paper they develop the idea that if the objective of the model is to forecast at a long-term forecast horizon, then it should be optimized to do this. Since standard methods of estimation and the form of information criteria are based on one step-ahead errors, it would not be surprising that such models were sub-optimal in terms of, say, 50-step ahead forecasts.

6.4 Models with short-run dynamics induced by expectations

A number of papers have considered the issue of estimating the linear quadratic cost of adjustment models under the type of dependence associated with cointegration (Dunne and Hunter 1998; Hunter 1989; and Engsted and Haldrup 1997). It should be understood that other forms of dependence might lead to similar types of problems. However, none of these are insurmountable. One issue which has been much discussed in the literature is the question of identification. As much of the analysis to date has concerned single equations, then the identification of the discount rate is of concern (Hendry et al. 1983; or Sargan 1982a). In general identification of parameters in structural or quasi-structural relationships is feasible (Arellano et al. 1999; Hunter 1989, 1992; and Pesaran 1981, 1987). A significant issue, as far as identification of forward-looking behaviour is concerned, is that both the IV and GMM estimators do not bind the solution based on the minimum of the optimization problem to the restrictions associated with the terminal condition (Nickel 1985; Hunter and Ioannidis 2000). Tests of over-identifying restrictions do not impose burdensome conditions on the estimator, and satisfaction of the necessary conditions follows without difficulty with the exception of highly non-persistent processes (Stock, Wright and Yogo 2002).

This section considers the impact of cointegration amongst endogenous and exogenous variables on rational expectations solutions and reveals a

computationally efficient estimation procedure that can readily be adapted to incorporate dependent I(1) processes either in the endogenous or the exogenous variables. The necessary and sufficient conditions for separation into two forms of long-run process is discussed in Hunter (1989, 1990), in terms of the types of condition discussed under cointegrating exogeneity in chapter 5. Otherwise efficient estimation of the long run requires the existence of a number of weakly exogenous variables either for the system or a sub-system for which behaviour is predominantly forward looking. This is intimately related to the notion of super exogeneity which may negate the practical use value of the Lucas critique (Lucas 1976; Hendry 1988; and Hendry and Favero 1992).

6.4.1 Linear quadratic adjustment cost models

Consider the following objective function based on Kollintzas (1985), though for ease of exposition the interaction between y_t and $(y_t - v_t)$ is excluded here:

$$E(\Im_t \Omega_t) = \sum_{t=o}^{T^*} E\{\delta^t (\Delta y_t' K \Delta y_t + (y_t - v_t)' H (y_t - v_t)) | \Omega_t)\}. \tag{6.37}$$

Let (6.37) define a control problem (Chow 1978), y_t is an n_1 vector of endogenous variables, v_t an n_1 vector of unobserved targets, that can be defined as a linear function of n_2 exogenous variables, z_t, where $v_t = Az_t + w_t$, A is a matrix of long-run multipliers, $w_t = z_t - E(z_t|\Omega_t)$ is a n_1 vector of white noise innovations and δ is the discount rate. With fixed initial conditions $y_0 = \bar{y}$, then from Kollintzas (1985) the Lagrange–Euler first-order condition after substituting out for v_t is:

$$E(\delta^t Q_0 y_t - \delta^{t+1} Q_1 y_{t+1} - \delta^t Q_1' y_{t-1} - \delta^t H A z_t | \Omega_t) = 0, \tag{6.38}$$

where $Q_0 = (1 + \delta) K + H$ and $Q_1 = K$.

Consider the process when it approaches its terminal value (at $T^* = T + N$):

$$E(\delta^{T^*} Q_0 y_{T^*} - \delta^{T^*+1} Q_1 y_{T^*+1} - \delta^{T^*} Q_1' y_{T^*-1} - \delta^{T^*} H A z_{T^*} | \Omega_t) = 0. \tag{6.39}$$

Stationarity is one precondition traditionally accepted for the transversality condition to be satisfied (Pesaran 1987), but when the structure includes a discount factor this assumption is too strong. In general all that is required is for (6.39) to be bounded as $T^* \to \infty$.

To reveal a standard symmetric solution to the forward-looking problem, (6.39) is scaled by $\delta^{-\frac{1}{2}(T^*+1)}$:

$$\begin{aligned} &E(\delta^{-\frac{1}{2}(T^*+1)} \delta^{T^*} Q_0 y_{T^*} - \delta^{-\frac{1}{2}(T^*+1)} Q_1 y_{T^*+1} - \\ &\delta^{-\frac{1}{2}(T^*+1)} \delta^{T^*} Q_1' y_{T^*-1} - \delta^{-\frac{1}{2}(T^*+1)} \delta^{T^*} H A z_{T^*} | \Omega_t = 0. \end{aligned} \tag{6.40}$$

Simplifying (6.40):

$$E(\delta^{-\frac{1}{2}}Q_0\delta^{-\frac{1}{2}(T^*)}y_{T^*} - \delta^{\frac{1}{2}(T^*+1)}Q_1 y_{T^*+1} - \delta^{\frac{1}{2}(T^*+1)}Q_1' y_{T^*-1} - \delta^{-\frac{1}{2}}HA\delta^{\frac{1}{2}T^*}z_{T^*}|\Omega_t) = 0. \tag{6.41}$$

Re-defining (6.41) in terms of $y^*_{T^*} = \delta^{-\frac{1}{2}(T^*)}\, y_{T^*}$ and $z^*_{T^*} = \delta^{-\frac{1}{2}(T^*)}\, z_{T^*}$ gives rise to the symmetric solution:

$$E(\delta^{-\frac{1}{2}}Q_0 y^*_{T^*} - Q_1 y^*_{T^*+1} - Q_1' y^*_{T^*-1} - \delta^{-\frac{1}{2}}HAz^*_{T^*}|\Omega_t) = 0. \tag{6.42}$$

In the limit (6.42) is bounded when the roots of the processes driving z_t and y_t are of mean order less than $\delta^{-\frac{1}{2}}$ as:

$$\underset{T^*\to\infty}{Lim}\, E(y^*_{T^*+1}|\Omega_t) \to 0 \;\; and \;\; \underset{T^*\to 0}{Lim}\, E(z^*_{T^*+1}|\Omega_t) \to 0.$$

Notice that (6.42) is bounded even when y and z have univariate time series representations that are non-stationary. Now consider the cointegration case. Dividing (6.38) by δ^t and transforming yields an error correction representation:

$$E(-\delta K\Delta y_{t+1} + K\Delta y_t + H(y_t - Az_t)|\Omega_t) = 0. \tag{6.43}$$

It follows that (6.43) is bounded in the limit when:

$$\underset{T^*\to\infty}{Lim}\{-\delta K\Delta y_{T^*+1} + K\Delta y_{T^*} + H(y_{T^*} - Az_{T^*})\} \to 0. \tag{6.44}$$

From the above discussion, a regular solution (see Pesaran 1987) to (6.42) exists, if and only if: (a) Q_o is symmetric; (b) K is non-singular; and (c) $\lambda < \delta^{-\frac{1}{2}}$. Dividing through (6.38) by δ^t yields the following difference equation:

$$E(Q_o y_t - \delta Q_1 y_{t+1} - Q_1 y_{t-1} - HAz_t|\Omega_t) = 0. \tag{6.45}$$

Redefining (6.45) using the forward (L^{-1}) and backward (L) lag operators:

$$Q(L)E(y_t|\Omega_t) = HAE(z_t|\Omega_t). \tag{6.46}$$

Now $Q(L) = (Q_o I - Q_1 L^{-1} - Q_1' L)$ has the following factorization:

$$Q_1 Q(L) = (I - G_1 L^{-1})(I - FL),$$

where $G_1 = \delta F$, $F = P\Lambda P^{-1}$ and Λ is a matrix whose diagonal elements are the stable eigenroots of the system. Therefore:

$$(I - G_1 L^{-1})(I - FL)E(y_t|\Omega_t) = K^{-1}HAE(z_t|\Omega_t). \tag{6.47}$$

It follows that the solution of the system can be written as:

$$y_t - Fy_{t-1} = \sum_{s=0}^{\infty}(G_1)^s F\, E(R_o Az_{t+s}|\Omega_t) + (G_1)^{-t}M_t + u_t \tag{6.48}$$

(Sargent (1978). Where $R_o = (\delta(F - I) + F^{-1} - I)$ and M_t satisfies the martingale property $E(M_{t+1}|\Omega_t) = (G_1)\, M_t$ (Pesaran 1987).

Reversing the transformation and applying it to (6.48):

$$\begin{aligned}
&(I - G_1L^{-1})(y_t - Fy_{t-1} - u_t)\\
&= (I - G_1L^{-1})(\sum_{s=0}^{\infty}(G_1)^s F\,E(R_oAz_{t+s}|\Omega_t) + (G_1)^{-t}M_t)\\
&= \sum_{s=0}^{\infty}(G_1)^s F\,E(R_oAz_{t+s}|\Omega_t) - G_1\sum_{s=0}^{\infty}(G_1)^s F\,E(R_oAz_{t+s}|\Omega_{t+1})\\
&\quad + (I - G_1L^{-1})(G_1)^{-t}M_t.
\end{aligned}$$

The first two terms on the right-hand side simplify, while the Koyck operator annihilates the bubble behaviour. Therefore:

$$\begin{aligned}
&(I - G_1L^{-1})(y_t - Fy_{t-1} - u_t)\\
&= FR_oAz_t + \sum_{s=1}^{\infty}(G_1)^s F\,E(R_oAz_{t+s}|\Omega_t) - \sum_{s=1}^{\infty}(G_1)^s F\,E(R_oAz_{t+s}|\Omega_{t+1})\\
&= FR_oAz_t + \sum_{s=1}^{\infty}(G_1)^s F(E(R_oAz_{t+s}|\Omega_t) - E(R_oAz_{t+s}|\Omega_{t+1})).
\end{aligned}$$

Assuming that there are no bubbles and a forcing process $z_t = B(L)w_t$ (w_t is white noise), then:

$$E(z_{t+s}|\Omega_t) - E(z_{t+s}|\Omega_{t+1}) = -B_{s-1}w_{t+1}$$

and

$$\begin{aligned}
(I - G_1L^{-1})(y_t - Fy_{t-1} - u_t) &= FR_oAz_t - \sum_{s=1}^{\infty}(G_1)^s F(B_{s-1}w_{t+1})\\
&= FR_oAz_t - FR_o(\sum_{s=1}^{\infty}(G_1)^s AB_{s-1})w_{t+1}.
\end{aligned}$$

Now reversing the Koyck lead and setting $(\sum_{s=1}^{\infty}(G_1)^s AB_{s-1}) = D$ gives rise to a forward-looking representation, which depends on future values of z_t. Therefore:

$$\begin{aligned}
(y_t - Fy_{t-1} - u_t) &= (I - G_1L^{-1})^{-1}(FR_oAz_t - FR_oDw_{t+1})\\
&= \sum_{s=1}^{\infty}(G_1)^s F(R_oAz_{t+s} - R_oDw_{t+s+1}).
\end{aligned}$$

It is possible to estimate the above model by FIML using the following recursion:

$$\begin{aligned}
(y_t - Fy_{t-1} - u_t) &= h_t \qquad (6.49)\\
h_t &= FR_oAz_t - FR_oDw_{t+1} + G_1h_{t+1}.
\end{aligned}$$

A fixed initial condition can be handled by recursively de-meaning the dependent variable (Taylor 1999), the problem of selecting an appropriate terminal

condition is solved by introducing a large enough future horizon or setting $(G_1)^s h_{t+s+1} = 0$.

Alternatively, the solution has the following backward representation, by substituting terms of the form $E(z_{t+s}|\Omega_t)$ using the Wiener–Kolmogorov prediction formula, which gives rise to the reduced form:

$$y_t - Fy_{t-1} = \Xi(L)z_t + u_t, \tag{6.50}$$

where $\Xi(L) = (\Xi_0 + \Xi_1 L + \ldots \Xi_{s-1}L^{s-1})$ is a function of δ, H, K, A and $\Theta(L) = (I + \Theta_1 L + \ldots \Theta_s L^s)$. However, this is a more complex set of non-linear relations to deal with (Hunter 1995; or Johansen and Swensen 1999).

It is also possible to give (6.49) a recursive structural form as long as K^{-1} exists. Notice that $R_o = K^{-1} H$ and:

$$\begin{aligned}(y_t - Fy_{t-1} - u_t) &= R_o FAz_t - R_o FDw_{t+1} + G_1 h_{t+1} \\ K(y_t - Fy_{t-1} - u_t) &= HFAz_t - HFDw_{t+1} + KG_1 h_{t+1}.\end{aligned} \tag{6.51}$$

As in a conventional system (Sargan 1988), to identify K, H and F, then $n_1 - 1$ additional restrictions are required (Hunter 1992a). Subject to knowledge of K and F, then H can be calculated from the following restriction $K\, R_o = K(\delta(F - I) + F^{-1} - I) = H$ as $R_o = K^{-1} H$ commutes. Essentially, identification of K follows from the additional restrictions, while identification of H follows from F, given knowledge of K and any additional restrictions to the system.

6.4.2 Models with forward behaviour and n_2 weakly exogenous I(1) variables.

If one considers the backward-looking form of the forward-looking model, then this is a VAR. The cointegrating VAR takes the from

$$\Delta y_t = [\alpha_{11} : \alpha_{12}] \begin{bmatrix}\beta_{1.} \\ \beta_{2.}\end{bmatrix} \begin{bmatrix}y_{t-1} \\ z_{t-1}\end{bmatrix} + \epsilon_{1t} \tag{6.52}$$

$$\Delta z_t = [\alpha_{21} : \alpha_{22}] \begin{bmatrix}\beta_{1.} \\ \beta_{2.}\end{bmatrix} \begin{bmatrix}y_{t-1} \\ z_{t-1}\end{bmatrix} + \epsilon_{2t}, \tag{6.53}$$

where any further dynamic can be incorporated in an appropriate time series representation of the error process. It follows for weak exogeneity relative to the long run, that $[\alpha_{21} : \alpha_{22}] = [0 : 0]$. As a result:

$$\Delta y_t = [\alpha_{11} : \alpha_{12}] \begin{bmatrix}\beta_{1.} \\ \beta_{2.}\end{bmatrix} \begin{bmatrix}y_{t-1} \\ z_{t-1}\end{bmatrix} + \epsilon_{1t} \tag{6.54}$$

$$\Delta z_t = \epsilon_{2t}, \tag{6.55}$$

where $\beta = [\beta'_{1.} \beta'_{2.}]$ and $\epsilon_{2t} = C(L)w_t$. Notice that inference on the short-run parameters is not appropriate as the coefficients of the ARMA error process forcing Δy_t depend on the MA process forcing ϵ_{1t}. It follows that the cointegrating relations are defined in the equations for Δy_t. Now consider the

solution to the forward-looking relationship given above, then the long-run behaviour that is important applies to the equation for y_t.

$$y_t - Fy_{t-1} = R_o F \sum_{s=0}^{\infty} (G_1)^s E(Az_{t+s}|\Omega_t) + u_t$$

where $R_oF = (\delta(F^2 - F) + I - F) = (\delta F(F - I) + I - F) = (I - F)(I - \delta F)$. It follows that:

$$\begin{aligned} & y_t - Fy_{t-1} - u_t \\ &= (I - F)(I - \delta F) \sum_{s=0}^{\infty} (G_1)^s E(Az_{t+s}|\Omega_t) \\ &= (I - F) \left(\sum_{s=0}^{\infty} (G_1)^s E(Az_{t+s}|\Omega_t) - \sum_{s=0}^{\infty} (G_1)^{s+1} E(Az_{t+s}|\Omega_t) \right) \\ &= (I - F)\{FAz_{t-1} + \sum_{s=0}^{\infty} (G_1)^s E(A\Delta z_{t+s}|\Omega_t)\}. \end{aligned} \tag{6.56}$$

Now it follows from the results in Engsted and Haldrup (1997) that (6.56) has an error correction type representation in differences and levels. Furthermore:

$$\Delta y_t + (I - F)y_{t-1} - u_t = (I - F)\{Az_{t-1} + \sum_{s=0}^{\infty} (G_1)^s E(A\Delta z_{t+s}|\Omega_t)\}$$

$$\Delta y_t + (I - F)\{y_{t-1} - Az_{t-1}\} - u_t = \sum_{s=0}^{\infty} (G_1)^s E(A\Delta z_{t+s}|\Omega_t).$$

In the error correction form $\alpha_1 = (I - F)$ and the cointegrating relations are normalized with respect of n_2 weakly exogenous variables as follows, $\beta = (I : A)$. The representations in Dolado et al. (1991) and Engsted and Haldrup (1997) rely on the existence of exactly n_2 weakly exogenous variables for the long-run to be estimated from the equations on y_t alone. It then follows that the above system can be estimated in two steps. Firstly the long run might be estimated using a regression or the Johansen Procedure, and then the short run relationship is estimated. There is no separate long-run relationship amongst the endogenous variables. Alternatively, consider a solved form similar to the one dealt with in sections 6.4.1:

$$y_t - Fy_{t-1} - (I - F)Az_{t-1} - u_t = \sum_{s=0}^{\infty} (G_1)^s E(A\Delta z_{t+s}|\Omega_t)\}$$

Reversing the Koyck transformation:

$$\begin{aligned} & (I - G_1 L^{-1})(y_t - Fy_{t-1} - (I - F)Az_{t-1} - u_t) \\ &= (I - G_1 L^{-1})(I - F)(\sum_{s=0}^{\infty} (G_1)^s E(A\Delta z_{t+s}|\Omega_t)) \\ &= (I - F)\{\sum_{s=0}^{\infty} (G_1)^s E(A\Delta z_{t+s}|\Omega_t) - G_1 \sum_{s=0}^{\infty} (G_1)^s E(\Delta Az_{t+s}|\Omega_{t+1})\} \end{aligned}$$

or

$$(I - G_1L^{-1})(y_t - Fy_{t-1} - (I - F)Az_{t-1} - u_t)$$
$$= (I - F)A\Delta z_t + (I - F)\sum_{s=1}^{\infty}(G_1)^s(E(A\Delta z_{t+s}|\Omega_t) - E(A\Delta z_{t+s}|\Omega_{t+1})).$$

It follows from the Granger representation theorem that Δz_t has the following Wold form $\Delta z_t = C(L)w_t$ and

$$E(\Delta z_{t+s}|\Omega_t) - E(\Delta z_{t+s}|\Omega_{t+1}) = -C_{s-1}w_{t+1}.$$

Substituting back into the forward-looking model:

$$(I - G_1L^{-1})(y_t - Fy_{t-1} - (I - F)Az_{t-1} - u_t)$$
$$= (I - F)A\Delta z_t - (I - F)\sum_{s=1}^{\infty}(G_1)^s(AC_{s-1}w_{t+1})$$
$$= (I - F)A\Delta z_t - (I - F)(\sum_{s=1}^{\infty}(G_1)^s AC_{s-1})w_{t+1}.$$

Now reversing the Koyck lead and setting $(\sum_{s=1}^{\infty}(G_1)^s AC_{s-1}) = D^*$, gives rise to a forward-looking representation, which depends on future values of Δz_t:

$$y_t - Fy_{t-1} - (I - F)Az_{t-1} - u_t$$
$$= (I - G_1L^{-1})^{-1}((I - F)A\Delta z_t - (I - F)D^*w_{t+1})$$
$$= \sum_{s=0}^{\infty}(G_1)^s((I - F)A\Delta z_{t+s} - (I - F)D^*w_{t+s+1}).$$

Now decompose the last relationship as follows:

$$y_t - Fy_{t-1} - (I - F)Az_{t-1} - u_t$$
$$= \sum_{s=0}^{\infty}(G_1)^s((I - F)A\Delta z_{t+s} - \sum_{s=0}^{\infty}(G_1)^s(I - F)D^*w_{t+s+1})$$
$$= \sum_{s=0}^{\infty}(G_1)^s(I - F)Az_{t+s} - \sum_{s=0}^{\infty}(G_1)^s(I - F)Az_{t+s-1}$$
$$- \sum_{s=0}^{\infty}(G_1)^s(I - F)D^*w_{t+s+1}.$$

Therefore:

$$y_t - Fy_{t-1} - u_t = \sum_{s=0}^{\infty}(G_1)^s(I - F)Ax_{t+s} + (I - F)Ax_{t-1} - (I - F)Ax_{t-1}$$
$$- \sum_{s=1}^{\infty}(G_1)^s(I - F)Ax_{t+s-1} - \sum_{s=0}^{\infty}(G_1)^s(I - F)D^*w_{t+s+1}.$$

Re-writing the above into an equation purely in levels:

$$y_t - Fy_t - u_t = (I-F)\{\sum_{s=0}^{\infty}(G_1)^s Az_{t+s} - G_1\sum_{s=1}^{\infty}(G_1)^{s-1}(I-F)Az_{t+s-1}\}$$
$$-\sum_{s=0}^{\infty}(G_1)^s(I-F)D^*w_{t+s+1}.$$

Re-indexing the second sum and gathering terms, yields a levels relationship:

$$y_t - Fy_{t-1} - u_t = (I-F)(I-G_1)\sum_{s=0}^{\infty}(G_1)^s Az_{t+s} - \sum_{s=0}^{\infty}(G_1)^s(I-F)D^*w_{t+s+1}.$$

It is possible to estimate the above model by FIML using the following recursion:

$$y_t - Fy_{t-1} - u_t = h_t \qquad (6.57)$$
$$h_t = FR_oAz_t - (I-F)D^*w_{t+1} + G_1h_{t+1}.$$

In such circumstances the above relationship has the same forward recursion as was considered before, except the transversality condition relies on the existence of cointegration. Decompose (6.44) as follows:

$$\underset{T^*\to\infty}{Lim}\{-\delta K\Delta y_{T^*+1} + K\Delta y_{T^*} + H(y_{T^*} - Az_{T^*})\} =$$
$$\underset{T^*\to\infty}{Lim}\{-\delta K\Delta y_{T^*+1} + K\Delta y_{T^*}\} + \underset{T^*\to\infty}{Lim}\{H(y_{T^*} - Az_{T^*})\} \to 0.$$

The conditions for cointegration (Engle and Granger 1987) are sufficient for this to be satisfied. That is $y_t \sim I(1)$ and $(y_t - Az_t) \sim I(0)$, y_t and z_t cointegrate. Furthermore, (6.57) has an error correction form:

$$\Delta y_t - (I-F)(y_{t-1} - Az_{t-1}) - u_t = h_t$$
$$h_t = (I-F)A\Delta z_t - (I-F)D^*w_{t+1} + G_1h_{t+1}$$

In the next section the case with dependence amongst the endogenous variables is considered.

6.4.3 Models with forward behaviour and unit roots in the process driving y_t

There are a number of reasons for finding dependence amongst the endogenous processes, one of which would be cointegration, the other would be the type of dependence that exists amongst series that might satisfy an adding up type constraint. In the former case the cause of rank failure is the existence of a unit root and it can be shown that the original objective function can be solved in the usual way (Hunter 1989a).

Consider the loss function

$$E(\Im_t|\Omega_t)=\sum_{t=o}^{T^\star}E\{\delta^t(\Delta y_t'K\Delta y_t+(y_t-v_t)'H(y_t-v_t))|\Omega_t)\} \tag{6.58}$$

where the *rank*(*H*) = r_1 As a result, the following decomposition exists: $H = E'E$ and *rank*(*E*) = r_1. Now define *M* such that the matrix $[E' : M']$ has full rank. Now we can redefine the loss function in terms of new variables:

$$E(\Im_t|\Omega_t)=\sum_{t=o}^{T^\star}E\{\delta^t(\Delta y_t^{*'}K^*\Delta y_t^*+(y_t^*-v_t^*)'H(y_t^*-v_t^*))|\Omega_t)\} \tag{6.59}$$

where $y_t^{*'}=\left[y_{1t}^{*'}\quad y_{2t}^{*'}\right]=y_t'\left[E':M'\right], K^\star=\left[E':M'\right]^{-1}K\begin{bmatrix}E\\M\end{bmatrix}^{-1}$ and $v_t^\star$ conformable with $y_t^\star$. It follows that the loss function has the following form:

$$\begin{aligned}E(\Im_t|\Omega_t)=\sum_{t=o}^{T^\star}E\{&\delta^t\Delta y_{1t}^{*'}K_{11}^*\Delta y_{1t}^*+2\Delta y_{1t}^{*'}K_{12}^*\Delta y_{2t}^*+\\&\Delta y_{2t}^{*'}K_{22}^*\Delta y_{2t}^*+(y_{1t}^*-v_{1t}^*)'(y_{1t}^*-v_{1t}^*))|\Omega_t)\}.\end{aligned} \tag{6.60}$$

Re-writing the above relationship in terms of a new set of stationary variables, then $y_t^{+\prime}=[y_{1t}^{\star\prime}\quad \Delta y_{2t}^{\star\prime}]$ and here it is assumed that the long-run target for $v_{2t}^\star = 0$ and $\Delta\Delta y_{2t} = 0$. Therefore:

$$\begin{aligned}E(\Im_t|\Omega_t)=\sum_{t=o}^{T^\star}E\{&\delta^t(\Delta y_{1t}^{+'}K_{11}^*\Delta y_{1t}^++2\Delta y_{1t}^{+'}K_{12}^*\Delta y_{2t}^+ +\\&y_{2t}^{+'}K_{22}^*y_{2t}^++(y_{1t}^+-v_{1t}^+)'(y_{1t}^+-v_{1t}^+))|\Omega_t)\}.\end{aligned}$$

Now differentiating with respect to y_{1t}^+ gives rise to the following first-order condition:

$$E(\delta^tK_{11}^*\Delta y_{1t}^+-\delta^{t+1}K_{11}^*\Delta y_{1t+1}^+-\delta^t(y_{1t}^+-v_{1t}^+)-2\delta^tK_{12}^*(y_{2t}^+-\delta y_{2t+1}^+)|\Omega_t)=0, \tag{6.61}$$

and with respect of y_{2t}^+:

$$E(\delta^tK_{21}^*\Delta y_{1t}^++\delta^tK_{22}^*y_{2t}^+|\Omega_t)=0.$$

Subtracting the above equation from its forward value and re-writing:

$$E(\delta^tK_{21}^*(\Delta y_{1t}^*-\delta\Delta y_{1t+1}^*)+\delta^tK_{22}^*(\Delta y_{2t}^*-\delta\Delta y_{2t+1}^*)|\Omega_t)=0.$$

Now consider the system:

$$E(\delta^t(K^*(\Delta y_t^*-\delta\Delta y_{t+1}^*)+\begin{bmatrix}I_r&0\\0&0\end{bmatrix}(y_t^*-v_t^*|\Omega_t)=0.$$

Now divide through by δ^t and reverse the transformation:

$$E(K(\Delta y_t-\delta\Delta y_{t+1})+H(y_t-z_t)|\Omega_t)=0.$$

Hence, irrespective of the existence of cointegration, the same first-order condition exists as does the solution dealt with before, except that H is rank deficient. Therefore $R_0 = K^{-1}H$ is rank deficient, F has $n_1 - r_1$ unit roots and $R_0 = (I - F)(I - \delta F)F^{-1}$ is rank deficient as can be observed from the following decomposition:

$$\begin{aligned}(I - F)(I - \delta F) &= (I - P\Lambda P^{-1})(I - \delta P\Lambda P^{-1}) \\ &= P(I - \Lambda)(I - \delta\Lambda)P^{-1}.\end{aligned}$$

Where the $rank((I - F)(I - \delta F)) = r_1$, when there are $n_1 - r_1$ unit roots. Hence the rank of the matrix H determines the number of unit roots. Now it is probably better to consider the recursive representation (6.57):

$$\begin{aligned}\Delta y_t - (I - F)(y_{t-1} - Az_{t-1}) - u_t &= h_t \\ h_t = (I - F)A\Delta x_t - (I - F)D * w_{t+1} + G_1 h_{t+1}&.\end{aligned}$$

If $I - F$ is rank deficient, then there is also the possibility of cointegration amongst the endogenous and exogenous variables. Notice the dependence also feeds forward into the relations in differences.

6.4.4 Estimation and inference

The benefit of the above approach is that it reduces the dimension of the estimation problem when forward-looking behaviour needs to be considered. Especially in terms of the need to estimate and store future predictions. However, the downside is that inference is made more complicated.

As far as estimation is concerned, then the usual likelihood function applies, where:

$$LogL((\delta, H, K, A), \Sigma|.) = -Tn\log(2\pi) - \frac{1}{2}T\log|\Sigma| - \frac{1}{2}tr(\Sigma^{-1}\sum_{t=1}^{T} u_t u_t')$$

and $u_t = y_t - Fy_{t-1} - h_t$. Now concentrating out Σ yields the quasi-likelihood:

$$LogL_c((\delta, H, K, A)|.) = C - \log|\Sigma|$$

where $S = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$ is a consistent estimate of Σ. The likelihood is maximized using a Quasi-Newton algorithm such as Gill, Murray and Pitfield (see Sargan 1988) or an equivalent method. The method due to Gill, Murray and Pitfield has the advantage of using the Cholesky factors from the inverse of the Hessian. They are then bounded to be positive definite subject to an appropriately conditioned Hessian matrix.

However, the conventional estimates of the parameter variance based on the information matrix are not valid, even when the model for the endogenous equations is estimated as a system. The correct estimate needs to take account of the generated regressors and their parameter estimates. The following algorithm is suggested to do this. Initial estimates of the exogenous

variables are estimated as a VAR, then the residuals are saved. The VMA representation is estimated by OLS using the method described by Spliid (1983). In state space form:

$$\Delta z = W\varsigma + w.$$

$$\text{where } \Delta z = \begin{bmatrix} \Delta z_1 \\ \Delta z_2 \\ \cdot \\ \cdot \\ \cdot \\ \Delta z_T \end{bmatrix}, W = \begin{bmatrix} w_{-1} & w_{-2} & \cdots & w_{-p} \end{bmatrix}, w = \begin{bmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ \cdot \\ w_T \end{bmatrix}$$

$$\text{and } \varsigma = \begin{bmatrix} C_1 \\ C_2 \\ \cdot \\ \cdot \\ \cdot \\ C_p \end{bmatrix}$$

Hence, the OLS estimator of the parameters is given by:

$$\Psi^{(0)} = (W'_{(o)}W_{(o)})^{-1}W_{(o)}\Delta z,$$

where $W_{(o)}$ contains the initial estimates of the surprises, unobserved values of the residual are set to zero and $\Psi^{(0)}$ are the initial estimates of the parameters. Once the system has been estimated, then the likelihood is re-estimated based on $B = 200$, bootstrap re-samplings of the original residuals vector w, where each iteration reallocates a block of residuals w_i by the new residual set $w_{(b)}$ used to provide new estimates of the VMA parameters ($\Psi^{(b)}$ for $b = 1, \ldots, B$). Then given the maximum likelihood estimates of the parameters (δ, H, K, A) an empirical distribution for the estimated test statistics are generated from the bootstrap re-sampling regime. A sample of 400 is created by the use of antithetic variance technique, providing at each bootstrap replication a pair of residuals $w_{(b)}$ and $-w_{(b)}$ (see Hendry 1995). Then percentiles of the empirical distribution can be used to determine critical values for the estimated parameters.

6.5 Conclusion

In this chapter a number of more advanced issues have been addressed: cointegration amongst series with different orders of integration; forecasting with cointegrating relationships; and cointegration combined with short-run structure defined by rational expectations.

With orders of integration in excess of I(1), inference is similar to the I(1) case except that there are now three types of process that evolve to generate the data. Cointegration not only occurs in the usual way amongst the levels, but may also occur between levels and differenced series, there are I(1) common trends and also I(2) trends. However, identification is a fundamental problem for the estimation of long-run behaviour in the I(2) case as three sets of parameters are potentially ill-defined.

When the order of integration is less than 1, then series are not likely to have the same fractional order of differencing. One approach is to consider the average non-integer order of differencing for a group of series. Estimation of the cointegrating vectors can be undertaken in a similar way as that for I(1) series when a non-parametric approach is considered (Robinson and Marinucci 1998), but testing is more complex (Robinson and Yajima 2002). It is relatively straightforward to compare the order of difference between series and to calculate the cointegrating rank, but there is no conventional procedure for inference.

Forecasting in cointegrated systems occurs at two levels – the short run and the long run and cointegration influences both of these. Short-run forecasts are less influenced by cointegration, but long-run forecasts may be strongly influenced. The literature is unclear as to whether gains in forecast accuracy depend on the restrictions that cointegration imposes on the long-run process or the interrelationship that cointegration imposes on the long-run forecasts. It appears that there is little difference between long-run forecasts derived from models that imposed the long-run restrictions as the forecasr evolves when they are compared with forecasts that ex-post have the cointegrating restriction imposed on them. This might suggest that the benefits to long-run forecasting associated with cointegration follow from the imposition of the restriction rather than cointegration per se. This would appear to be an issue for further investigation, though the authors would conjecture that cointegration has a role in the accuracy of long-run forecasts.

Estimation of the structural parameters of optimizing models has become enormously popular. It has become common practice to suggest that the VAR is a solution to a forward-looking model, but then not to consider the relation between the long-run and the short-run behaviour of the model. However, both the Engle–Granger and the Johansen procedure have been applied to models with forward-looking behaviour. The final section of this chapter considered the impact of unit root processes in the endogenous and exogenous variables on the solution and estimation of forward-looking models with rational expectations. Inference is significantly more complicated in these cases and has thus far had to derive from the proposition that series are cointegrated.