6

Econometric methods: A review

In this chapter, we provide an overview of the econometric methods used in long-run structural macroeconometric modelling. The aim is to place in context the methods employed, to describe the steps taken in the estimation and development of the model, to help explain the various econometric tools used in interpreting the empirical results, and to explore some of the ways in which a long-run structural macroeconometric model can be used.

The long-run structural VARX modelling approach adopted in our work is described in Pesaran and Shin (2002) and Pesaran, Shin and Smith (2000), jointly denoted PSS, and is based on a modified and generalised version of Johansen's (1988, 1991, 1995) maximum likelihood approach to the problem of estimation and hypothesis testing in the context of augmented vector autoregressive error correction models. Of course, the analysis of economic time series containing unit roots has a long history, traceable to Yule's seminal (1926) paper on the potential pitfalls of interpreting regressions based on such data.[1] Granger and Newbold (1974) revived the issue when they showed that spurious regressions could result from the regression of one independent random walk on another.[2] The theoretical rationale behind the Granger–Newbold spurious regression result was set out in Phillips (1986) who showed that the R^2 of the regressions involving $I(1)$ variables tend to one and the t-ratios grow without bound as the sample size increases, even if the underlying $I(1)$ variables

[1] Excellent surveys of the literature on cointegration are provided in Banerjee *et al.* (1993), Watson (1994), Hamilton (1994), and in the papers in the *Special Issue of the Journal of Economic Surveys* edited by Oxley and McAleer (1998). The material of this chapter draws on Pesaran and Smith (1998) in that *Special Issue*.

[2] The problem of spurious regression in the case of stationary but *highly* serially correlated regressors was demonstrated earlier by Champernowne (1960), also using Monte Carlo techniques.

are statistically independent. The possibility of spurious regression and the growing availability of tests for unit roots, *e.g.* Dickey and Fuller (1979), led to a proliferation of testing for the order of integration of economic time series in the 1980s. The classic study is Nelson and Plosser (1982) who raised the possibility that the null hypothesis of a unit root could not be rejected for most US economic time series. At the same time, Granger (1981, 1986) and Engle and Granger (1987) were developing the analysis of cointegrated systems, explaining the links with the (relatively well-established) error correction models used for example in Sargan (1964) and subsequently popularised through the work of Davidson *et al.* (1978). Johansen's maximum likelihood approach popularised the use of cointegration analysis, allowing for symmetric treatment of all the variables in the cointegrated system and for an analysis of the number of cointegrating relations. Our own approach, elaborated in PSS, builds on this to allow economic theory to motivate the exact and over-identifying restrictions studied in the cointegration analysis in place of the type of statistical identification used by Johansen. PSS also develop the econometric analysis of vector error correction models with weakly exogenous $I(1)$ variables.

In what follows, we provide a brief statement of the econometric issues involved in the modelling approach advanced in PSS. We start by describing a general structural VARX model, allowing for the possibility of drawing a distinction between endogenous and exogenous variables. We use this general model to place in context the identification issues raised in Chapter 3 and to introduce the ideas behind impulse response analysis. We then turn our attention to cointegrating VARX models, contrasting the PSS approach to the Johansen approach, commenting on the small sample properties of some of the test statistics and broadening the discussion of the impulse response analysis to a more general analysis of system dynamics in the cointegrated VARX context. We end the chapter with comments on the small sample properties of some of the test statistics discussed in the chapter and on the distributional properties of the impulse response function. These statistical properties can be readily investigated through simulation methods and we explain how simulation methods can be used in this regard. This sets the scene for the use of structural VARX models in forecasting discussed in Chapter 7. Throughout the chapter, our description of the econometric techniques is informed by how they are used in practice and we relate the discussion to the choices that an applied econometrician has to make in the practical application of cointegrating VARX techniques.

6.1 Augmented VAR or VARX models

6.1.1 *The structural VARX model*

The general structural VARX model for an $m_y \times 1$ vector of endogenous variables $\mathbf{y}_t$, is given by:[3]

$$\mathbf{A}\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{B}_0\mathbf{x}_t + \mathbf{B}_1\mathbf{x}_{t-1} + \cdots + \mathbf{B}_p\mathbf{x}_{t-p} + \mathbf{D}\mathbf{d}_t + \boldsymbol{\varepsilon}_t, \tag{6.1}$$

for $t = 1, 2, \ldots, T$, where $\mathbf{d}_t$ is a $q \times 1$ vector of deterministic variables (*e.g.* intercept, trend and seasonal variables), $\mathbf{x}_t$ is an $m_x \times 1$ vector of exogenous variables, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{m_y t})'$ is an $m_y \times 1$ vector of serially uncorrelated errors distributed independently of $\mathbf{x}_t$ with a zero mean and a constant positive definite variance–covariance matrix, $\boldsymbol{\Omega} = (\omega_{ij})$, where ω_{ij} is the (i, j)th element of $\boldsymbol{\Omega}$. For given values of $\mathbf{d}_t$ and $\mathbf{x}_t$, the above dynamic system is stable if all the roots of the determinantal equation

$$\left|\mathbf{A} - \mathbf{A}_1\lambda - \mathbf{A}_2\lambda^2 - \cdots - \mathbf{A}_p\lambda^p\right| = 0, \tag{6.2}$$

lie strictly outside the unit circle. This stability condition ensures the existence of long-run relationships between $\mathbf{y}_t$ and $\mathbf{x}_t$, which will be cointegrating when one or more elements of $\mathbf{x}_t$ are integrated, namely contain unit roots. The assumption, however, rules out the possibility that the endogenous variables, $\mathbf{y}_t$, will themselves be cointegrating when the model contains no exogenous variables.

The above VARX model is structural in the sense that it explicitly allows for instantaneous interactions between the endogenous variables through the contemporaneous coefficient matrix, $\mathbf{A}$. It can also be written as

$$\mathbf{A}(L)\mathbf{y}_t = \mathbf{B}(L)\mathbf{x}_t + \mathbf{D}\mathbf{d}_t + \boldsymbol{\varepsilon}_t, \tag{6.3}$$

where L is the lag operator such that $L\mathbf{y}_t = \mathbf{y}_{t-1}$, and

$$\mathbf{A}(L) = \mathbf{A} - \mathbf{A}_1L - \cdots - \mathbf{A}_pL^p; \quad \mathbf{B}(L) = \mathbf{B}_0 + \mathbf{B}_1L + \cdots + \mathbf{B}_pL^p.$$

Of particular interest are the system long-run effects of the exogenous variables which are given by:

$$\mathbf{A}(1)^{-1}\mathbf{B}(1) = \left(\mathbf{A} - \sum_{i=1}^{p}\mathbf{A}_i\right)^{-1}\sum_{i=0}^{p}\mathbf{B}_i.$$

[3] In general, different orders can be assumed for the distributed lag functions associated with the endogenous and exogenous variables. Alternatively, p can be viewed as the maximum lag order of the distributed lag functions on $\mathbf{y}_t$ and $\mathbf{x}_t$.

Notice that, since all the roots of (6.2) fall outside the unit circle by assumption, the inverse of $\mathbf{A}(1)$, which we denote by $\mathbf{A}(1)^{-1}$, exists.

INITIAL MODELLING CHOICES

The decision to work with a model of the type described above presents the applied econometrician with a number of important choices, namely:

1. The number and list of the endogenous variables to be included, $(m_y, \mathbf{y}_t)$.
2. The number and list of the exogenous variables (if any) to be included, $(m_x, \mathbf{x}_t)$.
3. The nature of the deterministic variables (intercepts, trends, seasonals) and whether the intercepts and/or the trend coefficients need to be restricted.
4. The lag orders of the VARX (the lag order of the $\mathbf{y}_t$ and $\mathbf{x}_t$ components of the VARX need not be the same).
5. The order of integration of the variables.

These choices change the maximised value of the log-likelihood (MLL) so that, in principle, they could be made on the basis of either hypothesis testing exercises or by means of model selection criteria such as the Akaike Information Criterion (AIC), or the Schwarz Bayesian Criterion (SBC). However, different significance levels, different forms of the tests and different model selection criteria invariably can lead to different model specifications. In many cases, little is known about the small sample properties of these procedures and what is known is often not reassuring. Little is also known about the properties of the tests or model selection criteria when the range of models considered does not include the data generation process. These choices are often closely related and the outcomes are sensitive to initial choices. The combination of these choices gives us a very large space of possible models and there is no reason to expect a series of sequential choices (*e.g.* fix m_y and m_x, then choose p conditional on m_y and m_x, *etc.*) to adequately explore the possible model space. Joint tests may lead to different inferences from a sequence of individual tests. Sequential procedures are likely to suffer from pre-test bias, while general to specific searches face the difficulty that the unrestricted models are profligate with parameters.

While data-dependent decision procedures are extremely important, they have to be supplemented with other considerations given the complexity of most applied modelling problems. In particular, choices will be

informed by the purpose of the exercise and by prior information from economic theory; theory being interpreted widely. In principle, this combination could be done formally by embodying the purpose of the exercise in an explicit loss function and the theory information in a prior probability distribution for the parameters, and then applying Bayesian techniques. In practice, the difficulty of formalising the loss function and prior probability distributions makes a formal use of these other considerations attractive only for relatively simple problems. Often, the applied econometrician will make use of a range of informal procedures for integrating economic and statistical information. For example, statistically insignificant variables may be retained when they are economically important, and statistically significant variables may be deleted when they are likely to be economically unimportant, since misleading statistical significance can arise for many reasons. For example, chance correlations with omitted variables, like cold winters or policy announcements, can make variables significant. It is a matter of judgement whether these variables or lags are regarded as economically important.

Given the size of the potential model space, defined by the choices discussed above, it is important to investigate a range of specifications and allow for model uncertainty in forecasting and policy analysis. At present full exploration of the model space is likely to be highly data-intensive and computationally burdensome, if not infeasible. Even much simpler problems, like determining the lag order in a single-equation autoregressive distributed lag model, as discussed in Pesaran and Shin (1999), require many hundred regressions. As full exploration is not feasible, organised sensitivity analysis plays an important role. This sensitivity analysis should investigate both the statistical significance and the economic importance of the restrictions.

6.1.2 *The reduced form VARX model*

The reduced form of the structural model (6.1), which expresses the endogenous variables in terms of the predetermined and exogenous variables, is given by

$$\mathbf{y}_t = \mathbf{\Phi}_1\mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_p\mathbf{y}_{t-p} + \mathbf{\Psi}_0\mathbf{x}_t + \mathbf{\Psi}_1\mathbf{x}_{t-1} + \cdots + \mathbf{\Psi}_p\mathbf{x}_{t-p} + \mathbf{\Upsilon}\mathbf{d}_t + \mathbf{u}_t, \tag{6.4}$$

where $\mathbf{\Phi}_i = \mathbf{A}^{-1}\mathbf{A}_i$, $\mathbf{\Psi}_i = \mathbf{A}^{-1}\mathbf{B}_i$, $\mathbf{\Upsilon} = \mathbf{A}^{-1}\mathbf{D}$, $\mathbf{u}_t = \mathbf{A}^{-1}\boldsymbol{\varepsilon}_t$ is *i.i.d.* $(0, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \mathbf{A}^{-1}\mathbf{\Omega}\mathbf{A}'^{-1} = (\sigma_{ij})$. The classical identification problem is how to recover the structural form parameters

$$(\mathbf{A}, \mathbf{A}_{i+1}, \mathbf{B}_i, i = 0, 1, \ldots, p;\ \mathbf{D} \text{ and } \boldsymbol{\Omega}),$$

from the reduced form parameters,

$$(\boldsymbol{\Phi}_i, \boldsymbol{\Psi}_i, i = 0, 1, \ldots, p, \boldsymbol{\Upsilon}, \text{ and } \boldsymbol{\Sigma}).$$

This is the identification issue raised in the discussion of Section 3.1, and Section 3.1.2 in particular. The resolution of this identification problem formed the basis of the Cowles Commission approach to structural modelling in econometrics. Exact identification of the structural parameters requires m_y^2 *a priori* restrictions, of which m_y restrictions would be provided by normalisation conditions. The restrictions typically involve setting certain elements of the structural coefficient matrices to zero. These were the *a priori* restrictions criticised by Sims (1980), particularly when such identifying restrictions were obtained by restricting the short-run dynamics. Most of the traditional macromodels were heavily over-identified and while, in principle, these over-identifying restrictions could be tested, in practice the number of exogenous and predetermined variables was so large that it was impossible to estimate the reduced form. There are a variety of other ways of imposing identifying restrictions. For instance, if after a suitable ordering, it is assumed that $\mathbf{A}$ is triangular *and* $\boldsymbol{\Omega}$ diagonal (though there is no general theoretical reason to expect it to be so), the structural system becomes a recursive causal chain, each equation of which can be consistently estimated by OLS. The assumptions that $\mathbf{A}$ is triangular and $\boldsymbol{\Omega}$ is diagonal each provide $m_y(m_y - 1)/2 + m_y(m_y - 1)/2$ restrictions respectively, which together with the m_y normalisation restrictions just identify the system. As we shall see below these assumptions are also equivalent to the use of the Choleski decomposition of $\boldsymbol{\Sigma}$ originally advocated by Sims for identification of impulse responses.

6.1.3 *Impulse response analysis*

One of the main features of the traditional macromodels was their dynamic multipliers, which measured the effect of a shock to an exogenous variable, *e.g.* a policy change, or a shock to one of the structural errors, $\boldsymbol{\varepsilon}_t$, on the (expected) future values of the endogenous variables. Here, we shall briefly review how one can measure the dynamic effects of shocks or impulse response functions.

Under the stability assumption (namely that the roots of (6.2) lie strictly outside the unit circle), $\mathbf{A}(L)$ is invertible and the time profile of the

effect of a shock can be calculated from the 'final form' of the structural model:

$$\mathbf{y}_t = \mathbf{A}(L)^{-1}\mathbf{B}(L)\mathbf{x}_t + \mathbf{A}(L)^{-1}\mathbf{D}\mathbf{d}_t + \mathbf{A}(L)^{-1}\boldsymbol{\varepsilon}_t. \tag{6.5}$$

This expresses each endogenous variable in terms of an infinite distributed lag on the exogenous variables and an infinite moving average process on the structural errors. Notice that the dynamic multipliers, the effects of a shock to $\mathbf{x}_t$, can be derived from the reduced form coefficients, but to measure the dynamic effect of a shock to the structural errors we have to identify the structural coefficients. The equivalent final form representation of the reduced form model is:

$$\mathbf{y}_t = \boldsymbol{\Phi}(L)^{-1}\boldsymbol{\Psi}(L)\mathbf{x}_t + \boldsymbol{\Phi}(L)^{-1}\boldsymbol{\Upsilon}\mathbf{d}_t + \boldsymbol{\Phi}(L)^{-1}\mathbf{u}_t, \tag{6.6}$$

where[4]

$$\boldsymbol{\Phi}(L) = \mathbf{I}_{m_y} - \boldsymbol{\Phi}_1 L - \cdots - \boldsymbol{\Phi}_p L^p, \quad \boldsymbol{\Psi}(L) = \boldsymbol{\Psi}_0 + \boldsymbol{\Psi}_1 L + \cdots + \boldsymbol{\Psi}_p L^p,$$

and $\mathbf{I}_{m_y}$ is an identity matrix of order m_y. Since $\boldsymbol{\Phi}(L)$ is invertible, we have the following moving-average representation of the structural errors:

$$\boldsymbol{\Phi}(L)^{-1}\mathbf{u}_t = \sum_{i=0}^{\infty}\boldsymbol{\Theta}_i\mathbf{u}_{t-i} = \sum_{i=0}^{\infty}\boldsymbol{\Theta}_i\mathbf{A}^{-1}\boldsymbol{\varepsilon}_{t-i}, \tag{6.7}$$

where the $\boldsymbol{\Theta}_i$'s can be calculated from the following recursive relations:

$$\boldsymbol{\Theta}_i = \boldsymbol{\Phi}_1\boldsymbol{\Theta}_{i-1} + \boldsymbol{\Phi}_2\boldsymbol{\Theta}_{i-2} + \cdots + \boldsymbol{\Phi}_p\boldsymbol{\Theta}_{i-p}, \quad \text{for } i = 0, 1, 2, \ldots, \tag{6.8}$$

where $\boldsymbol{\Theta}_i = \mathbf{0}$, for $i < 0$ and $\boldsymbol{\Theta}_0 = \mathbf{I}_{m_y}$.

Although this *infinite* moving average representation exists only when the model is stable, it turns out that similar results can be obtained even in the unstable case where one or more roots of (6.2) are on the unit circle. Irrespective of whether the model is stationary or contains unit roots, one can derive impulse response functions for the responses of the endogenous variables to a 'unit' displacement in the particular elements of either the exogenous variables, $\mathbf{x}_t$, or the errors ($\mathbf{u}_t$ or $\boldsymbol{\varepsilon}_t$). The former represents the time profile of the response of the system to changes in the observed forcing variables of the system, while the latter examines the responses of

[4] Since $\mathbf{A}$ is non-singular and the roots of $|\mathbf{A} - \mathbf{A}_1\lambda - \mathbf{A}_2\lambda^2 - \cdots - \mathbf{A}_p\lambda^p| = 0$ are assumed to fall outside the unit circle, it follows that the roots of $|\mathbf{I}_{m_y} - \boldsymbol{\Phi}_1\lambda - \boldsymbol{\Phi}_2\lambda^2 - \cdots - \boldsymbol{\Phi}_p\lambda^p| = 0$ will also fall outside the unit circle.

the system to changes in the unobserved forcing variables. The impulse response functions for the errors can be defined either with respect to the 'structural' errors, $\boldsymbol{\varepsilon}_t$, or with respect to the reduced form errors, $\mathbf{u}_t$. All these impulse responses can be obtained using the generalised impulse response approach advanced in Koop *et al.* (1996) for non-linear models and discussed in more detail for linear models in Pesaran and Shin (1998). The generalised impulse response function GIRF measures the change to the n period ahead forecast of each of the variables that would be caused by a shock to the exogenous variable, structural or reduced form disturbance.

GENERALISED IMPULSE RESPONSE FUNCTIONS

To formally define the generalised impulse response functions, denote the information set containing current and all lagged values of $\mathbf{y}_t$ and $\mathbf{x}_t$ by $\mathfrak{I}_t = (\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots; \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots)$. Consider a shock to the ith structural error, ε_{it}, and let $\mathfrak{g}(n, \mathbf{z} : \varepsilon_i)$ be the generalised impulse responses of $\mathbf{z}_{t+n} = (\mathbf{y}'_{t+n}, \mathbf{x}'_{t+n})'$ to a unit change in ε_{it}, measured by one standard deviation, namely $\sqrt{\omega_{ii}}$. At horizon n the GIRF is defined by the point forecast of $\mathbf{z}_{t+n}$ conditional on the information $\mathfrak{I}_{t-1}$ and the one standard error shock of the ith structural error, ε_{it}, relative to the baseline conditional forecasts. Namely,

$$\mathfrak{g}(n, \mathbf{z} : \varepsilon_i) = E(\mathbf{z}_{t+n} \mid \varepsilon_{it} = \sqrt{\omega_{ii}}, \mathfrak{I}_{t-1}) - E(\mathbf{z}_{t+n} \mid \mathfrak{I}_{t-1}).$$

Clearly, since the $\mathbf{x}_t$ are assumed to be strictly exogenous, the effects of shocking ε_{it} on $\mathbf{x}_{t+h}$ will be zero, *i.e.* $\mathfrak{g}(n, \mathbf{x} : \varepsilon_i) = \mathbf{0}$ for all n and i.[5] Since the ε_{it} are serially uncorrelated then their impulse response functions are non-zero only at horizon zero when $\mathfrak{g}(n, \boldsymbol{\varepsilon} : \varepsilon_i) = E(\boldsymbol{\varepsilon}_t \mid \varepsilon_{it} = \sqrt{\omega_{ii}})$ for $n = 0$, but for all other horizons $n > 0$ we have $\mathfrak{g}(n, \boldsymbol{\varepsilon} : \varepsilon_i) = 0$.

If the structural errors are correlated, a shock to one error will be associated with changes in the other errors. As shown by Koop *et al.* (1996), in the Gaussian case where $\boldsymbol{\varepsilon}_t \backsim i.i.d.N(\mathbf{0}, \boldsymbol{\Omega})$, $(\boldsymbol{\varepsilon}_t, \varepsilon_{it})$ are also normally distributed

$$\begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \varepsilon_{it} \end{pmatrix} \sim i.i.d.N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Omega} & \mathrm{Cov}(\boldsymbol{\varepsilon}_t, \varepsilon_{it}) \\ \mathrm{Cov}(\varepsilon_{it}, \boldsymbol{\varepsilon}_t) & V(\varepsilon_{it}) \end{pmatrix}\right],$$

[5] This would not of course be the case if $\mathbf{x}_t$ was only weakly exogenous.

then noting that $V(\varepsilon_{it}) = \omega_{ii}$

$$\begin{aligned} E(\boldsymbol{\varepsilon}_t | \varepsilon_{it} = \sqrt{\omega_{ii}}) &= E(\boldsymbol{\varepsilon}_t) + \frac{\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \varepsilon_{it})}{V(\varepsilon_{it})}(\sqrt{\omega_{ii}} - 0) \\ &= \frac{1}{\sqrt{\omega_{ii}}}\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \varepsilon_{it}) \\ &= \frac{1}{\sqrt{\omega_{ii}}}\begin{pmatrix} \omega_{1i} \\ \omega_{2i} \\ \vdots \\ \omega_{m_y i} \end{pmatrix}. \end{aligned} \quad (6.9)$$

which can be written more compactly as

$$E(\boldsymbol{\varepsilon}_t | \varepsilon_{it} = \sqrt{\omega_{ii}}) = \left(\frac{1}{\sqrt{\omega_{ii}}}\right)\boldsymbol{\Omega}\mathbf{e}_i,$$

where $\mathbf{e}_i$ is an $m_y \times 1$ selection vector of zeros except for its ith element which is set to unity.[6] This gives the predicted shocks in each structural error given a shock to ε_{it}, based on the typical correlation observed historically between the structural errors. In the special case where the structural errors are orthogonal, the shock only changes the ith error and we have

$$E\left(\boldsymbol{\varepsilon}_t \mid \varepsilon_{it} = \sqrt{\omega_{ii}}\right) = \sqrt{\omega_{ii}}\mathbf{e}_i.$$

Application of the generalised impulse response analysis to the VARX specification, (6.1), now yields

$$\mathbf{A}\mathfrak{g}(n, \mathbf{y} : \varepsilon_i) = \mathbf{A}_1\mathfrak{g}(n-1, \mathbf{y} : \varepsilon_i) + \cdots + \mathbf{A}_p\mathfrak{g}(n-p, \mathbf{y} : \varepsilon_i) + \mathfrak{g}(n, \boldsymbol{\varepsilon} : \varepsilon_i),$$

for $n = 0, 1, 2, \ldots$, with the initial values $\mathfrak{g}(n, \mathbf{y} : \varepsilon_i) = 0$ for $n < 0$ and as we saw above the last term is non-zero only for $n = 0$. The identification of $\mathfrak{g}(n, \mathbf{y} : \varepsilon_i)$ requires the identification of the structural coefficients $\mathbf{A}$ and $\mathbf{A}_i$, $i = 1, \ldots, p$, and the covariance matrix $\boldsymbol{\Omega}$. It is also possible to identify $\mathfrak{g}(n, \mathbf{y} : \varepsilon_i)$ by a mixture of identification restrictions on $\mathbf{A}$ and $\boldsymbol{\Omega}$. To see this we premultiply both sides of the above relationship by $\mathbf{A}^{-1}$ and obtain

$$\begin{aligned} \mathfrak{g}(n, \mathbf{y} : \varepsilon_i) = \boldsymbol{\Phi}_1\mathfrak{g}(n-1, \mathbf{y} : \varepsilon_i) + \cdots + \boldsymbol{\Phi}_p\mathfrak{g}(n-p, \mathbf{y} : \varepsilon_i) \\ + \mathbf{A}^{-1}\mathfrak{g}(n, \boldsymbol{\varepsilon} : \varepsilon_i), \end{aligned} \quad (6.10)$$

[6] This result also holds in non-Gaussian but linear settings where the conditional expectations $E(\boldsymbol{\varepsilon}_t \mid \varepsilon_{it} = \sqrt{\omega_{ii}})$ can be assumed to be linear.

where as before $\boldsymbol{\Phi}_i = \mathbf{A}^{-1}\mathbf{A}_i$ $i = 1, 2, \ldots, p$, and the last term is non-zero only for $n = 0$. The $\boldsymbol{\Phi}_i$ can be estimated from the reduced form, thus the indeterminacy is confined to the contemporaneous interaction of the structural errors through the expression $\mathbf{A}^{-1}g\,(0, \boldsymbol{\varepsilon} : \varepsilon_i)$, and is resolved up to a scalar multiplication if $\mathbf{A}^{-1}\boldsymbol{\Omega}$ can be estimated consistently. However, to identify (or consistently estimate) $\mathbf{A}^{-1}\boldsymbol{\Omega}$ involves the imposition of m_y^2 *a priori* restrictions on the elements of $\mathbf{A}$ and/or $\boldsymbol{\Omega}$. Evidently, the identification of the structural impulse responses does not require $\mathbf{A}$ and $\boldsymbol{\Omega}$ to be separately identified, and it is possible to trade off restrictions across $\mathbf{A}$ and $\boldsymbol{\Omega}$. But in cases where there are no *a priori* grounds for restricting $\boldsymbol{\Omega}$, since $\mathbf{A}^{-1}\boldsymbol{\Omega}\mathbf{A}'^{-1} = \Sigma$, then $\mathbf{A}^{-1}\boldsymbol{\Omega} = \boldsymbol{\Sigma}\mathbf{A}'$, and the identification of the impulse responses with respect to structural errors requires complete knowledge of the contemporaneous effects, $\mathbf{A}$.

ORTHOGONALISED IMPULSE RESPONSES

The standard approach to deriving impulse response functions is to start from the moving average representations of the final form, (6.6). The reduced form disturbances are correlated and the covariance matrix of $\mathbf{u}_t$, which can be consistently estimated, is given by $\boldsymbol{\Sigma} = \mathbf{A}^{-1}\boldsymbol{\Omega}\mathbf{A}'^{-1}$. Orthogonalised impulse response function advanced by Sims (1980) makes use of the Choleski decomposition of $\boldsymbol{\Sigma} = \mathbf{P}\mathbf{P}'$, where $\mathbf{P}$ is a lower triangular matrix. This can be used to create a new sequence of errors, $\mathbf{u}_t^* = \mathbf{P}^{-1}\mathbf{u}_t$, $t = 1, 2, \ldots, T$, which are orthogonal to each other contemporaneously with unit standard errors, namely $E\left(\mathbf{u}_t^*\mathbf{u}_t^{*\prime}\right) = \mathbf{I}_{m_y}$. Thus the effect of a shock to one of these orthogonalised errors, $\mathbf{u}_t^* = \left(u_{1t}^*, u_{2t}^*, \ldots, u_{m_yt}^*\right)'$, say u_{1t}^*, on the remaining shocks is unambiguous, because it is not correlated with the other orthogonalised errors. The impulse response analysis is also often supplemented by the forecast error variance decomposition where the error variance of forecasting the ith variable n periods ahead is decomposed into the components accounted for by innovations in different variables in the VAR.

There are two problems with orthogonalised impulse response functions and the forecast error variance decomposition. *First*, the impulse responses obtained refer to the effects on the endogenous variables, y_{it}, of a unit displacement (measured by one standard error) in the orthogonalised error, u_{jt}^*, and not in the structural or even the reduced form errors, ε_{jt} and u_{jt}. *Second*, notice that the choice of $\mathbf{P}$ is unique only for a particular ordering of the variables in the VAR. Unless $\boldsymbol{\Sigma}$ is diagonal, or close to diagonal, different orderings of the variables will give different estimates of the impulse response functions. In fact, the particular ordering of the variables in the VAR and the Choleski decomposition procedure used constitute an implicit identification assumption, equivalent to the recursive identifying restrictions discussed in Section 3.2.3. Orthogonalised impulse response functions, therefore, actually employ traditional identification assumptions, typically motivated by what we termed 'tentative' theory on contemporaneous relations. Other identification schemes based on similarly tentative theory were discussed in Section 3.2.3 in the context of the Structural VAR models. The interpretation of the impulse responses obtained on the basis of these is only as robust as the underlying identifying assumptions and, as the discussion of Section 3.2.3 showed, our view is that economic theory only rarely provides justification for robust short-run identifying restrictions (although it is more capable of providing justification for identifying restrictions on the long-run coefficients).

When plausible *a priori* information to identify the effects of structural shocks is not available, it would still be of some interest to examine the effect of shocks to the reduced form errors, $\mathbf{u}_t = \mathbf{A}^{-1}\boldsymbol{\varepsilon}_t$. The generalised impulse response function provides a natural way to do this since it measures the effect on the endogenous variables of a typical shock to the system, based on the estimated covariances of the reduced form shocks computed using the historical data. Recall from (6.9) that the generalised impulse responses of $\mathbf{y}_{t+n}$ with respect to u_{it} (the ith element of $\mathbf{u}_t$) are given by

$$g\,(n, \mathbf{y} : u_i) = \boldsymbol{\Phi}_1 g\,(n-1, \mathbf{y} : u_i) + \cdots + \boldsymbol{\Phi}_p g\,(n-p, \mathbf{y} : u_i) + g\,(n, \mathbf{u} : u_i)\,, \qquad (6.11)$$

where the last term is non-zero only for $n = 0$, when it is

$$g(n, \mathbf{u} : u_i) = \left(\frac{1}{\sqrt{\sigma_{ii}}}\right)\boldsymbol{\Sigma}\mathbf{e}_i \text{ for } n = 0. \qquad (6.12)$$

These impulse responses can be uniquely estimated from the parameters of the reduced form and unlike the orthogonalised impulse responses are invariant to the ordering of the variables in the VAR. One can also construct a comparable forecast error variance decomposition.

In the case of stationary variables the generalised impulse response function, as defined by (6.10) or (6.11), will tend to zero as n tends to infinity.

In the case of $I(1)$ variables it will tend to a non-zero constant as n goes to infinity. When the variables are $I(1)$ and cointegrated, there will be linear combinations of the generalised impulse response function that tend to zero and we discuss this further below.[7]

Note that an alternative methodology used to investigate the dynamic properties of large-scale systems, often employed by macromodellers, is to consider the effect of a displacement in the intercept of one of the model's equations. This is equivalent to shocking the innovation in the equation and implicitly assumes that changes in one equation's intercept has no effect on the intercepts of the other equations in the system. Of course, this is one possible counter-factual exercise that might be of interest. But in interrelated systems, it is not likely that one could change the parameters of one part of the system without initiating changes elsewhere. The interpretation of dynamics based on innovations of the type captured by generalised impulse responses is, in our opinion, a much more plausible type of counter-factual than the *ad hoc* once-and-for-all changes in parameter values considered by many macromodellers.

PERSISTENCE PROFILES

The above impulse responses consider the effect of a shock to a *particular* exogenous variable, x_{it}, or an error term, ε_{it} or u_{it}. An alternative approach, developed in Lee and Pesaran (1993), would be to consider the effect of system-wide shocks at time t on the evolution of the system at time $t+n$. Under this approach, the generalised impulse responses are derived with respect to the whole vector of shocks, $\boldsymbol{\varepsilon}_t$ or $\mathbf{u}_t$, and viewed as random variables. The probability distribution function of these random variables is then examined as a function of n. In the case where $\boldsymbol{\varepsilon}_t$ (or $\mathbf{u}_t$) are Gaussian, the generalised impulse responses with respect to the system-wide shocks are also Gaussian with a zero mean and the covariance matrix $\boldsymbol{\Theta}_n\boldsymbol{\Sigma}\boldsymbol{\Theta}_n'$ (see (6.8)). The diagonal elements of $\boldsymbol{\Theta}_n\boldsymbol{\Sigma}\boldsymbol{\Theta}_n'$ (appropriately scaled) are called the persistence profiles by Lee and Pesaran (1993). It is easily seen that the same persistence profiles are obtained for the structural as well as the reduced form errors. For a stationary VAR, the persistence profiles tend to zero as $n \to \infty$. For VARs with unit roots, the persistence profiles tend to the spectral density function (apart from a scalar constant) of $\Delta\mathbf{y}_t$ at zero frequency.

[7] The relationships between the generalised impulse response functions and the orthogonalised impulse responses are discussed in Pesaran and Shin (1998).

6.2 Cointegrating VAR models

Much of the econometric analysis of cointegration has been done in the context of a VAR(p), where all the variables are regarded as endogenous. Initially, we follow the literature and assume that the VAR model *only* contains endogenous $I(1)$ variables and linear deterministic trends. Setting $\mathbf{B}_i = \mathbf{0}$ in (6.1), we have:

$$\mathbf{y}_t = \boldsymbol{\Phi}_1\mathbf{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p\mathbf{y}_{t-p} + \mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{u}_t, \qquad (6.13)$$

where $\mathbf{a}_0$ and $\mathbf{a}_1$ are $m \times 1$ vectors of unknown coefficients.[8] To cover the unit root case we allow for the roots of

$$\left|\mathbf{I}_m - \boldsymbol{\Phi}_1\lambda - \boldsymbol{\Phi}_2\lambda^2 - \cdots - \boldsymbol{\Phi}_p\lambda^p\right| = 0, \qquad (6.14)$$

to fall on and/or outside the unit circle, but rule out the possibility that one or more elements of $\mathbf{y}_t$ be $I(2)$.[9] We shall return to the case where the model also contains exogenous $I(1)$ variables below. The model can be re-parameterised as a Vector Error Correction Model (VECM)

$$\Delta\mathbf{y}_t = -\boldsymbol{\Pi}\mathbf{y}_{t-1} + \sum_{i=1}^{p-1}\boldsymbol{\Gamma}_i\Delta\mathbf{y}_{t-i} + \mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{u}_t, \qquad (6.15)$$

where

$$\boldsymbol{\Pi} = \mathbf{I}_m - \sum_{i=1}^{p}\boldsymbol{\Phi}_i, \quad \boldsymbol{\Gamma}_i = -\sum_{j=i+1}^{p}\boldsymbol{\Phi}_j, \; i = 1, \ldots, p-1. \qquad (6.16)$$

If the elements of $\mathbf{y}_t$ were $I(0)$, $\boldsymbol{\Pi}$ will be a full rank $m \times m$ matrix. If the elements of $\mathbf{y}_t$ are $I(1)$ and not cointegrated then it must be that $\boldsymbol{\Pi} = \mathbf{0}$ and a VAR model in first differences will be appropriate. If the elements of $\mathbf{y}_t$ are $I(1)$ and cointegrated with rank($\boldsymbol{\Pi}$) $= r$, then $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $m \times r$ full column rank matrices, and there will be $r < m$ linear combinations of $\mathbf{y}_t$, the cointegrating relations, $\boldsymbol{\xi}_t = \boldsymbol{\beta}'\mathbf{y}_t$, which are $I(0)$. The variables $\boldsymbol{\xi}_t$ are often interpreted as the deviations from equilibrium, an interpretation that is at the heart of the long-run structural modelling strategy elaborated in Section 3.1.3.

Under cointegration, (6.15) can be written as:

$$\Delta\mathbf{y}_t = -\boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \sum_{i=1}^{p-1}\boldsymbol{\Gamma}_i\Delta\mathbf{y}_{t-i} + \mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{u}_t, \qquad (6.17)$$

[8] To simplify the notations in this section we denote the dimension of $\mathbf{y}_t$ by $m_y = m$.
[9] A review of the econometric analysis of $I(2)$ variables is provided in Haldrup (1998).

where α is the matrix of adjustment or feedback coefficients, which measure how strongly the deviations from equilibrium, the r stationary variables $\beta'\mathbf{y}_{t-1}$, feedback onto the system. If there are $0 < r < m$ cointegrating vectors, then some of the elements of α must be non-zero, *i.e.* there must be some Granger causality involving the levels of the variables in the system to keep the elements of $\mathbf{y}_t$ from diverging.

The unrestricted estimate of Π can be obtained using (6.15). In the restricted model, (6.17), which accommodates $r < m$ cointegrating vectors, we need to estimate the two $m \times r$ coefficient matrices, α and β. This rank reduction therefore imposes $m^2 - 2mr$ restrictions to be imposed on Π. Further, as noted in Section 3.1.1, α and β are not separately identified without some additional restrictions since, for any non-singular matrix Q, we have $\Pi = \alpha QQ^{-1}\beta'$, and the new coefficient matrices $\alpha^* = \alpha Q$ and $\beta^{*\prime} = Q^{-1}\beta'$ would be *observationally equivalent* to using α and β' respectively. Put differently, any linear combination of the $I(0)$ variables, $\xi_t = \beta'\mathbf{y}_t$, are also $I(0)$ variables. To avoid this indeterminacy, we require r independent restrictions on each of the r cointegrating relations, a total of r^2 further restrictions (r of which are provided by normalisation conditions). Thus in the restricted model, we impose $(m^2 - 2mr) + r^2 = (m-r)^2$, namely m^2-2mr restrictions imposed by the rank restrictions on Π, and r^2 exact identifying restrictions.

6.2.1 *Treatment of the deterministic components*

If there are unrestricted linear trends in the unrestricted VAR, in general there will be quadratic trends in the level of the variables when the model contains unit roots. To avoid quadratic trends, the linear trend coefficients must be restricted. As shown, for example, in Pesaran, Shin and Smith (2000), using (6.13), $\Delta\mathbf{y}_t$ can be represented by an infinite moving average[10]

$$\Delta\mathbf{y}_t = \mathbf{C}(L)\left(\mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{u}_t\right), \tag{6.18}$$

where

$$\mathbf{C}(L) = \sum_{j=0}^{\infty} \mathbf{C}_j L^j = \mathbf{C}(1) + (1-L)\mathbf{C}^*(L), \tag{6.19}$$

[10] This 'first-difference MA representation' was originally given in Engle and Granger (1987) for VAR models without linear trends.

$$\mathbf{C}^*(L) = \sum_{j=0}^{\infty} \mathbf{C}_j^* L^j,\ \mathbf{C}_j^* = \mathbf{C}_{j-1}^* + \mathbf{C}_j, \text{ or } \mathbf{C}_j^* = -\sum_{i=j+1}^{\infty} \mathbf{C}_i. \tag{6.20}$$

Consider now the relationship between $\mathbf{C}_i$ and $\Phi_1, \Phi_2, \ldots, \Phi_p$, the parameter matrices of the underlying VAR specification in (6.13), and note that since $\mathbf{C}(L)$ is invertible, we must also have[11]

$$\mathbf{C}^{-1}(L)\Delta\mathbf{y}_t = \Phi(L)\mathbf{y}_t,$$

where as before

$$\Phi(L) = \mathbf{I}_m - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p.$$

Hence, we also have

$$\left[\mathbf{C}^{-1}(L)(1-L) - \Phi(L)\right]\mathbf{y}_t = \mathbf{0},$$

or

$$(1-L)\,\mathbf{I}_m = \Phi(L)\mathbf{C}(L) = \mathbf{C}(L)\Phi(L).$$

Therefore

$$\mathbf{C}_i = \Phi_1\mathbf{C}_{i-1} + \Phi_2\mathbf{C}_{i-2} + \cdots + \Phi_p\mathbf{C}_{i-p}, \quad \text{for } i = 2, 3, \ldots, \tag{6.21}$$

where $\mathbf{C}_0 = \mathbf{I}_m$, $\mathbf{C}_1 = \Phi_1 - \mathbf{I}_m$, and $\mathbf{C}_i = \mathbf{0}$, for $i < 0$.

Using (6.19) in (6.18), it is now easily seen that

$$\Delta\mathbf{y}_t = \mathbf{b}_0 + \mathbf{b}_1 t + \mathbf{C}(1)\mathbf{u}_t + \mathbf{C}^*(L)\Delta\mathbf{u}_t, \tag{6.22}$$

where

$$\mathbf{b}_0 = \mathbf{C}(1)\mathbf{a}_0 + \mathbf{C}^*(1)\mathbf{a}_1, \quad \mathbf{b}_1 = \mathbf{C}(1)\mathbf{a}_1.$$

Cumulating (6.22) forward, we obtain the 'level MA representation'

$$\mathbf{y}_t = \mathbf{y}_0 + \mathbf{b}_0 t + \mathbf{b}_1\frac{t(t+1)}{2} + \mathbf{C}(1)\mathbf{s}_t + \mathbf{C}^*(L)(\mathbf{u}_t - \mathbf{u}_0),$$

where $\mathbf{s}_t$ denotes the partial sum $\mathbf{s}_t = \sum_{j=1}^{t}\mathbf{u}_j$, $t = 1, 2, \ldots,$ and $\text{rank}\,[\mathbf{C}(1)] = m - r$. It is immediately seen that, since $\mathbf{b}_1 = \mathbf{C}(1)\mathbf{a}_1$, in general $\mathbf{y}_t$ will contain m different linear deterministic trends, $\mathbf{b}_0 t$, $m - r$ different (independent) deterministic quadratic trends given by $\frac{1}{2}t(t+1)\mathbf{C}(1)\mathbf{a}_1$, $m - r$ unit root (or permanent) components given by

[11] Recall that by assumption $\Delta\mathbf{y}_t$ is covariance stationary.

$C(1)s_t$, and m stationary components given by $C^*(L)(u_t - u_0)$.[12] With a_1 unrestricted, the quadratic trend term disappears only in the full rank stationary case where there are no unit roots, namely if rank$(\Pi) = m$.

To remove the quadratic trends and ensure that the trend in the deterministic part of y_t is linear for all values of r, we need to restrict the trend coefficients such that

$$a_1 = \Pi\gamma, \tag{6.23}$$

where γ is an arbitrary $m \times 1$ vector of fixed constants. Note that γ is unrestricted only if Π is full rank. In this case $\gamma = \Pi^{-1}a_1$. But where Π is rank deficient, all elements of γ can be estimated from the reduced form coefficients. In this case the reduced form trend coefficients are restricted.

For the above choice of a_1, it is easily seen that $b_1 = C(1)\Pi\gamma = 0$.[13] Under this restriction on the trend coefficients, we have

$$y_t = y_0 + b_0 t + C(1)s_t + C^*(L)(u_t - u_0), \tag{6.24}$$

and its associated vector error correction formulation is given by

$$\begin{aligned}\Delta y_t &= -\alpha\beta'(y_{t-1} - \gamma t) + \sum_{i=1}^{p-1}\Gamma_i\Delta y_{t-i} + a_0 + u_t \\ &= -\Pi_* y^*_{t-1} + \sum_{i=1}^{p-1}\Gamma_i\Delta y_{t-i} + a_0 + u_t,\end{aligned} \tag{6.25}$$

where $\Pi_* = \alpha\beta'_*$, $\beta'_* = (\beta', -\beta'\gamma)$, $y^*_{t-1} = (y'_{t-1}, t)'$, and the deterministic trend is now specified to be a part of the cointegrating relations, $\beta'(y_{t-1} - \gamma t) = \beta'_* y^*_{t-1}$. This ensures that the y_t contains only linear and not quadratic deterministic trends. This result also shows that in general the cointegration relations could contain linear trends if y_t is trended. In the absence of a time trend term in the cointegrating relations we must have $\beta'\gamma = 0$. These provide r further restrictions, known as 'co-trending' restrictions which are testable.

A similar conclusion also follows from the 'level MA representation', (6.24). Premultiplying both sides of (6.24) by β' we have

$$\beta' y_t = \beta' y_0 + (\beta' b_0)t + \beta' C(1)s_t + \beta' C^*(L)(u_t - u_0).$$

12 This decomposition of the stochastic part of y_t into permanent and transitory components is not unique and raises a number of identification problems discussed by Levtchenkova *et al.* (1998).

13 Notice from (6.13) and (6.18) that since $C(L)\Phi(L) = (1 - L)I_m$, then $C(1)\Phi(1) = C(1)\Pi = 0$.

But $\beta' C(1) = 0$, and it is also easily established that

$$\beta' b_0 = \beta' C(1)a_0 + \beta' C^*(1)a_1 = \beta' C^*(1)\Pi\gamma = \beta'\gamma.$$

Hence

$$\beta' y_t = \beta' y_0 + (\beta'\gamma)t + \beta' C^*(L)(u_t - u_0), \tag{6.26}$$

and in the case of VAR models with linear trends, the cointegrating relations will generally contain deterministic trends, unless the co-trending restrictions $\beta'\gamma = 0$ are imposed. The hypothesis of co-trending is particularly relevant in the output convergence literature where 'convergence' involves both cointegration and co-trending. See, for example, Pesaran (2004a) for a pairwise approach to testing for output and growth convergence.

So far we have focused on cointegrating VAR models with linear deterministic trends. A similar consideration also applies to cointegrating VAR models that contain intercepts only. Once again to ensure that the level variables do not contain different numbers of independent linear deterministic trends as the cointegrating rank changes, the intercepts in these models must be restricted accordingly. It is also possible that different elements of y_t may have different trend characteristics. For example, output and interest rates are often included in the same VAR, while it is clear that these variables have different trend characteristics. Although there are a large number of possible treatments of the deterministic elements, it will be convenient to distinguish between five different cases often encountered in practice:

- **Case I:** (No intercepts; no trends.) $a_0 = 0$ and $a_1 = 0$. Hence, the VECM (6.17) becomes

$$\Delta y_t = -\Pi y_{t-1} + \sum_{i=1}^{p-1}\Gamma_i\Delta y_{t-i} + u_t. \tag{6.27}$$

- **Case II:** (Restricted intercepts; no trends.) $a_0 = \Pi\mu$ and $a_1 = 0$. The VECM (6.17) is

$$\Delta y_t = \Pi\mu - \Pi y_{t-1} + \sum_{i=1}^{p-1}\Gamma_i\Delta y_{t-i} + u_t. \tag{6.28}$$

- **Case III:** (Unrestricted intercepts; no trends.) $a_0 \neq 0$ and $a_1 = 0$. In this case, the intercept restriction $a_0 = \Pi\mu$ is ignored and the structural

VECM estimated is

$$\Delta \mathbf{y}_t = \mathbf{a}_0 - \boldsymbol{\Pi}\mathbf{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \mathbf{y}_{t-i} + \mathbf{u}_t. \tag{6.29}$$

- **Case IV:** (Unrestricted intercepts; restricted trends.) $\mathbf{a}_0 \neq \mathbf{0}$ and $\mathbf{a}_1 = \boldsymbol{\Pi}\boldsymbol{\gamma}$. Thus

$$\Delta \mathbf{y}_t = \mathbf{a}_0 + (\boldsymbol{\Pi}\boldsymbol{\gamma})t - \boldsymbol{\Pi}\mathbf{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \mathbf{y}_{t-i} + \mathbf{u}_t. \tag{6.30}$$

- **Case V:** (Unrestricted intercepts; unrestricted trends.) $\mathbf{a}_0 \neq \mathbf{0}$ and $\mathbf{a}_1 \neq \mathbf{0}$. Here, the VECM estimated is

$$\Delta \mathbf{y}_t = \mathbf{a}_0 + \mathbf{a}_1 t - \boldsymbol{\Pi}\mathbf{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \mathbf{y}_{t-i} + \mathbf{u}_t. \tag{6.31}$$

It should be emphasised that the DGPs for Cases II and III are identical as are those for Cases IV and V. However, as in the test for a unit root proposed by Dickey and Fuller (1979) compared with that of Dickey and Fuller (1981) for univariate models, estimation and hypothesis testing in Cases III and V proceed ignoring the constraints linking, respectively, the intercept and trend coefficient vectors, $\mathbf{a}_0$ and $\mathbf{a}_1$, to the parameter matrix $\boldsymbol{\Pi}$ whereas Cases II and IV fully incorporate these restrictions. As argued in Pesaran, Shin and Smith (2000), Cases II and IV are likely to be particularly relevant in practice and are preferable to the corresponding unrestricted Cases III and V.

6.2.2 *Trace and maximum eigenvalue tests of cointegration*

If the sole purpose of the cointegration analysis is simply to test for cointegration (or select the appropriate number of cointegrating relations), the nature of the r^2 restrictions employed to ensure there are r identified cointegrating relations is not important since the maximised log-likelihood values will be invariant to how the long-run relations are exactly identified. This was shown by Johansen (1988, 1991) who established an algorithm for maximising the likelihood of (6.25) subject to the constraint that $\boldsymbol{\Pi}_* = \boldsymbol{\alpha}\boldsymbol{\beta}'_*$ and under the assumption that the disturbances are Gaussian. The algorithm involves two steps. In the first, $\Delta \mathbf{y}_t$ and $\mathbf{y}^*_{t-1} = (\mathbf{y}'_{t-1}, \mathbf{t})'$ are regressed in turn on $\Delta \mathbf{y}_{t-1}, \Delta \mathbf{y}_{t-2}, \ldots, \Delta \mathbf{y}_{t-p+1}$ and 1 to generate residuals $\mathbf{r}_{0t}$ and $\mathbf{r}_{1t}$, respectively. Then, defining

$$\mathbf{S}_{00} = T^{-1}\sum_{t=1}^{T} \mathbf{r}_{0t}\mathbf{r}'_{0t}, \ \mathbf{S}_{01} = T^{-1}\sum_{t=1}^{T} \mathbf{r}_{0t}\mathbf{r}'_{1t}, \ \mathbf{S}_{11} = T^{-1}\sum_{t=1}^{T} \mathbf{r}_{1t}\mathbf{r}'_{1t}, \tag{6.32}$$

the m ordered eigenvalues of $\mathbf{S}_{11}^{-1}\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}$ namely $\lambda_1 > \lambda_2 > \cdots > \lambda_m$, are computed. The maximum value of the log-likelihood function subject to the constraint that there are r cointegrating relations is given by

$$\ell_T(\widehat{\boldsymbol{\beta}}_T) = -\frac{Tm}{2}\left(1 + \log(2\pi)\right) - \frac{T}{2}\log|\mathbf{S}_{00}| - \frac{T}{2}\sum_{i=1}^{r}\log\left(1 - \lambda_i\right), \tag{6.33}$$

where $\widehat{\boldsymbol{\beta}}_T$ is the ML estimate of the $m \times r$ cointegrating coefficient matrix (see also the discussion below). This expression can be calculated irrespective of the form of the r^2 independent restrictions on the cointegrating relations. In fact it is easily established that $\ell_T\left(\widehat{\boldsymbol{\beta}}'_T\right) = \ell_T\left(\mathbf{Q}\widehat{\boldsymbol{\beta}}'_T\right)$, for any choice of a $r \times r$ non-singular matrix, $\mathbf{Q}$.

If the applied econometrician is simply interested in testing the null hypothesis of r cointegrating relations in (6.15):

$$H_0 : \text{Rank}(\boldsymbol{\Pi}) = r, \tag{6.34}$$

there are two types of the log-likelihood ratio statistics that can be used for this purpose. The 'trace' statistic is intended for testing the null hypothesis (6.34) against the full rank hypothesis,

$$H_{1a} : \text{Rank}(\boldsymbol{\Pi}) = m, \tag{6.35}$$

and the 'maximum eigenvalue' statistic is intended for testing the null against

$$H_{1b} : \text{Rank}(\boldsymbol{\Pi}) = r + 1. \tag{6.36}$$

These statistics are computed as

$$\lambda_{\text{trace}} = -T\sum_{i=r+1}^{m}\ln(1 - \lambda_i), \tag{6.37}$$

$$\lambda_{\max} = -T\ln(1 - \lambda_{r+1}). \tag{6.38}$$

Given the presence of unit roots, the asymptotic distributions of both statistics are non-standard (and depend on the nature of the deterministic processes involved), but Johansen (1991) provided the appropriate critical values based on Monte Carlo simulations, and Pesaran, Shin and Smith (2000) provided the corresponding statistics under Cases I–V above.

6.2.3 *Identifying long-run relationships in a cointegrating VAR*

Typically, the applied econometrician will be interested not only in the number of cointegrating relations that might exist among the variables but also the specification of the identifying (and possibly over-identifying) restrictions on the cointegrating relations. Indeed, Johansen (1988, 1991) have provided procedures for estimating $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, using 'statistical' over-identifying restrictions. He computed the ML estimates of $\boldsymbol{\beta}$ as the r eigenvectors corresponding to the first r largest eigenvalues of the canonical correlation matrix, $\mathbf{S}_{11}^{-1}\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}$, where $\mathbf{S}_{00}$, $\mathbf{S}_{01}$, and $\mathbf{S}_{11}$ are defined in (6.32). These are often referred to as 'empirical' or 'statistical' identifying restrictions, and together impose the r^2 restrictions needed for exact identification of $\boldsymbol{\beta}$. However, while mathematically natural given the statistical structure of the problem, these restrictions have no economic meaning since there is no reason to expect economic cointegrating vectors to be orthogonal. When $r > 1$, the economic interpretation of the Johansen estimates of the cointegrating vectors is almost impossible. See also Phillips (1991) for an alternative non-economic identification adopting a triangular structure.

The more satisfactory approach promoted in Pesaran and Shin (2002) is to estimate the cointegrating relations under a general set of structural long-run restrictions provided by *a priori* economic theory. Suppose that we are considering an example of a model with unrestricted intercepts and restricted trends (Case IV), and the cointegrating vectors, $\boldsymbol{\beta}_*$, are subject to the following k general linear restrictions, including cross-equation restrictions:[14]

$$\mathbf{R}\,\text{vec}(\boldsymbol{\beta}_*) = \mathbf{f}, \tag{6.39}$$

where $\mathbf{R}$ and $\mathbf{f}$ are a $k \times (m+1)r$ matrix of full row rank and a $k \times 1$ vector of known constants, respectively, and $\text{vec}(\boldsymbol{\beta}_*)$ is the $(m+1)r \times 1$ vector of long-run coefficients, which stacks the r columns of $\boldsymbol{\beta}_*$ into a vector. Three cases can be distinguished: (i) $k < r^2$: the under-identified case; (ii) $k = r^2$: the exactly identified case; and (iii) $k > r^2$: the over-identified case.

ESTIMATION OF THE LONG-RUN COINTEGRATING VECTORS SUBJECT TO IDENTIFYING RESTRICTIONS

Following Pesaran and Shin (2002), we will describe the ML estimation of the long-run parameters $\boldsymbol{\beta}$ of the *VEC* model (6.25) subject to the k identifying restrictions on $\boldsymbol{\beta}$ given by (6.39). We first note that the concentrated

[14] Pesaran and Shin (2002) also consider the more general case where the restrictions on the cointegrating coefficients may be non-linear.

log-likelihood function for the cointegrated model is given by:[15]

$$\ell_T(\boldsymbol{\beta}) = \text{constant} - \frac{T}{2}\left\{\ln|\boldsymbol{\beta}'\mathbf{A}_T\boldsymbol{\beta}| - \ln|\boldsymbol{\beta}'\mathbf{B}_T\boldsymbol{\beta}|\right\}. \tag{6.40}$$

where

$$\mathbf{A}_T = \mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}, \quad \mathbf{B}_T = \mathbf{S}_{11}, \tag{6.41}$$

and $\mathbf{S}_{00}$, $\mathbf{S}_{10}$, $\mathbf{S}_{11}$ are defined in (6.32). Then, the ML estimator of $\boldsymbol{\theta} = \text{vec}(\boldsymbol{\beta})$ is obtained by solving

$$\max_{\boldsymbol{\theta},\boldsymbol{\lambda}} \Lambda(\boldsymbol{\theta},\boldsymbol{\lambda}),$$

where $\Lambda(\boldsymbol{\theta},\boldsymbol{\lambda})$ is the Lagrangian function for this constrained ML estimation problem and given by

$$\begin{aligned}\Lambda(\boldsymbol{\theta},\boldsymbol{\lambda}) &= \frac{1}{T}\ell_T(\boldsymbol{\theta}) - \frac{1}{2}\boldsymbol{\lambda}'\mathbf{h}(\boldsymbol{\theta}) \\ &= \text{constant} - \frac{1}{2}\left\{\ln\left|\boldsymbol{\beta}'\mathbf{A}_T\boldsymbol{\beta}\right| - \ln\left|\boldsymbol{\beta}'\mathbf{B}_T\boldsymbol{\beta}\right| - \boldsymbol{\lambda}'(\mathbf{R}\boldsymbol{\theta} - \mathbf{f})\right\},\end{aligned} \tag{6.42}$$

where $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{R}\boldsymbol{\theta} - \mathbf{f}$ and $\boldsymbol{\lambda}$ is a $k \times 1$ vector of Lagrange multipliers.

We distinguish between two cases: when the cointegrating vectors are exactly identified ($k = r^2$), and when they are subject to over-identifying restrictions ($k > r^2$). In both cases it is convenient to start with the exactly identified ML estimator of $\boldsymbol{\beta}$ obtained by Johansen's eigenvalue routine, *i.e.* the r eigenvectors corresponding to the first r largest eigenvalues of $\mathbf{S}_{11}^{-1}\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}$, which we denote by $\hat{\boldsymbol{\beta}}_J$.

Exactly identified case ($k = r^2$)

In the exactly identified case, the ML estimator of $\boldsymbol{\beta}$ is obtained simply by

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{I}_r \otimes \hat{\boldsymbol{\beta}}_J\right)\left[\mathbf{R}\left(\mathbf{I}_r \otimes \hat{\boldsymbol{\beta}}_J\right)\right]^{-1}\mathbf{f}, \tag{6.43}$$

where $\mathbf{I}_r$ is an $r \times r$ identity matrix and $\mathbf{R}$ and $\mathbf{f}$ are defined by (6.39). Note that by construction $\hat{\boldsymbol{\beta}}_J'\mathbf{B}_T\hat{\boldsymbol{\beta}}_J = \mathbf{I}_r$ and $\hat{\boldsymbol{\beta}}_{Ji}'(\mathbf{B}_T - \mathbf{A}_T)\hat{\boldsymbol{\beta}}_{Jj} = 0$ for $i \neq j$, $i,j = 1,2,\ldots,r$ and $\hat{\boldsymbol{\beta}}_{Ji}$ is the ith column of $\hat{\boldsymbol{\beta}}_J$.

[15] Since the main focus is on the long-run parameters, $\boldsymbol{\beta}$, we can concentrate out all the short-run parameters from the log-likelihood function.

Over-identified case ($k > r^2$)

Now there are $k - r^2$ additional restrictions that need to be taken into account at the estimation stage. This can be done by explicitly maximising the Lagrangian function (6.42). We assume that the normalisation restrictions on each of the r cointegrating vectors are also included in $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{R}\boldsymbol{\theta} - \mathbf{f} = \mathbf{0}$.

The first-order conditions are given by

$$\mathbf{d}_T(\hat{\boldsymbol{\theta}}) - \mathbf{R}'\hat{\boldsymbol{\lambda}} = \mathbf{0}, \tag{6.44}$$

$$\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{f} = \mathbf{0}, \tag{6.45}$$

where $\mathbf{d}_T(\boldsymbol{\theta}) = T^{-1}[\partial \ell_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}]$. Let $\hat{\boldsymbol{\theta}}^{(0)}$ and $\hat{\boldsymbol{\lambda}}^{(0)}$ be the initial estimates of the ML estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$. Taking the Taylor series expansion of (6.44) and (6.45) around $\hat{\boldsymbol{\theta}}^{(0)}$ and $\hat{\boldsymbol{\lambda}}^{(0)}$, we obtain[16]

$$\begin{bmatrix} \mathbf{G}_T\left(\hat{\boldsymbol{\theta}}^{(0)}\right) & \mathbf{R}' \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \begin{bmatrix} T\left(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{(0)}\right) \\ \hat{\boldsymbol{\lambda}} - \hat{\boldsymbol{\lambda}}^{(0)} \end{bmatrix} = \begin{bmatrix} \mathbf{d}_T\left(\hat{\boldsymbol{\theta}}^{(0)}\right) - \mathbf{R}'\hat{\boldsymbol{\lambda}}^{(0)} \\ -T\left(\mathbf{R}\hat{\boldsymbol{\theta}}^{(0)} - \mathbf{f}\right) \end{bmatrix} + o_p(1), \tag{6.46}$$

where $\mathbf{G}_T(\hat{\boldsymbol{\theta}}) = T^{-2}[-\partial^2 \ell_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}']$. To deal with the singularity of the normalised Hessian matrix, $\mathbf{G}_T(\hat{\boldsymbol{\theta}})$ in the case of cointegration, we let

$$\mathbf{J}_T(\boldsymbol{\theta}) = \mathbf{G}_T(\boldsymbol{\theta}) + \mathbf{R}'\mathbf{R}.$$

Then, the solution of (6.46) using a generalised inverse based on $\mathbf{J}_T(\hat{\boldsymbol{\theta}})$ is given by

$$\begin{bmatrix} T\left(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{(0)}\right) \\ \hat{\boldsymbol{\lambda}} - \hat{\boldsymbol{\lambda}}^{(0)} \end{bmatrix} = \begin{bmatrix} \mathbf{V}_{\theta\theta}\left(\hat{\boldsymbol{\theta}}^{(0)}\right) & \mathbf{V}_{\theta\lambda}\left(\hat{\boldsymbol{\theta}}^{(0)}\right) \\ \mathbf{V}'_{\theta\lambda}\left(\hat{\boldsymbol{\theta}}^{(0)}\right) & \mathbf{V}_{\lambda\lambda}\left(\hat{\boldsymbol{\theta}}^{(0)}\right) \end{bmatrix} \times \begin{bmatrix} \mathbf{d}_T\left(\hat{\boldsymbol{\theta}}^{(0)}\right) - \mathbf{R}'\hat{\boldsymbol{\lambda}}^{(0)} \\ -T\left(\mathbf{R}\hat{\boldsymbol{\theta}}^{(0)} - \mathbf{f}\right) \end{bmatrix} + o_p(1), \tag{6.47}$$

[16] The detailed derivations for $\mathbf{d}_T(\boldsymbol{\theta})$ and $\mathbf{G}_T(\boldsymbol{\theta})$ can be found in the DAE Working Paper version of Pesaran and Shin (2002).

where

$$\mathbf{V}_{\theta\theta}(\hat{\boldsymbol{\theta}}) = \mathbf{J}_T^{-1}(\hat{\boldsymbol{\theta}}) - \mathbf{J}_T^{-1}(\hat{\boldsymbol{\theta}})\mathbf{R}'\left[\mathbf{R}\mathbf{J}_T^{-1}(\hat{\boldsymbol{\theta}})\mathbf{R}'\right]^{-1}\mathbf{R}\mathbf{J}_T^{-1}(\hat{\boldsymbol{\theta}}), \tag{6.48}$$

$$\mathbf{V}_{\theta\lambda}(\hat{\boldsymbol{\theta}}) = \mathbf{J}_T^{-1}(\hat{\boldsymbol{\theta}})\mathbf{R}'\left[\mathbf{R}\mathbf{J}_T^{-1}(\hat{\boldsymbol{\theta}})\mathbf{R}'\right]^{-1}, \quad \mathbf{V}_{\lambda\lambda}(\hat{\boldsymbol{\theta}}) = \left[\mathbf{R}\mathbf{J}_T^{-1}(\hat{\boldsymbol{\theta}})\mathbf{R}'\right]^{-1}.$$

Hence, we obtain the following generalised version of the Newton–Raphson algorithm:[17]

$$\begin{pmatrix} \hat{\boldsymbol{\theta}}^{(i)} \\ \hat{\boldsymbol{\lambda}}^{(i)} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\theta}}^{(i-1)} \\ \hat{\boldsymbol{\lambda}}^{(i-1)} \end{pmatrix} \begin{bmatrix} \mathbf{V}_{\theta\theta}\left(\hat{\boldsymbol{\theta}}^{(i-1)}\right) & \mathbf{V}_{\theta\lambda}\left(\hat{\boldsymbol{\theta}}^{(i-1)}\right) \\ \mathbf{V}'_{\theta\lambda}\left(\hat{\boldsymbol{\theta}}^{(i-1)}\right) & \mathbf{V}_{\lambda\lambda}\left(\hat{\boldsymbol{\theta}}^{(i-1)}\right) \end{bmatrix} \times \begin{bmatrix} T^{-1}\left\{\mathbf{d}_T\left(\hat{\boldsymbol{\theta}}^{(i-1)}\right) - \mathbf{R}'\hat{\boldsymbol{\lambda}}^{(i-1)}\right\} \\ -T\left(\mathbf{R}\hat{\boldsymbol{\theta}}^{(i-1)} - \mathbf{f}\right) \end{bmatrix}. \tag{6.49}$$

From (6.47) we also find that

$$T\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \overset{a}{\sim} MN\left[0, \mathbf{V}_{\theta\theta}(\boldsymbol{\theta})\right], \tag{6.50}$$

which shows that the cointegrating parameters are super-consistent and have an asymptotic (mixture) normal distribution. It also shows that a consistent estimator of the asymptotic variance of $\hat{\boldsymbol{\theta}}$ is given by (6.48). See also Pesaran and Shin (2002) for a proof, and Pesaran and Pesaran's (1997) *Microfit 4.0* for more details of the numerical algorithms and other computational considerations.

For the initial estimates, $\hat{\boldsymbol{\theta}}^{(0)}$, we suggest using the linearised exactly identified estimators given by (6.43). One important aspect of this methodology is the fact that we begin with the exactly identifying restrictions from economic theory, rather than the type of statistical identification favoured by Johansen. This is particularly important for models with a relatively large number of long-run relations. Clearly, without some guidance from theory it would be extremely difficult to advance an exactly identified model with meaningful and understandable properties.

[17] See Magnus and Neudecker (1988, pp. 57–60) for the algebra about the bordered Gramian matrix.

TESTING THE VALIDITY OF OVER-IDENTIFYING RESTRICTIONS

Consider the general $k \geq r^2$ restrictions on $\boldsymbol{\theta}$ given by (6.39), and partition these restrictions as

$$\begin{pmatrix} \mathbf{R}_A\boldsymbol{\theta} \\ \mathbf{R}_B\boldsymbol{\theta} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_A \\ \mathbf{f}_B \end{pmatrix}, \tag{6.51}$$

where $\mathbf{R}_A$, $\mathbf{R}_B$ are $r^2 \times (m+1)r$, and $(k-r^2) \times (m+1)r$ known matrices, and $\mathbf{f}_A$, $\mathbf{f}_B$ are $r^2 \times 1$ and $(k-r^2) \times 1$ known vectors, respectively. Since we need r^2 independent restrictions to just identify $\boldsymbol{\theta}$, without loss of generality, $\mathbf{R}_A\boldsymbol{\theta} = \mathbf{f}_A$ can be regarded as such r^2 just-identifying restrictions. The remaining restrictions, $\mathbf{R}_B\boldsymbol{\theta} = \mathbf{f}_B$, will then constitute the $k-r^2$ over-identifying restrictions.

Let $\hat{\boldsymbol{\theta}}_1$ be the (unrestricted) ML estimators of $\boldsymbol{\theta}$ obtained subject to the r^2 exactly identifying restrictions (say, $\mathbf{R}_A\boldsymbol{\theta} = \mathbf{f}_A$), and $\hat{\boldsymbol{\theta}}_0$ be the restricted ML estimators of $\boldsymbol{\theta}$ obtained subject to the full k restrictions (namely, $\mathbf{R}\boldsymbol{\theta} = \mathbf{f}$), respectively. Then, the $k-r^2$ over-identifying restrictions on $\boldsymbol{\theta}$ can be tested using the log-likelihood ratio statistic given by

$$LR = 2\left[\ell_T\left(\hat{\boldsymbol{\theta}}_1\right) - \ell_T\left(\hat{\boldsymbol{\theta}}_0\right)\right], \tag{6.52}$$

where $\ell_T\left(\hat{\boldsymbol{\theta}}_1\right)$ and $\ell_T\left(\hat{\boldsymbol{\theta}}_0\right)$ represent the maximised values of the log-likelihood function obtained under $\mathbf{R}_A\boldsymbol{\theta} = \mathbf{f}_A$ and $\mathbf{R}\boldsymbol{\theta} = \mathbf{f}$, respectively. Pesaran and Shin (2002) prove that the log-likelihood ratio statistic for testing $\mathbf{R}\boldsymbol{\theta} = \mathbf{f}$ given by (6.52) has a χ^2 distribution with $k-r^2$ degrees of freedom, asymptotically. Small sample properties of the tests of over-identifying restrictions on the cointegrating vectors are described in Section 6.4 below.

6.2.4 *Estimation of the short-run parameters of the conditional VEC model*

Having computed the ML estimates of the cointegrating vectors $\hat{\boldsymbol{\beta}}_*' = \left(\hat{\boldsymbol{\beta}}', -\hat{\boldsymbol{\beta}}\boldsymbol{\gamma}'\right)'$, obtained under the exact and/or over-identifying restrictions given by (6.39), the ML estimates of the short-run parameters $(\boldsymbol{\alpha}, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{p-1}, \mathbf{a}_0)$ in (6.25) can be computed by the OLS regressions of $\Delta\mathbf{y}_t$ on

$$\hat{\boldsymbol{\xi}}_t^*, \Delta\mathbf{y}_{t-1}, \ldots, \Delta\mathbf{y}_{t-p+1} \text{ and } 1,$$

where $\hat{\boldsymbol{\xi}}_t^* = \hat{\boldsymbol{\beta}}_*'\mathbf{y}_{t-1}^*$ is the ML estimate of $\boldsymbol{\xi}_t^* = \boldsymbol{\beta}_*'\mathbf{y}_{t-1}^*$. Notice that $\hat{\boldsymbol{\beta}}$ is super-consistent, while the ML estimators of the short-run parameters are $\sqrt{T}$-consistent. The ML estimate of the (restricted) trend coefficients are then obtained by $\hat{\mathbf{a}}_1 = \hat{\boldsymbol{\alpha}}_y\hat{\boldsymbol{\beta}}'\boldsymbol{\gamma}$.

It is worth emphasising that, having established the form of the long-run relations, then standard OLS regression methods and standard testing procedures can be applied. All of the right-hand side variables in the error correction regression models are stationary and are either dated at time $t-1$ or earlier. In these circumstances, OLS is the appropriate estimation procedure and diagnostic statistics for residual serial correlation, normality, heteroscedasticity and functional form misspecifications can be readily computed, based on these OLS regressions, in the usual manner.[18] This is an important observation because it simplifies estimation and diagnostic testing procedures. Moreover, it makes it clear that the modelling procedure is robust to uncertainties surrounding the order of integration of particular variables. It is frequently difficult to establish the order of integration of particular variables using the techniques and samples of data which are available, and it would be worrying if the modelling procedure relied on assumptions that variables were integrated of a particular order. However, the observations above indicate that, so long as the $\hat{\boldsymbol{\xi}}_t^* = \hat{\boldsymbol{\beta}}_*'\mathbf{y}_{t-1}^*$ are stationary, the conditional VEC model, estimated and interpreted in the usual manner, will be valid even if it turns out that some or all of the variables in $\mathbf{y}_{t-1}^*$ are $I(0)$ and not $I(1)$ after all. A related discussion with mathematical proofs is given in Pesaran and Shin (1999) for cases where $r = 1$.

6.2.5 *Analysis of stability of the cointegrated system*

Having estimated the system of equations in the cointegrating VAR, we will typically need to check on the stability of the system as a whole, and more particularly to check that the disequilibria from the cointegrating relations are in fact mean-reverting. Although such a mean-reverting property is intrinsic to the modelling framework when the cointegration restrictions are not rejected, it is possible that the estimated model does not display this property in practice or that, if it does, the speed with which the system

[18] Further discussion of the validity of standard diagnostic test procedures when different estimation procedures are adopted in models involving unit roots and cointegrating relations is provided in Gerrard and Godfrey (1998), and the importance of the use of predicted values in the tests is discussed in Pesaran and Taylor (1999).

reverts back to its equilibrium is very slow. Summary statistics that shed light on the convergence property of the error correction terms, $\hat{\xi}_t$, will therefore be of some interest.

In the empirical applications of cointegration analysis where $r = 1$, the rate of convergence of $\hat{\xi}_t$ to its equilibrium is ascertained from the estimates of the error correction coefficients, $\boldsymbol{\alpha}$. However, as we shall demonstrate below, this procedure is not generally applicable. Consider the simple two variable error correction model

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = -\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (\beta_1 y_{1,t-1} + \beta_2 y_{2,t-1}) + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}, \quad (6.53)$$

in which the variables y_{1t} and y_{2t} are cointegrated with cointegrating vector $\boldsymbol{\beta} = (\beta_1, \beta_2)'$. Denoting $\xi_{t+1} = \beta_1 y_{1t} + \beta_2 y_{2t}$, and premultiplying both sides of (6.53) by $\boldsymbol{\beta}'$, we obtain

$$\Delta\xi_{t+1} = -(\boldsymbol{\beta}'\boldsymbol{\alpha})\xi_t + \boldsymbol{\beta}'\mathbf{u}_t,$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$ and $\mathbf{u}_t = (u_{1t}, u_{2t})'$, or

$$\xi_{t+1} = (1 - \boldsymbol{\beta}'\boldsymbol{\alpha})\xi_t + \boldsymbol{\beta}'\mathbf{u}_t. \quad (6.54)$$

Since $\boldsymbol{\beta}'\mathbf{u}_t$ is I(0), then, the stability of this equation requires $|1 - \boldsymbol{\beta}'\boldsymbol{\alpha}| = |1 - \beta_1\alpha_1 - \beta_2\alpha_2| < 1$, or $\beta_1\alpha_1 + \beta_2\alpha_2 > 0$, and $\beta_1\alpha_1 + \beta_2\alpha_2 < 2$. It is clear that these conditions depend on the adjustment parameters from both equations (α_1 and α_2) as well as the parameters of the cointegrating vector, and the estimate of α_1 alone will not allow us to sign the expressions $\beta_1\alpha_1 + \beta_2\alpha_2$ and $\beta_1\alpha_1 + \beta_2\alpha_2 - 2$. Hence, for example, restricting α_1 to lie in the range (0, 2) ensures the stability of (6.54) only under the normalisation $\beta_1 = 1$, and in the simple case where $\alpha_2 = 0$.[19]

More generally, we can rewrite (6.17) as an infinite order difference equation in an $r \times 1$ vector of (stochastic) disequilibrium terms, $\boldsymbol{\xi}_t = \boldsymbol{\beta}'\mathbf{y}_{t-1}$. Under our assumption that all the variables in $\mathbf{y}_t$ are $I(1)$, and all the roots of $\left|\mathbf{I}_m - \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \mathbf{y}^i\right| = 0$ fall outside the unit circle, we have the following expression for $\Delta\mathbf{y}_t$:

$$\Delta\mathbf{y}_t = \boldsymbol{\Gamma}(L)^{-1}\left(-\boldsymbol{\alpha}\boldsymbol{\xi}_t + \mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{u}_t\right), \quad t = 1, 2, \ldots, T, \quad (6.55)$$

where $\boldsymbol{\Gamma}(L) = \mathbf{I}_m - \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i L^i$. Defining $\boldsymbol{\Theta}(L) = \boldsymbol{\Gamma}(L)^{-1} = \sum_{i=0}^{\infty} \boldsymbol{\Theta}_i L^i$, then it is easily seen that the following recursive relations hold:

$$\boldsymbol{\Theta}_n = \boldsymbol{\Gamma}_1\boldsymbol{\Theta}_{n-1} + \boldsymbol{\Gamma}_2\boldsymbol{\Theta}_{n-2} + \cdots + \boldsymbol{\Gamma}_{p-1}\boldsymbol{\Theta}_{n-p+1}, \quad n = 1, 2, \ldots,$$

[19] When $\alpha_2 = 0$, y_{2t} is said to be 'long-run forcing' for y_{1t}.

where $\boldsymbol{\Theta}_0 = \mathbf{I}_m$, and $\boldsymbol{\Theta}_n = \mathbf{0}$ for $n < 0$. Premultiplying (6.55) by $\boldsymbol{\beta}'$, then we have

$$\Delta\boldsymbol{\xi}_{t+1} = -\boldsymbol{\beta}'\left(\mathbf{I}_m + \sum_{i=1}^{\infty} \boldsymbol{\Theta}_i L^i\right)\boldsymbol{\alpha}\boldsymbol{\xi}_t + \boldsymbol{\beta}'\left(\mathbf{I}_m + \sum_{i=1}^{\infty} \boldsymbol{\Theta}_i L^i\right)(\mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{u}_t), \quad (6.56)$$

or

$$\boldsymbol{\xi}_{t+1} = \left[(\mathbf{I}_r - \boldsymbol{\beta}'\boldsymbol{\alpha}) - \sum_{i=1}^{\infty} (\boldsymbol{\beta}'\boldsymbol{\Theta}_i\boldsymbol{\alpha}) L^i\right]\boldsymbol{\xi}_t + \left(\boldsymbol{\beta}' + \sum_{i=1}^{\infty} \boldsymbol{\beta}'\boldsymbol{\Theta}_i L^i\right)(\mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{u}_t). \quad (6.57)$$

This shows that, in general, when $p \geq 2$, the error correction variables, $\boldsymbol{\xi}_{t+1}$, follow infinite order VARMA processes, and there exists no simple rule involving $\boldsymbol{\alpha}$ alone that could ensure the stability of the dynamic processes in $\boldsymbol{\xi}_{t+1}$.[20]

However, given the assumption that none of the roots of $\left|\mathbf{I}_m - \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i z^i\right| = 0$ fall on or inside the unit circle, it is easily seen that the matrices $\boldsymbol{\Theta}_i$, $i = 0, 1, 2, \ldots$ are absolutely summable,[21] and therefore a suitably truncated version of $\sum_{i=1}^{\infty} (\boldsymbol{\beta}'\boldsymbol{\Theta}_i\boldsymbol{\alpha}) L^i$ can provide us with an adequate approximation in practice. Using an ℓ-order truncation we have

$$\boldsymbol{\xi}_{t+1} \approx \sum_{i=1}^{\ell} \mathbf{D}_i\boldsymbol{\xi}_{t-i+1} + \mathbf{v}_t, \quad t = 1, 2, \ldots, T, \quad (6.58)$$

where

$$\mathbf{D}_1 = \mathbf{I}_r - \boldsymbol{\beta}'\boldsymbol{\alpha}, \quad \mathbf{D}_i = -\boldsymbol{\beta}'\boldsymbol{\Theta}_{i-1}\boldsymbol{\alpha}, \ i = 2, 3, \ldots, \ell, \quad (6.59)$$

$$\mathbf{v}_t = \left(\boldsymbol{\beta}' + \sum_{i=1}^{\ell} \boldsymbol{\beta}'\boldsymbol{\Theta}_i L^i\right)(\mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{u}_t).$$

To explicitly evaluate the stability of the cointegrated system, we rewrite (6.58) more compactly as

$$\check{\boldsymbol{\xi}}_{t+1} = \boldsymbol{D}\check{\boldsymbol{\xi}}_t + \check{\mathbf{v}}_t, \quad t = 1, 2, \ldots, T, \quad (6.60)$$

[20] This result also highlights the deficiency of residual-based approaches to testing for cointegration, where finite-order ADF regressions are fitted to the residuals even if the order of the underlying VAR is 2 or more.

[21] The matrix sequence, $\{\boldsymbol{\Theta}_i, i = 0, 1, 2, \ldots\}$ is said to be absolutely summable if $\sum_{i=0}^{\infty} \left[tr(\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i')\right]^{1/2} < \infty$, which is satisfied since $\boldsymbol{\Gamma}(L)$ is invertible. See, for example, Lütkepohl (1991), Section C3, pp. 488–491.

where

$$\underset{r\ell\times 1}{\check{\boldsymbol{\xi}}_t} = \begin{pmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \vdots \\ \boldsymbol{\xi}_{t-\ell+1} \end{pmatrix},$$

$$\underset{r\ell\times r\ell}{\mathbf{D}} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{D}_2 & \mathbf{D}_3 & \cdots & \mathbf{D}_{\ell-1} & \mathbf{D}_\ell \\ \mathbf{I}_r & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_r & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ & & & \vdots & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_r & \mathbf{0} \end{pmatrix},$$

$$\underset{r\ell\times 1}{\check{\mathbf{v}}_t} = \begin{pmatrix} \mathbf{v}_t \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}. \qquad (6.61)$$

The above cointegrated system is stable if all the roots of

$$\left|\mathbf{I}_r - \mathbf{D}_1 z - \cdots - \mathbf{D}_\ell z^\ell\right| = 0$$

lie outside the unit circle, or if all the eigenvalues of $\mathbf{D}$ have modulus less than unity.[22]

6.2.6 *Impulse response analysis in cointegrating VARs*

Using the level MA representation, (6.24), generalised impulse response functions can be calculated for the cointegrating VAR model (6.25) in a way similar to the VAR discussed above. Now, it is easily seen that the effect of a unit shock to the ith reduced form error, u_{it}, is given by[23]

$$g\left(n, \mathbf{y} : u_i\right) = \frac{1}{\sqrt{\sigma_{ii}}}\tilde{\mathbf{C}}_n \boldsymbol{\Sigma} \mathbf{e}_i, \quad n = 0, 1, \ldots, \quad i = 1, \ldots, m, \qquad (6.62)$$

where $\mathbf{u}_t$ is *i.i.d.* $(\mathbf{0}, \boldsymbol{\Sigma})$, $\tilde{\mathbf{C}}_n = \sum_{j=0}^{n} \mathbf{C}_j$, $\mathbf{C}_j$'s are given by the recursive relations (6.21), and $\mathbf{e}_i$ is a selection vector of zeros with unity as its ith

[22] Notice that the stability analysis is not affected by the presence of deterministic and stationary exogenous variables in the system.

[23] Combining (6.11) and (6.24) together, we obtain $g\left(n, \mathbf{y} : u_i\right) = \sigma_{ii}^{-1/2}\left\{\mathbf{C}(1) + \mathbf{C}_n^*\right\}\boldsymbol{\Sigma}\mathbf{e}_i$. Then, using (6.20), we find that $\mathbf{C}(1) + \mathbf{C}_n^* = \sum_{j=0}^{n} \mathbf{C}_j = \tilde{\mathbf{C}}_n$.

element. For the effect of a unit shock to the ith structural form error, ε_{it}, we notice that (6.24) can be written as

$$\mathbf{y}_t = \mathbf{y}_0 + \mathbf{b}_0 t + \mathbf{C}(1)\mathbf{s}_t + \mathbf{C}^*(L)\,\mathbf{A}^{-1}\left(\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_0\right), \qquad (6.63)$$

where we use $\mathbf{u}_t = \mathbf{A}^{-1}\boldsymbol{\varepsilon}_t$, and therefore we have

$$g\left(n, \mathbf{y} : \varepsilon_i\right) = \frac{1}{\sqrt{\omega_{ii}}}\tilde{\mathbf{C}}_n \mathbf{A}^{-1}\boldsymbol{\Omega}\mathbf{e}_i, \quad n = 0, 1, \ldots, \quad i = 1, \ldots, m, \qquad (6.64)$$

where $\boldsymbol{\varepsilon}_t$ is *i.i.d.* $(\mathbf{0}, \boldsymbol{\Omega})$ with $\boldsymbol{\Sigma} = \mathbf{A}^{-1}\boldsymbol{\Omega}\mathbf{A}'^{-1}$. In particular,

$$g\left(\infty, \mathbf{y} : \varepsilon_i\right) = \omega_{ii}^{-1/2}\mathbf{C}(1)\mathbf{A}^{-1}\boldsymbol{\Omega}\mathbf{e}_i, \; g\left(\infty, \mathbf{y} : u_i\right) = \sigma_{ii}^{-1/2}\mathbf{C}(1)\boldsymbol{\Sigma}\mathbf{e}_i,$$

which shows that shocks will have permanent effects on the $I(1)$ variables, unlike the stationary case.

Shocks will have only a temporary effect on the cointegrating relations though. Hence, the generalised impulse response function for the cointegrating relations $\boldsymbol{\xi}_t = \boldsymbol{\beta}'\mathbf{y}_{t-1}$ with respect to a unit shock to the structural errors is given by

$$g\left(n, \boldsymbol{\xi} : \varepsilon_i\right) = \frac{1}{\sqrt{\omega_{ii}}}\boldsymbol{\beta}'\tilde{\mathbf{C}}_n \mathbf{A}^{-1}\boldsymbol{\Omega}\mathbf{e}_i, \quad n = 0, 1, \ldots, \quad i = 1, \ldots, m. \qquad (6.65)$$

Since $\boldsymbol{\beta}'\tilde{\mathbf{C}}_\infty = \boldsymbol{\beta}'\mathbf{C}(1) = \mathbf{0}$, it follows that $g\left(\infty, \boldsymbol{\xi} : \varepsilon_i\right) = \mathbf{0}$, and ultimately the effects of shocks on the cointegrating relations will disappear. Nevertheless, estimation of $g\left(n, \boldsymbol{\xi} : \varepsilon_i\right)$ for a finite n still requires *a priori* identification of $\mathbf{A}^{-1}\boldsymbol{\Omega}$. Once again, a variety of identification schemes can be used for this purpose. Alternatively, we could focus on the impulse response functions of $\boldsymbol{\xi}_t = \boldsymbol{\beta}'\mathbf{y}_{t-1}$ with respect to the ith reduced form shock, u_{it}. In this case

$$g\left(n, \boldsymbol{\xi} : u_i\right) = \frac{1}{\sqrt{\sigma_{ii}}}\boldsymbol{\beta}'\tilde{\mathbf{C}}_n \boldsymbol{\Sigma}\mathbf{e}_i, \quad n = 0, 1, \ldots, \quad i = 1, \ldots, m. \qquad (6.66)$$

which is uniquely determined from the knowledge of the reduced form parameters.

Furthermore, generalised forecast error variance decompositions for the cointegrating VAR model (6.25) can be computed as follow:

$$\psi_{ij,n} = \frac{\sigma_{ii}^{-1}\sum_{\ell=0}^{n}\left(\mathbf{e}_i'\tilde{\mathbf{C}}_\ell\boldsymbol{\Sigma}\mathbf{e}_j\right)^2}{\sum_{\ell=0}^{n}\mathbf{e}_i'\tilde{\mathbf{C}}_\ell\boldsymbol{\Sigma}\tilde{\mathbf{C}}_\ell'\mathbf{e}_i}, \quad n = 0, 1, \ldots; \text{ and } i, j = 1, \ldots, m. \qquad (6.67)$$

The $\psi_{ij,n}$ in (6.67) measures the proportion of the n-step ahead forecast error variance of variable i accounted for by the reduced form error in the jth equation in the system unlike the orthogonalised forecast error variance decomposition, which due to non-zero correlations across the shocks, cause the different proportions not necessarily to add up to unity.

Corresponding orthogonalised impulse response functions and forecast error variance decompositions for the cointegrating VAR model (6.25) are given by:

$$o\left(n, \mathbf{y} : u_i^*\right) = \tilde{\mathbf{C}}_n \mathbf{P} \mathbf{e}_i, \quad n = 0, 1, \ldots, \quad i = 1, \ldots, m, \tag{6.68}$$

$$o\left(n, \boldsymbol{\xi} : u_i^*\right) = \boldsymbol{\beta}' \tilde{\mathbf{C}}_n \mathbf{P} \mathbf{e}_i, \quad n = 0, 1, \ldots, \quad i = 1, \ldots, m, \tag{6.69}$$

$$o_{ij,n} = \frac{\sigma_{ii}^{-1} \sum_{\ell=0}^{n} \left(\mathbf{e}_i' \tilde{\mathbf{C}}_\ell \mathbf{P} \mathbf{e}_j\right)^2}{\sum_{\ell=0}^{n} \mathbf{e}_i' \tilde{\mathbf{C}}_\ell \boldsymbol{\Sigma} \tilde{\mathbf{C}}_\ell' \mathbf{e}_i}, \quad n = 0, 1, 2, \ldots, \quad i, j = 1, \ldots, m, \tag{6.70}$$

where u_{it}^* is an orthogonalised residual and $\mathbf{P}$ is a lower triangular matrix obtained by the Choleski decomposition of $\boldsymbol{\Sigma} = \mathbf{P}\mathbf{P}'$.

Finally, we could examine the effect of system-wide shocks on the cointegrating relations using the persistence profiles discussed above in Section 6.1.3. Pesaran and Shin (1996) suggest using the persistence profiles to measure the speed of convergence of the cointegrating relations to equilibrium. The scaled persistence profiles of the jth cointegrating relation is given by

$$h\left(\boldsymbol{\beta}_j' \mathbf{y}_t, n\right) = \frac{\boldsymbol{\beta}_j' \tilde{\mathbf{C}}_n \boldsymbol{\Sigma} \tilde{\mathbf{C}}_n' \boldsymbol{\beta}_j}{\boldsymbol{\beta}_j' \boldsymbol{\Sigma} \boldsymbol{\beta}_j}, \quad n = 0, 1, \ldots, \quad j = 1, \ldots, r, \tag{6.71}$$

which is scaled to have a value of unity on impact. The profiles tend to zero as $n \to \infty$, and provide a useful graphical representation of the extent to which the cointegrating (equilibrium) relations adjust to system-wide shocks. Once again, the main attraction of persistence profiles lies in the fact that they are uniquely determined from the reduced form parameters and do not depend on the nature of the system-wide shocks considered. Using (6.26), the cointegrating relations in terms of the structural errors may be written as

$$\boldsymbol{\beta}_j' \mathbf{y}_t = \boldsymbol{\beta}_j' \mathbf{y}_0 + \left(\boldsymbol{\beta}_j' \boldsymbol{\gamma}\right) t + \boldsymbol{\beta}_j' \mathbf{C}^*(L) \mathbf{A}^{-1} \left(\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_0\right),$$

and the persistence profile of $\boldsymbol{\beta}_j' \mathbf{y}_t$ with respect to the structural errors, $\boldsymbol{\varepsilon}_t$, is given by

$$\frac{\boldsymbol{\beta}_j' \left(\tilde{\mathbf{C}}_n \mathbf{A}^{-1}\right) \boldsymbol{\Omega} \left(\tilde{\mathbf{C}}_n \mathbf{A}^{-1}\right)' \boldsymbol{\beta}_j}{\boldsymbol{\beta}_j' \mathbf{A}^{-1} \boldsymbol{\Omega} \mathbf{A}'^{-1} \boldsymbol{\beta}_j}, \quad n = 0, 1, \ldots, \quad j = 1, \ldots, r.$$

But, since $\boldsymbol{\Sigma} = \mathbf{A}^{-1} \boldsymbol{\Omega} \mathbf{A}'^{-1}$, this persistence profile is in fact identical to the one derived using the reduced form errors, $\mathbf{u}_t$, given by (6.71).

6.3 The cointegrated VAR model with $I(1)$ exogenous variables

The most complete econometric model that we might wish to consider is the case in which there are both endogenous and exogenous variables and linear deterministic trends. This is the model discussed in Pesaran, Shin and Smith (2000), where we distinguish between an $m_y \times 1$ vector of endogenous variables $\mathbf{y}_t$ and an $m_x \times 1$ vector of exogenous $I(1)$ variables $\mathbf{x}_t$ among the core variables in $\mathbf{z}_t = (\mathbf{y}_t', \mathbf{x}_t')'$, with $m = m_y + m_x$.

We begin with the extended *vector error correction* model (VECM) in $\mathbf{z}_t$ (*cf.* (6.17)),

$$\Delta \mathbf{z}_t = -\boldsymbol{\Pi} \mathbf{z}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \mathbf{z}_{t-i} + \mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{u}_t, \tag{6.72}$$

where the short-run response matrices $\{\boldsymbol{\Gamma}_i\}_{i=1}^{p-1}$ and the long-run multiplier matrix $\boldsymbol{\Pi}$ are similarly defined to those below (6.17).

By partitioning the error term $\mathbf{u}_t$ conformably with $\mathbf{z}_t = (\mathbf{y}_t', \mathbf{x}_t')'$ as $\mathbf{u}_t = \left(\mathbf{u}_{yt}', \mathbf{u}_{xt}'\right)'$ and its variance matrix as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{pmatrix},$$

we are able to express $\mathbf{u}_{yt}$ conditionally in terms of $\mathbf{u}_{xt}$ as

$$\mathbf{u}_{yt} = \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} \mathbf{u}_{xt} + \boldsymbol{v}_t, \tag{6.73}$$

where $\boldsymbol{v}_t \sim i.i.d.\,(\mathbf{0}, \boldsymbol{\Sigma}_{vv})$, $\boldsymbol{\Sigma}_{vv} \equiv \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xy}$ and $\boldsymbol{v}_t$ is uncorrelated with $\mathbf{u}_{xt}$ by construction. Substitution of (6.73) into (6.72) together with a similar partitioning of the parameter vectors and matrices $\mathbf{a}_0 = \left(\mathbf{a}_{y0}', \mathbf{a}_{x0}'\right)'$,

$\mathbf{a}_1 = \left(\mathbf{a}'_{y1}, \mathbf{a}'_{x1}\right)'$, $\boldsymbol{\Pi} = \left(\boldsymbol{\Pi}'_y, \boldsymbol{\Pi}'_x\right)'$, $\boldsymbol{\Gamma}_i = \left(\boldsymbol{\Gamma}'_{yi}, \boldsymbol{\Gamma}'_{xi}\right)'$, $i = 1, \ldots, p-1$, provides a conditional model for $\Delta\mathbf{y}_t$ in terms of $\mathbf{z}_{t-1}, \Delta\mathbf{x}_t, \Delta\mathbf{z}_{t-1}, \Delta\mathbf{z}_{t-2}, \ldots$; *viz*.

$$\Delta\mathbf{y}_t = -\boldsymbol{\Pi}_{yy.x}\mathbf{z}_{t-1} + \boldsymbol{\Lambda}\Delta\mathbf{x}_t + \sum_{i=1}^{p-1} \boldsymbol{\Psi}_i\Delta\mathbf{z}_{t-i} + \mathbf{c}_0 + \mathbf{c}_1 t + \boldsymbol{v}_t, \tag{6.74}$$

where $\boldsymbol{\Pi}_{yy.x} \equiv \boldsymbol{\Pi}_y - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Pi}_x$, $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}$, $\boldsymbol{\Psi}_i \equiv \boldsymbol{\Gamma}_{yi} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Gamma}_{xi}$, $i = 1, \ldots, p-1$, $\mathbf{c}_0 \equiv \mathbf{a}_{y0} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{a}_{x0}$ and $\mathbf{c}_1 \equiv \mathbf{a}_{y1} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{a}_{x1}$,

Following Johansen (1995), we assume that the process $\{\mathbf{x}_t\}_{t=1}^{\infty}$ is weakly exogenous with respect to the matrix of long-run multiplier parameters $\boldsymbol{\Pi}$, namely,

$$\boldsymbol{\Pi}_x = \mathbf{0}. \tag{6.75}$$

Therefore,

$$\boldsymbol{\Pi}_{yy.x} = \boldsymbol{\Pi}_y. \tag{6.76}$$

Consequently, from (6.72) and (6.74), the system of equations is rendered as

$$\Delta\mathbf{y}_t = -\boldsymbol{\Pi}_y\mathbf{z}_{t-1} + \boldsymbol{\Lambda}\Delta\mathbf{x}_t + \sum_{i=1}^{p-1} \boldsymbol{\Psi}_i\Delta\mathbf{z}_{t-i} + \mathbf{c}_0 + \mathbf{c}_1 t + \boldsymbol{v}_t, \tag{6.77}$$

$$\Delta\mathbf{x}_t = \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_{xi}\Delta\mathbf{z}_{t-i} + \mathbf{a}_{x0} + \mathbf{u}_{xt}, \tag{6.78}$$

where now the restrictions on trend coefficients (6.23) are modified to

$$\mathbf{c}_1 = \boldsymbol{\Pi}_y\boldsymbol{\gamma}. \tag{6.79}$$

The restriction $\boldsymbol{\Pi}_x = \mathbf{0}$ in (6.75) implies that the elements of the vector process $\{\mathbf{x}_t\}_{t=1}^{\infty}$ are not cointegrated among themselves as is evident from (6.78). Moreover, the information available from the differenced VAR($p-1$) model (6.78) for $\{\mathbf{x}_t\}_{t=1}^{\infty}$ is redundant for efficient conditional estimation and inference concerning the long-run parameters $\boldsymbol{\Pi}_y$ as well as the deterministic and short-run parameters $\mathbf{c}_0$, $\mathbf{c}_1$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}_i$, $i = 1, \ldots, p-1$, of (6.77).[24] Furthermore, we may regard $\{\mathbf{x}_t\}_{t=1}^{\infty}$ as *long-run forcing* for $\{\mathbf{y}_t\}_{t=1}^{\infty}$; see Granger and Lin (1995). Note that this restriction does not preclude $\{\mathbf{y}_t\}_{t=1}^{\infty}$ being *Granger-causal* for $\{\mathbf{x}_t\}_{t=1}^{\infty}$ in the *short run*.

[24] In general the variance of $\boldsymbol{v}_t$ will be smaller than that of $\mathbf{u}_{yt}$ because it is easily seen that

$$\boldsymbol{\Sigma}_{vv} - \boldsymbol{\Sigma}_{yy} = -\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} \leq 0.$$

When there are r cointegrating relations among $\mathbf{z}_t$, then we may express

$$\boldsymbol{\Pi}_y = \boldsymbol{\alpha}_y\boldsymbol{\beta}', \tag{6.80}$$

where $\boldsymbol{\alpha}_y$ $(m_y \times r)$ and $\boldsymbol{\beta}$ $(m \times r)$ are matrices of error correction coefficients and of the long-run (or cointegrating) coefficients, both of which are of full column rank, r. For the purpose of empirical analysis, we assume that the lag order p is large enough so that $\mathbf{u}_t$ and $\boldsymbol{v}_t$ are serially uncorrelated, and have zero mean and positive definite covariance matrices, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{vv}$, respectively. For the purpose of the ML estimation, we also assume that $\mathbf{u}_t$ and $\boldsymbol{v}_t$ are normally distributed, although this is not binding if the number of the time series observations available is large enough.[25]

To a large extent, the analysis of a cointegrated VAR model containing exogenous $I(1)$ variables follows very similar lines to that described in Section 6.2 above. Hence, to avoid the unsatisfactory possibility that there exist quadratic trends in the level solution of the data generating process for $\mathbf{z}_t$ when there is no cointegration, we can again assume that there are restrictions on the intercepts and/or time trends corresponding to Cases I–V in Section 6.2.1 above. We delineate five cases of interest; *viz*.

- **Case I:** (No intercepts; no trends.) $\mathbf{c}_0 = \mathbf{0}$ and $\mathbf{c}_1 = \mathbf{0}$. Hence, the structural VECM (6.77) becomes

$$\Delta\mathbf{y}_t = -\boldsymbol{\Pi}_y\mathbf{z}_{t-1} + \boldsymbol{\Lambda}\Delta\mathbf{x}_t + \sum_{i=1}^{p-1} \boldsymbol{\Psi}_i\Delta\mathbf{z}_{t-i} + \boldsymbol{v}_t. \tag{6.81}$$

- **Case II:** (Restricted intercepts; no trends.) $\mathbf{c}_0 = \boldsymbol{\Pi}_y\boldsymbol{\mu}$ and $\mathbf{c}_1 = \mathbf{0}$. The structural VECM (6.77) is

$$\Delta\mathbf{y}_t = \boldsymbol{\Pi}_y\boldsymbol{\mu} - \boldsymbol{\Pi}_y\mathbf{z}_{t-1} + \boldsymbol{\Lambda}\Delta\mathbf{x}_t + \sum_{i=1}^{p-1} \boldsymbol{\Psi}_i\Delta\mathbf{z}_{t-i} + \boldsymbol{v}_t. \tag{6.82}$$

- **Case III:** (Unrestricted intercepts; no trends.) $\mathbf{c}_0 \neq \mathbf{0}$ and $\mathbf{c}_1 = \mathbf{0}$. In this case, the intercept restriction $\mathbf{c}_0 = \boldsymbol{\Pi}_y\boldsymbol{\mu}$ is ignored and the structural

Therefore, the parameters in the conditional model (6.77) are likely to be estimated more precisely than the parameters of the unconditional model. Whether this is an advantage depends on what the economic parameters of interest are. If the parameters of interest are $\boldsymbol{\Pi}_y = (\boldsymbol{\Pi}_{yy}, \boldsymbol{\Pi}_{yx})$, it is clear from the above equation that $\Delta\mathbf{x}_t$ will be weakly exogenous for $\boldsymbol{\Pi}_y$ only if either $\boldsymbol{\Sigma}_{yx} = \mathbf{0}$ so that $\boldsymbol{\Theta} = 0$ or if $\boldsymbol{\Pi}_x = (\boldsymbol{\Pi}_{xy}, \boldsymbol{\Pi}_{xx}) = \mathbf{0}$. In either of these cases the coefficient matrix on $(\mathbf{y}_{t-1}, \mathbf{x}_{t-1})$ in the conditional model will provide an estimate of $\boldsymbol{\Pi}_y$. In other cases the economic parameter of interest may be simply the long-run effects of $\mathbf{x}_t$ on $\mathbf{y}_t$ so one might be interested in $\boldsymbol{\Pi}_y - \boldsymbol{\Theta}\boldsymbol{\Pi}_x$ directly, in which case the model conditional on $\mathbf{x}_t$ is appropriate whether or not $\boldsymbol{\Pi}_x = 0$.

[25] For a more precise statement of these assumptions see Johansen (1995), and Pesaran, Shin and Smith (2000).

VECM estimated is

$$\Delta \mathbf{y}_t = \mathbf{c}_0 - \mathbf{\Pi}_y \mathbf{z}_{t-1} + \mathbf{\Lambda} \Delta \mathbf{x}_t + \sum_{i=1}^{p-1} \mathbf{\Psi}_i \Delta \mathbf{z}_{t-i} + \boldsymbol{v}_t. \quad (6.83)$$

- **Case IV:** (Unrestricted intercepts; restricted trends.) $\mathbf{c}_0 \neq \mathbf{0}$ and $\mathbf{c}_1 = \mathbf{\Pi}_y \boldsymbol{\gamma}$. Thus,

$$\Delta \mathbf{y}_t = \mathbf{c}_0 + (\mathbf{\Pi}_y \boldsymbol{\gamma})\, t - \mathbf{\Pi}_y \mathbf{z}_{t-1} + \mathbf{\Lambda} \Delta \mathbf{x}_t + \sum_{i=1}^{p-1} \mathbf{\Psi}_i \Delta \mathbf{z}_{t-i} + \boldsymbol{v}_t. \quad (6.84)$$

- **Case V:** (Unrestricted intercepts; unrestricted trends.) $\mathbf{c}_0 \neq \mathbf{0}$ and $\mathbf{c}_1 \neq \mathbf{0}$. Here, the deterministic trend restriction $\mathbf{c}_1 = \mathbf{\Pi}_y \boldsymbol{\gamma}$ is ignored and the structural VECM estimated is

$$\Delta \mathbf{y}_t = \mathbf{c}_0 + \mathbf{c}_1 t - \mathbf{\Pi}_y \mathbf{z}_{t-1} + \mathbf{\Lambda} \Delta \mathbf{x}_t + \sum_{i=1}^{p-1} \mathbf{\Psi}_i \Delta \mathbf{z}_{t-i} + \boldsymbol{v}_t. \quad (6.85)$$

Tests of the cointegrating rank are obtained along exactly the same lines as those in Section 6.2.2, with the first step in the algorithm generating residuals $\mathbf{r}_{0t}$ and $\mathbf{r}_{1t}$ from the regression of, in turn, $\Delta \mathbf{y}_t$ and $\mathbf{z}^*_{t-1} = \left(\mathbf{z}'_{t-1}, t\right)'$ on $\Delta \mathbf{x}_t, \Delta \mathbf{z}_{t-1}, \Delta \mathbf{z}_{t-2}, \ldots, \Delta \mathbf{z}_{t-p+1}$ and 1.[26] Estimation of the VECM subject to exactly and over-identifying long-run restrictions can be carried out using maximum likelihood methods as in Section 6.2.3 applied to (6.77) subject to the appropriate restrictions on the intercepts and trends, subject to Rank$(\mathbf{\Pi}_y) = r$, and subject to k general linear restrictions of the form in (6.39). And, having computed ML estimates of the cointegrating vectors, estimation of the short-run parameters of the conditional VECM can be computed using OLS regressions exactly as in Section 6.2.4.

The investigation of the dynamic properties of the system including exogenous $I(1)$ variables does require a little care, however. For this, we require the full-system VECM, obtained by augmenting the conditional model for $\Delta \mathbf{y}_t$, (6.77), with the marginal model for $\Delta \mathbf{x}_t$, (6.78). This is written as

$$\Delta \mathbf{z}_t = -\boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{z}_{t-1} + \sum_{i=1}^{p-1} \mathbf{\Gamma}_i \Delta \mathbf{z}_{t-i} + \mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{H}\boldsymbol{\zeta}_t, \quad (6.86)$$

[26] Asymptotic distributions of the trace and maximum eigenvalue statistics are again non-standard, and depend on whether the intercepts and/or the coefficients on the deterministic trends are restricted or unrestricted. Pesaran, Shin and Smith (2000) have tabulated the upper 5% and 10% quantiles of the asymptotic critical values of both statistics via stochastic simulations with $T = 500$ and 10,000 replications. See also Mackinnon (1996).

where $\boldsymbol{\beta}$ is defined by (6.80),

$$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_y \\ \mathbf{0} \end{pmatrix}, \; \mathbf{\Gamma}_i = \begin{pmatrix} \mathbf{\Psi}_i + \mathbf{\Lambda}\mathbf{\Gamma}_{xi} \\ \mathbf{\Gamma}_{xi} \end{pmatrix}, \; \mathbf{a}_0 = \begin{pmatrix} \mathbf{c}_0 + \mathbf{\Lambda}\mathbf{a}_{x0} \\ \mathbf{a}_{x0} \end{pmatrix}, \; \mathbf{a}_1 = \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{0} \end{pmatrix}, \quad (6.87)$$

$$\boldsymbol{\zeta}_t = \begin{pmatrix} \boldsymbol{v}_t \\ \mathbf{u}_{xt} \end{pmatrix}, \; \mathbf{H} = \begin{pmatrix} \mathbf{I}_{m_y} & \mathbf{\Lambda} \\ \mathbf{0} & \mathbf{I}_{m_x} \end{pmatrix}, \; \text{Cov}(\boldsymbol{\zeta}_t) = \mathbf{\Sigma}_{\zeta\zeta} = \begin{pmatrix} \mathbf{\Sigma}_{vv} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{xx} \end{pmatrix}. \quad (6.88)$$

Analysis of the stability of the cointegrated system follows the arguments of Section 6.2.5, and impulse response analysis follows the arguments in Section 6.2.6, but *applied to the full system in* (6.86). While efficient conditional estimation of, and inference on, the parameters of (6.77) can be conducted without reference to the marginal model (6.78), the dynamic properties of the system have to accommodate the influence of the processes driving the exogenous variables.

This last point is worth emphasising and applies to any analysis involving counter-factuals, including impulse response analysis and forecasting exercises, for example. Macromodellers frequently consider the dynamic response of a system to a change in an exogenous variable by considering the effects of a once-and-for-all increase in the variable.[27] This (implicitly) imposes restrictions on the processes generating the exogenous variable, assuming that there is no serial correlation in the variable and that a shock to one exogenous variable can be considered without having to take into account changes in other exogenous variables. These counter-factual exercises might be of interest. But, generally speaking, one needs to take into account the possibility that changes in one exogenous variable will have an impact on other exogenous variables and that these effects might continue and interact over time. This requires an explicit analysis of the dynamic processes driving the exogenous variables, as captured by the marginal model in (6.78). The whole point of the approach to investigating model dynamics reflected in the model of (6.86) and incorporated in the idea of generalised impulse response analysis is to explicitly allow for the conditional correlation structure in errors and the interactions between endogenous and exogenous variables to provide a 'realistic' counter-factual exercise based on the contemporaneous covariances and interactions observed historically in the data.

[27] This corresponds to our earlier discussion of the dynamic impact of a once-and-for-all shock to an equation in a system captured as an intercept shift.

6.4 Small sample properties of test statistics

The distributions of the trace and maximal eigenvalue statistics used to test the number of cointegrating relationships (see (6.37) and (6.38)) and of the log-likelihood ratio statistic used to test the validity of the over-identifying restrictions (see (6.52)) are appropriate only asymptotically. Moreover, recent work has shown that the asymptotic results are valid only when relatively large samples of data are available if the cointegrating VAR model is of even modest size (in terms of the number of parameters involved); that is, when the order of the VAR or the number of variables in the VAR exceeds three or four, say.[28] This suggests that care should be taken in interpreting the test statistics obtained.

In some cases, it is possible to undertake bootstrapping exercises to investigate directly the small sample properties of the estimated statistics. For example, suppose that the VEC model of (6.77) and (6.78) has been estimated subject to the just- or over-identifying restrictions suggested by economic theory. Using the observed initial values for each variable, it is possible to generate S new samples of data (of the same size as the original) under the hypothesis that the estimated version of (6.77) and (6.78) is the true data generating process. For each of the S replications of the data, the tests of the cointegrating rank and of the over-identifying restrictions can be carried out and, hence, distributions of the test statistics are obtained which take into account the small sample of data available when calculating the statistics. Working at the α% level of significance, critical values which take into account the small sample properties of the tests can be obtained by observing, from the right tails of the simulated distributions, the value of the statistics which would ensure that the probability that the null is not rejected when it is true is $(1-\alpha)$.

More specifically, suppose that the model in (6.77) has been estimated under the exactly or over-identifying restrictions given by (6.39). We therefore have estimates of the cointegrating vectors, $\hat{\boldsymbol{\beta}}_*$, of the short-run parameters, $\left(\hat{\boldsymbol{\alpha}}_y, \hat{\boldsymbol{\Psi}}_1, \ldots, \hat{\boldsymbol{\Psi}}_{p-1}, \hat{\boldsymbol{\Lambda}}, \hat{\mathbf{c}}_0\right)$, and of the covariance matrix, $\hat{\boldsymbol{\Sigma}}_{vv}$. Taking the observed values of $\Delta\mathbf{x}_t$ as fixed or re-sampled using (6.78) over the whole sample, and taking the p lagged values of the $\mathbf{y}_t$ observed just prior to the sample as fixed also, for the sth replication, we can recursively

[28] See, for example, Abadir *et al.* (1999), Gonzalo (1994) and Muscatelli and Hurn (1995).

simulate the values of $\Delta\mathbf{y}_t^{(s)}$, $s = 1, 2, \ldots, S$, using

$$\Delta\mathbf{y}_t^{(s)} = -\hat{\boldsymbol{\alpha}}_y \hat{\boldsymbol{\beta}}_*' \mathbf{z}_{t-1}^{*(s)} + \sum_{i=1}^{p-1} \hat{\boldsymbol{\Psi}}_i \Delta\mathbf{z}_{t-i}^{(s)} + \hat{\boldsymbol{\Lambda}}\Delta\mathbf{x}_t + \hat{\mathbf{c}}_0 + \boldsymbol{v}_t^{(s)}, \quad t = 1, 2, \ldots, T. \tag{6.89}$$

To obtain $\boldsymbol{v}_t^{(s)}$, allowing for the observed correlation of shocks across the $\Delta\mathbf{y}_t$, we can generate draws from a multivariate normal distribution chosen to match the observed correlation of the estimated reduced form errors, $\hat{\boldsymbol{\Sigma}}_{vv}$, (termed a *parametric* bootstrap) or we can re-sample with replacement from the estimated residuals (a *non-parametric* bootstrap).[29]

Having generated the $\Delta\mathbf{y}_t^{(s)}$, $t = 1, \ldots, T$, and making use of the observed $\Delta\mathbf{x}_t$, it is straightforward to estimate the VECM of (6.77) subject to just-identifying restrictions and then subject to the over-identifying restrictions of (6.39) to obtain a sequence of log-likelihood ratio test statistics, $LR^{(s)}$, each testing the validity of the over-identifying restrictions in the s-th simulated dataset, $s = 1, \ldots, S$.[30] These statistics can be sorted into ascending order and, given that the data has been generated by the model at (6.77) incorporating the over-identifying restrictions of $\hat{\boldsymbol{\beta}}_*$, critical values can be identified which are relevant to this particular model and which take into account the sample size. Hence, for example, the value of $LR^{(s)}$ which exceeds 95% of the observed statistics represents the appropriate 95% critical value for the test of the validity of the over-identifying restrictions.[31]

6.5 Empirical distribution of impulse response functions and persistence profiles

The simulation methods described above are relatively easy to implement in the context of a VAR and can be applied in various contexts.

[29] More detailed discussion on generating simulated errors in bootstrap procedures is provided in Section 7.3.3.

[30] The maximum likelihood estimation of the VECM can be time-consuming, especially if one is to be sure that all of the estimates relate to global and not local maxima. Practically, the choice of an optimisation algorithm is likely to be important in this exercise, and the simulated annealing algorithm discussed in Goffe *et al.* (1994) can prove useful in this respect.

[31] Simulation here is used to find the probability of rejection for one point in the space covered by the null (that the over-identifying restrictions are valid). The classical significance level is the *maximum* of the rejection probabilities over the null space. By using a single point, the observed critical values potentially understate the true rejection level.

An important example is in examining the distributional properties of the various statistics used to investigate the dynamic properties of the estimated models we have discussed in the chapter. Specifically, in this section, we describe the steps involved in the calculation of empirical distribution of generalised (orthogonalised) impulse response functions and persistence profiles based on a vector error correction model using stochastic simulation techniques.

Consider the underlying vector error correction model, (6.86), which can be rewritten as

$$\mathbf{z}_t = \sum_{i=1}^{p} \mathbf{\Phi}_i \mathbf{z}_{t-i} + \mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{H}\boldsymbol{\zeta}_t, \quad t = 1, 2, \ldots, T, \tag{6.90}$$

where $\mathbf{\Phi}_1 = \mathbf{I}_m - \boldsymbol{\alpha\beta}' + \mathbf{\Gamma}_1$, $\mathbf{\Phi}_i = \mathbf{\Gamma}_i - \mathbf{\Gamma}_{i-1}$, $i = 2, \ldots, p-1$, $\mathbf{\Phi}_p = -\mathbf{\Gamma}_{p-1}$. In what follows, we take into account parameter uncertainty and describe how to evaluate the empirical distributions of generalised (orthogonalised) impulse response functions of both individual variables and cointegrating relations and persistence profiles. In the presence of exogenous $I(1)$ variables, they are given, respectively, by

$$\mathfrak{g}\left(n, \mathbf{z} : \zeta_i\right) = \frac{1}{\sqrt{\sigma_{\zeta,ii}}} \tilde{\mathbf{C}}_n \mathbf{H} \mathbf{\Sigma}_{\zeta\zeta} \mathbf{e}_i, \quad n = 0, 1, \ldots, \quad i = 1, \ldots, m, \tag{6.91}$$

$$\mathfrak{g}\left(n, \boldsymbol{\xi} : \zeta_i\right) = \frac{1}{\sqrt{\sigma_{\zeta,ii}}} \boldsymbol{\beta}' \tilde{\mathbf{C}}_n \mathbf{H} \mathbf{\Sigma}_{\zeta\zeta} \mathbf{e}_i, \quad n = 0, 1, \ldots, \quad i = 1, \ldots, m, \tag{6.92}$$

$$\mathfrak{o}\left(n, \mathbf{z} : \zeta_i^*\right) = \tilde{\mathbf{C}}_n \mathbf{H} \mathbf{P}_\zeta \mathbf{e}_i, \quad n = 0, 1, \ldots, \quad i = 1, \ldots, m, \tag{6.93}$$

$$\mathfrak{o}\left(n, \boldsymbol{\xi} : \zeta_i^*\right) = \boldsymbol{\beta}' \tilde{\mathbf{C}}_n \mathbf{H} \mathbf{P}_\zeta \mathbf{e}_i, \quad n = 0, 1, \ldots, \quad i = 1, \ldots, m, \tag{6.94}$$

$$h\left(\boldsymbol{\beta}_j' \mathbf{z}, n\right) = \frac{\boldsymbol{\beta}_j' \tilde{\mathbf{C}}_n \mathbf{H} \mathbf{\Sigma}_{\zeta\zeta} \mathbf{H}' \tilde{\mathbf{C}}_n' \boldsymbol{\beta}_j}{\boldsymbol{\beta}_j' \mathbf{H} \mathbf{\Sigma}_{\zeta\zeta} \mathbf{H}' \boldsymbol{\beta}_j}, \quad n = 0, 1, \ldots, \quad j = 1, \ldots, r, \tag{6.95}$$

where $\boldsymbol{\zeta}_t$ is *i.i.d.* $\left(\mathbf{0}, \mathbf{\Sigma}_{\zeta\zeta}\right)$, $\sigma_{\zeta,ij}$ is (i, j)th element of $\mathbf{\Sigma}_{\zeta\zeta}$, $\tilde{\mathbf{C}}_n = \Sigma_{j=0}^{n} \mathbf{C}_j$, with $\mathbf{C}_j$'s given by the recursive relations (6.21), $\mathbf{H}$ and $\mathbf{\Sigma}_{\zeta\zeta}$ are given in (6.88), $\boldsymbol{\xi}_t = \boldsymbol{\beta}' \mathbf{z}_{t-1}$, $\mathbf{e}_i$ is a selection vector of zeros with unity as its ith element, $\mathbf{P}_\zeta$ is a lower triangular matrix obtained by the Choleski decomposition of $\mathbf{\Sigma}_{\zeta\zeta} = \mathbf{P}_\zeta \mathbf{P}_\zeta'$, and $m = m_x + m_y$.

Suppose that the ML estimators of $\mathbf{\Phi}_i$, $i = 1, \ldots, p$, $\mathbf{a}_0$, $\mathbf{a}_1$, $\mathbf{H}$ and $\mathbf{\Sigma}_{\zeta\zeta}$ are given and denoted by $\hat{\mathbf{\Phi}}_i$, $i = 1, \ldots, p$, $\hat{\mathbf{a}}_0$, $\hat{\mathbf{a}}_1$, $\hat{\mathbf{H}}$ and $\hat{\mathbf{\Sigma}}_{\zeta\zeta}$, respectively.

To allow for parameter uncertainty, we use the bootstrap procedure and simulate S (*in-sample*) values of $\mathbf{z}_t$, $t = 1, 2, \ldots, T$, denoted by $\mathbf{z}_t^{(s)}$, $s = 1, \ldots, S$, where

$$\mathbf{z}_t^{(s)} = \sum_{i=1}^{p} \hat{\mathbf{\Phi}}_i \mathbf{z}_{t-i}^{(s)} + \hat{\mathbf{a}}_0 + \hat{\mathbf{a}}_1 t + \hat{\mathbf{H}} \zeta_t^{(s)}, \quad t = 1, 2, \ldots, T, \tag{6.96}$$

realisations are used for the initial values, $\mathbf{z}_{-1}, \ldots, \mathbf{z}_{-p}$, and $\boldsymbol{\zeta}_t^{(s)}$'s can be drawn either by parametric or non-parametric methods (see 7.3.3).

Having obtained the S sets of simulated in-sample values,

$$\left(\mathbf{z}_1^{(s)}, \mathbf{z}_2^{(s)}, \ldots, \mathbf{z}_T^{(s)}\right),$$

the VAR(p) model, (6.90), is re-estimated S times to obtain the ML estimates, $\hat{\mathbf{\Phi}}_i^{(s)}$, $\hat{\mathbf{a}}_0^{(s)}$, $\hat{\mathbf{a}}_1^{(s)}$, $\hat{\mathbf{H}}^{(s)}$ and $\hat{\mathbf{\Sigma}}_{\zeta\zeta}^{(s)}$, for $i = 1, 2, \ldots, p$, and $s = 1, 2, \ldots, S$. For each of these bootstrap replications, we then obtain the estimates of $\mathfrak{g}^{(s)}\left(n, \mathbf{z}^{(s)} : \zeta_i^{(s)}\right)$, $\mathfrak{g}^{(s)}\left(n, \boldsymbol{\xi}^{(s)} : \zeta_i^{(s)}\right)$, $\mathfrak{o}^{(s)}\left(n, \mathbf{z}^{(s)} : \zeta_i^{*(s)}\right)$, $\mathfrak{o}^{(s)}\left(n, \boldsymbol{\xi}^{(s)} : \zeta_i^{*(s)}\right)$, $h^{(s)}\left(\boldsymbol{\beta}_j' \mathbf{z}^{(s)}, n\right)$. Therefore, using these S sets of simulated estimates, we will obtain both empirical mean and confidence intervals of impulse response functions and persistence profiles.