Contents

Chapter 0

Introduction

In one of the first attempts to apply regression techniques to economic data, Moore (1914) estimated the "law of demand" for various commodities. In his application the percentage change in the price per unit is explained by a linear or cubic function of the percentage change of the produced quantities. His results are summarized as follows:

> *"The statistical laws of demand for the commodities corn, hay, oats, and potatoes present the fundamental characteristic which, in the classical treatment of demand, has been assumed to belong to all demand curves, namely, they are all negatively inclined".*

(Moore 1914, p. 76). Along with his encouraging results, Moore (1914) estimated the demand curve for raw steel (pig-iron). To his surprise he found a positively sloped demand curve and he claimed he have found a brand-new type of demand curve. Lehfeldt (1915), Wright (1915) and Working (1927) argued, however, that Moore has actually estimated a supply curve because the data indicated a moving demand curve that is shifted during the business cycle, whereas the supply curve appears relatively stable.

This was probably the first thorough discussion of the famous identification problem in econometrics. Although the arguments of Wright (1915) come close to a modern treatment of the problem, it took another 30 years until Haavelmo (1944) suggested a formal framework to resolve the identification problem. His elegant

probabilistic framework has become the dominating approach in subsequent years and was refined technically by Fisher (1966), Rothenberg (1971), Theil (1971) and Zellner (1971), among others.

Moore's (1914) estimates of "demand curves" demonstrate the importance of prior information for appropriate inference from estimated economic systems. This is a typical problem when collected data are used instead of experimental data that are produced under controlled conditions. Observed data for prices and quantities result from an interaction of demand and supply so that any regression between such variables require further assumptions to disentangle the effects of shifts in the demand and supply schedules.

This ambiguity is removed by using prior assumptions on the underlying economic structure. A structure is defined as a complete specification of the probability distribution function of the data. The set of all possible structures $\mathcal{S}$ is called a model. If the structures are distinguished by the values of the parameter vector θ that is involved by the probability distribution function, then the identification problem is equivalent to the problem of distinguishing between parameter points (see Hsiao 1983, p. 226). To select a unique structure as a probabilistic representation of the data, we have to verify that there is no other structure in $\mathcal{S}$ that leads to the same probability distribution function. In other words, an identified structure implies that there is no observationally equivalent structure in $\mathcal{S}$. In this case we say that the structure is identified (e.g. Judge et al. 1988, Chapter 14).

In this thesis I consider techniques that enables structural inference (that is estimation and tests in identified structural models) by focusing on a particular class of dynamic linear models that has become important in recent years. Since the books of Box and Jenkins (1970) and Granger and Newbold (1977), time series techniques have become popular for analysing the dynamic relationship between time series. Among the general class of the multivariate ARIMA (AutoRegressive Integrated Moving Average) model, the Vector Autoregressive (VAR) model turns out to be particularly convenient for empirical work. Although there are important reasons to allow also for moving average errors (e.g. Lütkepohl 1991, 1999), the

VAR model has become the dominant work horse in the analysis of multivariate time series. Furthermore, Engle and Granger (1987) show that the VAR model is an attractive starting point to study the long run relationship between time series that are stationary in first differences. Since Johansen's (1988) seminal paper, the cointegrated VAR model has become very popular in empirical macroeconomics.

An important drawback of the cointegrated VAR approach is that it takes the form of a "reduced form representation", that is, its parameters do not admit a structural interpretation. In this thesis, I review and supplement recent work that intends to bridge the gap between such reduced form VAR representations and structural models in the tradition of Haavelmo (1944). To do this, I first discuss in Chapter 1 aspects of the reduced form model that are fundamental for the subsequent structural analysis as well. In Chapter 2 I consider structural models that take the form of a linear set of simultaneous equations advocated by the influential *Cowles Commission*. An alternative kind of structural models are known as "Structural VAR models" or "Identified VAR models". These models are considered in Chapter 3. Problems due to the temporal aggregation of time series are studied in Chapter 4 and Chapter 5 deals with some new approaches to analyze nonlinear models. Chapter 6 concludes and makes suggestions for future work.

Chapter 1

The reduced form

Since Haavelmo (1944) it is common in econometrics to distinguish a structural model from the reduced form of an economic system. The reduced form provides a data admissible statistical representation of the economic system and the structural form can be seen as a reformulation of the reduced form in order to impose a particular view suggested by economic theory. Therefore, it is important to specify both the reduced and structural representation appropriately.

In this chapter the vector autoregressive (VAR) model is used as a convenient statistical representation of the reduced form relationship between the variables. Zellner and Palm (1974) and Wallis (1977) argue that under certain conditions the reduced (or final) form of a set of linear simultaneous equations can be represented as a VARMA (Vector-Autoregressive-Moving-Average) process. Here it is assumed that such a VARMA representation can be approximated by a VAR model with a sufficient lag order. A similar framework is used by Monfort and Rabemananjara (1990), Spanos (1990), Clemens and Mizon (1991), Juselius (1993) inter alia.

The reduced form model is represented by a conditional density function of the vector of time series y_t conditional on $\mathcal{I}_t$ denoted by $f(y_t|\mathcal{I}_t;\theta)$, where θ is a finite dimensional parameter vector (e.g. Hendry and Mizon 1983). Here we let $\mathcal{I}_t = \{y_{t-1}, y_{t-2}, \ldots\}$ and it is usually assumed that $f(\cdot|\cdot\,;\theta)$ is the normal density. Sometimes the conditioning set includes a vector of "exogenous variables". How-

ever, the distinction between endogenous and exogenous variables is considered as a structural problem and will be discussed in Chapter 2.

The specification of an appropriate VAR model as a statistical representation of the reduced form involves the following problems:

- The choice of the model variables.
- The choice of an appropriate variable transformation (if necessary).
- The selection of the lag order.
- The specification of the deterministic variables (dummy variables, time trend etc.)
- The selection of the cointegration rank.

This chapter contributes mainly to the last issue, that is, the selection of the cointegration rank. Problems involved by deterministic variables are only touched occasionally and the choice of an appropriate variable transformation is considered only in the sense that the choice of the cointegration rank may suggest that (some of) the variables must be differenced to obtain a stationary VAR representation. We do not discuss the choice of the lag order because there already exists an extensive literature dealing with this problem (cf. Lütkepohl 1991, Lütkepohl and Breitung 1997, and the references therein). Furthermore, it is assumed that the variables of the system are selected guided by to economic theory.

If the reduced form VAR model is specified, it can be estimated by using a maximum likelihood approach. For completeness I restate in Section 1.1 some well-known results on the estimation of stationary VAR models that are enhanced in Section 1.3 by introducing deterministic terms. Some useful representations of cointegrated VAR models are considered Section 1.3. Section 1.4 suggests a unifying approach for the estimation of the cointegration vectors and Section 1.5 discusses different approaches for testing the cointegration rank.

1.1 The stationary VAR model

Assume that the $n \times 1$ times series vector y_t is stationary with $E(y_t) = 0$ and $E(y_t y_{t+j}') = \Gamma_j$ such that there exists a Wold representation of the form:

$$y_t = \varepsilon_t^* + B_1\varepsilon_{t-1}^* + B_2\varepsilon_{t-2}^* + \cdots \quad (1.1)$$

$$= B(L)\varepsilon_t^* \ , \quad (1.2)$$

where $B(L) = I_n + B_1 L + B_2 L^2 + \cdots$ is a (possibly infinite) $n \times n$ lag polynomial and ε_t^* is a vector of white noise errors with positive definite covariance matrix $E(\varepsilon_t^* \varepsilon_t^{*\prime}) = \Sigma^*$. Furthermore, it is assumed that the matrix polynomial $|B(z)| \neq 0$ for all $|z| \leq 1$. If in addition the coefficient matrices $B_1, B_2, \ldots$ obey $\sum_{j=1}^{\infty} j^{1/2}||B_j|| < \infty$, where $||B_j|| = [tr(B_j B_j')]^{1/2}$, then there exists a VAR representation of the form

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + \varepsilon_t^* \ .$$

In practice this infinite VAR representation is approximated by a finite order VAR[p] model:

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t \ , \quad (1.3)$$

where $\varepsilon_t = \varepsilon_t^* + A_{p+1} y_{t-p-1} + A_{p+2} y_{t-p-2} + \cdots$ and, thus, the error vector ε_t includes the approximation error $\eta_t^p = A_{p+1} y_{t-p-1} + A_{p+2} y_{t-p-2} + \cdots$. In what follows it is assumed that the approximation error is "small" relative to the innovation ε_t^* and so I am able to neglect the term η_t^p. With respect to the consistency and asymptotic normality of the least-squares estimator, Lewis and Reinsel (1985) have shown that the approximation error is asymptotically negligible if for $\to \infty$ and $p \to \infty$

$$\sqrt{T} \sum_{j=p+1}^{\infty} ||A_j|| \to 0 \ . \quad (1.4)$$

In many cases this condition is satisfied if p increases with the sample size T but at a smaller rate than T. For example, if y_t is generated by a finite order MA process, then $p(T) = T^{1/\delta}$ with $\delta > 3$ is sufficient for (1.4) to hold (see Lütkepohl 1991, p. 307).

Unfortunately, such asymptotic conditions are of limited use in practice. First, there is usually a wide range of valid rates for $p(T)$. For MA models we may use $p(T) = T^{1/3.01}$ as well as $p(T) = T^{1/100}$. Obviously, both possible rules will render quite different model orders. Second, a factor c may be introduced such that $p(T) = cT^{1/\delta}$. For asymptotic considerations the factor c is negligible as long as $c > 0$. However, in small samples it can make a big difference if $c = 0.1$ or $c = 20$, for example. In practice it is therefore useful to employ selection criteria for the choice of the autoregressive order p (see Lütkepohl 1991, Chapter 4).

For later reference I now summarize the basic assumptions of the VAR model used in the subsequent sections.

Assumption 1.1 *(Stationary VAR[p] model). Let $y_t = [y_{1t}, \ldots, y_{nt}]'$ be an $n \times 1$ vector of stationary time series with the VAR[p] representation*

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t \ , \tag{1.5}$$

where $\{\varepsilon_t\}$ is white noise with $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_t') = \Sigma$ and Σ is a positive definite $n \times n$ matrix.

Usually, the coefficient matrices are unknown and can be estimated by multivariate least-squares. Let $x_t = [y_{t-1}', \ldots, y_{t-p}']'$ and $A = [A_1, \ldots, A_p]$ so that the VAR[p] model can be written as $y_t = Ax_t + \varepsilon_t$. Then the least-squares estimator is given by

$$\widehat{A} = \sum_{t=p+1}^{T} y_t x_t' \left(\sum_{t=p+1}^{T} x_t x_t' \right)^{-1} .$$

Under Assumption 1.1 the least-squares estimator is consistent and asymptotically normally distributed with

$$\sqrt{T}\ \mathrm{vec}(\widehat{A} - A) \xrightarrow{d} N(0, V_{\widehat{A}}) \ ,$$

where

$$V_{\widehat{A}} = [E(x_t x_t')]^{-1} \otimes \Sigma \ .$$

If in addition it is assumed that ε_t is normally distributed, then the least-squares estimator is asymptotically equivalent to the maximum likelihood estimator and, hence, the least-squares estimator is asymptotically efficient.

The covariance matrix Σ can be consistently estimated using

$$\widehat{\Sigma} = \frac{1}{T-p}\sum_{t=p+1}^{T}(y_t - \widehat{A}x_t)(y_t - \widehat{A}x_t)' \ . \tag{1.6}$$

1.2 Deterministic terms

So far I have assumed that $E(y_t) = 0$. In most applications, however, y_t has a nonzero mean such as a constant or a linear time trend. Assume that the mean is a linear function of the $k \times 1$ vector d_t so that

$$E(y_t) = Cd_t \ . \tag{1.7}$$

For example, the elements of the matrix d_t may be the terms of a polynomial time trend or dummy variables.

Another possibility to accommodate a nonzero mean is to include deterministic terms in the autoregressive representation

$$y_t = C^*d_t^* + A_1y_{t-1} + \cdots + A_py_{t-p} + \varepsilon_t \ . \tag{1.8}$$

The relationship between the mean function implied by (1.7) and (1.8) is found from solving the difference equation

$$Cd_t - A_1Cd_{t-1} - \cdots - A_pCd_{t-p} = C^*d_t^* \ .$$

If the elements of d_t can be represented as t^k for $k = \{0, 1, 2, \ldots\}$, then $d_t^* = d_t$. However in other cases $C \neq C^*$, in general.

The matrix C^* in (1.8) can be asymptotically efficiently estimated by OLS. The mean function in (1.7) is asymptotically efficiently estimated by applying a GLS procedure to

$$y_t = Cd_t + u_t \ , \tag{1.9}$$

where

$$u_t = A_1u_{t-1} + \cdots + A_pu_{t-p} + \varepsilon_t \ .$$

The GLS estimator of C results as

$$\widetilde{C} = \sum_{t=p+1}^{T}\tilde{y}_t\tilde{d}_t'\left(\sum_{t=p+1}^{T}\tilde{d}_t\tilde{d}_t'\right)^{-1} \ , \tag{1.10}$$

where

$$\begin{aligned} \tilde{y}_t &= y_t - A_1 y_{t-1} - \cdots - A_p y_{t-p} \\ \tilde{d}_t &= d_t - A_1 d_{t-1} - \cdots - A_p d_{t-p} \ . \end{aligned}$$

To obtain a feasible GLS procedure, the unknown matrices $A_1, \ldots, A_p$ must be replaced by consistent estimates.

As shown by Grenander and Rosenblatt (1957, Sec. 7) there are important cases where the OLS estimator of C is as efficient as the GLS estimator. For example, this is the case if the elements of d_t are the terms of a polynomial trend regression, i.e., $d_t = (t^j)_{j=0,\ldots,k}$. Another example are seasonal dummy variables, which can be estimated efficiently by OLS (cf Grenander and Rosenblatt 1957, p. 246).

Besides trend polynomials and seasonal dummies the deterministic term often includes "impulse-dummies" and "step-dummies". Since such terms are not considered by Grenander and Rosenblatt (1957), the following theorem states that for step-dummies a similar result applies while for an impulse-dummy the OLS estimate has a different limiting distribution than the GLS estimate. As in Grenander and Rosenblatt (1957) I consider a univariate process but the generalization to a vector process is straightforward.

THEOREM 1.1 *Let d_t^p and d_t^s denote an impulse-dummy and a step-dummy defined as*

$$d_t^p(\lambda) = \begin{cases} 1 & \text{for } t = T_0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad d_t^s(\lambda) = \begin{cases} 0 & \text{for } t \leq T_0 \\ 1 & \text{for } t > T_0 \end{cases}$$

where $T_0 = [\lambda T]$, $0 < \lambda < 1$, and $[\cdot]$ indicates the integer part of the argument.
(i) For the regression model $y_t = c^s d_t^s(\lambda) + u_t$, where $u_t = \alpha_1 u_{t-1} + \ldots + \alpha_p u_{t-p} + \varepsilon_t$ is a stationary AR[p] process, the OLS and GLS estimates have the same limiting distribution.
(ii) The respective estimates for the model $y_t = c^p d_t^p(\lambda) + u_t$ have a different distribution as long as u_t is different from a white noise process.

PROOF: (i) In the model with a step-dummy $d_t^s(\lambda)$ we have

$$T^{-1/2}\sum_{t=1}^{T} d_t^s(\lambda)u_t = T^{-1/2}\sum_{t=T_0}^{T} u_t \xrightarrow{d} N\left(0, \frac{\sigma^2(1-\lambda)}{(1-\alpha_1-\cdots-\alpha_p)^2}\right) .$$

Furthermore, $T^{-1}\sum_{t=1}^{T} d_t^s(\lambda)^2 = (1-\lambda)$. It follows that the OLS estimator of c^s is asymptotically distributed as

$$N\left(0, \frac{\sigma^2}{(1-\lambda)(1-\alpha_1-\cdots-\alpha_p)^2}\right) .$$

To derive the limiting distribution of the GLS estimator, let

$$\widetilde{d}_t^s(\lambda) = d_t^s(\lambda) - \widehat{\alpha}_1 d_{t-1}^s(\lambda) - \cdots - \widehat{\alpha}_p d_{t-p}^s(\lambda) .$$

Using $\widetilde{d}_t^s(\lambda) = 1 - \widehat{\alpha}_1 - \ldots - \widehat{\alpha}_p$ for $t > T_0 + p$ we obtain

$$\operatorname*{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T}[\widetilde{d}_t^s(\lambda)]^2 = (1-\lambda)(1-\alpha_1-\cdots-\alpha_p)^2$$

and

$$T^{-1/2}\sum_{t=1}^{T}\widetilde{d}_t^s(\lambda)\varepsilon_t \xrightarrow{d} N\left(0, \sigma^2(1-\lambda)(1-\alpha_1-\cdots-\alpha_p)^2\right) .$$

Combining these results it follows that

$$\sqrt{T}\,\frac{\sum_{t=1}^{T}\widetilde{d}_t^s(\lambda)\varepsilon_t}{\sum_{t=1}^{T}[\widetilde{d}_t^s(\lambda)]^2} \xrightarrow{d} N\left(0, \frac{\sigma^2}{(1-\lambda)(1-\alpha_1-\cdots-\alpha_p)^2}\right),$$

and, thus, the GLS estimator has the same asymptotic distribution as the OLS estimator.

(ii) For the model with an impulse-dummy $d_t^p(\lambda)$ we have for the OLS estimator $\widehat{c}^p = y_{T_0}$ so that

$$\widehat{c}^p - c^p \xrightarrow{d} N(0, \sigma_u^2) ,$$

where σ_u^2 denotes the variance of u_t. For the GLS estimator we have $\widetilde{c}^p = y_{T_0} - \widehat{\alpha}_1 y_{T_0-1} - \cdots - \widehat{\alpha}_p y_{T_0-p}$ and, thus,

$$\widetilde{c}^p - c^p \xrightarrow{d} N(0, \sigma_\varepsilon^2) .$$

Unless $\sigma_u^2 = \sigma_\varepsilon^2$, that is u_t is white noise, the limiting distributions for the estimators of c^p are different, in general. ■

Since the least-squares estimation of a VAR system is equivalent to the separate estimation of the equations, it is straightforward to show that this result also holds for a multivariate estimation of the VAR system. Furthermore it can be shown by using the same techniques as in Theorem 1.1 that in a regression with a polynomial trend dummy defined as $d_t^j = d_t^1(\lambda)t^j$ the OLS and GLS estimates have the same limiting distribution as well.

The Grenander-Rosenblatt theorem and its extension to step dummies in Theorem 1.1 implies that for estimating the parameters of a VAR process the estimation method (OLS or GLS) is irrelevant for the asymptotic properties.[1] Furthermore the invariance of the ML estimation implies that the ML estimation of λ is identical to $\widetilde{\lambda} = g(\widehat{\theta})$, where $g(\cdot)$ is a matrix function $\mathcal{R}^k \to \mathcal{R}^k$ with a regular matrix of first derivatives and θ, λ are $k \times 1$ vectors. Since there exists a one-to-one relationship between C and C^* it therefore follows that asymptotically the estimates of $A_1, \ldots, A_p$ and Σ are not affected whether the process is demeaned by estimating the mean in (1.7) or in (1.8). Thus I present only the limiting distributions for the case of an OLS based on (1.8).

THEOREM 1.2 *Let $y_t - Cd_t$ be a stationary $n \times 1$ vector generated by a VAR[p] as in Assumption 1.1. Furthermore assume that there exits a diagonal matrix $\Upsilon_T = diag[T^{\delta_1}, \ldots, T^{\delta_k}]$ with $\delta_r > 0$ for $r = 1, \ldots, k$ such that the limiting matrix*

$$\overline{\Gamma}_d = \lim_{T\to\infty} \Upsilon_T^{-1} \sum_{t=p+1}^{T} \widetilde{d}_t \widetilde{d}_t' \qquad \text{where } \widetilde{d}_t = d_t - A_1 d_{t-1} - \cdots - A_p d_{t-p}$$

exists and is positive definite. Let $\widehat{a} = vec(\widehat{A})$, $\widehat{\sigma} = vech(\widehat{\Sigma})$ and $\widehat{c} = vec(\widehat{C})$, where $vec(\widehat{A})$ stacks the columns of $\widehat{A}$ into a vector, $vech(\widehat{\Sigma})$ stacks the non-redundant

[1] Notice that in Grenander and Rosenblatt (1957) as well as in Theorem 1.1 it is assumed that $y_t - E(y_t)$ is stationary. The results do not apply if the process has one or more roots on the unit circle (see Lütkepohl and Saikkonen 2000).

elements of the columns of $\widehat{\Sigma}$ *into a* $n(n+1)/2$ *vector and*

$$\begin{aligned}\widehat{A} &= \sum_{t=p+1}^{T} \bar{y}_t \bar{x}_t' \left(\sum_{t=p+1}^{T} \bar{x}_t \bar{x}_t' \right)^{-1} \\ \widehat{\Sigma} &= T^{-1} \sum_{t=p+1}^{T} (\bar{y}_t - \widehat{A}\bar{x}_t)(\bar{y}_t - \widehat{A}\bar{x}_t)' \ ,\end{aligned}$$

$\bar{y}_{t-j} = y_{t-j} - \widehat{C} d_t$, $\bar{x}_t = [\bar{y}_{t-1}', \ldots, \bar{y}_{t-p}']'$. *As* $T \to \infty$

$$\begin{bmatrix} \sqrt{T}(\widehat{a} - a) \\ \sqrt{T}(\widehat{\sigma} - \sigma) \\ \Upsilon_T^{1/2}(\widehat{c} - c) \end{bmatrix} \xrightarrow{d} N(0, diag[V_a, V_\sigma, V_c]) \ ,$$

where

$$\begin{aligned} V_a &= \Gamma_x^{-1} \otimes \Sigma \\ V_\sigma &= 2D_n^+ (\Sigma \otimes \Sigma) D_n^{+\prime} \\ V_c &= \Gamma_d^{-1} \otimes \Sigma \ . \end{aligned}$$

where $D_n^+ = (D_n' D_n)^{-1} D_n'$ *is the Moore-Penrose generalized inverse of the* $n^2 \times n(n+1)/2$ *duplication matrix* D_n *(cf. Lütkepohl 1991, p. 84).*

PROOF: The proof is a straightforward extension of the proof in Lütkepohl (1991, Sec. 3.4.3). ■

Since the asymptotic covariance matrix is block diagonal, it follows that any other consistent estimator for C besides $\widehat{C}$ can be used without affecting the asymptotic properties of the other estimators. Thus, even if a mean function is used where the Grenander-Rosenblatt theorem does not apply, the limiting distributions of $A_1, \ldots, A_p$ and Σ are not affected by the estimator of C as long as C is estimated consistently. Furthermore a possible overspecification of the deterministic terms does not affect the asymptotic properties of the estimators of $A_1, \ldots, A_p$ and Σ.

1.3 Alternative representations of cointegrated VARs

As already observed by Box and Jenkins (1970), many economic variables must be differenced to become stationary. They introduced the notation that a (mean-adjusted) variable is called $I(d)$ (integrated of order d) if at least d differences are necessary to achieve a stationary series. Modeling integrated time series in a multivariate system raises a number of important problems and since the late 80s various inference procedures were suggested to deal with such problems. It is not the intention to give a detailed account of all developments in this area.[2] Rather, I focus on the most important developments as well as on my own work in this area.

Consider the VAR[p] model

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t \ , \tag{1.11}$$

where for convenience we leave out deterministic terms like constants, time trends and dummy variables. As noted in Section 1.1, the process is stationary if the polynomial $A(L) = I_n - A_1 L - \cdots - A_p L^p$ has all roots outside the unit circle, that is, if

$$|I_n - A_1 z - \cdots - A_p z^p| \neq 0 \quad \text{for all } |z| \leq 1 \ .$$

On the other hand, if $|A(z_j)| = 0$ for $|z_j| = 1$ and $j = 1, 2, \ldots, q$, we say that the process has q unit roots. In what follows, I will focus on unit roots "at frequency zero", i.e., $z_j = 1$ for $j = 1, 2, \ldots, q$. Complex unit roots are important in the analysis of the seasonal behavior of the time series but are left out here for ease of exposition.

To assess the properties of the process, it is not sufficient to consider merely the number of unit roots. For example, assume that the process for $y_t = [y_{1t}, y_{2t}, y_{3t}]'$ has two unit roots. This may be due to fact that $[\Delta y_{1t}, \Delta y_{2t}, y_{3t}]$ is stationary, where $\Delta = 1 - L$ denotes the difference operator. Another possibility is that

[2] For recent surveys see, e.g., Hamilton (1994), Watson (1994), Mills (1998), Lütkepohl (1999a).

$[\Delta^2 y_{1t}, y_{2t}, y_{3t}]$ is stationary, i.e., y_{1t} is $I(2)$ in the terminology of Box and Jenkins (1970). Finally the unit roots may be due to the fact that $[\Delta y_{1t}, \Delta y_{2t}, y_{3t} - by_{1t}]$ is stationary. In this case y_{3t} and y_{1t} are integrated but there exists a linear combination $y_{3t} - by_{1t}$ that is stationary. In this case we say that the variables y_{3t} and y_{1t} are *cointegrated*.

To facilitate the analysis, it is convenient to rule out that components of y_t are integrated with a degree larger than one. The analysis of $I(2)$ variables is considerably more complicated than the analysis of $I(1)$ variables (see, e.g., Stock and Watson 1993, Johansen 1995c), and in empirical practice the case with $I(1)$ variables is more important. We therefore make the following assumption:

Assumption 1.2 *The vector Δy_t is stationary.*

The VECM representation. Following Engle and Granger (1997) it is convenient to reformulate the VAR system as a "vector error correction model" (VECM) given by

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + \varepsilon_t \ , \tag{1.12}$$

where $\Pi = \sum_{j=1}^{p} A_j - I_n$ and $\Gamma_j = -\sum_{i=j+1}^{p} A_i$. This representation can be used to define cointegration in a VAR system.

DEFINITION 1.1 *(Cointegration). A VAR[p] system as defined in Assumption 1.1 is called cointegrated with rank r, if $r = rk(\Pi)$ with $0 < r < n$.*

If Π has a reduced rank then there exists a factorisation $\Pi = \alpha\beta'$ such that α and β are $n \times r$ matrices. Furthermore, from Assumption 1.2 and (1.12) it follows that $\Pi y_{t-1} = \alpha\beta' y_{t-1}$ is stationary. Since α is a matrix of constants, $\beta' y_t$ defines r stationary linear combinations of y_t. Furthermore, it follows that Δy_t has a MA representation of the form

$$\begin{aligned} \Delta y_t &= \varepsilon_t + C_1 \varepsilon_{t-1} + C_2 \varepsilon_{t-1} + \cdots \\ \Delta y_t &= C(L)\varepsilon_t \ . \end{aligned}$$

As shown by Johansen (1991), the MA representation can be reformulated as

$$\Delta y_t = C(1)\varepsilon_t + C^*(L)\Delta\varepsilon_t \ , \tag{1.13}$$

where $C^*(L) = C_0^* + C_1^* L + C_2^* L^2 + \cdots$ has all roots outside the complex unit circle,

$$C(1) = \beta_\perp [\alpha_\perp' \Gamma(1) \beta_\perp]^{-1} \alpha_\perp' \ , \tag{1.14}$$

and $\Gamma(1) = I + \sum_{j=1}^{p-1} \Gamma_j$ (Johansen 1991, Theorem 4.1). Assumption 1.2 implies that the matrix $[\alpha_\perp' \Gamma(1) \beta_\perp]$ is invertible.

A canonical representation. The VECM representation used by Engle and Granger (1987), Johansen (1995a) and many others is not the only way to represent a cointegrated system. Phillips (1991) uses a "triangular representation" resulting from the partitioning $y_t = [y_{1t}', y_{2t}']'$, where y_{1t} and y_{2t} are $r \times 1$ and $(n-r) \times 1$ subvectors. In the subsequent sections it will be convenient to use another representation that is based on the following theorem.

THEOREM 1.3 *Let y_t be a $n \times 1$ vector of cointegrated variables with $0 < r < n$ and Δy_t is stationary. Then there exists an invertible matrix $Q = [\beta^*, \gamma^*]'$, where β^* is an $n \times r$ cointegration matrix and γ^* is an $n \times (n-r)$ matrix linearly independent of β^* such that*

$$\begin{aligned} x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} &= Q y_t = \begin{bmatrix} \beta^{*\prime} y_t \\ \gamma^{*\prime} y_t \end{bmatrix} \\ T^{-1/2} \sum_{i=1}^{[aT]} x_{1i} &\Rightarrow W_r(a) \\ T^{-1/2} x_{2,[aT]} &\Rightarrow W_{n-r}(a), \end{aligned}$$

where $[aT]$ signifies the integer part of aT and $W_r(a)$, $W_{n-r}(a)$ are uncorrelated r and $(n-r)$ dimensional Brownian motions with unit covariance matrix.

PROOF: From the MA representation (1.13) we have

$$\begin{aligned} \beta' y_t &= \beta' C^*(L) \varepsilon_t \\ &= \beta' C^*(1) \varepsilon_t + \beta' C^{**}(L) \Delta \varepsilon_t \\ \gamma' y_t &= \gamma' C(1) \sum_{i=1}^{t} \varepsilon_i + C^*(L) \varepsilon_t \ , \end{aligned}$$

where γ is an $n \times (n-r)$ matrix linearly independent of β and $C^{**}(L) = [C^*(L) - C^*(1)](1-L)^{-1}$ has all roots outside the complex unit circle. The expression $(1-L)^{-1}$ is equivalent to the polynomial $1 + L + L^2 + L^3 + \ldots$. Let R be a lower block diagonal matrix such that

$$R = \begin{bmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{bmatrix}$$

and

$$\begin{bmatrix} \beta' C^*(1) \\ \gamma' C \end{bmatrix} \Sigma \begin{bmatrix} \beta' C^*(1) \\ \gamma' C \end{bmatrix}' = RR'.$$

Then, by using

$$Q = R^{-1} \begin{bmatrix} \beta' \\ \gamma' \end{bmatrix} = \begin{bmatrix} \beta^{*\prime} \\ \gamma^{*\prime} \end{bmatrix}$$

it follows that $T^{-1/2} \sum_{i=1}^{[aT]} x_{1i}$ and $T^{-1/2} x_{2,[aT]}$ converge weakly to the standard Brownian motions W_r and W_{n-r}, respectively (e.g. Phillips and Durlauf 1986). ■

This representation is called "canonical" since it transforms the system into r asymptotically independent stationary and $n-r$ nonstationary components with uncorrelated limiting processes. Since this representation separates the stationary and non-stationary components from the system it is convenient for the analysis of the asymptotic properties of the system. Furthermore, the representation is related to Phillips' (1991) triangular representation given by

$$y_{1t} = B y_{2t} + u_t \tag{1.15}$$

$$\Delta y_{2t} = v_t \ , \tag{1.16}$$

where u_t and v_t are $I(0)$. However, (1.15) implies the normalization $\beta = [I_r, -B]'$ that is not assumed in the former representation.

The SE representation. Another convenient reformulation of the system is the representation in the form of a traditional system of **S**imultaneous **E**quations (SE). This representation imposes r^2 normalization restrictions on the loading

matrix α. Specifically, we let

$$\alpha_* = \begin{bmatrix} \phi' \\ I_r \end{bmatrix}, \tag{1.17}$$

where ϕ is an unrestricted $r \times (n-r)$ matrix. Obviously, $\phi' = \alpha_1 \alpha_2^{-1}$, where $\alpha = [\alpha_1', \alpha_2']'$ and α_2 is an invertible $r \times r$ matrix. Note that the variables in y_t can always be arranged such that α_2 is invertible.

The system (1.12) is transformed by using the matrix

$$C_0 = \begin{bmatrix} I_{n-r} & -\phi' \\ 0 & I_r \end{bmatrix}$$

so that

$$C_0 \Delta y_t = \Pi^* y_{t-1} + \Gamma_1^* \Delta y_{t-1} + \cdots + \Gamma_{p-1}^* \Delta y_{t-p+1} + u_t^* \ , \tag{1.18}$$

where $\Pi^* = C_0 \alpha \beta' = [0, \pi_2]'$, $\Gamma_j^* = C_0 \Gamma$ and $\pi_2' = \alpha_2 \beta'$ is the lower $r \times n$ block of $\Pi = \alpha \beta'$. Let $y_t = [y_{1t}', y_{2t}']'$, then (1.18) can be represented by the two subsystems:

$$\Delta y_{1t} = \phi' \Delta y_{2t} + \text{lags} + w_{1t} \tag{1.19}$$

$$\Delta y_{2t} = \pi_2' y_{t-1} + \text{lags} + w_{2t} \tag{1.20}$$

where "lags" represent the terms due to $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$. Although, the system (1.18) is written as a structural model considered by Hsiao (1997), it is not a "structural" system in the usual sense. It should further be noticed that in (1.19) the rank restrictions show up in the form of $(n-r)^2$ linear over-identifying restrictions. The remaining r equations in (1.20) are just identified. The SE representation turns out to be useful for imposing restrictions on the parameters (see Chapter 2).

1.4 Weak exogeneity in stationary VARs

An important structural assumption is the distinction between exogenous and endogenous variables. Let $z_t' = [y_t', x_t']$, where y_t and x_t are $m \times 1$ and $k \times 1$ vectors of time series, respectively. Furthermore we define the increasing sigma-field $\mathcal{Z}_t = \{z_t, z_{t-1}, z_{t-2}, \ldots\}$. Then, according to Engle et al. (1983) the variable

x_t is (weakly) exogenous if we can factorize the joint density of z_t with parameter vector $\theta = [\theta_1', \theta_2']'$ as

$$f(z_t|\mathcal{Z}_{t-1}, ; \theta) = f_1(y_t|x_t, \mathcal{Z}_{t-1}; \theta_1) \cdot f_2(x_t|\mathcal{Z}_{t-1}; \theta_2)$$

such that the parameter vector θ_1 of the conditional density $f_1(\cdot|\cdot\ ; \theta_1)$ does not depend on the parameter vector θ_2 of the conditional density $f_2(\cdot|\cdot\ ; \theta_2)$, and θ_1 and θ_2 are variation free, that is, a change in θ_2 has no effect on θ_1 (cf Engle et al. 1983).

In the dynamic structural model given in (**??**) the parameter θ_1 comprises the elements of the matrices $\Gamma_0, \ldots, \Gamma_p, B_0, \ldots, B_p$ and the non-redundant elements of Σ. To embed the structural form in a corresponding form derived from the VAR representation of the system we define the matrix

$$Q = \begin{bmatrix} I_m & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_k \end{bmatrix} ,$$

where the covariance matrix of the VAR innovations $\Sigma = E(\varepsilon_t\varepsilon_t')$ is decomposed as

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

such that Σ_{11} is the covariance matrix of the innovations of y_t and Σ_{22} is the covariance matrix of the innovations of x_t. Multiplying the VAR system (1.5) by Q yields a block recursive system of the form

$$\Phi_0 z_t = \Phi_1 z_{t-1} + \cdots + \Phi_p z_{t-p} + v_t \tag{1.21}$$

or for the first m equations

$$y_t = \Phi_{12} x_t + \Phi_{1,1} z_{t-1} + \cdots + \Phi_{1,p} z_{t-p} + v_{1t} , \tag{1.22}$$

where $\Phi_0 = Q$, $\Phi_{12} = \Sigma_{12}\Sigma_{22}^{-1}$ and the matrix $\Phi_{1,j}$ denotes the upper $m \times n$ block of the matrix $\Phi_j = QA_j$ for $j = 1, \ldots, p$. Similarly, v_t is partitioned as $Q\varepsilon_t = v_t = [v_{1t}', v_{2t}']'$, where the covariance matrix of v_t is block diagonal with respect to v_1 and v_2.

In many applications, economic theory does not imply restrictions on the short run dynamics of the system.[3] Thus we follow Monfort and Rabemananjara (1990) and assume that there are no restrictions on the matrices $\Gamma_1, \Gamma_2, \ldots, \Gamma_p$. Premultiplying (1.22) by B_0 and comparing the result with (**??**) gives rise to the following characterization of a vector of weakly exogenous variables.

DEFINITION 1.2 *Let $z_t = [y_t', x_t']'$ be an $n \times 1$ time series vector with a stationary VAR[p] representation as given in Assumption 1.1 and $\varepsilon_t \sim N(0, \Sigma)$. The subvector x_t is weakly exogenous for the parameters of the structural form (**??**), iff*

$$B_0 \Phi_{12} = \Gamma_0 \ . \tag{1.23}$$

It is straightforward to show that this definition is indeed equivalent to the definition of weak exogeneity suggested by Engle et al. (1983). From (1.22) it follows that

$$E(y_t | x_t, z_{t-1}, \ldots, z_{t-p}) = \Phi_{12} x_t + \Phi_{1,1} z_{t-1} + \cdots + \Phi_{1,p} z_{t-p}$$

Accordingly, if x_t is predetermined, the parameters of the structural form result as functions from the parameters of the conditional mean and variance of y_t given $x_t, z_{t-1}, \ldots, z_{t-p}$. Under normality it follows that the vector of structural parameters θ_1 in $f_1(y_t | x_t, \mathcal{Z}_{t-1}; \theta_1)$ does not depend on θ_2 in $f_2(x_t | \mathcal{Z}_{t-1}; \theta_2)$.

If there are (cross-equation) restrictions on the matrices $B_1, \ldots, B_p$ some extra conditions are needed to ensure that x_t is weakly exogenous (see Monfort and Rabemananjara 1990). An important example for such restrictions are rank restrictions in cointegrated systems.

Assume that the structural analog of a cointegrated system can be represented as

$$C_0 z_t = C_1 z_{t-1} + C_2 z_{t-2} + \cdots + C_p z_{t-p} + e_t \ , \tag{1.24}$$

where $z_t = [y_t', x_t']'$ is partitioned such that

$$C_0 = \begin{bmatrix} B_0 & -\Gamma_0 \\ C_{xx} & C_{xy} \end{bmatrix}$$

[3] A notable exception are models based on dynamic maximization assuming rational expectations (e.g. Wickens 1982).

and the upper $m \times n$ block of C_j $(j = 1, \ldots, p)$ is equal to $[B_j, \Gamma_j]$. The error vector $e_t = [u_t', w_t']'$ is white noise. Accordingly, the upper m equations of the system yield a traditional structural form as given in (**??**). The structural system as given in (1.24) is obtained from the reduced form VAR representation (1.5) by a pre-multiplication with the matrix C_0.

Premultiplying the reduced form VECM (1.12) by C_0 the structural form of the cointegrated system is obtained (cf Johansen and Juselius 1994)

$$B_0 \Delta y_t = \alpha_1^* \beta' z_{t-1} + \Gamma_0 \Delta x_t + \Gamma_1^* \Delta z_{t-1} + \cdots + \Gamma_{p-1}^* \Delta z_{t-p+1} + u_t \ , \tag{1.25}$$

where Γ_j^* is the upper $m \times n$ block of the matrix $C_0 \Gamma_j$ and $\alpha_1^* = [\Gamma_0, B_0]\alpha$. Without additional restrictions both expectations $E(y_t | x_t, z_{t-1}, \ldots, z_{t-p})$ and $E(y_t | z_{t-1}, \ldots, z_{t-p})$ depend on the error correction term $\beta' z_{t-1}$, in general. It follows that the parameter vectors θ_1 in $f_1(y_t | x_t, \mathcal{Z}_{t-1}; \theta_1)$ and θ_2 in $f_2(x_t | \mathcal{Z}_{t-1}; \theta_2)$ depend on β and, hence, x_t is not weakly exogenous in the sense of Engle et al. (1983). However, if the lower $k \times n$ block of α (resp. Π) is a zero matrix, that is, the error correction term does not enter the "marginal model", then the vector θ_1 does not depend on β (see Boswijk and Urbain (1997) and the references therein).

As before let

$$E(y_t | x_t, z_{t-1}, \ldots, z_{t-p}) = \Phi_{12} x_t + \Phi_{1,1} z_{t-1} + \cdots + \Phi_{1,p} z_{t-p} \ .$$

If there are no restrictions on $\Gamma_1^*, \ldots, \Gamma_{p-1}^*$, Definition 1.2 can be straightforwardly adapted to the case of weak exogeneity in a cointegrated system.

DEFINITION 1.3 *Let $z_t' = [y_t', x_t']$ be a $(m + k) \times 1$ time series vector with a cointegrated VAR[p] representation given in (1.12) and $\varepsilon_t \sim N(0, \Sigma)$. The subvector x_t is weakly exogenous with respect to the structural VECM given in (1.25), iff*

$$\begin{aligned} & (i) \qquad B_0 \Phi_{12} = \Gamma_0 \\ \textit{and} \quad & (ii) \qquad \alpha_2 = 0, \end{aligned}$$

where α_2 is the lower $k \times r$ block of the matrix α.

This definition of weak exogeneity is more general than the definition suggested by Johansen (1992b), who assumes that $B_0 = I$ and Boswijk and Urbain (1997), who assume that the matrix B_0 is block triangular. In the latter case , the condition (i) of Definition 1.3 can be replaced by the condition (i') $E(u_t w_t') = 0$, where $e_t = [u_t', w_t']$ is the vector of disturbances in (1.24).

If x_t is weakly exogenous for the structural parameters $B_0, \Gamma_0, \Gamma_1^*, \ldots, \Gamma_{p-1}^*$, then the partial system (1.25) can be estimated efficiently without involving the marginal model for x_t (Johansen 1992b). In particular, if $m = 1$, the parameters can be estimated efficiently by OLS on the single equations. Dolado (1992) shows that condition (ii) in Definition 1.3 is not necessary to establish the efficiency of the OLS estimator. The reason is that for an efficient OLS estimator it is required that

$$\lim_{T\to\infty} E(\Delta x_T u_T') = 0 \ .$$

This condition is satisfied by imposing $\alpha_2 = 0$ but it may also be fulfilled by imposing restrictions on $\beta_\perp$ (cf. Dolado 1992).

1.5 Identifying restrictions

Consider the structural VECM model given by (1.25). To achieve a unique identification of the structural form, restrictions on the parameters are required. Following Hsiao (1997) I first make the following assumption:

Assumption 1.3 *It is assumed that* $|B_0| \neq 0$ *and* $T^{-2}\sum_{t=1}^{T} x_t x_t'$ *converges in distribution to a nonsingular random matrix.*

Hsiao (1997) shows that this assumption implies that the roots of the polynomial $B_0 + B_1 L + \cdots + B_p L^p$ lie outside the unit circle and, thus, the usual stability condition for dynamic systems (e.g. Davidson and Hall 1991) is satisfied. An important property of the stable dynamic system is that the distribution of y_t conditional on x_t does not depend on initial conditions.

Johansen and Juselius (1994) distinguish four kinds of identifying assumptions:

(i) Linear restrictions on the contemporary relationships:

$$R_0 \text{vec}(B_0, \Gamma_0) = r_0 \ . \tag{1.26}$$

(ii) Restrictions on the short run dynamics:

$$R_1 \text{vec}(\Gamma_1^*, \ldots, \Gamma_{p-1}^*) = r_1 \ . \tag{1.27}$$

(iii) Restrictions on the long run relationships:

$$R_\beta \text{vec}(\beta) = r_\beta \ . \tag{1.28}$$

(iv) Restrictions on the "loading matrix":

$$R_\alpha \text{vec}(\alpha_1^*) = r_\alpha \ . \tag{1.29}$$

In principle we may also include restrictions on the covariance matrix Σ in the list of identifying assumptions. However, in the traditional Cowles-Commission type of structural models such restrictions are not very common. In contrast, the "structural VAR approach" which is considered in Chapter 3 relies heavily on covariance restrictions.

To identify the parameters of the structural form, a sufficient number of restrictions is required. Hsiao (1997) calls the matrix $\Pi_1^* = \alpha_1^* \beta'$ "long run relation matrices". He assumes that linear restrictions are imposed on the matrix $A^* = [B_0, \Gamma_0, \Gamma_1^*, \ldots, \Gamma_{p-1}^*, \Pi_1^*]$ so that for the g'th equation the restriction can concisely be written as $R_g^* a_g^* = 0$, where a_g^* is the g'th column of $A^{*\prime}$ and R_g^* is a known matrix. In this case the rank condition is

$$\text{rk}(R_g^* A^{*\prime}) = m - 1 \ .$$

Hsiao (1997) emphasize that this rank condition is equivalent to the usual rank condition in the SE model and, thus, cointegration does not imply additional complications to the identification problem. However, this is only true if Π_1^* is considered as the long run parameter matrix. In Johansen's (1995b) framework the long run parameters are represented by the matrix β and the nonlinearity implied by the product $\alpha_1^* \beta'$ indeed imply additional problems for the identification

of the system. Specifically Johansen (1995b) points out that the rank condition must be checked for every possible value of β. He suggests a sequential procedure to verify a more restrictive concept labeled as "generic identification".

In practice, identification is often checked by applying the so-called *order condition*, which is a necessary condition for identification. The application of these criteria for restrictions of the form (i) and (ii) is well documented in many econometric text books (e.g. Judge et al. 1988, Hamilton 1994) and there is no need to repeat the discussion here. Rather I will concentrate on the structural form of a cointegrated system given in (1.25).

First, I consider the identification of the cointegration matrix β. Johansen and Juselius (1990, 1992) consider restrictions of the form

$$R\beta = 0, \quad j = 1, \ldots, r \tag{1.30}$$

$$\text{or} \qquad \beta = H\varphi, \tag{1.31}$$

where R is a given $(n-q) \times n$ matrix and H is a $n \times q$ matrix obeying $RH = 0$ and $q \leq n - r$. Comparing this restriction with (1.28) reveals two differences. First, the restriction (1.30) assumes $r_\beta = 0$. This specification excludes the restriction of cointegration parameters to prespecified values. Since the cointegration property is invariant to a scale transformation of the cointegration vector, such constants are not identified.[4] Second, all r cointegration vectors are assumed to satisfy the same linear restriction $R\beta_j = 0$, where β_j is the j'th row of β. Of course, this is a serious limitation of such type of restrictions. Nevertheless, in many empirical applications, the restrictions on the cointegration vectors can be written as in (1.30) (e.g. Johansen and Juselius 1990, 1992, Hoffman and Rasche 1996).

Of course, if there is only one cointegration vector, then this kind of restriction does not imply a loss of generality. Another important class of restrictions covered by (1.30) is the case that the basis of the cointegration space is known. As in King et al. (1991) assume that $y_t = [c_t, i_t, o_t]'$, where c_t denotes the log of consumption,

[4] To facilitate the interpretation, the cointegration vectors are often normalized so that one of the coefficients is unity. However such a normalization does not restrict the cointegration space and is therefore not testable.

i_t is the log of investment, and o_t denotes the log output. Suppose that $c_t - o_t$ and $i_t - o_t$ are stationary. Accordingly, the cointegration space can be represented as

$$\beta = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} \varphi \ .$$

Another important special case are "separable restrictions", that is, restrictions that apply to different parts of the cointegration vectors. An example is the restriction:

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \beta = 0 \ .$$

Notwithstanding these cases encountered frequently in practice, the restriction (1.30) rules out important cases. Johansen and Juselius (1994) therefore consider more general restrictions given by

$$\text{diag}(R_1, \ldots, R_r)\text{vec}(\beta) = 0$$

or

$$\beta = [H_1\varphi_1, \ldots, H_r\varphi_r] \ , \tag{1.32}$$

where H_j is a $n \times q_j$ matrix and φ_j is a $q_j \times 1$ vector with $q_j \leq n - r$. This set of restrictions is more general than (1.30), since it allows for different linear restrictions on the cointegration vectors. However, no restrictions across cointegration vectors are accommodated. Such restrictions between different cointegration vectors, however, do not seem to be important in practice.

To identify the cointegration vector β_j it is required that no other cointegration vector (or a linear combination thereof) satisfy the restriction for β_j. Accordingly, the rank condition results as

$$\text{rk}(R_j\beta_1, \ldots, R_j\beta_r) = r - 1$$

(cf. Johansen and Juselius 1994). The problem with the application of such a rank condition is that it depends on the (unknown) parameter values $\beta_1, \ldots, \beta_r$. To overcome the difficulties Johansen and Juselius (1994) suggest a criterion to check for *generic* identification. Inserting (1.32) gives

$$rk(R_jH_1\varphi_1, \ldots, R_jH_r\varphi_r) = r - 1 \ .$$

From this rank condition Johansen (1995) derives a sequence of rank criteria which can be used to check the identification for "almost all" possible vectors β. Furthermore a simple order condition can be derived. Since $(R_j\beta)$ is a $q_j \times r$ matrix, $q_j \geq r-1$ restrictions are needed to identify β_j in addition to a normalization restriction.

Davidson (1998) suggests an "atheoretical" approach to achieve unique cointegration vectors that are identified up to a scale transformation. A cointegration vector is called irreducible if no variable can be omitted from the cointegration relationship without loss of the cointegration property. Such an irreducible cointegration vector is unique up to a scale transformation. Davidson (1998) provides a program that allows to determine the irreducible cointegration vectors from an estimated cointegration matrix.

Whenever the long run parameters β are properly identified, the short run parameters can be identified in the usual way. Letting $w_t = \beta' y_t$, the structural form of the VECM can be written as

$$B_0 \Delta y_t = \alpha_1^* w_{t-1} + \Gamma_0 \Delta x_t + \Gamma_1^* \Delta z_{t-1} + \cdots + \Gamma_{p-1}^* \Delta z_{t-p+1} + u_t \ , \tag{1.33}$$

which takes the form of a traditional linear system of simultaneous equations. It follows that the "short run parameters" $B_0, \Gamma_0, \alpha_1^*, \Gamma_1^*, \ldots, \Gamma_{p-1}^*$ can be identified by applying the traditional rank or order conditions (e.g. Judge et al. 1988, Hsiao 1997).

As in Johansen and Juselius (1994) it is assumed that the long run and short run parameters were identified separately. However, as pointed out by Boswijk (1995), it is possible to identify β by using restrictions on α. For example, assume that α is restricted to have a form similar to $\alpha = [I_r, \alpha_2]'$. Then β is identified and can be computed from the reduced form as the upper block of the matrix $\Pi = \alpha\beta'$. This identification is used in the SE representation of a cointegrated system which is discussed in Section 1.3. The mixed case using restrictions on α and β together to identify β is more complicated and does not seem important for the empirical practice. See Boswijk (1995) for more details.

As in the usual SE model, the identifying assumptions are derived from economic theory. The assumptions on the long run relationships often result from

equilibrium conditions on markets for goods and services, whereas the short run restrictions are more difficult to motivate. An important source of short run restrictions is the theory of rational expectations. Unfortunately, the resulting restrictions usually imply highly nonlinear cross equation restrictions that are difficult to impose on the SE systems. Therefore, the short run restrictions are often imposed by making informal ("plausible") assumptions or by testing the coefficients against zero (the simplification stage of Hendry's methodology). Examples are Juselius (1998) and Lütkepohl and Wolters (1998). Similarly, Garratt et al. (1999) advocate a different treatment of long and short run restrictions. They derive the long run relationships from (steady state) economic theory and impose these restrictions on the cointegration vectors of a cointegrated VAR. The resulting model for the long run relationship is called the "core model":

$$\beta' z_t - c_0 - c_1 t = w_t \ ,$$

where the vector $z_t = [y_t', x_t']'$ comprises the endogenous and exogenous variables of the system and β is subject to linear restrictions given in (1.28). At the second stage, the short run response is represented in the model of the usual form (1.25). The lag length of the adjustment model is selected by using conventional information criteria like AIC or the BIC (cf. Lütkepohl 1991). Furthermore, coefficients may be set to zero whenever they turn out to be insignificant with respect to a prespecified significance level.

1.6 Estimation under long run restrictions

First the estimation of cointegrated VAR models with restrictions on the cointegration vectors is considered. Since we assume that all other parameters are unrestricted, the model can be estimated in its concentrated form:

$$\Delta \widetilde{y}_t = \alpha \beta' \widetilde{y}_{t-1} + \widetilde{\varepsilon}_t \ , \tag{1.34}$$

where $\Delta \widetilde{y}_t$ and $\widetilde{y}_{t-1}$ are residual vectors from a regression of Δy_t and y_{t-1} on $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$ and possible deterministic terms. The concentrated form is

equivalent to a cointegrated VAR[1] model. In what follows we therefore drop the tildes for notational convenience.

In the case that the restrictions on β take the form as in (1.31), Johansen and Juselius (1990, 1992) suggest a simple ML estimation procedure. The restriction is inserted in the VECM format (1.34) yielding

$$\begin{aligned} \Delta y_t &= \alpha\phi' H' y_{t-1} + \varepsilon_t \\ &= \alpha^* y^*_{t-1} + \varepsilon_t \ , \end{aligned} \tag{1.35}$$

where $\alpha^* = \alpha\phi'$ and $y^*_{t-1} = H' y_{t-1}$. The restricted cointegration vectors can easily be estimated from a reduced rank regression of Δy_t on y^*_{t-1} (cf Johansen and Juselius 1992).

To estimate the model under the more general set of restrictions given in (1.32) no such simple reformulation of the model is available. Inserting the restriction for the j'th cointegration vector in (1.34) gives

$$\Delta y_t = \sum_{j=1}^{r} \alpha_j \varphi_j' H_j' y_{t-1} + \varepsilon_t \ . \tag{1.36}$$

Assume that we want to estimate the parameters of the first cointegration vector φ_1. Equation (1.36) can then be reformulated as

$$\Delta y_t = \alpha_1 \varphi_1' H_1' y_{t-1} + \vartheta_2' y_{t-1} + \varepsilon_t \ , \tag{1.37}$$

where $\vartheta_2 = [\beta_2, \ldots, \beta_r]$. The idea of the switching algorithm suggested by Johansen (1995) is to estimate $\alpha_1^* = \alpha_1 \varphi_1'$ conditional on an initial estimate of the remaining cointegration vectors stacked in ϑ_2. In other words the system is estimated by treating the additional variables $z_{2t} = \vartheta_2' y_{t-1}$ as given. With the resulting estimate of β_1 a new set of variables is formed that are treated as given for the estimation of the second cointegration vector. Therefore, the procedure employs updated cointegration vectors on every estimation stage and proceeds until the estimates have converged.

Johansen (1995b) was not able to show that his "switching algorithm" indeed converges to the global maximum of the likelihood function. Nevertheless, his method is computationally convenient and seems to have reasonable properties in

practice. It is implemented in the PcGive 9.0 software of Doornik and Hendry (1996).

Pesaran and Shin (1995) consider the ML estimation of the restricted likelihood function which is equivalent to maximizing the function

$$S^*(\beta,\lambda) = \log|\beta' A_T \beta| - \log|\beta' B_T \beta| + 2\lambda' H_\beta \mathrm{vec}(\beta) \ , \tag{1.38}$$

where $A_T = S_{11} - S_{01}S_{11}S_{01}'$ and $B_T = S_{11}$ with S_{ij} as defined in Section 1.4 and restrictions of the general form (1.28) with $r_\beta = 0$. The derivative is

$$\frac{\partial S^*(\beta,\lambda)}{\mathrm{vec}(\beta)} = \Big([(\beta' A_T \beta)^{-1} \otimes A_T] - [(\beta' B_T \beta)^{-1} \otimes B_T]\Big)\mathrm{vec}(\beta) + H_\beta' \lambda \ .$$

From this derivative and $\partial S^*(\beta,\lambda)/\lambda = H_\beta \mathrm{vec}(\beta)$, Pesaran and Shin (1995) derive a first order condition which can be written as $\mathrm{vec}(\beta) = f(\beta)$, where $f(\beta)$ is a complicated nonlinear function. Based on this first order condition they suggest an iterative scheme, where the updated estimate $\beta^{(1)}$ results from the preliminary estimate $\beta^{(0)}$ as $f(\beta^{(0)})$. An important problem with such a procedure is, however, that it is unknown whether it converges to a maximum. Pesaran and Shin (1995) therefore suggest a "generalized Newton Raphson procedure" based on the first and second derivatives of $S^*(\beta,\lambda)$ given in (1.38). This estimator turns out to be quite complicated but can be implemented by using numerical techniques (cf Pesaran and Shin (1995) for more details).

Hsiao (1997) argues that structural models can be estimated in the usual way (e.g. using 2SLS, 3SLS or FIML) from a structural version of the VECM model. However, this is only possible if the long run restrictions can be written as linear restrictions on the matrix $\Pi_1^* = \alpha_1^* \beta'$. Unfortunately, the matrix Π^* mixes short and long run parameters so that a linear restriction on β must be translated into linear restriction on Π^*. A simple way to do this is suggested in Breitung (1995b). As in Section 1.3 we reformulate the system using $\alpha_* = \alpha\alpha_2^{-1} = [\phi, I_r]'$ and $\phi = \alpha_1\alpha_2^{-1}$. Furthermore, we define $\pi_2 = \beta\alpha_2'$ so that $\alpha_*\pi_2' = \alpha\beta'$.

The reduced form VECM is multiplied by the matrix

$$C_0 = \begin{bmatrix} I_{n-r} & -\phi' \\ 0 & I_r \end{bmatrix}$$

so that the resulting system can be written as

$$\Delta y_{1t} = \phi' \Delta y_{2t} + w_{1t} \tag{1.39}$$

$$\Delta y_{2t} = \pi_2' y_{t-1} + w_{2t} \ , \tag{1.40}$$

where $y_t = [\Delta y_{1t}', \Delta y_{2t}']'$. For more details on this representation see Section 1.3.

The restriction for the j'th cointegration vector can be formulated as

$$\beta_j = \pi_2 a_j = H_j \varphi_j \ , \quad j = 1, \ldots, r \ ,$$

where a_j is the j'th row of α_2^{-1}. Inserting this restriction into the subsystem (1.40) gives

$$\begin{aligned} a_j' \Delta y_{2t} &= a_j' \pi_2' y_{t-1} + w_{2t}^* \\ &= \varphi_j H_j' y_{t-1} + w_{2t}^* \\ &= \varphi_j y_{j,t-1}^* + w_{2t}^* \ , \end{aligned} \tag{1.41}$$

where $w_{2t}^* = a_j' w_{2t}$ and $y_{j,t-1}^* = H_j' y_{t-1}$. Accordingly, the lower subsystem of the structural model (1.40) is replaced by equations of the form (1.41), where the vector $y_{j,t-1}^*$ is a vector of transformed variables.

An example may help to illustrate the approach. To highlight the key features of the transformation, consider the following example. Let $y_t = [Y_t, R_t, r_t, M_t]'$, where Y_t is the log of output, R_t and r_t are a long term and a short term interest rates, and M_t is the log of real money balances. Economic theory gives rise to two cointegrating relationships, namely, a money demand relationship and the term structure of interest rates. Accordingly, the cointegration space can be represented as $b_1 M_t - b_1 Y_t + b_2 R_t + b_3 r_t \sim I(0)$ and $b_4 R_t - b_4 r_t \sim I(0)$ (see, e.g., Hoffman and Rasche, 1996, p. 194). Hence, under this hypothesis the cointegration space is given by the matrix

$$\beta = \begin{bmatrix} -b_1 & 0 \\ b_2 & -b_4 \\ b_3 & b_4 \\ b_1 & 0 \end{bmatrix} = [H_1 \varphi_1, H_2 \varphi_2] \ , \tag{1.42}$$

where

$$H_1 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \varphi_1 = [\, b_1, b_2, b_3 \,], \quad H_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \quad \varphi_2 = b_4 \; .$$

Imposing these restrictions gives the following structural form:

$$\begin{aligned} \Delta Y_t &= \phi_{11}\Delta r_t + \phi_{12}\Delta M_t + w_{1t} \\ \Delta R_t &= \phi_{21}\Delta r_t + \phi_{22}\Delta M_t + w_{2t} \\ a_{11}\Delta r_t &= -a_{12}\Delta M_t + \varphi_{11}(M_{t-1} - Y_{t-1}) + \varphi_{12}R_{t-1} + \varphi_{13}r_{t-1} + w^*_{1t} \\ a_{22}\Delta M_t &= -a_{21}\Delta r_t + \varphi_{21}(R_{t-1} - r_{t-1}) + w^*_{2t} \; . \end{aligned}$$

To estimate this system, the third and fourth equation must be divided by a_{11} and a_{22}, respectively. The resulting system can be estimated with conventional system estimators such as the 3SLS or the FIML estimator and no additional complications arise by the cointegration properties of the system (cf. Hsiao 1997). The asymptotic properties of the 2SLS and 3SLS estimator are given in

THEOREM 1.4 *Let y_t be generated by a cointegrated VAR[1] with $0 < r < n$. Furthermore, J_1 and J_2 are known matrices satisfying $rk(J_1\alpha) = s \leq r$ and $rk(\beta' J_2) \leq s$. Then:*

(i) The 2SLS and 3SLS estimates of ϕ in (1.39) are identical.

(ii) The 2SLS and the 3SLS estimates of $J_1\pi' J_2$ are $\sqrt{T}$–consistent and asymptotically normally distributed with the same non-singular covariance matrix.

PROOF: (i) As has been shown by Zellner and Theil (1962) the 2SLS and 3SLS estimates of an over-identified subsystem are identical if the remaining equations are just identified.
(ii) The model can be re-written as

$$\Delta y_t = \begin{bmatrix} \phi' & 0 \\ 0 & \pi_2' \end{bmatrix} \begin{bmatrix} \Delta y_{2t} \\ y_{t-1} \end{bmatrix} + w_t$$

$$
\begin{aligned}
&= \begin{bmatrix} \phi' & 0 & 0 \\ 0 & \alpha_2 & \tau \end{bmatrix} \begin{bmatrix} \Delta y_{2t} \\ \beta' y_{t-1} \\ \beta_\perp' y_{t-1} \end{bmatrix} + w_t \\
&= X_t\theta + w_t \ ,
\end{aligned}
$$

where $\tau = 0$ and

$$
\theta = \mathrm{vec}(\phi, \pi_2, \alpha_2'), \quad X_t = \begin{bmatrix} I_{n-r} \otimes \Delta y_{2t}' & 0 \\ 0 & I_r \otimes y_{t-1}'\beta \end{bmatrix} .
$$

In this representation the subvector θ_1 in $\theta = [\theta_1', \theta_2']'$ comprises the parameters attached to stationary variables, whereas θ_2 contains the parameters attached to the nonstationary variables $\beta_\perp' y_{t-1}$. In this representation $[\alpha_2, \tau] = \pi_2' Q^{-1}$, where $Q = [\beta, \beta_\perp]'$.

Stacking the observations for $t = 2, \ldots, T$ into matrices such as $X = [X_{12}', \ldots, X_{1T}']'$, $y = [\Delta y_2', \ldots, \Delta y_T']'$, and $w = [w_2', \ldots, w_T']'$ the model is written as $y = X\theta + w$. The matrix of instruments is defined as $Z_{1t} = (I_n \otimes y_{t-1}')$ and $Z = [Z_{12}', \ldots, Z_{1T}']'$. The IV estimator of θ is given by

$$
\widehat{\theta}_{iv} = [X'Z(Z'\Omega Z)^{-1}Z'X]^{-1}X'Z(Z'\Omega Z)^{-1}Z'y \ .
$$

For the 2SLS estimate $\Omega = I$ and for the 3SLS estimate $\Omega = (I_{T-1} \otimes C_0 \Sigma C_0')$.

Let $\Upsilon_T = diag\{T^{-1/2}I, T^{-1}I\}$. Using $Q = [\beta, \beta_\perp]'$ we get

$$
\begin{aligned}
\Upsilon_T X'ZQ\Upsilon_T(\Upsilon_T Q'Z'\Omega ZQ\Upsilon_T)^{-1}\Upsilon_T Q'Z'X\Upsilon_T \ \Rightarrow \ & \begin{bmatrix} A_1' & 0 \\ 0 & B_1' \end{bmatrix} \begin{bmatrix} A_2 & 0 \\ 0 & B_2 \end{bmatrix}^{-1} \begin{bmatrix} A_1 & 0 \\ 0 & B_1 \end{bmatrix} \\
& = \begin{bmatrix} A_1'A_2^{-1}A_1 & 0 \\ 0 & B_1'B_2^{-1}B_1 \end{bmatrix}
\end{aligned}
$$

and

$$
\begin{aligned}
\Upsilon_T X'ZQ\Upsilon_T(\Upsilon_T Q'Z'\Omega ZQ\Upsilon_T)^{-1}\Upsilon_2 Q'Z'w \ \Rightarrow \ & \begin{bmatrix} A_1' & 0 \\ 0 & B_2' \end{bmatrix} \begin{bmatrix} A_2 & 0 \\ 0 & B_2 \end{bmatrix}^{-1} \begin{bmatrix} B_3 \\ B_4 \end{bmatrix} \\
& = \begin{bmatrix} A_1'A_2^{-1}B_3 \\ B_1'B_2^{-1}B_4 \end{bmatrix},
\end{aligned}
$$

where A_i $(i = 1, 2)$ are fixed matrices and B_i $(i = 1, \ldots, 4)$ are stochastic matrices which can be represented as functionals of Brownian motions. Note that only the

matrices A_2 and B_3 depend on the covariance matrix Ω. With these results we obtain:

$$\Upsilon_T(\widehat{\theta}_{iv} - \theta) \;\Rightarrow\; \begin{bmatrix} (A_1' A_2^{-1} A_1)^{-1} A_1' A_2^{-1} B_3 \\ (B_1' B_2^{-1} B_1)^{-1} B_1' B_2^{-1} B_4 \end{bmatrix}.$$

Since A_1 and B_2 are square matrices we get

$$\Upsilon_T(\widehat{\theta}_{iv} - \theta) \;\Rightarrow\; \left[\begin{pmatrix} A_1^{-1} B_3 \\ B_1^{-1} B_4 \end{pmatrix} \right].$$

which does not depend on Ω. Thus, the limiting distributions of the 2SLS and the 3SLS estimators are the same.

(iii) Since B_3 is normally distributed it is seen from (ii) that the IV estimator for θ_1 is asymptotically normal. Furthermore, B_1 is the limit of $T^{-2} \sum \beta_\perp' y_{t-1} y_{t-1}' \beta_\perp$ and B_4 is the limit of $T^{-1} \sum \beta_\perp' y_{t-1}$ so that $B_1^{-1} B_4$ is mixed normal. The IV estimate of the matrix π_2 is equivalent to the product of IV estimates $\widehat{\alpha}_{2,iv} \widehat{\beta}_{iv}'$. Since $\widehat{\beta}$ is super-consistent, the asymptotic behaviour is similar to $\widehat{\alpha}_{2,iv} \beta'$. Therefore a necessary condition for $J_1 \widehat{\pi}_{2,iv}'$ to have a regular normal limiting distribution is that the matrix J_1 has rank $s_1 \leq r$ rows. Similarly, it is easy to show that a second necessary condition is that $\widehat{\pi}_{2,iv}' J_2 = \widehat{\alpha}_{2,iv} \beta' J_2$ whenever $\mathrm{rk}(\beta' J_2) \leq s$ because otherwise the rank of the covariance matrix is singular. ■

Remark A: It is important to notice that the cointegration parameters are not estimated super-consistently but have the usual rate for coefficients attached to stationary variables. The reason is that in the SE system the matrix $\pi_2 = \beta \alpha_2'$ is a product of short and long run parameters so that the properties implied by the short run parameters dominate the asymptotic properties of the estimate of π_2.

Remark B: Since the system (1.39) – (1.40) is a linear transformation of the VECM system, the FIML estimate $\widehat{\pi}_2$ is identical to $(\hat{\beta} \hat{\alpha}_2')$, where $\widehat{\beta}$ and $\widehat{\alpha}_2$ denote Johansen's (1988) ML estimators. Accordingly, $T-$consistent estimates of the cointegration vectors can be obtained by post-multiplying $\widehat{\pi}_2$ with the inverse of $\widehat{\alpha}_2'$. It will be shown below that if the cointegration vectors are identified by using sufficient long run restrictions, the associated parameters can be estimated $T-$consistently.

Remark C: The matrix of coefficients attached to the lagged levels admits the expansion $\widehat{\pi}_2 = \widehat{\alpha}_2\beta' + O_p(T^{-1})$, where $\widehat{\alpha}_2$ is the least-squares estimate of α_2 in the regression $\Delta y_{2t} = \alpha_2 z_{t-1} + w_{2t}$ and $z = \beta' y_t$. Thus, for any fixed matrices J_1 and J_2 the estimates are $\sqrt{T}-$consistent and asymptotically normal. To obtain a nonsingular covariance matrix of $J_1\pi_2' J_2$, rank conditions on the matrices J_1 and J_2 are required.

Next, we consider the Full Information Maximum Likelihood (FIML) estimator. Using the SE representation (1.39) and (1.40) the following lemma gives simple expressions for the scores of the likelihood function.

LEMMA 1.1 *(i) Let $\widehat{B}_1(z_t)$ and $\widehat{B}_2(z_t)$ denote the least-squares estimates of B_1 and B_2 in a regression*

$$z_t = B_1 w_{1t} + B_2 w_{2t} + e_t \ ,$$

where w_{1t}, w_{2t} are $(n-r)\times 1$ and $r\times 1$ subvectors such that $w_t = [w_{1t}', w_{2t}']' = C_0\varepsilon_t$. Then, the scores of the likelihood function for the SE model given in (1.39) – (1.40) can be written as

$$\begin{aligned} \frac{\partial \mathcal{L}(\phi, \pi_2)}{\partial \phi'} &= \widehat{B}_1(\Delta y_{2t}) \\ \frac{\partial \mathcal{L}(\phi, \pi_2)}{\partial \pi_2'} &= \widehat{B}_2(y_{t-1}), \end{aligned}$$

where $\mathcal{L}(\cdot)$ denotes the (conditional) log-likelihood function.

PROOF: (i) For convenience we first orthogonalize the system given by (1.39) and (1.40) so that it is written in a recursive form. Let $w_{1t} = H_1 w_{2t} + v_{1t}$, $E(w_{2t}w_{2t}') = \Sigma_{22}$, $H_1 = [E(w_{2t}w_{2t}')]^{-1}E(w_{2t}w_{1t}')$ such that v_{1t} is orthogonal to w_{2t} and $\Sigma_{1|2}$ denotes the covariance matrix of v_{1t}. Then, the log-likelihood function is given by

$$\begin{aligned} 2\mathcal{L}(\cdot) &= \text{const} \ - T\ln|\Sigma_{22}| - T\ln|\Sigma_{1|2}| \\ &- \sum w_{2t}'\Sigma_{22}^{-1}w_{t2} - \sum (w_{1t} - H_1' w_{2t})'\Sigma_{1|2}^{-1}(w_{1t} - H_1' w_{2t}). \end{aligned}$$

It is easy to obtain the derivatives with respect to $\Sigma^* = E(w_t w_t')$ as

$$\frac{\partial \mathcal{L}}{\partial(\Sigma^*)} = \frac{T}{2}\Sigma^* - \frac{1}{2}W'W \ ,$$

where $W = [W_1, W_2]$, $W_1 = [w_{12}, \ldots, w_{1T}]'$, $W_2 = [w_{22}, \ldots, w_{2T}]'$. Differentiating with respect to ϕ' and inserting estimates for H_1 and $\Sigma_{1|2}$ gives

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \phi'} &= \Delta Y_2'(W_1 - W_2 H_1)\Sigma_{1|2}^{-1} \\ &= \Delta Y_2' P_2 W_1 (W_1' P_2 W_1) \\ &= \widehat{B}_1(y_{2t})\end{aligned}$$

where $P_2 = I - W_2(W_2'W_2)^{-1}W_2'$.

Concentrating with respect to Σ_{22} and $\Sigma_{1|2}$ we get

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \pi_2} &= Y_{-1}'[I - V_1(V_1'V_1)^{-1}W_1']W_2(W_2'W_2)^{-1} \\ &= [Y_{-1}' - \widehat{B}_1(y_{t-1})W_1']W_2(W_2'W_2)^{-1} \\ &= Y_{-1}' P_1 W_2 (W_2' P_1 W_2)^{-1} \qquad (1.43) \\ &= \widehat{B}_2(y_{t-1}) \ , \qquad (1.44)\end{aligned}$$

where $P_1 = I - W_1(W_1'W_1)^{-1}W_1'$. ■

Using the expressions for the scores in this lemma, a simple scoring algorithm can be constructed by replacing the conditional expectations by least-squares coefficients from multivariate regressions of the respective vectors on the residuals $\widehat{w}_{1t}, \widehat{w}_{2t}$ of the previous iteration.

It is interesting to know whether the asymptotic equivalence of the estimators is also reflected in small samples. To this end a small Monte Carlo experiment is performed. The data are generated according to the model

$$\begin{aligned}y_{1t} &= y_{1,t-1} + \varepsilon_{1t} \qquad (1.45) \\ y_{2t} &= \pi_{11} y_{1,t-1} + \varepsilon_{2t} \ , \qquad (1.46)\end{aligned}$$

where

$$\varepsilon_t \sim N\left(0, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right).$$

For this model $\phi = 0$ and $\pi_2' = [\pi_{21}, -1]$. The FIML estimator is computed using the scoring algorithm suggested in Theorem 1.1. Table 2.1 presents the

Table 2.1: Standardized RMSE for different estimators

	T=100			T=1000		
$\sqrt{T}$·RMSE of	2SLS	3SLS	FIML	2SLS	3SLS	FIML
	π_{21}=1					
$\widehat{\phi}$	1.02	1.02	0.94	1.01	1.01	1.00
$\widehat{\pi}_{21}$	1.50	1.45	1.41	1.43	1.44	1.43
$\widehat{\pi}_{22}$	1.44	1.44	1.39	1.43	1.43	1.43
	π_{21}=0.5					
$\widehat{\phi}$	0.92	0.92	0.87	0.90	0.90	0.89
$\widehat{\pi}_{21}$	0.79	0.69	0.68	0.64	0.63	0.63
$\widehat{\pi}_{22}$	1.30	1.29	1.25	1.25	1.25	1.25
	π_{21}=0.2					
$\widehat{\phi}$	0.80	0.80	0.77	0.79	0.79	0.78
$\widehat{\pi}_{21}$	0.49	0.34	0.34	0.26	0.23	0.23
$\widehat{\pi}_{22}$	1.14	1.13	1.11	1.09	1.09	1.09

Note: This table presents the standardized root mean squared errors for alternative estimators computed from 1000 Monte Carlo replications of the model (1.45) and (1.46) with $\phi = 0$ and $\pi_2' = [\pi_{21}, -1]$.

standardized root mean squared error computed as

$$\sqrt{T} \cdot \text{RMSE}(\hat{\theta}) = \sqrt{\frac{T}{M} \sum_{i=1}^{M} (\hat{\theta}^i - \theta)^2} \ , \tag{1.47}$$

where $\hat{\theta}^i$ is the i'th realization $(i = 1, \ldots, M)$ of the estimator $\hat{\theta}$ for θ. In case of $\sqrt{T}$–consistent estimates the standardized RMSE should converge to the limiting value as $T \to \infty$. In our experiment we let $\pi_{21} \in \{1, 0.5, 0.2\}$, $T \in \{100, 1000\}$ and $M = 1000$.

As can be seen from Table 2.1 the alternative estimators perform roughly similar in samples as large as $T = 1000$. Moreover, the standardized RMSE are of the same magnitude confirming our theoretical result that all estimates are $\sqrt{T}$–

consistent. In small samples, however, the performance of the estimators depends crucially on the parameter π_{21}. This parameter determines the importance of the random walk component in y_{2t} and thus affects the validity of the asymptotic approximation in small samples. In effect, if π_{21} is small, the dynamic properties of the series y_{2t} are dominated by the stationary term ε_{2t}. For stationary variables the 3SLS (and FIML) estimates are more efficient than the 2SLS estimates, so that a gain in efficiency is observed for $\pi_{21} = 0.2$. For a more important random walk component in y_{2t} we observe that all estimators perform similarly.

An important problem is the normalization of the equation (1.41). Usually the matrix C_0 is normalized to have unit elements on the leading diagonal. This normalization implies that the variable with a unit coefficient is the dependent variable in the equation. For this normalization, all parameter estimates are asymptotically normal with the usual convergence rate of $\sqrt{T}$. The reason is that the ML estimate of $\pi_2' = \alpha_2\beta'$ is identical to $\widehat{\alpha}_2\widehat{\beta}'$, where $\widehat{\alpha}_2$ and $\widehat{\beta}$ denote the ML estimates using Johansen's approach. Since $\widehat{\alpha}_2$ is $\sqrt{T}$-consistent and asymptotically normal, the asymptotic properties of $\widehat{\pi}_2$ are dominated by the properties of $\widehat{\alpha}_2$.

If one is interested in super-consistent estimates of the cointegration parameters, a normalization with respect to the cointegration parameters is required. A possibility is to normalize the cointegration vectors as in Phillips (1991). This is achieved by letting $\widehat{\beta}_P = \widehat{\pi}_2(\widehat{\pi}_{21})^{-1}$, where $\widehat{\pi}_{21}$ is the upper $r \times r$ block of $\widehat{\pi}_2$. The resulting estimator is $T-$consistent and has the same asymptotic properties as Phillips' (1991) estimator.

1.7 Restrictions on short run parameters

Following Johansen and Juselius (1994) and Hsiao (1997), the parameters α_1^*, Γ_0, $\Gamma_1^*, \ldots, \Gamma_{p-1}^*$ in (1.25) are classified as "short run parameters". Usually, economic theory is silent about the short run parameters $\Gamma_1^*, \ldots, \Gamma_{p-1}^*$. Therefore, these parameters are left unrestricted and, thus, these parameters can be "partialled out" for convenience. In contrast, economic theory often motivates hypotheses on

the loading matrix α (or α_1^*).

Hypotheses on α. Johansen (1991) and Johansen and Juselius (1992) consider the null hypothesis

$$R_\alpha \alpha = 0 \quad \text{or} \quad \alpha = A\varphi_\alpha \ , \tag{1.48}$$

where R_α is a known $(n-q) \times n$ matrix and A is an $n \times q$ matrix satisfying $R_\alpha A = 0$. Note that q cannot be smaller than r because otherwise the rank of α is smaller than r.

To estimate the system under restriction (1.48) we consider again a VAR[1] model and assume that no other restrictions are imposed. Following Johansen (1995a, p. 124) the system is multiplied by the matrices $\bar{A}$ and $A_\perp$ with the properties that $\bar{A}'A = I$ and $A_\perp' A = 0$ so that

$$\begin{aligned} \bar{A}'\Delta y_t &= \varphi_\alpha \beta' y_{t-1} + \bar{A}'\varepsilon_t \\ A_\perp'\Delta y_t &= A_\perp'\varepsilon_t \ . \end{aligned}$$

The restricted eigenvalue problem results as

$$|\lambda S_{11.a_p erp} - S'_{a1.a_p erp} S^{-1}_{aa.a_\perp} S_{a1.a_\perp}| = 0 \ , \tag{1.49}$$

where

$$\begin{aligned} S_{11.a_\perp} &= S_{11} - S_{10}A_\perp (A_\perp' S_{00} A_\perp)^{-1} A_\perp' S_{10}' \\ S_{a1.a_p erp} &= \bar{A}' S_{11} - \bar{A}' S_{10} A_\perp (A_\perp' S_{00} A_\perp)^{-1} A_\perp' S_{10}' \\ S_{aa.a_\perp} &= \bar{A}' S_{11} \bar{A} - \bar{A}' S_{10} A_\perp (A_\perp' S_{00} A_\perp)^{-1} A_\perp' S_{10}' \bar{A} \end{aligned}$$

(see Johansen, 1995a, Theorem 8.2). Estimates for the cointegration vectors are obtained as the eigenvectors of the eigenvalue problem (1.49).

Since this approach is fairly complicated, it is attractive to consider the corresponding procedure in the SE approach. Premultiplying the (concentrated) VECM format with R_α gives

$$R_\alpha \Delta y_t = \varepsilon_t^*$$

where $\varepsilon_t^* = R_\alpha \varepsilon_t$.

An important special case of linear restrictions on α is the hypothesis that a subset of variables is weakly exogenous. Let $z_t = [y_t', x_t']'$, where y_t and x_t are $m \times 1$ and $k \times 1$ subvectors. Then, the conditional model for y_t given x_t can be estimated efficiently if x_t is weakly exogenous for the parameters of the conditional model in the sense of Engle, Hendry and Richard (1983). From Definition 1.3 it follows that this is the case if the error correction term $\beta' y_{t-1}$ does not enter the marginal system for x_t and, thus, the corresponding block α_2 is equal to zero.

This restriction can easily be imposed using the simultaneous equation approach given in (1.39) – (1.40). The null hypothesis (1.48) implies that the rows of R_α lie in the space spanned by the columns of the orthogonal complement $\alpha_\perp$. Accordingly, the null hypothesis can be imposed by replacing the upper $n-q$ rows of transformation matrix C_0 by the first $n-q$ rows of R_α.

As an example, assume that we want to test whether the variable i_t in the cointegrated system $y_t = [Y_t, i_t, M_t]'$ is weakly exogenous. If we assume a single cointegration relationship, this hypothesis implies that the second element of the vector α is zero. Accordingly $R_\alpha = [0, 1, 0]$ and the model becomes

$$\Delta i_t = w_{1t} \tag{1.50}$$

$$\Delta Y_t = a_1 \Delta M_t + w_{2t} \tag{1.51}$$

$$\Delta M_t = b_1 Y_{t-1} + b_2 i_{t-1} + b_3 M_{t-1} + w_{3t} \; . \tag{1.52}$$

The additional over-identifying restriction in (1.50) can be tested, e.g., using an LR test procedure, for example. In our example it is also possible to test the restriction by testing whether the additional inclusion of ΔM_t in equation (1.50) yields a significant coefficient.

Hypotheses on α may also be motivated by a structural model for the permanent shocks defined as $\tau_t = \alpha_\perp' \varepsilon_t$ (cf. Johansen 1995, p. 74). For example, a real business cycle framework implies that the permanent shocks are related to the technical progress (or other supply side factors) and, thus, it may be asserted that innovations of monetary variables do not enter the permanent shock. To identify the first permanent shock of the system (1.50) — (1.52) as a real shock, we may therefore set the coefficient a_1 equal to zero, which is of course identical to assuming that Y_t is weakly exogenous.

The information that some variables in y_t are exogenous may also be used to improve the power of the LR test for the cointegration rank (Harbo et al. 1998 and Pesaran et al. 1999). Consider a VAR[1] process for the vector $z_t = [y_t', x_t']'$, where y_t is a vector of m endogenous variables and x_t is the vector of k exogenous variables. Following Harbo et al. (1998) assume that the structural VECM can be written as in (1.25) where $B_0 = I_m$ and the matrices Γ_0 and Γ_j^* $(j = 1, \ldots, p-1)$ are assumed to be unrestricted. Hence, the only restriction, which is used is that x_t is exogenous and, thus, the respective submatrix of α in the reduced form VECM is equal to zero. The concentrated model results as

$$\Delta \widetilde{y}_t = \Pi_1^* \widetilde{z}_{t-1} + \widetilde{u}_t \ ,$$

where $\widetilde{y}_t$ and $\widetilde{z}_{t-1}$ are the residuals from a regression of y_t and $z_{t-1} = [y_{t-1}', x_{t-1}']'$ on $\Delta x_t, \ldots, \Delta x_{t-p+1}$. The LR test of the hypothesis $r = r_0$ results as the sum of the eigenvalues from a reduced rank regression of $\Delta \widetilde{y}_t$ on $\widetilde{z}_{t-1}$. As shown by Harbo et al. (1998) the resulting LR (trace) statistic is asymptotically distributed as

$$\Lambda(r_0) \ \Rightarrow \ tr\Big\{ \int_0^1 dW_{m-r}(a) W_{n-r}(a)' \left(\int_0^1 W_{n-r}(a) W_{n-r}(a)' da \right)^{-1} \int_0^1 W_{n-r}(a) dW_{m-r}(a)' \Big\}$$

and, thus, the table for the critical values must account for the dimensions m and r (see Harbo et al. 1998).

Rahbek and Mosconi (1999) consider tests of the cointegration rank in the presence of weakly exogenous I(0) regressors. Let x_t denote a stationary weakly exogenous variable. Rahbek and Mosconi (1999) suggest forming the partial sum process $x_t^* = x_1 + \cdots + x_t$ and include x_t^* instead of x_t in the system. The resulting cointegration vector is restricted not to include the variables in x_t^* and the usual LR test procedures can be applied in the system $z_t^* = [y_t', x_t^{*\prime}]'$.

Testing causality. Another important hypothesis concerning the short run parameters of the system is that some variables are not causal for other variables of the system. Let y_t be partitioned as $y_t = [y_{1t}', y_{2t}', y_{3t}']'$, where the three subvectors are of dimension n_1, n_2 and n_3, respectively. If y_{2t} is not causal for y_{1t} according to the definition of Granger (1969), then $E(y_{1,t+1}|\mathcal{I}_t) = E(y_{1,t+1}|\mathcal{I}_t^*)$, where

$\mathcal{I}_t = \{y_t, y_{t-1}, \ldots\}$ and $\mathcal{I}_t^* = \{y_{1t}, y_{3t}, y_{1,t-1}, y_{3,t-1}, \ldots\}$. Although this hypothesis imposes restrictions on the matrices $\Pi = \alpha\beta'$ and $\Gamma_1, \ldots, \Gamma_{p-1}$ in (1.12), we concentrate our discussion to the restrictions on the matrix Π. Restrictions on the matrices $\Gamma_1, \ldots, \Gamma_{p-1}$ do not imply additional complications.

Toda and Phillips (1993, 1994) consider two approaches to test the causality hypothesis. First, the hypothesis may be tested by using an unrestricted VAR in levels. In a VAR[1] system4 this amounts to testing whether the estimated matrix $\widetilde{\Pi}_{12}$ from the partitioning

$$\widetilde{\Pi} = \begin{bmatrix} \widetilde{\Pi}_{11} & \widetilde{\Pi}_{12} & \widetilde{\Pi}_{13} \\ \widetilde{\Pi}_{21} & \widetilde{\Pi}_{22} & \widetilde{\Pi}_{23} \\ \widetilde{\Pi}_{11} & \widetilde{\Pi}_{12} & \widetilde{\Pi}_{13} \end{bmatrix} = \left(\sum_{t=2}^{T} y_t y_{t-1}' \right) \left(\sum_{t=2}^{T} y_{t-1} y_{t-1}' \right)^{-1}$$

is significantly different from the zero matrix. Since this estimator has asymptotic representation $\widetilde{\Pi}_{12} = \widehat{\alpha}_1 \beta_2' + O_p(T^{-1})$, where α_j and β_j $(j = 1, 2, 3)$ are submatrices of α and β partitioned according to the subvectors of y_t. To obtain an asymptotically normal estimator for Π_{12} with a nonsingular covariance matrix we need to assume that $\mathrm{rk}(\beta_2) = r_2$. This condition is called "sufficient cointegration" by Toda and Phillips (1994). Of course this condition may fail in practice and for this case, Toda and Phillips (1993) show that the asymptotic distribution of the Wald statistic is nonstandard and depends on nuisance parameters. Assume for example that $n = 3$ and $r = 1$ with a cointegration vector $\beta = [\beta_1, 0, \beta_3]'$ so that $\mathrm{rk}(\beta_2) = 0$. In this case the Wald statistic fails to yield an asymptotically χ^2 distributed test statistic.

The second approach is to estimate α and β by using Johansen's ML estimator and testing whether $\widehat{\alpha}_1 \widehat{\beta}_2'$ is significantly different from zero. This approach suffers from a similar rank problem and sufficient conditions for χ^2 distributed test statistics are the rank conditions (i) $\mathrm{rk}(\beta_2) = r_2$ or (ii) $\mathrm{rk}(\alpha_1) = n_1$ (cf Toda and Phillips 1993).

Dolado and Lütkepohl (1996) and Toda and Yamamoto (1995) suggest an elegant approach to resolve these problems. They show that the Wald test has the usual χ^2 distribution whenever the VAR model is estimated with an additional lag which is, however, not included in the null hypothesis. This test is called the

"lag-augmentation test" for causality. Assume that we estimate a VAR[2] model instead of the true VAR[1] process. Then, the estimator of Π_{12} results from replacing y_{1t} and $y_{2,t-1}$ by the residual vectors from the regressions of y_{1t} and $y_{2,t-1}$ on y_{t-2}. In contrast to the original variables, these residuals are stationary and, therefore, the resulting estimate for Π_{12} has standard asymptotic properties and the Wald statistic is valid in any case. Of course this approach implies a loss of power because the model is augmented with unnecessary regressors.

The causality hypothesis can also be implemented in the SE approach. Assume that the system can be arranged so that the last $n_1 < r$ equations have y_{1t} as the set of dependent variables. If there is no causality from y_{2t}, then the respective coefficients of the matrix π_2' in the lower n_1 equations are zero. Of course, the dimension of the vector must have a dimension lower than r which is also assumed in the work of Toda and Phillips (1993, 1994). Furthermore from Theorem 1.4 (iii) rank conditions for α_1 and β_2 can be deduced that are the same as in Toda and Phillips (1993, 1994). Therefore, the SE approach suffers from the same problems as the other approaches.

1.8 Deterministic terms

In this section the treatment of deterministic terms is discussed. Following Johansen (1994) we will confine ourselves to the case of a constant and linear trend. Models with dummy variables like intervention dummies or seasonal dummies can be handled in a similar manner.

For convenience we consider again a cointegrated VAR[1] model. Since we can concentrate out any unrestricted higher order dynamics this does not imply a loss of generality. Introducing a linear time trend in the VECM representation yields

$$\Delta y_t = \mu_0^* + \mu_1^* t + \Pi y_{t-1} + \varepsilon_t \ . \tag{1.53}$$

Since Δy_t is allowed to have a linear time trend, y_t may have a quadratic time trend, in general. Accordingly, the *demeaned* VECM can be written as

$$\Delta y_t - \mu_0 - \mu_1 t = \Pi \left[y_{t-1} - \mu_0 - \mu_1 (t-1) - \mu_2 (t-1)^2 \right] + \varepsilon_t \ . \tag{1.54}$$

Table 2.2: Hypotheses on the deterministic trends

	mean function	Restrictions in SE	Explanation
$H(r)$	$\mu_0^* + \mu_1^* t$	—	no restrictions
$H^*(r)$	$\mu_0^* + \alpha b_1 t$	$c_1^* = 0$	no quadratic trend in data
$H_1(r)$	μ_0	$c_1^* = 0,\ \ \mu_{21}^* = 0$	no trend in $\beta' y_t$
$H_1^*(r)$	αb_0	$c_0^* = 0,\ \ c_1^* = 0,\ \ \mu_{21}^* = 0$	no trend in data
H_2	0	no deterministics	

Using (1.53) we get

$$\begin{aligned} \mu_0^* + \mu_1^* t &= E(\Delta y_t) - E(\Pi y_{t-1}) \\ &= \mu_0 + \mu_1 t - \Pi(\mu_0 - \mu_1 + \mu_2) - \Pi(\mu_1 - 2\mu_2)t - \Pi\mu_2 t^2 \ . \end{aligned}$$

To match both sides of the equation we obtain $\Pi\mu_2 = 0$ or, equivalently, $\beta'\mu_2 = 0$.

The vectors μ_i^* $(i = 0, 1)$ can be projected onto the subspaces spanned by the columns of α and $\alpha_\perp$:

$$\mu_i^* = \alpha b_i + \alpha_\perp c_i \ , \tag{1.55}$$

where $b_i = (\alpha'\alpha)^{-1}\alpha'\mu_i^*$ and $c_i = (\alpha_\perp'\alpha_\perp)^{-1}\alpha_\perp'\mu_i^*$ (see Johansen, 1994, p. 208). Accordingly we define

$$\begin{aligned} \alpha_\perp'\mu_i^* &= \alpha_\perp'\alpha_\perp c_i \\ &\equiv c_i^* \ . \end{aligned}$$

Using $\alpha_\perp' = [I, -\phi']$, the SE system (1.39) and (1.40) can accommodate deterministic terms:

$$\Delta y_{1t} = c_0^* + c_1^* t + \phi'\Delta y_{t2} + w_{1t} \tag{1.56}$$

$$\Delta y_{2t} = \mu_{20}^* + \mu_{21}^* t + \pi_2' y_{t-1} + w_{2t} \ , \tag{1.57}$$

where μ_{2j}^* denotes the lower $r \times 1$ subvector of μ_j^*.

Johansen (1994) considers 5 hypotheses of interest which are included in Table 2.2. Using the representation given in (1.56) and (1.57) these hypotheses can

be formulated by using restrictions on c_0^*, c_1^*, μ_{20}^* and μ_{21}^* (see Table 2.2). Thus, Johansen's hypotheses about the deterministic part of the model can conveniently be tested using LR tests for restrictions on the constant and the drift. Since the tests are based on a linear transformation of the model, the LR statistics have the same asymptotic χ^2 distributions as in Johansen (1994).

1.9 An empirical example

To illustrate the estimation and test procedures discussed in this chapter we consider a four variable system $y_t = [Y_t, R_t, r_t, M_t]'$ as in Hoffman and Rasche (1996). We use quarterly U.S. data running from 1970(i) through 1994(iv). Output (Y_t) is measured by the log of real GNP, R_t is the ten year government bond yield, the short term interest rate r_t is measured by the 3-month LIBOR and M_t is the log of the money base M3 adjusted by the implicit price deflator of GNP. The data for Y_t and M_t are seasonally adjusted and taken form the Main Economic Indicator data base of the OECD. The interest rates are taken from the IMF data base.

We start with an appropriate deterministic specification. For this purpose an unrestricted linear trend is included and the lag order of the VAR is determined. The BIC and Hannan-Quinn criteria are minimized by a VAR[2] specification so a VAR[2] model is used for the following analysis. The likelihood ratio statistics cannot reject the null hypothesis that there is no trend term in the VAR representation ($H_1(r)$). Since the hypothesis $H_1^*(r) : c_0^* = 0$ is rejected but $H^*(r) : c_1^* = 0$ cannot be rejected, I conclude that the unconditional mean of the variables can be represented by a linear time trend.

Next, the cointegration rank is selected. This can be done by using Johansen's test based on eigenvalues or by using a likelihood ratio test on the restrictions in the SE representation (1.39) – (1.40). The results of both procedures are presented in Table 2.3. As expected, both approaches yield identical results and suggest that the cointegration rank is $r = 2$.

Economic theory suggests (see, e.g., Hoffman and Rasche 1996) that the two

Table 2.3: LR tests of the cointegration rank

(a) Johansen's LR test statistics

H_0 :	$r_0 = 4$	$r_0 = 3$	$r_0 = 2$	$r_0 = 1$	$r_0 = 0$
max. EV	—	0.887	3.846	27.77	46.71
trace	—	0.887	4.732	32.50	79.21

(b) Simultaneous equation representation

	$r_0 = 4$	$r_0 = 3$	$r_0 = 2$	$r_0 = 1$	$r_0 = 0$
$L(r_0)$	1052.746	1052.302	1050.379	1036.496	1013.141
$-2[L(r_0) - L(r_0 + 1)]$	—	0.887	3.846	27.77	46.71
$-2[L(r_0) - L(n)]$	—	0.887	4.732	32.50	79.21

(c) 0.05 Critical values

H_0 :	$r_0 = 4$	$r_0 = 3$	$r_0 = 2$	$r_0 = 1$	$r_0 = 0$
max. EV ($\mu_0 \neq 0$)	—	3.962	14.04	20.78	27.17
trace ($\mu_0 \neq 0$)	—	3.962	15.20	29.51	47.18

cointegration relationships are a money demand relation

$$M_t = Y_t + \beta_1 R_t - \beta_2 r_t + u_{1t} \tag{1.58}$$

and the term structure relation

$$R_t = r_t + u_{2t} \ , \tag{1.59}$$

where u_{1t} and u_{2t} are stationary error terms. Although the order condition for identification is fulfilled for these two cointegration vectors, the rank condition is violated. This can be seen by subtracting $\delta(R_t - r_t - u_{2t}) = 0$ from the right hand side of (1.58) yielding

$$M_t = Y_t + (\beta_1 - \delta)R_t - (\beta_2 - \delta)r_t + (u_{1t} + \delta u_{2t}) \ .$$

For any value for δ this gives a new cointegrated relationship and, therefore, β_1 and β_2 are not jointly identified. However, the *difference* between the coefficients of the interest rates *is* identified. In practice one may therefore normalize one of the coefficients to be equal to zero by letting $\delta = \beta_2$, for example. In this case the interest rate r_t cancels from the money demand relation and the coefficient of R_t measures the difference $\beta_1 - \beta_2$. For a detailed discussion of this and alternative identification procedures see Hoffman and Rasche (1996, 194f).

The corresponding system of simultaneous equations is estimated as

$$\begin{aligned}
\Delta\widetilde{Y}_t &= 0.030\Delta\widetilde{r}_t + 2.693\Delta\widetilde{M}_t + \widetilde{w}_{1t} \\
\Delta\widetilde{R}_t &= -0.773\Delta\widetilde{r}_t - 138.4\Delta\widetilde{M}_t + \widetilde{w}_{2t} \\
\Delta\widetilde{r}_t &= -110.2\Delta\widetilde{M}_t + 0.072(\widetilde{R}_{t-1} - \widetilde{r}_{t-1}) + \widetilde{w}_{3t} \\
\Delta\widetilde{M}_t &= -0.0007\Delta\widetilde{r}_t - 0.025(\widetilde{M}_{t-1} - \widetilde{Y}_{t-1}) - 0.0007R_{t-1} + \widetilde{w}_{4t} \ ,
\end{aligned}$$

where the constants of the equations are suppressed. From these estimates, the money demand relation results as $M_t = Y_t - 0.026R_t + \hat{u}_{1t}$, where 0.026 is the estimated difference between β_1 and β_2 in (1.58). The LR test statistic for the over-identifying restriction in the term structure equation (1.59) is 0.9905 which is not significant with respect to the critical values of an asymptotic χ^2-distribution with one degree of freedom. Exactly the same value of the test statistic results if the test statistic is computed by using the LR procedure of Johansen and Juselius (1994).

Finally, we test whether the variables Y_t and R_t are weakly exogenous. Applying the simultaneous equations approach suggested in Section 1.3 we test that the corresponding row of the matrix α is zero. The LR statistic for the hypothesis that Y_t is weakly exogenous is 24.349. This value is highly significant with respect to an asymptotic χ^2 distribution with two degrees of freedom. In contrast, the LR statistic for the hypothesis that R_t is exogenous is only 1.340 and, therefore, implies that this variable may be treated as (weakly) exogenous. These results correspond well with conventional wisdom suggesting that the long run interest rate is determined outside the system on the international capital markets (e.g. Nautz and Wolters 1999).

Chapter 2

Structural VARs

Until the late 70th, the simultaneous equation approach, advocated by the influential *Cowles Commission* clearly dominated the empirical research in econometrics. However, the initial optimism about the potential of the simultaneous equation model was not fulfilled and the inability of large macroeconomic models to compete with "atheoretic" Box-Jenkins ARIMA models on predictive grounds led to an increased adoption of time series techniques. In particular, the seminal paper by Sims (1980) prepared the ground for the ultimate success of vector autoregressions in econometrics. However, as argued forcefully by, e.g., Cooley and LeRoy (1985), vector autoregressions have the status of "reduced form" and, thus, are merely vehicles to summarize the dynamic properties of the data. Without referring to a specific economic structure such reduced form VAR models are difficult to understand. For example, it is often difficult to draw any conclusion from the large number of coefficient estimates in a VAR system. As long as such parameters are not related to "deep" structural parameters characterizing preferences, technologies, and optimization behaviour, the parameters do not have an economic meaning and are subject to the so-called "Lucas critique".

The new research program known as Real Business Cycle (RBC) agenda employs dynamic stochastic general equilibrium models that are driven by real technology shocks. In later work (e.g. Christiano and Eichenbaum 1992) further "shocks" like demand shocks resulting from public expenditures or the supply of

money were included to represent other aspects of the economic system. Kydland and Prescott (1982) acknowledge from the outset that their models, like all models, are false and they recognize that traditional econometric estimation procedures such as Gaussian maximum likelihood may be inappropriate. Therefore, they advocate to apply less structured "calibration" methods, that is, the parameters underlying the simulated model economy are typically set to values that are considered to be "reasonable" or in agreement with earlier estimates.

At the same time, Sims (1981, 1986), Bernanke (1986) and Shapiro and Watson (1988) put forward a new kinds of econometric model that is now known as "structural VAR" (SVAR) or "identified VAR" approach. There are several features of this approach that make the SVAR model an attractive model for empirical work in the spirit of the RBC agenda. First, the deviations from the steady state of a RBC model can usually be represented by a (vector) ARMA model that can be conveniently approximated by a vector autoregression. Second, the driving forces of an RBC type model are different kinds of exogenous shocks. This parallels the structural shocks identified from a typical SVAR model.

The similarity of the RBC and the SVAR approach initiated several studies comparing the outcomes of calibrated RBC models with the corresponding findings from estimating a SVAR model. Cogley and Nason (1995) are able to produce a good correspondence between both approaches, whereas Cooley and Dwyer (1998) observe that results from SVAR models are sensitive to the identifying assumption and may produce outcomes that are at odds with the original RBC style model. Using German data, Breitung and Heinemann (1998) find that both approaches yield qualitatively similar findings but differ substantially in detail.

Another important motivation for the development of SVAR models was the paradigm of "rational expectations" (cf Sims 1980). In Section 3.1 the relationship between rational expectation models and the identification of shocks is discussed. Different approaches to identify the structural shocks are discussed in Section 3.2. A general class of structural VAR models is presented in Section 3.3 and Section 3.4 discusses alternative approaches to estimate the structural parameters.

Section 3.5 suggests a latent variable framework that can be used for estimation and inference. Long run restrictions are considered in Section 3.6 and inference on impulse response function is considered in Section 3.7. Section 3.8 gives three empirical applications and Section 3.9 discusses problems of the structural VAR approach.

2.1 Rational expectations

The theory of rational expectations implies that the effects of expected policy actions are generally different from the effects of an unexpected policy. This can be demonstrated using the money demand model of Cuthbertson and Taylor (1989).[1]

Assume that (real) money demand M_t is decomposed into a planned component M_t^p and an unplanned component ε_t^M. It is assumed that agents choose short run money balances to minimize the expected present value of a quadratic loss function. This gives rise to the planned money demand of the form

$$M_t^p = \beta_1 M_{t-1} + \beta_1(1-\beta_2)\sum_{j=0}^{\infty}\beta_2^j E_{t-1}(M_{t+j}^*) \ ,$$

where $0 < \beta_1, \beta_2 < 1$ are parameters implied by the quadratic loss function, E_{t-1} denotes expectation with respect to the information available at time $t-1$ and

$$M_t^* = \gamma_0 + \gamma_1 Y_t + \gamma_2 i_t$$

denotes the desired long run money balances which are obtained by ignoring adjustment costs (cf. Cuthbertson and Taylor 1989). The variable Y_t and i_t represent (logged) output and the relevant interest rate.

The unplanned component is represented as

$$\varepsilon_t^M = M_t - E_{t-1}M_t = b_1\varepsilon_t^Y + b_2\varepsilon_t^i + u_t^M, \tag{2.1}$$

i.e., the ε_t^M is a weighted average of the prediction errors $\varepsilon_t^Y = Y_t - E_{t-1}Y_t$ and $\varepsilon_t^i = i_t - E_{t-1}i_t$ and an additional error u_t^M.

[1] The relationship between rational expectations and SVAR models is discussed in Dhrymes and Thomakos (1998).

We are now in the position to compare the effects of an expected and an unexpected monetary policy using the interest rates as instruments. A predictable change in i_t affects the observed money demand via the expected long run money demand $E_{t-1}(M^*_{t+j})$, whereas an unexpected change in the interest rate enters the unplanned component with coefficient γ_1. Accordingly, the effects of expected and unexpected changes in the interest rates may be quite different (see also McCallum 1999). The SVAR approach focuses on the effects of unexpected variation in the policy instrument by using a linear relationship similar to (2.1).

If expectations of M^*_{t+j} are linear with respect to the information set $\{x_{t-1}, x_{t-2}, \ldots\}$, where $x_t = [M_t, Y_t, i_t]'$, we obtain

$$M_t \;=\; \psi_0 + \sum_{j=1}^{\infty} x'_{t-j}\psi_j + b_1\varepsilon_t^Y + b_2\varepsilon_t^i + u_t^M \;.$$

This equation contains all typical features of an SVAR model. The conditional expectation with respect to the past of the process is $\mu^M_{t-1} = \psi_0 + \sum_{j=1}^{\infty} x'_{t-j}\psi_j$. In an SVAR model this expectation is approximated by a finite number of lags in all variables of the system. Furthermore, possible (nonlinear) restrictions on the parameters $\psi_1, \psi_2, \ldots$ resulting from the particular rational expectation model are ignored in order to keep the estimation procedure "sophisticatedly simple" (Zellner 1992). The terms $b_1\varepsilon_t^Y$ and $b_2\varepsilon_t^i$ represent the part of the prediction error of M_t that is due to other variables like Y_t and i_t and, thus, the remaining error u_t^M represents the "autonomous" shock associated with real money balances.

To sum up, it is seen that models involving rational expectation can be represented as an SVAR model. The resulting structural shocks represent the "news" in the variable that cannot be explained by other variables of the system and, therefore, measure the autonomous contribution to the respective variables. It is quite natural to focus on the dynamic effects of such shocks when assessing the impact of political instruments on the target variables.

2.2 The identification of shocks

As already noted, structural shocks are the central quantities in an SVAR model. These shocks are unpredictable with respect to the past of the process and are the input of a linear dynamic system producing the n-dimensional time series vector y_t. These shocks are attached with an economic meaning such as an oil price shock, reunification shock, exchange rate shock or a monetary shock. It should be noted, however, that these shocks do not only represent disastrous singular events. In contrast, we assume that the economy is hit regularly by such shocks and the size of the shock is usually small.

Since the shocks are not directly observed, assumptions are needed to identify them. There seems to be a consensus that structural shocks should be mutually uncorrelated (and thus "orthogonal"). This assumption is required to consider the dynamic impact of an isolated shock. If the shocks were correlated, we would have to take into account the relationship between the shocks. Moreover, the decomposition into orthogonal components has a long tradition in statistical analysis, and is also used in factor analysis, for example. Another possibility is to consider the shocks as bits of information about the state of the economic system. Hence, the news of the system is projected into a particular orthogonal space, where the axes represent different aspects of reality.

The assumption of orthogonal shocks is, however, not sufficient to achieve identification. For an n dimensional system, $n(n-1)/2$ additional restrictions are necessary. These restrictions can be obtained from a "timing scheme" for the shocks. Such an identification scheme assumes that the shocks may affect a subset of variables directly within the current time period, whereas another subset of variables is affected with a time lag only. An example of such an identification scheme is the triangular (or recursive) identification suggested by Sims (1980). In this model the shocks enter the equation successively so that the shock of the second equation does not affect the variable explained by the first equation in the same period. Similarly, the third shock does not affect the variables explained by the first and second equation in the current time period. Such a scheme is called a "Wold causal chain system" (Wold 1960) and is often associated with a causal

chain from the first to the last variable in the system.

An alternative approach to the identification of the shocks is to formulate structural equations for the errors of the system. In this case it is convenient to think of the equations as an IS curve or a money demand relation, for example, but with the difference that the equations apply to the unexpected part of the variables (the "innovations") instead of the variables themselves. If the equations are valid for the system variables, then they also apply for the unexpected part of the variables. Therefore, the identification using a set of simultaneous equations is appealing with respect to the traditional approach advocated by the Cowles Commission and it is not surprising that this kind of identification has become very popular in empirical work using SVAR models.

In recent work, the identification of shocks using restrictions on the long run effects of structural shocks has become popular. In many cases economic theory suggests that the effects of some shocks are zero in the long run, that is, the shocks are transitory with respect to particular variables. For example, classical economic theory implies that the effect of nominal shocks on real variables like output or unemployment vanishes as time goes by. Such assumptions give rise to nonlinear restrictions on the parameters and may be used to identify the structure of the system. However, recent work demonstrates (e.g. Faust and Leeper 1997) that such long run restrictions may be problematical in practice.

There are several important differences between the identification of an SVAR model and the identification of a simultaneous equation model. First, the latter models are usually identified by linear (exclusion) restrictions. In contrast, SVAR models assume orthogonal shocks so that the structure is identified using also restrictions on the covariance matrix of the errors. This complicates the estimation of such systems considerably.

Second, traditional simultaneous equation models usually employ many more restrictions than necessary to identify the system, that is, these models are often highly over-identified. In his famous critique, Sims (1980) qualifies these overly restricted models as "incredible". SVAR proponents therefore try to avoid to over-simplifying the structure and impose just enough restrictions needed to identify

the structure. Accordingly, most SVAR models are just identified. However, it is important to notice that just identified models are merely a convenient reformulation of the reduced form. Therefore, as long as the reduced form is correctly specified, any just identified structure is also correctly specified in a statistical sense and it is not possible to decide between alternative identified structures on empirical grounds. To quote Dhrymes and Tomakos (1998, p. 190):

> *"Thus, two just identifying sets of conditions, which have diametrically opposed economic implications may well have the same empirical justification, viz. the estimates of the reduced form parameters in the context of which they are applied. As such they do not add anything further to our understanding of the economic phenomenon in question over and above what was conveyed by the reduced form, and they are both equally defensible or equally subject to severe criticism."*

SVAR models are used to quantify prior views of the economy and to assess the plausibility of the outcomes. For example, we may use a Keynesian structure to investigate the long run effects of a monetary shock on unemployment. A monetarist economist would in contrast favour an SVAR model with the restriction that the long run effect of a monetary shock on real variables vanishes. Since both structures are based on the same reduced form, both structures are admissible and, thus, it is not possible to decide which model "is true".

However, it is possible to assess the plausibility of the outcomes. If, for example, a Keynesian economist finds that an expansive monetary shock leads to increasing unemployment and rising prices, then such result would be highly implausible with respect to a Keynesian framework. If, on the other hand, our Keynesian economist would find that the effects of a monetary shock are similar to the effects in a monetarist type of model, then this result suggests that the observed effect of a monetary policy is not sensitive with respect to the different orthodox perspectives (e.g. Dolado, Lopez-Salido and Vega 1999).

2.3 A class of structural VARs

Assume that the $n \times 1$ time series vector y_t admits a stationary VAR[p] representation as defined in Assumption 1.1. This VAR model is referred to as the *reduced form* of the system. Associated with the reduced form is a *structural form* resulting from the set of structural equations:

$$B\varepsilon_t = Ru_t \ , \tag{2.2}$$

so that

$$By_t = BA_1y_{t-1} + \cdots + BA_py_{t-p} + Ru_t \ . \tag{2.3}$$

The matrices B and R are assumed to be invertible and u_t is an $n \times 1$ vector of *structural shocks* with covariance matrix $E(u_tu_t') = \Omega$. This includes all models considered by Amisano and Giannini (1997).

The dynamic effect of the structural shocks is analysed by considering the moving average representation

$$\begin{aligned} y_t &= \varepsilon_t + \Phi_1\varepsilon_{t-1} + \Phi_2\varepsilon_{t-2} + \cdots \equiv \Phi(L)\varepsilon_t \\ &= B^{-1}Ru_t + \Phi_1B^{-1}Ru_{t-1} + \Phi_2B^{-1}Ru_{t-2} + \cdots = \Psi(L)u_t \ , \end{aligned}$$

where $\Phi(L) = A(L)^{-1}$ and $\Psi(L) = A(L)^{-1}B^{-1}R$. If y_t is measured in first differences, then the matrix

$$\Psi \equiv \Psi(1) = (I_n - A_1 - \cdots - A_p)^{-1}B^{-1}R \tag{2.4}$$

measures the long run impact of u_t on the levels of y_t.

To identify the parameters, restrictions on the parameter matrices $B, R, A_1, \ldots, A_p, \Omega$ are necessary. In empirical applications such restrictions are suggested by economic theory or are imposed just for convenience. The most popular kinds of restrictions can be classified as follows:

(i) The structural errors are assumed to be mutually uncorrelated such that $\Omega = E(u_tu_t')$ is a diagonal matrix. In some applications, the shocks are normalized to have a unit variance, i.e., $\Omega = I_n$.

(ii) $R = I_n$. The vector of innovations (ε_t) is modeled as an interdependent system of linear equations such that $B\varepsilon_t = u_t$, where B is normalized to have ones on the leading diagonal and linear restrictions of the form $Q_1\text{vec}(B) = h_1$ are imposed. Empirical examples include Sims (1986), Bernanke (1986) and Shapiro and Watson (1988).

(iii) $B = I_n$. In this case the model for the innovations is $\varepsilon_t = Ru_t$ and to exclude some (linear combinations of the) structural shocks in particular equations, restrictions of the form $Q_2\text{vec}(R) = h_2$ are imposed. Empirical examples can be found in Blanchard and Quah (1989), Roberts (1993) and Ahmed et al. (1993).

(iv) The AB-model of Amisano and Giannini (1997) combines (ii) and (iii) such that the model for the innovations is $B\varepsilon_t = Ru_t$. Accordingly, the two sets of restrictions $Q_1\text{vec}(B) = h_1$ and $Q_2\text{vec}(R) = h_2$ are used to identify the system. Empirical examples are provided by Blanchard (1989) and Gali (1992).

(v) If (some of) the variables are measured in first differences, there may be prior information on the long run effect of the shocks. The long run effect of the shocks are measured by the matrix Ψ which is defined in (2.13). If a shock is assumed to have no permanent effect on the elements of y_t, then the respective elements of Ψ are zero. Such linear restrictions can be written as $Q_3\text{vec}(\Psi) = h_3$ (see, e.g. Lütkepohl and Breitung 1997). Empirical examples are given by Blanchard and Quah (1989), King et al. (1991) and Roberts (1993).

2.4 Estimation

The estimation of the SVAR model is equivalent to the problem of estimating a simultaneous equation model with covariance restrictions. Using (2.2) the structural model can be reformulated as

$$B\Delta y_t = \Pi^* y_{t-1} + \Gamma_1^* \Delta y_{t-1} + \cdots + \Gamma_{p-1}^* \Delta y_{t-p+1} + Ru_t \ , \tag{2.5}$$

where $\Pi^* = -B(I_n - \sum_{j=1}^{p} A_j) =$ and $\Gamma_k^* = -\sum_{j=k+1}^{p} BA_j$ for $k = 1, \ldots, p-1$. In this formulation the long run behaviour of the shocks is determined by the matrix Π^*, while the short run dynamics are governed by the matrices $\Gamma_1^*, \ldots, \Gamma_{p-1}^*$.

Let $\Delta\widetilde{y}_t$ and $\widetilde{y}_{t-1}$ denote the residual vectors from the regressions of Δy_t and y_{t-1} on $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$. If there are no restrictions on the short run parameters $\Gamma_1^*, \ldots, \Gamma_{p-1}^*$ we may concentrate the system as

$$B\Delta\widetilde{y}_t = \Pi^*\widetilde{y}_{t-1} + R\widetilde{u}_t \ , \tag{2.6}$$

which results from replacing $\Gamma_1^*, \ldots, \Gamma_p^*$ by least-squares estimates.

For notational convenience we reformulate model (2.5) as

$$B\Delta y_t = \Pi^* y_{t-1} + Cx_{t-1} + Ru_t \ ,$$

where $C = [\Gamma_1^*, \ldots, \Gamma_{p-1}, C_{p-1}^*]$ and $x_{t-1} = [\Delta y_{t-1}', \ldots, \Delta y_{t-p+1}']'$. It is assumed that u_t is white noise with $u_t \sim N(0, I_n)$. The log-likelihood function of the model is

$$\begin{aligned} \mathcal{L}(B, C, R) = \text{const} + (T-p)\ln|\det(BR^{-1})| \qquad (2.7) \\ -\frac{1}{2}\sum_{t=p+1}^{T}(B\Delta y_t - \Pi^* y_{t-1} - Cx_{t-1})(RR')^{-1}(B\Delta y_t - \Pi^* y_{t-1} - Cx_{t-1})' \ , \end{aligned}$$

subject to the restrictions:

$$Q_1\text{vec}(B) = h_1 \tag{2.8}$$

$$Q_2\text{vec}(R) = h_2 \tag{2.9}$$

$$Q_3\text{vec}(\Psi) = h_3 \ . \tag{2.10}$$

where the (2.10) implies restrictions on the long run effect of the structural shocks.

In general, the maximization of (2.7) under the constraints (2.8) – (2.10) is a computationally demanding problem. Even for the very simple model of Blanchard and Quah (1989), there exist multiple solutions and there are points in the parameter space, where the log-likelihood function is undefined. In particular the latter problem occurs when the matrix B or R is singular and convergence problems can arise in the neigbourhood of singular matrices.

If there are no restrictions on the long run impact matrix Ψ, the maximation problem can be simplified by "concentrating out" the parameters in C. Setting the derivative with respect to C equal to zero gives

$$\widehat{C} = \sum_{t=p+1}^{T} (B\Delta - \Pi^* y_{t-1}) y_t x_{t-1}' \left(\sum_{t=p+1}^{T} x_{t-1} x_{t-1}' \right)^{-1}$$

and

$$\begin{aligned} B\Delta y_t - \Pi^* y_{t-1} - \widehat{C} x_{t-1} &= B\Delta y_t - \Pi^* y_{t-1} \\ &\quad - \Big(\sum_{t=p+1}^{T} (B\Delta y_t - \Pi^* y_{t-1}) x_{t-1}' \Big) \Big(\sum_{t=p+1}^{T} x_{t-1} x_{t-1}' \Big)^{-1} x_{t-1} \\ &= B\Delta \widetilde{y}_t - \Pi^* \widetilde{y}_{t-1} \ , \end{aligned}$$

Accordingly, the ML problem is equivalent to the maximation of the concentrated likelihood function

$$\begin{aligned} \mathcal{L}_C(B, \Pi^*, R) &= \text{const} + (T-p) \ln |\det(BR^{-1})| \\ &\quad - \frac{1}{2} \sum_{t=p+1}^{T} (\Delta \widetilde{y}_t - \Pi^* \widetilde{y}_{t-1})' B (RR')^{-1} (B\Delta \widetilde{y}_t - \Pi^* \widetilde{y}_{t-1}) \ . \end{aligned}$$

The likelihood function corresponds to a "concentrated model" given by

$$B\Delta \widetilde{y}_t = \Pi^* \widetilde{y}_{t-1} + R\widetilde{u}_t \ . \tag{2.11}$$

Accordingly, the ML estimation of the concentrated system is equivalent to the maximization of (2.7).

The concentration with respect to C is not possible in systems with restrictions on the long run effect on u_t and in other cases where C is restricted. In the next section, a convenient representation is suggested, which allows the inclusion of long run restrictions.

Instead of the ML estimation of the system, other estimation principles may be adopted. The *Minimum Distance* (MD) estimation first estimates the reduced form parameters $\theta = \text{vec}(A_1, \ldots, A_p, \Sigma)$ and then obtains structural estimates of the vector of structural parameters λ by solving the problem

$$\lambda_{MD} = \underset{\lambda}{\text{argmin}} \{ [\widehat{\theta} - F(\lambda)]' V_\theta [\widehat{\theta} - F(\lambda)] \} \ , \tag{2.12}$$

subject to the restrictions (2.8) – (2.10), where $\widehat{\theta}$ denotes the vector of the estimated reduced form parameters (cf. Watson 1995). The matrix V_θ denotes the asymptotic covariance matrix of the estimator $\widehat{\theta}$. The function $\theta = F(\lambda)$ relates the reduced form parameters to the structural parameters. This function is the vectorized analog of the two matrix functions

$$\begin{aligned} \Sigma &= B^{-1}RR'(B')^{-1} \\ \Psi &= (I_n - A_1 - \cdots - A_p)^{-1}B^{-1}R \end{aligned}$$

subject to the constraints (2.8) – (2.10). The estimation procedure based on (2.12) is asymptotically equivalent to the ML estimation.

Furthermore, a GMM procedure can be adopted. The assumption that the structural shocks are uncorrelated gives rise to the following $(n-1)n/2$ moment conditions:

$$E(u_{it}u_{jt}) = 0 \quad \text{for } i < j \ , \tag{2.13}$$

subject to the constraints (2.8) – (2.10), where $u_t = [u_{1t}, \ldots, u_{nt}]'$ is given by

$$u_t = B^{-1}R(y_t - A_1 y_{t-1} - \cdots - A_p y_{t-p}) \ .$$

The GMM estimate is obtained by minimizing the distance function

$$\theta_{\text{GMM}} = \underset{\lambda}{\text{argmin}} \left\{ \left(\sum_{t=p+1}^{T} m_t(\lambda)' \right) W \left(\sum_{t=p+1}^{T} m_t(\lambda) \right) \right\}$$

where $m_t(\lambda) = (u_{it}u_{jt})_{i<j}$, λ is the vector of structural parameters and W is a weight matrix given by

$$W = \left[\sum_{t=p+1}^{T} E[m_t(\lambda)m_t(\lambda)'] \right]^{-1} \ .$$

In practice this weight matrix is replaced by

$$W_T = \left[\sum_{t=p+1}^{T} m_t(\widehat{\lambda})m_t(\widehat{\lambda})' \right]^{-1} \ ,$$

where $m_t(\widehat{\lambda}) = (\hat{u}_{it}\hat{u}_{jt})_{i<j}$ and $\hat{u}_{it}$ denotes the estimated structural shock from a consistent initial estimate of the model. A possible initial estimate may be

obtained by replacing W by the identity matrix. If the errors are i.i.d. and normally distributed, the GMM estimator is asymptotically equivalent to the ML estimator. However, if the errors are non-normal or heteroscedastic, the GMM estimator is asymptotically more efficient than the (pseudo) ML estimator.

2.5 A latent variables framework

Assume that the matrices $\Gamma_1^*, \ldots, \Gamma_{p-1}^*$ in (2.5) are not subject to restrictions so that the model can be represented as

$$B\Delta\widetilde{y}_t = \Pi^*\widetilde{y}_{t-1} + Ru_t, \tag{2.14}$$

where $\Delta\widetilde{y}_t$ and $\widetilde{y}_{t-1}$ denote the residual vectors from the regressions of Δy_t and y_{t-1} on $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$. For convenience we drop the tildes and consider the VAR[1] model.

To derive the latent variable representation we pre-multiply (2.14) by R^{-1} yielding

$$\begin{aligned} R^{-1}B\Delta y_t &= R^{-1}\Pi^* y_{t-1} + u_t \\ &= -\Psi^{-1}y_{t-1} + u_t. \end{aligned}$$

Using $\Psi = -\Pi^{*-1}R$ the two sets of latent variables are defined as

$$\xi_t^y = R^{-1}B\Delta y_t, \tag{2.15}$$

$$\xi_t^x = -\Psi^{-1}y_{t-1}. \tag{2.16}$$

Using these variables the system can simply be written as

$$\xi_t^y = \xi_t^x + u_t. \tag{2.17}$$

The measurement equations (2.15) and (2.16) relate the latent variables to the observed variables, whereas the system equations (2.17) specify the relationship between the latent variables. The problem with (2.15) is that the equation is non-linear in the structural parameters. To avoid such complications it is convenient to introduce another set of latent variables defined by the measurement equation

$\Delta y_t = \xi_t^z$ and, thus, from (2.15) the corresponding system equation is $R\xi_t^y = B\xi_t^z$. The complete LV system results as

$$\textbf{Measurement equations:} \quad \begin{cases} \Delta y_t &= \xi_t^z \\ y_{t-1} &= -\Psi \xi_t^x \end{cases}$$

$$\textbf{System equations:} \quad \begin{bmatrix} \xi_t^z \\ \xi_t^y \end{bmatrix} = \begin{bmatrix} (I-B) & R \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \xi_t^z \\ \xi_t^y \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} \xi_t^x + \begin{bmatrix} 0 \\ u_t \end{bmatrix} .$$

This is the conventional formulation of Linear Structural Relations (LISREL) as suggested by Jöreskog (1969). The matrix B is normalized to have ones on the leading diagonal. Note that the vector of latent variables ξ_t^z is a linear combination of ξ_t^y and, therefore, no additional measurement equations are required.[2] Furthermore, if $B = I_n$ we can neglect ξ_t^z and a LV model can be formulated with (2.17) as the set of system equations.

The main advantage of formulating the structural VAR as a LV model of the LISREL type is that the system is formulated linearly in the parameters of interest B, R, Ψ. Such a system can conveniently be estimated and tested using widely used software like the LISREL package by Jöreskog and Sörbom (1993).

2.6 Imposing long run restrictions

So far we have assumed that the vector y_t has a stationary representation. In this section the case that the lag polynomial $A(L)$ has some roots on the unit circle is considered and, thus, the shocks may have a permanent effect on the time series.

The first model that introduces information on the long run behaviour is the often cited paper by Blanchard and Quah (1989). In this paper it is assumed that one variable (output) is $I(1)$, whereas the second variable (the unemployment rate)

[2] To implement such a system into the LISREL program, one specifies the first set of measurement equations as $\Delta y_t = \xi_t^z + K\xi_t^y$ and imposes the restriction $K = 0$.

is $I(0)$. For the identification of the shocks it is assumed that the demand shock has a transitory effect on output (that is its effect tends to zero with an increasing time horizon), whereas supply shocks may effect output permanently. Crowder (1995) demonstrates that this model can be written as a special cointegrated system with the cointegration vector $\beta = [0, 1]'$.[3]

The Blanchard-Quah model was extended to systems with more then two variables by King et al. (1991). Further developments of this kind of models are discussed in Levtchenkova, Pagan and Robertson (1998).

Contemporaneous identification of the shocks. While cointegration among the variables imposes restrictions on the long run relationship between the system variables, the first class of models assumes that the *structural shocks* are identified by restrictions on R and B. Accordingly, the structural shocks are identified by their contemporaneous relationship instead of their long run effects on the variables. An example is the empirical model of Clarida and Gertler (1996).

Assume that the cointegrating vectors are known so that the error correction terms can be constructed as $w_t = \beta' y_t$. The structural system is written as

$$B\Delta y_t = \alpha^* w_{t-1} + \sum_{k=1}^{p-1} \Gamma_k^* \Delta y_{t-k} + R u_t \ , \tag{2.18}$$

where $\alpha^* = B\alpha$. Since α is assumed to be unrestricted, so is α^*. Hence, we are able to "partial out" the unrestricted coefficient matrices α^* and $\Gamma_1^*, \ldots, \Gamma_{p-1}^*$ by regressing Δy_t on $w_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$. Let $\widetilde{\varepsilon}_t$ denote the corresponding vector of residuals. Then, the structural system is transformed into

$$B\widetilde{\varepsilon}_t = R\widetilde{u}_t \ . \tag{2.19}$$

It follows that the only difference from the original SVAR model is that the structural model is formulated by using the residuals from the VECM representation (1.12) instead of the unrestricted VAR residuals.

[3] However, Crowder (1995) is in error when claiming that $\alpha'_\perp$ is (a multiple of) $[0, 1]$ (p.236). Such a restriction would imply that no lagged U_t enters the output equation. This, however, is an additional restriction not entailed in the original model. Since the BQ-Model is just identified, there is no such over-identification restriction to be tested as suggested by Crowder.

Permanent and transitory shocks. To accommodate a structural interpretation of the long run behaviour of the system, the shocks are classified as "permanent" and "transitory". While a permanent shock has a long run effect on the future level of at least one variable, all impulse response functions with respect to a transitory shock die out as the lag horizon tends to infinity.

In a system with cointegration rank r, there are $n-r$ permanent and r transitory shocks. As shown, e.g., in Johansen (1994), the vector of permanent shocks can be represented as $u_t^p = \alpha_\perp' \varepsilon_t$. The vector of transitory (or "cyclical") shocks can be represented as $u_t^c = \gamma' \varepsilon_t$, where the columns of the $n \times r$ matrix γ are linearly independent of the columns of $\alpha_\perp$. In the structural form (2.2) the shocks are related to the innovations by

$$u_t = \begin{bmatrix} u_t^p \\ u_t^c \end{bmatrix} = R^{-1} B \varepsilon_t \ . \tag{2.20}$$

In order to impose the restriction that the first $n-r$ elements of u_t are permanent shocks, the upper $(n-r) \times n$ block of the matrix $(R^{-1}B)$ must be orthogonal to α. Furthermore, let R be lower block-triangular so that the upper right $(n-r) \times r$ block is zero. In this case, the permanent shocks result from setting the upper block of B equal to $\alpha_\perp$. Accordingly, shocks are restricted to have a permanent effect by setting the corresponding rows of $\Gamma = R^{-1}B$ equal to zero.

This gives rise to the following estimation procedure. Let $\Delta\widetilde{y}_t$ and $\widetilde{y}_{t-1}$ denote the residuals from a regression of Δy_t and y_{t-1} on $\Delta y_{t-1}, \ldots, \Delta y_{t-p}$, respectively. Letting $\xi_t^y = R^{-1}B\Delta\widetilde{y}_t = \Gamma\widetilde{w}_{t-1} + \widetilde{u}_t$ with $\widetilde{w}_{t-1} = \beta'\widetilde{y}_{t-1}$, the corresponding LV model is given by

Measurement equations:
$$\begin{cases} \Delta\widetilde{y}_t &= \xi_t^z \\ \widetilde{w}_{t-1} &= \xi_t^x \end{cases}$$

System equations:
$$\begin{bmatrix} \xi_t^z \\ \xi_t^y \end{bmatrix} = \begin{bmatrix} (I-B) & R \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \xi_t^z \\ \xi_t^y \end{bmatrix} + \begin{bmatrix} 0 \\ \Gamma \end{bmatrix} \xi_t^x + \begin{bmatrix} 0 \\ u_t \end{bmatrix} ,$$

where the permanent shocks result from setting the corresponding rows of Γ equal to zero.

This framework is more general than the estimation method suggested by King et al. (1991) that is based on a just identified system with a triangular identification of the transitory shocks (see Section 3.8.3 for more details).

Linear restrictions on the long run effect of the shocks. If the structural model implies restrictions on the *magnitude* of the long run effect, a LV representation of the LISREL type does not seem possible in the general case. However, for $B = I_n$ the LV framework can be used as follows.

Following Warne (1990) the cointegrated system is transformed to a stationary VAR using the vector of transformed variables $z_t = [y_t'\beta, \Delta y_t'\gamma]'$, where γ is some fixed $n \times (n-r)$ matrix linearly independent of β. The vector z_t has a stationary VAR(p) representation and, thus, the usual SVAR framework can be used. Let Ψ_z denote the long run impulse response of the integrated series $s_t = \sum_{j=1}^{t} z_j = [\sum_{j=1}^{t} y_j'\beta, y_t'\gamma]'$ with respect to the structural shocks. From the relationship between y_t and z_t we can deduce the long run response of s_t from the impact matrix Ψ_z. Let

$$D(L) = \begin{bmatrix} I_r & 0 \\ 0 & (1-L)I_{n-r} \end{bmatrix} \quad \text{and} \quad \overline{D}(L) = \begin{bmatrix} (1-L)I_r & 0 \\ 0 & I_{n-r} \end{bmatrix}$$

such that $D(L)\overline{D}(L) = (1-L)I_n$. Furthermore, let $Q = [\beta, \gamma]'$. The relationship between z_t and y_t is given by

$$z_t = D(L)Qy_t \ . \tag{2.21}$$

Multiplying with $Q^{-1}\overline{D}(L)$ gives

$$\Delta y_t = Q^{-1}\overline{D}(L)z_t \ . \tag{2.22}$$

It follows that the long run impulse response of y_t denoted by Ψ_y is given by

$$\Psi_y = Q^{-1}D(1)\Psi_z \tag{2.23}$$

and, thus, linear restrictions on Ψ_y imply linear restrictions on Ψ_z. That is, the structural form of the cointegrated VAR can be estimated by using the LV representation for z_t as considered in Section 3.5.

It is important to notice that the matrix Q is treated as fixed. Since β can be estimated super-consistently, one may use an estimate of β without affecting the asymptotic properties of the estimation procedure. The choice of γ is more problematical. Of course it is possible to select the values of γ arbitrarily as a block from the identity matrix, for example. However, if there is no prior information about the cointegrating space, it may be that some vectors of γ fall in the cointegrating space and, therefore, the matrix Q is not invertible. In practice one should therefore test whether the column vectors of γ fall in the cointegrating space by using, e.g., Johansen's (1991) LR tests.

2.7 Inference on impulse responses

The impulse response $\varphi_h(i,j)$ measures the effect of the j'th shock on the i'th variable h periods ahead (cf. Lütkepohl 1991, Sec. 2.3.2). Formally the impulse response function is defined as

$$\varphi_h(i,j) = \frac{\partial E(y_{i,t+h}|\mathcal{Y}_{t-1}, u_{jt})}{\partial u_{jt}} \ , \tag{2.24}$$

where $\mathcal{Y}_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$. The impulse responses can be computed from the MA representation of the system. It is convenient to introduce the companion form of the VAR model:

$$z_t = M z_{t-1} + \nu_t \ , \tag{2.25}$$

where

$$\begin{aligned} z_t &= [y_t', y_{t-1}', \ldots, y_{t-p+1}']' \\ \nu_t &= [\varepsilon_t', 0']' \\ M &= \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{bmatrix} . \end{aligned}$$

The matrices B and R are subject to the linear restrictions

$$Q_1 \text{vec}(B) = h_1 \tag{2.26}$$
$$Q_2 \text{vec}(R) = h_2. \tag{2.27}$$

These restrictions can be re-written as

$$\text{vec}(B) = h_1^* + Q_1^* b^* \tag{2.28}$$
$$\text{vec}(R) = h_2^* + Q_2^* r^* \ , \tag{2.29}$$

where b^* and r^* are unrestricted parameter vectors, Q_1^* and Q_2^* are known matrices with the property $Q_1 Q_1^* = 0$ and $Q_2 Q_2^* = 0$ and h_1, h_2 are vector of constants.

From this representation the original time series vector is obtained as $y_t = J' z_t$, where J is a selection matrix defined as $J' = [I_n, 0, \ldots, 0]$. Using the companion form, the MA representation can be written as

$$\begin{aligned} y_t &= \sum_{j=0}^{\infty} J' M^j J \varepsilon_{t-j} \\ &= \sum_{j=0}^{\infty} J' M^j J B R^{-1} u_{t-j} \ . \end{aligned}$$

Let Φ_h denote the impulse response matrix with typical element $\varphi_h(i,j)$. From the MA representation this matrix results as

$$\Phi_h = J' M^h J B^{-1} R \ . \tag{2.30}$$

The impulse response can therefore be represented as a nonlinear function of the structural parameters:

$$\varphi_h = \text{vec}(\Phi_h) = f_\varphi^h(\theta) \ , \tag{2.31}$$

where $\theta = \text{vec}([A_1, \ldots, A_p], B, R) \equiv [\theta_1', \theta_2', \theta_3']'$.

Lütkepohl (1990) derives the asymptotic distribution of the estimated impulse response function for the VAR model with triangular identification scheme by using the Delta method. This method can also be used to derive the asymptotic distribution of the estimated impulse responses in the structural model:

$$\widehat{\varphi}_h = f_\varphi^h(\widehat{\theta}) \ . \tag{2.32}$$

The Delta method is based on the mean value expansion:

$$\widehat{\varphi}_h = \varphi_h + \nabla f_\varphi^h(\bar{\theta})(\widehat{\theta} - \theta) \ , \tag{2.33}$$

where $\bar{\theta}$ lies on the intersection of θ and $\widehat{\theta}$, and $\nabla f_\varphi^h(\bar{\theta})$ denotes the derivative $\partial f_\varphi^h(\theta)/\partial\theta'$ evaluated at $\bar{\theta}$. To derive the asymptotic distribution of the estimated impulse responses, the derivatives must be computed. For the case $R = I_n$ and a structural matrix B with the restriction $Q_1 \text{vec}(B) = h_1$, the derivative is derived in Vlaar (1999). The following lemma provides the derivative of $f_\varphi^h(\theta)$ for the case that the variables are not cointegrated and no long run restrictions are used to identify the parameters.

LEMMA 2.1 *Let y_t be generated by the structural model given by (2.3) with the parameter vector $\theta = [\theta_1', b^{*\prime}, r^{*\prime}]'$ and $\theta_1 = vec(A_1, \ldots, A_p)$. Then, the derivative is given by*

$$\nabla f_\varphi^h(\theta) = [D_h^a, D_h^b, D_h^r] \ ,$$

where

$$\begin{aligned}
D_0^a &= 0 \quad \textit{for } h = 0 \\
D_h^a &= \sum_{j=0}^{h-1} (B^{-1}R)' J' (M')^{h-1-j} \otimes J' M^h J \quad \textit{for } h = 1, 2, \ldots \\
D_h^b &= \left(-R' B'^{-1} \otimes J' M^h J B^{-1} \right) Q_1^* \\
D_h^r &= \left(I \otimes J' M^h J B^{-1} \right) Q_2^* \ .
\end{aligned}$$

PROOF: The result follows from (2.30) and eq. (13) in Sec. 10.5.1 of Lütkepohl (1996) and is similar to the result given in Lütkepohl (1990, Prop. 1).

From eq. (1) in Sec. 10.6.1 of Lütkepohl (1996) it follows that

$$\frac{\partial f_\varphi^h(\widehat{\theta})}{\partial \text{vec}(B)'} = -R' B'^{-1} \otimes J' M^h J B^{-1} \ .$$

Using $\partial \text{vec}(B)/\partial b^{*\prime} = Q_1^*$ the matrix D_0^b follows immediately.

From eq. (3) in Sec. 10.4.1 of Lütkepohl (1996) it follows that

$$\frac{\partial f_\varphi^h(\widehat{\theta})}{\partial \text{vec}(R)'} = I \otimes J' M^h J B^{-1} \ .$$

Using $\partial \text{vec}(R)/\partial r^{*\prime} = Q_2^*$ the matrix D_0^r follows easily. ■

If the structural model is subject to long run restrictions, the derivative is much more complicated. The problem is that the respective restrictions are nonlinear:

$$\begin{aligned} & Q_3 \text{vec}\left[(I_n - A_1 - \cdots - A_p)^{-1} B\right] \\ & \quad = Q_3 \left[I_n \otimes (I_n - A_1 - \cdots - A_p)^{-1}\right] \text{vec}(B) = h_3 \ . \end{aligned}$$

For the estimated model this restriction takes the form:

$$\widehat{Q}_3 \text{vec} \widehat{B} = h_3 \ , \tag{2.34}$$

where $\widehat{Q}_3 = Q_3[I_n \otimes (I_n - \widehat{A}_1 - \cdots - \widehat{A}_p)^{-1}]$. Vlaar (1999) calls (2.34) a "stochastic restriction". He notes that the software packages MALCOLM in RATS (see Mosconi 1998) treats this restriction like a usual linear restriction by neglecting the stochastic nature of the matrix $\widehat{Q}_3$. Vlaar (1999) therefore derives the correct asymptotic distribution by taking into account the distribution of $\widehat{Q}_3$. The resulting formulae are however rather messy and it is therefore appealing to make use of the reformulation used to derive the latent variable representation.

Following Johansen (1988) the system can be represented as

$$\begin{aligned} \Delta y_t &= \Pi y_{t-p} + \widetilde{\Gamma}_1 \Delta y_{t-1} + \cdots + \widetilde{\Gamma}_{p-1} \Delta y_{t-p+1} + \varepsilon_t \\ &= -B^{-1} R \Psi^{-1} y_{t-p} + \widetilde{\Gamma}_1 \Delta y_{t-1} + \cdots + \widetilde{\Gamma}_{p-1} \Delta y_{t-p+1} + B^{-1} R u_t \ , \end{aligned}$$

where $\widetilde{\Gamma}_j = -I_n + A_1 + \cdots + A_j$ for $j = 1, \ldots, p-1$. The difference to the form used in (1.12) is that here the levels are lagged by p periods instead of a single lag. It is assumed that $\widetilde{\Gamma}_j$ are unrestricted matrices for $j = 1, \ldots, p-1$. This model can be written in companion form as in (2.25), where

$$\begin{aligned} z_t &= [\Delta y_t', \Delta y_{t-1}', \ldots, \Delta y_{t-p+2}', y_{t-p+1}'] \\ M &= \begin{bmatrix} \widetilde{\Gamma}_1 & \widetilde{\Gamma}_2 & \cdot & \widetilde{\Gamma}_{p-2} & \widetilde{\Gamma}_{p-1} & -B^{-1} R \Psi^{-1} \\ I_n & 0 & . & 0 & 0 & 0 \\ 0 & I_n & . & 0 & 0 & 0 \\ . & . & . & . & . & . \\ 0 & 0 & . & 0 & I_n & 0 \\ 0 & 0 & . & 0 & I_n & I_n \end{bmatrix} \end{aligned}$$

$$\nu_t = [(B^{-1}Ru_t)', 0]' .$$

The impulse response matrix with respect to h periods results as

$$\Phi_h = J'M^hJB^{-1}R .$$

The important advantage of this representation is that long run restrictions can be written as linear restrictions on Ψ:

$$Q_3\text{vec}(\Psi) = h_3 \quad \text{or} \quad \text{vec}(\Psi) = h_3^* + Q_3^*\psi^* .$$

The derivative can be obtained in the same manner as in the case without long run restrictions. The results are given in the following lemma.

LEMMA 2.2 *Let y_t be generated by the VAR[p] model with the structural form (2.2) and the parameter vector $\theta = [\theta_1', b^{*\prime}, r^{*\prime}, \psi^{*\prime}]'$, where $\theta_1 = vec(A_1, \ldots, A_p)$. Then, the derivative is given by*

$$\nabla f_\varphi^h(\theta) = [D_h^a, D_h^b, D_h^r, D_h^\psi] ,$$

where

$$\begin{aligned}
D_0^a &= 0 \quad \textit{for } h = 0 \\
D_h^a &= \sum_{j=0}^{h-1}(B^{-1}R)'J'(M')^{h-1-j} \otimes J'M^hJ \textit{ for } h = 1, 2, \ldots \\
D_h^b &= \left[F_M^h(I_{p+1} \otimes K_{n,p+1} \otimes I_n)(vec(\tau) \otimes \Psi'R'B'^{-1} \otimes B^{-1}) - R'^{-1} \otimes J'M^hJ\right] Q_1^* \\
D_h^r &= \left[-F_M^h(I_{p+1} \otimes K_{n,p+1} \otimes I_n)(vec(\tau) \otimes I_n \otimes B^{-1}R) + I_n \otimes J'M^hJB^{-1}\right] Q_2^* \\
D_h^\psi &= \left[-F_M^h(I_p \otimes K_{n,p} \otimes I_n)(vec(\tau) \otimes \Psi' \otimes B^{-1})\right] Q_3^*
\end{aligned}$$

where $K_{n,k}$ is a commutation matrix (e.g. Lütkepohl 1996, p. 9) and

$$\begin{aligned}
F_M^h &= \sum_{j=0}^{h-1}(B^{-1}R)'J'(M')^{h-1-j} \otimes J'M^h \\
\tau &= \begin{bmatrix} 0_{1\times p-1} & 1 \\ 0_{p-1\times p-1} & 0_{p-1\times 1} \end{bmatrix} .
\end{aligned}$$

PROOF: The derivative of $f_\varphi^h(\widehat{\theta})$ with respect to θ_1 has the same form as in the case of Lemma 2.1. For the remaining derivatives we have to account for the dependence of M on B, R and Ψ. Using the chain rule and the product rule we obtain

$$\frac{\partial f_\varphi^h(\widehat{\theta})}{\mathrm{vec}(B)'} = F_M^h \frac{\mathrm{vec}(M)}{\mathrm{vec}(B)'} - R'B'^{-1} \otimes J'M^hJB^{-1}$$
$$\frac{\partial f_\varphi^h(\widehat{\theta})}{\mathrm{vec}(R)'} = F_M^h \frac{\mathrm{vec}(M)}{\mathrm{vec}(R)'} + I_n \otimes J'M^hJB^{-1} \ ,$$

where $F_M^h = \partial f_\varphi^h(\widehat{\theta})/\mathrm{vec}(M)'$. To compute the derivatives of M it is useful to write

$$M = \begin{bmatrix} \widetilde{\Gamma}_1 & \widetilde{\Gamma}_2 & \cdot & \widetilde{\Gamma}_{p-2} & \widetilde{\Gamma}_{p-1} & 0 \\ I_n & 0 & . & 0 & 0 & 0 \\ 0 & I_n & . & 0 & 0 & 0 \\ . & . & . & . & . & . \\ 0 & 0 & . & 0 & I_n & 0 \\ 0 & 0 & . & 0 & I_n & I_n \end{bmatrix} + \tau \otimes -B^{-1}R\Psi$$

Using eq. (7b) of Sec. 10.5.5, eq. (3) of Sec. 10.4.1 and eq. (1) of Sec. 10.6.1 of of Lütkepohl (1996) we get

$$\begin{aligned} \frac{\partial \mathrm{vec}(M)}{\partial \mathrm{vec}(\Psi)'} &= -(I_{p+1} \otimes K_{n,p+1} \otimes I_n)(\mathrm{vec}(\tau) \otimes I_n \otimes B^{-1}R) \\ \frac{\partial \mathrm{vec}(M)}{\partial \mathrm{vec}(B)'} &= (I_p \otimes K_{n,p} \otimes I_n)(\mathrm{vec}(\tau) \otimes \Psi'R'B'^{-1} \otimes B^{-1}) \\ \frac{\partial \mathrm{vec}(M)}{\partial \mathrm{vec}(R)'} &= -(I_p \otimes K_{n,p} \otimes I_n)(\mathrm{vec}(\tau) \otimes \Psi' \otimes B^{-1}). \end{aligned}$$

Finally, using $\partial \mathrm{vec}(B)/\partial b^{*\prime} = Q_1^*$, $\partial \mathrm{vec}(R)/\partial r^{*\prime} = Q_2^*$ and $\partial \mathrm{vec}(\Psi)/\partial \psi^{*\prime} = Q_3^*$ the derivatives follow. ■

With these results for the derivative $\nabla f_\varphi^h(\theta)$ we are able to state the main result:

THEOREM 2.1 *Under the assumptions that*

$$(i) \qquad rk[\nabla f_\varphi^h(\theta)] = n^2$$
$$(ii) \qquad \sqrt{T}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma_\theta)$$

where Σ_θ is the asymptotic covariance matrix of $\widehat{\theta}$, it follows that

$$\sqrt{T}(\widehat{\varphi}_h - \varphi_h) \xrightarrow{d} N(0, \nabla f_\varphi^h(\theta)\Sigma_\theta \nabla f_\varphi^h(\theta)') \ ,$$

where $\nabla f_\varphi^h(\theta)$ is given in Lemma 2.1 or Lemma 2.2.

PROOF: The proof of this theorem immediately follows from the mean value expansion given in (2.33). ■

It is important to notice that the assumptions (i) and (ii) in Theorem 2.1 may fail in practice. As pointed out by Lütkepohl (1990), the derivative may be zero for some particular values of the structural parameters. This problem already occurs in a univariate model (cf. Benkwitz et al. 2000). In the simple AR(1) model $y_t = \alpha y_{t-1} + \varepsilon_t$ the estimated impulse response is $\widehat{\varphi}_h = \widehat{\alpha}^h$. The derivative is $\nabla f_\varphi^h(\theta) = h\alpha^{h-1}$. Obviously, the derivative is zero for $\alpha = 0$ and $h \geq 2$ and, thus, the impulse responses have a nonstandard limiting distribution and converge with the rate $T^{h/2}$ instead of $\sqrt{T}$. Benkwitz et al. (2000) discuss alternative methods to overcome this problem.

Assumption (ii) of Theorem 2.1 may be violated if the process has roots on or inside the unit circle. In this case some parameters have a nonstandard limiting distribution. Again it is useful to consider the simplest case of a univariate AR(1) model. If $\alpha = 1$, the least-squares estimator is distributed as

$$T(\widehat{\alpha} - 1) \ \Rightarrow \ \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr}$$

and, obviously, the impulse responses fail to be asymptotically normally distributed for all h. Phillips (1998) shows that for $h \to \infty$ the impulse responses are not even consistent as $h/T \to \lambda$ for $T \to \infty$ and $\lambda > 0$, that is, if the lag horizon of the impulse responses are large relative to the sample size. This problem can however be resolved by using the VECM representation instead of an unrestricted VAR in levels (Phillips 1998 and Vlaar 1999).

Sims and Zha (1994) argue hat classical (frequentist) confidence intervals can be poor measures of the ignorance about estimated impulse response functions in small samples. First, using bootstrap methods yields a confidence interval which

is based on "an average across informative and uninformative samples" (Sims and Zha 1994, p.6). In contrast, a Bayesian confidence interval always conditions on the actual sample and, thus, provides a measure of ignorance by using the data at hand. Second, in small samples confidence intervals often exhibit a substantial asymmetry. This important feature cannot be represented by using an asymptotic approach based on a normal limiting distribution.

Most of the empirical studies using SVAR models neglect possible information on the long run relationships among the variables. To address this question in more detail, assume that y_t is generated by a cointegrated model with a VECM representation given in (1.12). Engle and Yoo (1991) show that the information matrix of the likelihood function assuming normal innovations is block diagonal with respect to $\theta_1 = \text{vec}(\beta)$, $\theta_2 = \text{vec}(\alpha, \Gamma_1, \ldots, \Gamma_{p-1})$ and $\theta_3 = \text{vec}(\Sigma)$. Accordingly, the estimates $\widehat{\Sigma}$ (and therefore $\widehat{B}$ and $\widehat{R}$) are asymptotically independent of the estimates of $\Gamma_1, \ldots, \Gamma_{p-1}, \Pi$ and β. It follows that the structural parameters in B and R are not affected by imposing the rank restrictions on Π as long as Π is estimated consistently.

Since the impulse responses involve the parameter matrices B, R and Π, this result is not sufficient to conclude that the distribution of the impulse response function is not affected by the cointegration properties. Nevertheless, it can be shown that the impulse response matrix at a fixed lag horizon is indeed asymptotically the same whether an unrestricted VAR in levels or an appropriate VECM is used to estimate the impulse responses. This result is based on the following lemma (see also Phillips and Durlauf 1986 and Sims et al. (1990) for similar results).

LEMMA 2.3 *Assume that y_t is generated by a cointegrated VAR. Let $\widetilde{\Pi}$ denote the unrestricted least-squares estimator of Π and $\widehat{\Pi}$ is the ML estimator conditional on the true cointegration rank r. Then, $\widetilde{\Pi} = \widehat{\Pi} + o_p(T^{-1/2})$.*

PROOF: To facilitate the notation, a VAR(1) model is used to demonstrated the result. The extension to VAR(p) models is straightforward.

Let $\Upsilon = [T^{-1/2}\beta, T^{-1}\beta_\perp]'$. In a VAR(1) model the unrestricted least-squares

estimator for Π is given by:

$$\begin{aligned}\widetilde{\Pi} &= \left(\sum_{t=2}^{T}\Delta y_t y_{t-1}'\right)\left(\sum_{t=2}^{T} y_{t-1}y_{t-1}'\right)^{-1} \\ &= \left(\sum_{t=2}^{T}\Delta y_t y_{t-1}'\Upsilon'\right)\left(\Upsilon\sum_{t=2}^{T} y_{t-1}y_{t-1}'\Upsilon'\right)^{-1}\Upsilon.\end{aligned}$$

Denoting

$$\begin{aligned}\sum_{t=2}^{T}\Upsilon y_{t-1}y_{t-1}'\Upsilon' &= \begin{bmatrix} A_1 & o_p(1) \\ o_p(1) & A_2 \end{bmatrix} \\ \sum_{t=2}^{T}\varepsilon_t y_{t-1}'\Upsilon' &= \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix},\end{aligned}$$

where

$$\begin{aligned} A_1 &= T^{-1}\sum_{t=2}^{T}\beta' y_{t-1}y_{t-1}'\beta & A_2 &= T^{-2}\sum_{t=2}^{T}\beta_\perp' y_{t-1}y_{t-1}'\beta_\perp \\ B_1 &= T^{-1/2}\sum_{t=2}^{T}\varepsilon_t y_{t-1}'\beta & B_2 &= T^{-1/2}\sum_{t=2}^{T}\varepsilon_t y_{t-1}'\beta_\perp \\ B_3 &= T^{-1}\sum_{t=2}^{T}\varepsilon_t y_{t-1}'\beta & B_4 &= T^{-1}\sum_{t=2}^{T}\varepsilon_t y_{t-1}'\beta_\perp\ .\end{aligned}$$

It follows that

$$\sqrt{T}(\widetilde{\Pi}-\Pi) = \begin{bmatrix} B_1A_1^{-1}\beta' + T^{-1/2}B_2A_2^{-1}\beta_\perp' \\ B_3A_1^{-1}\beta' + T^{-1/2}B_4A_2^{-1}\beta_\perp' \end{bmatrix}.$$

Since the ML estimator $\widehat{\Pi}$ is asymptotically equivalent to the least-squares estimator of Δy_t on $\beta' y_{t-1}$ we have

$$\sqrt{T}(\widehat{\Pi}-\Pi) = \begin{bmatrix} B_1A_1^{-1}\beta' \\ B_3A_1^{-1}\beta' \end{bmatrix} + o_p(1)$$

and, thus, the required result follows. ■

It may be surprising to learn that imposing the correct rank restriction on the matrix Π yields no gain in efficiency for estimating the impulse responses. The reason is that the limiting behaviour of the impulse responses at a finite lag horizon is dominated by the limiting behaviour of the short run parameters.

THEOREM 2.2 *Let y_t be generated by a cointegrated VAR with cointegration rank r and a VECM representation as in (1.12). Then, the asymptotic distribution of the estimated impulse response at a finite lag horizon h is the same whether the impulse response is estimated from an unrestricted VAR instead or a VECM with cointegration rank r.*

PROOF: The VECM model has a companion form according to (2.25) with

$$z_t = \begin{bmatrix} \Delta y_t \\ \vdots \\ \Delta y_{t-p+2} \\ y_t \end{bmatrix}, \quad M = \begin{bmatrix} \widetilde{\Gamma}_1 & \widetilde{\Gamma}_2 & . & \widetilde{\Gamma}_{p-2} & \widetilde{\Gamma}_{p-1} & \Pi \\ I_n & 0 & . & 0 & 0 & 0 \\ . & . & . & . & . & . \\ 0 & 0 & . & I_n & 0 & 0 \\ 0 & 0 & . & 0 & I_n & I_n \end{bmatrix}, \quad \nu_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The impulse responses result as

$$\Phi_h = J' M^h J B R^{-1}$$

or

$$\varphi_h = f_\varphi^h(\theta) \ ,$$

where θ =vec$(\widetilde{\Gamma}_1, \ldots, \widetilde{\Gamma}_{p-1}, \Pi, B, R)$. Using Lemma (2.3) it follows that $\widehat{\theta} = \widetilde{\theta} + o_p(T^{-1/2})$ so that under the conditions of Theorem 2.1

$$\begin{aligned} \widehat{\varphi}_h(i,j) &= f_\varphi(\widehat{\theta}) \\ &= f_\varphi(\widetilde{\theta}) + o_p(T^{-1/2}) \end{aligned}$$

and, thus, $\widehat{\varphi}_h(i,j)$ and $\widetilde{\varphi}_h(i,j)$ have the same limiting distributions. ■

This result implies that information on the long run relationship of the variables does not help to improve the short run impulse responses. However, Clarida and Gertler (1996) claim that *in small samples* the estimates may be improved by imposing the correct cointegration rank. Furthermore, as shown by Phillips (1998) the estimates of the impulse responses deteriorate with an increasing lag horizon and become insignificant if the lag horizon grows as fast as the sample size. Hence, in practice it may be advantageous to use a VECM representation with a proper cointegration rank to estimate the impulse responses.

2.8 Empirical applications

In this section we consider three structural models that can be seen to be "representative" for many recent applications using SVAR models. The first example is related to the early contributions by Sims (1981, 1986). The simple structure of the model allows the application of a convenient instrumental variable estimator. The model suggested by Blanchard and Quah (1989) achieves identification by imposing assumptions on the long run effect of the shocks. This model is extended by King et al. (1991) in order to identify $n - r$ permanent shocks in a cointegrated system. For the latter two models point estimates of the structural parameters can be obtained analytically by exploiting the relationship between the parameters of the structural and the reduced form. However, standard errors of the estimates are not available by using this approach.

The fourth example is an over-identified structural model which was recently suggested by Swanson and Granger (1997). This model is motivated using a causal interpretation based on graph theory. The LV framework provides efficient estimation and inference procedures for estimating and testing such kind of models.

2.8.1 A simple IS-LM model

To illustrate the "first generation" SVAR models advocated by Sims (1981), Bernanke (1986) and others, a traditional IS-LM model is estimated using quarterly U.S. data from 1970(i) to 1997(iv). The output Y_t is measured by real GDP, M_t is the monetary base as computed by the Federal Reserve Bank of St. Louis divided by the GDP deflator, and i_t is the 3-month interbank interest rate.

Let ε_t^Y, ε_t^i and ε_t^M denote the innovations of the VAR equations for Y_t, i_t and M_t. The model considered in Pagan (1995) is given by

$$\begin{aligned} \varepsilon_t^Y &= \alpha\, \varepsilon_t^i + u_t^{IS} && \text{(IS curve)} \\ \varepsilon_t^i &= \beta_1\, \varepsilon_t^Y + \beta_2\, \varepsilon_t^M + u_t^{LM} && \text{(inverted money demand)} \\ \varepsilon_t^M &= u_t^{MS} && \text{(money supply)} \end{aligned} \tag{2.35}$$

Accordingly, $R = I_3$ and

$$B = \begin{bmatrix} 1 & -\alpha & 0 \\ -\beta_1 & 1 & -\beta_2 \\ 0 & 0 & 1 \end{bmatrix} .$$

Estimating this model (2.35) by using the LISREL package yields the estimated equations:

$$\begin{aligned} \varepsilon_t^Y = \; & 0.04\,\varepsilon_t^i + \quad u_t^{IS} \\ & (0.26) \end{aligned}$$

$$\begin{aligned} \varepsilon_t^i = \; & 0.14\,\varepsilon_t^Y - \quad 0.73\,\varepsilon_t^M + u_t^{LM} \; , \\ & (0.51) \qquad (-4.99) \end{aligned}$$

where t-values are given in parentheses. The estimate of the coefficient α is insignificant and positive so that we do not obtain a reasonable estimate of the IS equation. The parameters of the LM curve have the correct sign but the estimate of β_1 is statistically insignificant.

Nevertheless, we compute the impulse response functions to analyse the impact of structural shocks on the system variables. The respective graphs are given in Figure 3.1. It should be mentioned that the LISREL package does not provide confidence intervals for the impulse responses so that it is difficult to assess the economic relevance of the results. With this limitation Figure 3.1 a) shows that in response to an upward shift of the IS curve output increases immediately but the increase dies out within six years. A positive money demand shock (LM) is equivalent to an increase in the interest rate and from Figure 3.1 b) it can be seen that output decreases gradually with a minimum after two years. The impulse response function shows that the small positive correlation between output and interest rates resulting from the estimated IS curve is merely a short run phenomenon and at longer horizons the relationship is strong and negative as expected from economic theory.

The negative response of output following a positive money supply shock is somewhat puzzling. From theory we expect that an increase in real money balances yields an expansive effect on output. However, the impulse response function

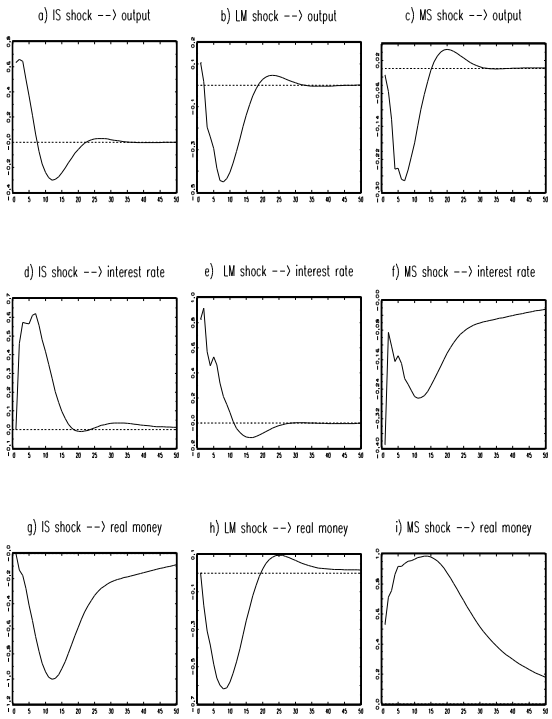


Figure 2.1: Impulse responses for the IS-LM model

Table 3.1: Variance decompositions

		IS	LM	MS
y_t	1	0.996	0.003	0.001
	4	0.915	0.055	0.030
	8	0.525	0.331	0.144
	12	0.133	0.648	0.219
i_t	1	0.010	0.810	0.180
	4	0.332	0.614	0.054
	8	0.533	0.404	0.063
	12	0.622	0.264	0.114
M_t	1	0.000	0.000	1.000
	4	0.050	0.068	0.882
	8	0.525	0.331	0.144
	12	0.133	0.648	0.219

The table presents the forecast variances decomposition with respect to the structural shocks. All numbers are given in decimales. The forecast horizon is measured in quarters.

given in Figure 3.1 c) implies that a positive money supply shock produces a drop in output after roughly 2 years. Although the size of the effect is moderate, this result suggests that monetary policy shocks are not measured appropriately by innovations in the monetary base. Indeed, recent empirical work demonstrates that monetary policy in the U.S. is more accurately measured by innovations in the Federal Funds Rate or a combination of a narrow monetary base and short-term interest rates (Bernanke and Mihov 1997).

A positive shift in the IS curve increases interest rates with a maximal response at two years (Figure 3.1 d), whereas real money decreases gradually with a minimum at three years (Figure 3.1 g). These effects are predicted by the IS-LM model. Similarly, a LM shock leads to an increase in interest rates and a

decrease in real money (see Figures 3.1 e and 3.1 h) which is also predicted by the theoretical model. Finally, a positive money supply shock leads to an immediate drop in interest rates (Figure 3.1 f) and a gradual increase in real money. This effect is known as the "liquidity effect" and is also an important consequence of the standard IS-LM model (e.g. Hamilton 1997).

To assess the importance of the three different shocks for the system variables, the forecast error variances of the variables are decomposed with respect to the shocks. Since the shocks are orthogonal, their contribution can be measured as a fraction of the total forecast error variance. The results for different forecast horizons are presented in Table 3.1. It turns out that the money supply shock contributes only a small fraction to the forecast error variance of output. Once again this result confirms that the innovation in money does not seem to be an accurate indicator for monetary policy. IS shocks clearly dominate the short run behaviour of the output series but with respect to a longer forecast horizon, IS shocks become less important. Finally, money demand shocks play a minor role in the short run. However with an increasing forecast horizon, the LM shocks become more and more important. The relative contribution of the shocks to the forecast error variance of interest rates and real money can be interpreted in a similar manner.

Summing up, the results demonstrate that even such a simple model for the U.S. economy is able to produce reasonable results although some findings are at variance with our preconceptions emerging from the basic version of an IS-LM model. In fact, it is often encountered in practice that some aspects of the SVAR model are inconsistent with theoretical reasoning. For example, in several studies for the U.S., monetary policy shocks produce results consistent with common priors about the qualitative effects of monetary policy on output and prices, but produce a so-called "liquidity puzzle", that is, expansive monetary policy increases interest rates (see, e.g., Fung and Kasumovich 1998). However, such a result does not demonstrate the general failure of the SVAR approach. Rather, such findings give an indication of the direction in which the model may be improved.

2.8.2 The Blanchard-Quah model

Based on a simple economic model, Blanchard and Quah (1989) identify supply shocks to have persistent effects on output whereas demand shocks are transitory. That is, in a VAR model with $y_t = [\Delta Y_t, U_t]'$, where Y_t denotes the log of output and U_t is the unemployment rate, the vector of structural shocks $u_t = [u_t^s, u_t^d]'$ is identified by setting the (1,2) element of the long run impact matrix Ψ equal to zero. Furthermore, we let $\Omega = I$, so that the shocks are normalized to have unit variance.

The structural form is related to the reduced form as

$$\Psi\Psi' = (I - A_1 - \cdots - A_p)^{-1}\Sigma(I - A_1' - \cdots - A_p')^{-1} \ .$$

Since Ψ is assumed to be lower triangular, it can be obtained from a Choleski decomposition of the matrix $(I - A_1 - \cdots - A_p)^{-1}\Sigma(I - A_1' - \cdots - A_p')^{-1}$. It is important to note, however, that this computation implies that the diagonal elements of Ψ are strictly positive. This additional identification restriction will be discussed below.

The Blanchard-Quah model can also be estimated by using the LV representation that has the simple form:

$$\textbf{Measurement equations:} \quad \begin{cases} \Delta\widetilde{y}_t &= R\xi_t^y \\ \widetilde{y}_{t-1} &= -\Psi\xi_t^x \end{cases}$$

$$\textbf{System equations:} \quad \xi_t^y = \xi_t^x + u_t \ ,$$

where $\Delta\widetilde{y}_t$ and $\widetilde{y}_{t-1}$ denote the first step residuals obtained from a regression of Δy_t and y_{t-1} on the lags of Δy_t and the deterministic terms of the system. The $(1,2)$ element of Ψ is set to zero.

With respect to the discussion in Section 3.6 it is also interesting to compare the findings with the corresponding results using a cointegration approach. Since U_t is assumed to be stationary, it follows that $\beta = [0,1]'$ and the error correction

form of the model is

$$\begin{aligned}\Delta Y_t &= \alpha_1 U_{t-1} + \sum_{j=1}^{8} c_{11,j}\Delta Y_{t-j} + \sum_{j=1}^{7} c_{12,j}\Delta U_{t-j} + \varepsilon_{1t} \\ \Delta U_t &= \alpha_2 U_{t-1} + \sum_{j=1}^{8} c_{21,j}\Delta Y_{t-j} + \sum_{j=1}^{7} c_{22,j}\Delta U_{t-j} + \varepsilon_{2t} \ .\end{aligned}$$

An estimate of $\alpha = [\alpha_1, \alpha_2]'$ is obtained by OLS and an orthogonal complement of this vector is $\alpha_\perp = [\alpha_2, -\alpha_1]'$. Accordingly, an estimator of the matrix Q defined in Section 3.6 can be constructed as

$$Q = \begin{bmatrix} 0 & 1 \\ \alpha_2 & -\alpha_1 \end{bmatrix}.$$

To compute R, let P be the Choleski factor of $(Q\Sigma Q')$. Then, as shown in Section 3.6, R can be obtained as $R = Q^{-1}P$.

To estimate the system, we first employ the software package LISREL 8.12a to estimate the latent variable representation. The maximum likelihood estimates result as

$$\widehat{R} = \begin{bmatrix} 0.070 & 0.876 \\ (0.237) & (15.75) \\ 0.207 & -0.196 \\ (3.067) & (2.841) \end{bmatrix}$$

and

$$\widehat{\Psi} = \begin{bmatrix} 0.489 & 0 \\ (3.067) & \\ 0.008 & -3.809 \\ (0.005) & (2.995) \end{bmatrix},$$

where the absolute t-values are given in parentheses. The estimates are numerically identical to the estimates obtained by applying the computational method suggested by Blanchard and Quah (1989). However, the t-statistics are not available by using the latter technique.

It may be interesting to note that the "supply shock" has an insignificant contemporaneous impact ($|t| = 0.237$) on output growth. Thus, the demand shock is approximately the innovation of the output growth equation.

Figure 2.2: Impulse responses for the Blanchard-Quah model

In Figure 3.2 the impulse response functions with respect to the structural shocks are presented. The impulse response function are very similar to the ones obtained by Blanchard and Quah (1989) and a careful economic interpretation of the figures is given in their paper. The 90% confidence intervals are computed by 1000 Monte Carlo draws of the estimated reduced form and normally distributed errors. Comparing our confidence sets with the bootstrap intervals presented in Blanchard and Quah (1989) reveals that their confidence bands are much more asymmetric than ours. As argued by Sims and Zha (1994), this is due to an improper bootstrap algorithm used by Blanchard and Quah (1989).

In practical applications, the cointegration approach yields different estimates of the structural parameters even if maximum likelihood techniques are used. This is due to the different model representation used in this framework. If some variables are differenced, the degree of the corresponding lag polynomials is reduced by one and, thus, the VAR equations have a different lag order. It is possible to account for such differences by using a VAR representation with different lag orders for the variables. The efficient estimation of such systems is, however, much more complicated than an unrestricted VAR so that we do not consider such modifications in what follows.

For the Blanchard-Quah model the cointegration approach yields the estimate

$$\widehat{R}_{CI} = \begin{bmatrix} 0.0746 & 0.9296 \\ 0.2198 & -0.2082 \end{bmatrix},$$

which is – in absolute values – not far away from the estimates using the LISREL approach.

2.8.3 The KPSW model

A simple example of a cointegrated system with a structural identification of the shocks is provided by King, Plosser, Stock and Watson (KPSW) (1991). Standard RBC models with permanent technology shocks imply that the logarithm of private output (q_t), consumption (c_t) and investment (i_t) are $I(1)$ processes with the cointegration relations $c_t - q_t \sim I(0)$ and $i_t - q_t \sim I(0)$. Since $\beta_\perp$ (the orthogonal complement of β) is a vector with identical elements, the common trend

representation according to (1.13) is found as

$$\Delta y_t = \begin{bmatrix} \Delta q_t \\ \Delta c_t \\ \Delta i_t \end{bmatrix} = \begin{bmatrix} a \\ a \\ a \end{bmatrix} (\alpha_\perp' \varepsilon_t) + \widetilde{C}(L)\Delta\varepsilon_t \ ,$$

where the polynomial $\widetilde{C}(L) = \widetilde{C}_0 + \widetilde{C}_1 L + \widetilde{C}_2 L^2 + \cdots$ has all roots outside the complex unit circle. It is seen that the permanent shock has an identical long run impact on the variables.

To achieve a structural identification, King et al. (1991) assume that the first shock (u_{1t}) is permanent and the remaining two shocks (u_{2t}, u_{3t}) are transitory. Furthermore, the shocks are mutually uncorrelated with unit variance, i.e., $E(u_t u_t') = I_3$. These assumptions are sufficient to identify the permanent shock but are not sufficient for the identification of the transitory shocks. The identification of u_{1t} follows immediately from the fact that in this case $\alpha_\perp$ is a 3×1 vector and, therefore, $u_{1t} = \alpha_\perp' \varepsilon_t / \sqrt{\alpha_\perp' \Sigma \alpha_\perp}$. However, to identify the transitory shocks an additional restriction is needed.

King et al. (1991) assume that the transitory shocks enter the system recursively, as in the triangular identification scheme implied by using a Choleski decomposition. Such a procedure is obtained by letting R be lower triangular and

$$B = \begin{bmatrix} 1 & b_1 & b_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} .$$

Since the first shock is assumed to be permanent, the first row $[1, b_1, b_2]$ must be orthogonal to the columns of the matrix α. Furthermore the (1,1) element of B is normalized to unity in order to achieve identification of the (1,1) element of R. The matrix $\Gamma = R^{-1}B\alpha$ is a 3×2 matrix with zeros in the first row (see Section 3.6). This model has $\{6, 2, 4\}$ unknown parameters for $\{R, B, \Gamma\}$ and, thus, the structural model has the same number of parameters as the reduced form.

For the empirical application we use U.S. real national income account variables for 1960(i) through 1988(iv), so that the data and the sample period are similar to King et al. (1991). The variable q_t is the per capita private Gross National

Product (GNP less government purchases of goods and services), c_t measures per capita Private Consumption Expenditures and i_t is measured by per capita Gross Private Fixed Investments.[4] As in King et al. (1991) all computations are based on a VAR(6) representation.

Table 3.2 presents the results of Johansen's (1991) trace test for the cointegration rank. An unrestricted linear time trend is included and the corresponding critical values of Perron and Campbell (1993) are applied. The hypothesis that the three series are not cointegrated ($r = 0$) can be rejected at a significance level of 0.10. Furthermore, the hypothesis that the cointegration rank is $r = 1$ or $r = 2$ cannot be rejected. Although this is only a weak empirical support we follow King et al. (1991) and choose the cointegration rank $r = 2$ in what follows. Using the LR test suggested by Johansen and Juselius (1992) the theoretical cointegration vectors $\beta_1 = [-1, 1, 0]'$ and $\beta_1 = [-1, 0, -1]'$ cannot be rejected at a critical value of 0.05.

Maximizing the likelihood function of the structural model, the following estimates are obtained:

$$\widehat{B} = \begin{bmatrix} 1 & 0.418 & -0.300 \\ & (0.09) & (0.50) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \widehat{\Gamma} = \begin{bmatrix} 0 & 0 \\ -0.103 & 0.065 \\ (1.01) & (1.75) \\ 0.079 & -0.029 \\ (0.62) & (0.37) \end{bmatrix}$$

and

$$\widehat{R} = \begin{bmatrix} 0.950 & 0 & 0 \\ (0.28) & & \\ 0.488 & -0.486 & 0 \\ (1.76) & (1.77) & \\ -2.003 & -2.781 & 1.794 \\ (1.92) & (0.90) & (0.33) \end{bmatrix}.$$

[4] Our data is different from the data used in KPSW. In our application the variables are taken from the data bank of St. Louis Federal Reserve Bank, except the population series which comes from the Citibase data tape.

Figure 2.3: Responses w.r.t. the permanent shock

In Figure 3.3 the impulse response functions with respect to the permanent shock are depicted. In all, they have a similar shape as those presented in King et al. (1991). However, the investment response is much more pronounced than the respective response presented in the original paper. This seems to be due to differences in the data construction.

An alternative way to formulate the structural model is to adopt a structural model for the transformed set of variables $z_t = [c_t - q_t, i_t - q_t, \Delta q_t]'$. The corre-

Table 3.2: Cointegration statistics

H_0	Trace	0.10 critical value	0.05 critical value
$r \leq 2$	4.481	9.75	11.62
$r \leq 1$	15.74	21.22	23.65
$r = 0$	37.17	36.52	39.67
$\beta_1 = [-1, 1, 0]'$	2.575	2.71	3.84
$\beta_2 = [-1, 0, 1]'$	0.160	2.71	3.84
joint(β_1, β_2)	2.642	4.60	5.99

Notes: This table presents LR statistic for the cointegration rank and for prespecified cointegration vectors. The critical values of Perron and Campbell (1993) are applied.

sponding (concentrated) latent variables representation is

$$\textbf{Measurement equations:} \quad \begin{cases} \Delta\widetilde{z}_t &= R\xi_t^z \\ \widetilde{z}_{t-1} &= -\Psi_z \xi_t^x \end{cases}$$

$$\textbf{System equations:} \quad \xi_t^z = \xi_t^x + u_t \ .$$

The matrix Ψ_z has the following structure:

$$\Psi_z = \begin{bmatrix} * & * & * \\ * & * & * \\ a & 0 & 0 \end{bmatrix},$$

where a is the long run impact of the permanent shock. The zeros in the last row of Ψ_z result from the assumption that the second and the third shock are transitory. In order to identify the transitory shocks we need an additional restriction. Again one may set the (2,3) element of R equal to zero. However, one should note that the matrix R in the representation for z_t is not identical to the matrix R of the former model for Δy_t.

Since the (concentrated) reduced form implies 15 parameters and $[\Psi_z, R]$ contains 7+9=16 parameters we need an additional restriction. In King et al. (1991) and the cointegration approach considered above the transitory shocks are uniquely obtained by using a Choleski decomposition. An equivalent identification is obtained using a matrix with a structure like

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & 0 \\ r_{31} & r_{32} & r_{33} \end{pmatrix} ,$$

where r_{22} and r_{33} are positive numbers. Here, identification is achieved by assuming that the third structural shock is uncorrelated with the second innovation. The kind of identification restriction for the transitory shocks does not, however, affect the construction of the permanent shock.

Maximizing the likelihood function of the LV model, the long run impact of the permanent shock is estimated as 0.855 with an estimated standard deviation of 0.181. The corresponding estimate from the cointegration approach (0.617, see above) is in the range of two times the standard deviation.

In the present application we use the transformation matrix

$$Q = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad Q^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} .$$

According to (2.22) the impulse response function (IRF) of q_t with respect to the permanent shocks is identical to the cumulated IRF of Δq_t with respect to the permanent shock. The IRF of c_t is equal to the sum of the cumulated IRF of $c_t - q_t$ with respect to the permanent shock and the IRF of $c_t - q_t$ with respect to the the first transitory shock. The IRF of i_t is equal to the sum of the cumulated IRF of $c_t - q_t$ with respect to the permanent shock and the IRF of $i_t - q_t$ with respect to the the second transitory shock.

From $\widehat{B}$ it is seen that the estimated permanent shock is proportional to $[1, 0.418, -0.3]\varepsilon_t$. However, the parameter estimates are not significantly different from zero. Excluding the investment innovation gives a permanent shock with the

estimate 1.81 and a t-value of 3.56 for the coefficient attached to the consumption innovation. From these results it is obvious that the parameter estimates in $\widehat{B}$ are highly correlated and a data consistent restriction can change the remaining parameter estimates dramatically.

In all, this application demonstrates that the simple neoclassical growth model seems to be perform well at least for the data used in King et al. (1991). The long run response of consumption, investment and output to a permanent (technology) shock is the same but the short run response of investment is much more cyclical than the short run response of consumption. This is also predicted by standard neoclassical models (e.g. Kydland and Prescott 1982).

2.8.4 The causal graph model of Swanson-Granger (1997)

In order to identify a causal structure for the innovations, Swanson and Granger (1997) adopt a graph theoretical approach. Assume that a vector of uncorrelated shocks $u_t = [u_{1t}, \ldots, u_{nt}]'$ can be found such that the innovations of the (stationary) VAR are arranged as $\varepsilon_{1t} = u_{1t}$; $\varepsilon_{2t} = a_2\varepsilon_{1t} + u_{2t}$; $\cdots$; $\varepsilon_{nt} = a_n\varepsilon_{n-1,t} + u_{nt}$. Such a structure can be represented by a causal graph as

$$\begin{array}{ccccccc} \varepsilon_{1t} & \longrightarrow & \varepsilon_{2t} & \longrightarrow & \cdots & \longrightarrow & \varepsilon_{nt} \\ \uparrow & & \uparrow & & \uparrow & & \uparrow \\ u_{1t} & & u_{2t} & & \cdots & & u_{nt} \end{array} \tag{2.36}$$

The arrows are assumed to indicate a simple form of causal directionality (cf. Swanson and Granger, 1997). It should be noted that this concept of causality is different from what is known as "Granger causality". The latter concept implies an ordering in time so that a cause must be prior to the effect. The causal graph approach is a way to formalize the notion of "instantaneous causality" discussed, e.g., in Lütkepohl (1993).

The causal ordering given in (2.36) implies that $E(\varepsilon_{it}\varepsilon_{kt}|\varepsilon_{jt}) = 0$ for any $i < j < k$. Accordingly, Swanson and Granger (1997) suggest testing the partial correlation between ε_{it} and ε_{kt} conditional on ε_{jt} in order to recover the causal ordering empirically. To motivate this concept consider the condition

$E(\varepsilon_{1t}\varepsilon_{3t}|\varepsilon_{2t}) = 0$. From $\varepsilon_{3t} = a_3\varepsilon_{2t} + u_{3t}$ and $E(\varepsilon_{2t}u_{3t}) = 0$ it follows that given ε_{2t} the variable ε_{1t} does not help to "predict" ε_{3t}. On the other hand, if $a_3 \neq 0$, then the partial correlation between ε_{3t} and ε_{2t} conditional on ε_{1t} is different from zero and, thus, ε_{2t} provide additional information to predict ε_{3t}. Hence, ε_{2t} may be seen as a cause of ε_{3t}. It is important to note that the causal graph $\varepsilon_{1t} \to \varepsilon_{2t} \to \varepsilon_{3t}$ implies the same restriction on the conditional expectation as $\varepsilon_{3t} \to \varepsilon_{2t} \to \varepsilon_{1t}$ and, therefore, the direction of the graph is not identified.

This causality concept can be represented as a structural VAR by letting

$$B = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -a_2 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -a_3 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -a_n & 1 \end{bmatrix} \tag{2.37}$$

and $R = I_n$. The corresponding LV representation is given by

Measurement equations: $\widetilde{\varepsilon}_t = \xi_t^y$

System equations: $\xi_t^y = (I - B)\xi_t^y + u_t$.

It is easy to see that such a structural model implies $(n-2)(n-1)/2$ over-identifying restrictions, which can be tested empirically by comparing the value of the log likelihood function with the corresponding value of the unrestricted VAR using the LR test statistic. This test avoids problems of the sequential test procedure suggested in Swanson and Granger (1997).

As an application of this approach we consider a four-variate system for the U.S. including the log of the money base M3 (M), log of real GNP (Y), the 3-month LIBOR interest rate (R) and the log of the implicit GNP deflator (P). The data for M, Y and P are seasonally adjusted and is taken from the Main Economic Indicator data base. The interest rate R is not seasonally adjusted and taken from the IMF data base. All data run from 1970(i) through 1994(iv).

Table 3.3: Tests for the cointegration rank of the reduced form

H_0	max. eigenvalue	crit. val.	trace	crit. val.
$r = 3$	1.473	3.962	1.473	3.962
$r = 2$	5.509	14.04	6.982	15.20
$r = 1$	33.31	20.78	40.29	29.51
$r = 0$	50.06	27.17	90.35	47.18

Note: "max. eigenvalue" and "trace" indicate Johansen's LR statistics for the cointegration rank including an unrestricted constant in VECM representation. The critical values correspond to a significance value of 0.05 and are taken from Osterwald-Lenum (1990).

The vector of time series is modeled using a cointegrated VAR including a vector of constants in the error correction representation. Model selection criteria like AIC and tests for serial correlation of the errors suggest that a VAR(2) model is sufficient to render white noise errors. Table 3.3 presents Johansen's (1991) trace and maximum eigenvalue test statistics for the cointegration rank. The results suggest that there are two cointegration relationships among the variables. Therefore, the residuals from the estimated VECM model are used for the structural model of the form $B\widetilde{\varepsilon}_t = \widetilde{u}_t$, where B has the form as in (2.37).

Table 3.4 presents the results of the LR test of the over-identifying restrictions implied by the corresponding causal graph. There exist $4! = 24$ possible graphs. However, since the direction of the graph is not identified, only 12 graphs are considered. The "mirror graphs" yield the same values of the LR statistic, so there is no point in reporting them. It turns out that only two causal graphs are accepted at a significance level of 0.05. Both graphs do not correspond well with the type of transmission mechanism of monetary policy discussed in the literature. The graphs imply that (real) effects on output are "causally prior" to the (nominal) effects on prices, while theory usually conjectures the reverse ordering. Since we only intend to give an illustration of the approach suggested by Swanson and Granger (1997), we do not attempt to reconcile this finding with the

Table 3.4: Tests of the causal graph model

graph	LR	p-value	graph	LR	p-value
$M \to Y \to R \to P$	3.32	0.34	$R \to M \to Y \to P$	15.5	< 0.01
$R \to Y \to M \to P$	5.77	0.12	$M \to P \to Y \to R$	15.9	< 0.01
$R \to P \to M \to Y$	10.7	0.01	$Y \to R \to M \to P$	16.1	< 0.01
$M \to Y \to P \to R$	11.8	0.01	$M \to R \to Y \to P$	17.1	< 0.01
$Y \to M \to R \to P$	12.0	< 0.01	$M \to R \to P \to Y$	22.1	< 0.01
$M \to P \to R \to Y$	12.3	< 0.01	$R \to M \to P \to Y$	24.5	< 0.01

Note: The entries of this table present the values of the LR statistic of the over-identifying restrictions of the Swanson-Granger model. Under the null hypothesis the statistic is χ^2 distributed with three degrees of freedom.

counterfactual evidence using alternative methods. A more thorough discussion of the merits and pitfalls of this new approach is left for future research.

2.9 Problems with the SVAR approach

Identification problems. In a critique of Blanchard and Quah (1989), Lippi and Reichlin (1993) point out that a stationary VAR has an infinite manifold of different MA representations. Among those, however, only one (the Wold or fundamental representation) has the property that the determinant of the MA polynomial has all its roots on or outside the unit circle. There are, however, non-fundamental representations with roots inside the unit circle. The associated white noise errors are linear combinations of current, past and future values of y_t. Lippi and Reichlin (1993) argue that such non-fundamental representations usually cannot be ruled out a priori. Specifically, nonfundamental representations can occur if the information space of the economic agents is different from the information space of the econometrician.

In a reply to this critique Blanchard and Quah (1993) admit that the limitation to fundamental representations is somewhat arbitrary but that it cannot

be considered less plausible than alternative non-fundamental representations. In fact, in most cases it is not possible to select the "correct" representation on empirical grounds and it seems thus natural to select the most convenient one. The Blanchard–Quah methodology formally decompose a set of time series into certain orthogonal shocks with some prespecified properties. Whether such quantities have the attached economic meaning of representing demand and supply shifts cannot be decided empirically. It merely provides one admissible interpretation of the data which is generally not refutable as long as just identified structures are considered.

Another important problem is that even if one confines oneself to the Wold representation, the structural model need not be unique. The Blanchard-Quah model provides a good example for illustrating the problems involved by the identification of structural VARs. The reduced form is a bivariate VAR modeling the unemployment rate and output growth and the two structural errors are interpreted as supply and demand shocks, where the effect of the demand shocks on the output measure is assumed to vanish in the long run (see Section 3.8.2).

Let r_{ij}, σ_{ij}, c_{ij} denote the (i,j) elements of R, $\Sigma = RR'$ and $C = \Pi^{-1}$. Then the set of restrictions for the structural model is:

$$r_{11}^2 + r_{12}^2 = \sigma_{11} \tag{2.38}$$

$$r_{21}^2 + r_{22}^2 = \sigma_{22} \tag{2.39}$$

$$r_{21}r_{11} + r_{22}r_{12} = \sigma_{12} \tag{2.40}$$

$$c_{11}r_{12} + c_{12}r_{22} = 0\ . \tag{2.41}$$

It is easily seen that these restrictions identify the absolute value of the parameters but are not sufficient to determine the sign of the parameter uniquely. Specifically, if the matrix $R_{(1)}$ obeys the restrictions then also the matrix $R_{(2)} = -R_{(1)}$ and

$$R_{(3)} = \begin{pmatrix} -r_{11} & -r_{12} \\ r_{21} & r_{22} \end{pmatrix}\ , \quad R_{(4)} = \begin{pmatrix} r_{11} & r_{12} \\ -r_{21} & -r_{22} \end{pmatrix}\ ,$$
$$R_{(5)} = \begin{pmatrix} -r_{11} & r_{12} \\ -r_{21} & r_{22} \end{pmatrix}\ , \quad R_{(6)} = \begin{pmatrix} r_{11} & -r_{12} \\ r_{21} & -r_{22} \end{pmatrix}\ ,$$

fulfill the above constraints. In this simple example it is obvious that a unique

representation is obtained if there are additional assumptions about the signs of the diagonal elements of R. Indeed, this is the case if the Choleski decomposition is applied for computing the structural parameters from the reduced form as in Blanchard and Quah (1989). In models using $R = I$, identification is usually achieved by setting the diagonal elements of B equal to one.

Waggoner and Zha (1997) show that the normalization used to identify the shocks is not only important for the interpretation of the model but also affects statistical inference in small samples. A similar point is made by Pagan and Robertson (1998) by using a different perspective. To illustrate the problem, we follow these authors and consider the model suggested by Gordon and Leeper (1994):

$$\begin{aligned} MS: &\quad m_t = b_{12} i_t + \gamma_1' x_{1t} + u_t^s \\ MD: &\quad i_t = b_{21} m_t + \gamma_2' x_{2t} + u_t^d \ , \end{aligned}$$

where u_t^s and u_t^d represent money supply and money demand shocks, respectively. The vectors x_{1t} and x_{2t} comprise further (exogenous) variables that are used to identify the two equations. These variables are, however, inessential for our reasoning and so we drop these variables for the ease of exposition.

The contemporary impulse responses are obtained as

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}^{-1} = \frac{1}{1 - b_{21} b_{12}} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} . \tag{2.42}$$

Economic theory implies that $b_{12} > 0$ and $b_{21} < 0$ so that $b_{21} b_{12}$ is negative. In finite samples there is a positive probability that for the estimates $\hat{b}_{21} \hat{b}_{12} > 1$, in particular, if b_{21} and b_{12} are estimated using "weak instruments" (Pagan and Robertson 1998). It follows that in finite samples there may be a substantial probability that the impulse responses swich their sign. The resulting distribution of the impulse responses therefore tends to have two different modes and the normal distribution yields a poor approximation to the actual distribution. Pagan and Robertson (1998) show that this problem is in fact relevant in empirical applications. Waggoner and Zha (1997) resolve the problem by using a data dependent normalization rule (the "ML distance normalization"). It is important

to note, however, that the problem disappears as T tends to infinity, because the denominator in (2.42) converges in probability to a positive constant under the maintained assumption.

Another identification problem arises when using long run restrictions. Faust and Leeper (1997) show that additional assumptions on the short run dynamics are required to enable reliable inference. An intuitive explanation is that for a VAR with an infinite lag order, a small change of the model parameters has a cumulative effect on the long run responses of the shocks. An obvious remedy of this problem is to assume that the VAR order is finite which is equivalent to the assumption that all higher autoregressive coefficients are equal to zero. Alternatively, one may assume that not only the long run effect is zero but also the responses beyond a lag of, say, 40 periods vanish (Faust and Leaper 1997, Sec. 2.2). Although this critique is certainly important from a theoretical perspective, it is difficult to assess the relevance for empirical practice. The message for practical work, which can be concluded from such kind of reasoning, is that it may be hazardous to rely on long run restrictions, when identifying the model, in particular, if the lag order of the VAR is large.

Finally it is interesting to consider the problem, that the dimension of the empirical VAR is smaller than the underlying dynamic system. In this case the estimated shocks can only be estimated in an aggregated form. The crucial question is, whether the aggregated shocks have the same properties than the original shocks. In the Blanchard-Quah model the demand shock are identified as the transitory shock, whereas the supply shock has a permanent effect on output. Now, assume that there are $m_1 > 1$ orthogonal demand shocks and $m_2 > 1$ orthogonal supply shocks with the same long run properties. The interesting question is, whether it is possible to identify a "joint" aggregated demand (supply) shock as a linear combination of the original demand (supply) shocks only, or whether the aggregated system mixes up both type of shocks. Faust and Leeper (1997) derive conditions that ensure that the aggregated demand and supply shock can be suitably separated. However, without knowledge of the complete dynamical process, the condition cannot be verified. Faust and Leeper (1997) present in-

direct evidence that the condition for an appropriate separation of demand and supply shocks is violated.

Robustness. In a number of recent papers, the SVAR approach is criticized for its lack of robustness under different structural assumptions. For example, Cochrane (1998) finds that output responses vary a lot as one changes identifying assumptions. Rudebusch (1998, p. 925) concludes the comparison of the shocks from three different SVAR specifications:

> *"Obviously, these three series give very different interpretations of the history of monetary policy surprises, and in several periods, the VAR series describe a stance for monetary policy that is greatly at variance with historical accounts."*

As a second example, Cooley and Dwyer (1998) summarize their comparison of different SVAR models:

> *"The findings of this paper suggest that conclusions about the importance of technology and other shocks based on simple SVARs are certainly not invariant to the identifying assumptions and may not be very reliable as vehicles for identifying the relative importance of shocks."*

Needless to say that this danger is always present, when structural econometric models are based on prior information derived from economic theory. There seems to be a consensus, however, that SVAR models are particularly sensitive to alternative identifying assumptions and that conclusions from an SVAR model are therefore "unreliable" and "fragile".[5] In a similar vein, it is claimed that economic theory is not informative enough to arrive at a unique SVAR specification and, hence, the researcher is free to choose among a large number of possible specifications. To quote Uhlig (1997):

> *"There is a danger here that we just get out what we have stuck in, albeit a bit more polished and with numbers attached ..."*

[5] It is interesting to note that the same arguments were put forward by Leamer (1983) to criticize the simultaneous equation approach.

To overcome this problem, different approaches were suggested to account for the uncertainty of model specification. King and Watson (1997) compute the impulse responses for all models with plausible identification assumptions. The resulting range of impulse response functions represents the possible effects on the variable of interest. Of course, whenever the resulting range of impulse responses is wide, little is learned from the estimation of SVAR models.

A similar approach is suggested by Uhlig (1997) and Faust (1998). They assume that economic theory is at best able to make predictions on the sign of the impulse response at various lag horizons. Uhlig (1997) uses a penalized likelihood approach that introduces a penalty term implying a sharp decrease of the maximization criterion whenever the inequality conditions on the impulse response functions are violated. Faust (1998) considers the range of models that results from a rotation of the system and satisfies the sign restrictions on the impulse responses. Within this range of models he considers the worst case and therefore draws a very cautious conclusion about the effect on the shocks.

A closely related approach employs a Bayesian methodology. Sims and Zha (1998) impose prior information on the autoregressive parameters, whereas Dwyer (1998) and Gordon and Boccanfuso (1998) directly specify prior distributions on the plausible shape of the impulse response functions. The latter approach allows the imposition of inequality restrictions on the impulse response functions by assigning zero weights to implausible values of the prior distributions of the impulse responses.

Uhlig (1997) adopts a related Bayesian approach. He uses Litterman's prior distribution on the parameters of the VAR model and computes an estimate of the structural parameters for each realization of the artificial model, simulated by using the posterior distribution of the VAR parameters. The parameter values which do not satisfy the structural assumptions (for example that a positive shock in interest rates leads to a decreases in the monetary base) are dropped and confidence intervals can be computed from the set of valid impulse responses among all simulated realizations.

These alternative approaches are suggested to overcome the lack of robustness

in the structural VAR analysis. It is however not clear, whether it is realistic to claim for such kind of robustness. In general, prior information is necessary because economic data are too uninformative about the underlying structure. Therefore, it is expected that prior assumptions are important for the analysis of economic data and, thus, different assumptions may produce different results. Similarly, we cannot hope to be able to refute economic theories definitely by using a structural approach (cf. Breitung et al. 1993). Therefore, these methods should be seen as a more or less useful device to recover structures behind the data. In other words, economic data is used to quantify prior beliefs about the economic system rather then to decide between alternative theories.

Chapter 3

Problems of temporal aggregation

This chapter addresses the question, how temporal aggregation affects structural inference. Indeed this is an important issue, since in practice the frequency of observation is usually different from what may be called the *natural* frequency of the underlying time series. For example, at financial markets, agents react very rapidly to news and, thus, the natural frequency of a model that describes the behaviour of financial agents is likely to be minutes rather the hours or days.

However, if the natural frequency of the underlying process is high relative to the *observed* frequency of the data, temporal aggregation may completely change the structural relations in the system. It is therefore interesting to investigate the effects of temporal aggregation on structural inference.

To do so, it is assumed that the data are generated by a discrete stochastic process with some causal ordering as suggested by Granger (1969). Given the causal structure of the data generating process, what kind of structural model results when the data is temporally aggregated? Does the observed structure correspond to the underlying causal relationship among the variables or is the observed structure different from the original causal structure?

In this chapter it is argued that SVAR models as considered in Chapter 3 can be motivated with respect to the underlying causal ordering by using Granger's original definition of a causal process. Conditions are given to ensure that sufficiently aggregated data are able to reflect the causal structure in the covariance

matrix of the VAR innovations. It does not come as a surprise, however, that temporal aggregation implies a loss of information. Specifically, the contemporaneous correlation of the innovations cannot identify the direction of causality. However, conditions can be given that rule out "spurious instantaneous causality" and, thus, we are able to verify that the observed structure is not an artifact of the underlying aggregation process.

Furthermore, I address the question what happens with a particular structure that can be represented by a SVAR model if the data is temporally aggregated. It is shown that in general the structural representation of the VAR may change substantially with the aggregation level of the data. For an empirical example I find that the general shape of the impulse responses does not change a lot by aggregating the data. However, temporal aggregation has a substantial effect on the *magnitude* of impulse responses.

There is already a rich literature that analyses the effects of temporal aggregation in a multiple time series framework. An early example is Quenouille (1957), where the temporal aggregation of ARMA processes is studied. Amemiya and Wu (1972), and Brewer (1973) refine and generalize Quenouille's result by including exogenous variables. Zellner and Montmarquette (1971) discuss the effects of temporal aggregation on estimation and testing. Engle (1969) and Wei (1978) analyze the effects of temporal aggregation on parameter estimation in a distributed lag model. More recently, Weiss (1984), Stram and Wei (1986), Lütkepohl (1987), and Marcellino (1999) provide detailed studies of the effects of temporal aggregation in a VARMA framework.

The results of these studies can be summarized by quoting Tiao (1999): *"So the causality issue is muddled once the data are aggregated. The problem is that if the data are observed at intervals when the dynamics are not working properly, then we may not get any kind of causality."* In this chapter it is argued that this statement is also true for inference based on structural VARs.

In Section 4.1 I discuss the concept of Granger causality under temporal aggregation. Some useful asymptotic results for large aggregation intervals are presented in Section 4.2. The consequences for causal inference are discussed in

Section 4.3 and the relevance in finite aggregation intervals is studied in Section 4.4 by using Monte Carlo simulations. In Section 4.5 presents an empirical application where the effects of temporal aggregation are studied by using US data of unemployment and inflation.

3.1 Granger causality

Following Granger (1969), consider a conditional distribution with respect to two information sets which are available at time t, say $\mathcal{I}_t$ and $\mathcal{I}_t^* = \{\mathcal{I}_t, x_t, x_{t-1}, \ldots\}$, where x_t denotes a (possibly causal) variable. As in Section 2.4, the variable x_t is defined to be Granger causal for the variable y_t if there exist an $h \in \{1, 2, \ldots\}$ such that

$$E(y_{t+h}|\mathcal{I}_t) \neq E(y_{t+h}|\mathcal{I}_t^*) \ . \tag{3.1}$$

If $\mathcal{I}_t = \{y_t, y_{t-1}, \ldots\}$, that is, when causality is investigated in a bivariate system, it is sufficient to consider $h = 1$ (e.g. Lütkepohl 1999).

An important problem with this definition is the choice of the sampling interval. It is possible that Granger causality is observed when data is measured at a daily basis, say, but at a monthly or quarterly frequency the causality may disappear. Thus, for inference on causality it is important that the data has the "correct" sampling frequency. There are two alternative approaches to deal with this problem. First, it may be assumed that the underlying time series has a continuous time Wold representation given by $y(t) = \int f(\tau)\varepsilon(t-\tau)\mathrm{d}\tau$, where $f(\tau)$ is a continuous vector function and $\varepsilon(t)$ is continuous white noise (e.g. Christiano and Eichenbaum (1987), Renault and Szafarz (1991)). In this framework the data should be observed at a (approximately) continuous basis to avoid misleading inference on causality. However, with the exception of some financial variables, economic data are usually observed at a monthly or quarterly frequency, so the relevance of such a continuous time framework is limited in practice.

An alternative approach is to assume that the data are generated by a *discrete* ARIMA process sampled at some given frequency (e.g. Tiao 1972, Lütkepohl 1987, Marcellino 1999). When investigating causality it seems quite obvious how

to choose the natural frequency. If we assume that cause and effect are ordered in time so that the causal event is observed in advance of the effect then the sampling frequency should be sufficient to distinguish the cause and the effect. Thus, at the natural frequency there should be no contemporaneous relationship between the cause and effect (see also Granger 1988). On the other hand, if the process is aggregated such that the cause and effect is (partly) observed at the same time, then the causal relationship implies a contemporary relationship between the variables.

It therefore seems useful to link the concept of Granger causality at the natural frequency with contemporaneous correlation for aggregated processes. Such a causality concept is provided by the following definition.

DEFINITION 3.1 *Let $\xi_t = [x_t, y_t, z_t']'$ be an $n \times 1$ vector of time series and z_t is an $(n-2) \times 1$ subvector. Further, define the information sets as $\mathcal{I}_t = \{z_t, y_t, z_{t-1}, y_{t-1}, \ldots\}$ and $\mathcal{I}_t^+ = \{\xi_t, \xi_{t-1}, \ldots\} = \{\mathcal{I}_t, x_t, x_{t-1}, \ldots\}$. Then x_t is said to be a cause of y_t with respect to the underlying sampling frequency if*

$$E(y_{t+1}|x_{t+1}, z_{t+1}, \mathcal{I}_t^+) \neq E(y_{t+1}|z_{t+1}, \mathcal{I}_t).$$

This definition combines traditional Granger causality with instantaneous causality as used in Lütkepohl (1991, p. 40f), for example. In a vector autoregressive system, causality according to Definition 3.1 can be tested by running the regression

$$y_t = \sum_{i=1}^{p} \alpha_i y_{t-j} + \sum_{j=0}^{p} \beta_j x_{t-j} + \sum_{j=0}^{p} \gamma_j' z_{t-j} + u_t$$

and testing the hypothesis $\beta_0 = \beta_1 = \cdots = \beta_p$.

If the process is observed at the natural sampling frequency, then there is no contemporaneous relationship between x_t and y_t conditional on z_t and the past of the process. Accordingly, we have $\beta_0 = 0$ at the natural frequency and our definition of causality is equivalent to the traditional (Granger) causality. If, on the other hand, the process is multivariate white noise, then x_t is an "instantaneous cause" of y_t whenever $\rho(y_t, x_t|z_t) \neq 0$, where $\rho(a, b|c)$ denotes the partial correlation between a and b conditional on c.

3.2 Asymptotics

In order to examine the relationship between contemporaneous correlation and Granger causality in the presence of temporally aggregated data, two different procedures are considered (e.g. see Lütkepohl 1987). For flow data, time series values are cumulated (or averaged) at k successive time periods

$$\bar{y}_t = k^{-1/2} \sum_{j=0}^{k-1} y_{t-j} \ ,$$

where the factor $k^{-1/2}$ is introduced to obtain a limiting process with a finite variance. The aggregated series results from applying *skip*-sampling of the form

$$\bar{Y}_N = \bar{y}_{kN} \ , \quad N = 1, 2, \dots \ .$$

Stock data are aggregated by directly applying the skip-sampling scheme to the data, so that $Y_N = y_{kN}$ for $N = 1, 2, \dots$.

It is known (see, e.g., Lütkepohl 1987 and Marcellino 1999) that if the original process is an ARMA process then the aggregated process also has an ARMA representation. In general, zero restrictions on the autoregressive representation of the original process are lost when aggregating the process, so that "spurious causality" may occur in aggregated time series.

In order to obtain some general results on causality in aggregated time series we consider an asymptotic theory for large aggregation intervals k. The following theorem summarizes the results for stationary variables.

THEOREM 3.1 *Let y_t be generated by an n dimensional linear process $y_t = C_0\varepsilon_t + C_1\varepsilon_{t-1} + C_2\varepsilon_{t-2} + \cdots$, where ε_t is white noise, $C_0 = I_n$, $E(\varepsilon_t\varepsilon_t') = \Omega$, and y_t is one-summable such that $\sum_{j=0}^{\infty} j||C_j|| < \infty$. As $k \to \infty$, the processes for the aggregated vectors Y_N and $\bar{Y}_N$ have the properties:*

stock variables:

$$(i) \qquad \lim_{k\to\infty} E(Y_N Y_N') = \sum_{j=0}^{\infty} C_j \Omega C_j'$$

$$(ii) \qquad \lim_{k\to\infty} E(Y_N Y_{N+j}') = 0 \quad \textit{for } j \geq 1,$$

flow variables:

$$(iii) \quad \lim_{k\to\infty} E(\bar{Y}_N \bar{Y}_N') = 2\pi f_y(0)$$

$$(iv) \quad \lim_{k\to\infty} k \cdot E(\bar{Y}_N \bar{Y}_{N+1}') = \sum_{j=1}^{\infty} \left(\sum_{i=0}^{j} C_i \right) \Omega \left(\sum_{i=j+1}^{\infty} C_i \right)'$$

$$(v) \quad \lim_{k\to\infty} k \cdot E(\bar{Y}_N \bar{Y}_{N+j}') = 0 \quad \textit{for } j \geq 2,$$

where $f_y(\omega)$ denotes the spectral density matrix of y_t at frequency ω.

PROOF: (i) From $Y_N = y_{kN}$ and assuming stationarity we have

$$E(Y_N Y_N') = E(y_t y_t') = \sum_{i=0}^{\infty} C_i \Omega C_i' \ .$$

(ii) Since the process is assumed to be ergodic we have

$$\lim_{k\to\infty} E(Y_N Y_{N+j}') = \lim_{k\to\infty} E(y_{kN} y_{kN+jk}') = 0$$

for all $j \geq 1$.

(iii) The vector of aggregated flow variables is given by

$$\bar{Y}_N = k^{-1/2} \sum_{j=0}^{k-1} y_{kN-j}$$

and therefore $\bar{Y}_N$ behaves as a (normalized) vector partial sum. For partial sums it is known that

$$\lim_{k\to\infty} E(\bar{Y}_N \bar{Y}_N') = \Omega + \Gamma + \Gamma',$$

where $\Gamma = \sum_{j=1}^{\infty} E(y_1 y_{1+j}')$. In the frequency domain this expression can be represented as

$$2\pi f_y(0) = \left(\sum_{j=0}^{\infty} C_j \right) \Omega \left(\sum_{j=0}^{\infty} C_j' \right).$$

(iv) Let

$$\begin{aligned} k^{1/2}\bar{Y}_N &= (1 + L + L^2 + L^{k-1})C(L)\varepsilon_t \\ &\equiv D(L)\varepsilon_t, \end{aligned}$$

where

$$D(L) = I_m + D_1 L + D_2 L^2 + \cdots$$

and

$$D_j = \sum_{i=0}^{\min(j,k-1)} C_{j-i} \ .$$

It is convenient to decompose $\bar{Y}_N$ as

$$\begin{aligned} k^{1/2}\bar{Y}_N &= D_0\varepsilon_t + D_k\varepsilon_{t-k} + D_{2k}\varepsilon_{t-2k} + \cdots \\ &\quad + D_1\varepsilon_{t-1} + D_{k+1}\varepsilon_{t-k-1} + D_{2k+1}\varepsilon_{t-2k-1} + \cdots \\ &\quad \vdots \\ &\quad + D_{k-1}\varepsilon_{t-k+1} + D_{2k-1}\varepsilon_{t-2k+1} + D_{3k-1}\varepsilon_{t-3k+1} + \cdots \\ &\equiv u_{0t} + \cdots + u_{k-1,t} \ , \end{aligned}$$

where

$$u_{jt} = D_j\varepsilon_{t-j} + D_{j+k}\varepsilon_{t-j-k} + \cdots$$

Note that $E(u_{it}u'_{jt}) = 0$ for $i \neq j$.

From

$$\begin{aligned} k^{1/2}\bar{Y}_N &= u_{0t} + \cdots + u_{k-1,t} \\ k^{1/2}\bar{Y}_{N+1} &= u_{0,t+k} + \cdots + u_{k-1,t+k} \\ k^{1/2}\bar{Y}_{N+2} &= u_{0,t+2k} + \cdots + u_{k-1,t+2k} \end{aligned}$$

we obtain:

$$k \cdot E(\bar{Y}_N\bar{Y}'_{N+1}) = \sum_{j=0}^{k-1} E(u_{jt}u'_{j,t+k}).$$

Consider

$$E(u_{0t}u'_{0,t+k}) = D_0\Omega D'_k + D_k\Omega D'_{2k} + \cdots.$$

For a summable sequence C_i we have

$$\lim_{k\to\infty} |D_{2k}| = \lim_{k\to\infty} |C_{k+1} + C_{k+2} + \cdots + C_{2k}| = 0$$

so that

$$\begin{aligned} \lim_{k\to\infty} E(u_{0t}u'_{0,t+k}) &= D_0\Omega D'_k \\ &= \Omega(C_1 + C_2 + \cdots + C_k)' \ . \end{aligned}$$

Similarly we get:

$$\begin{aligned} \lim_{k\to\infty} E(u_{1t}u'_{1,t+k}) &= D_1\Omega D'_{k+1} \\ &= (I_n + C_1)\Omega(C_2 + C_3 + \cdots + C_{k+1})' \\ \lim_{k\to\infty} E(u_{k-1,t}u'_{k-1,t-k}) &= (C_1 + \cdots + C_{k-1})\Omega(C_k + C_{k+1} + \cdots + C_{2k-1})'. \end{aligned}$$

Adding these expressions gives the desired result.

It remains to show that $\sum_{j=0}^{\infty}(\sum_{i=0}^{j} C_i)\Omega(\sum_{i=j+1}^{\infty} C_i)'$ is bounded.

This follows from

$$\left\| \sum_{j=0}^{\infty} \left(\sum_{i=0}^{j} C_i \right) \Omega \left(\sum_{i=j+1}^{\infty} C'_i \right) \right\| \leq \sum_{j=0}^{\infty} \left\| \sum_{i=0}^{j} C_i \right\| \|\Omega\| \sum_{i=j+1}^{\infty} j \, \|C_i\| \, ,$$

which is finite by assumption.

Consider

$$E(u_{0t}u'_{0,t-pk}) = D_0\Omega D'_{pk} + D_k\Omega D_{(p+1)k} + \cdots .$$

Since

$$\lim_{k\to\infty} D_{(p+j)k} = 0 \qquad \text{for } p \geq 2 \ \text{ and } \ j = 0, 1, \ldots$$

it follows that the autocovariances disappear for $p \geq 2$. ■

According to Theorem 3.1, it turns out that for $k \to \infty$, the aggregated processes are asymptotically white noise. Of course this result is not particularly surprising, since it is intuitively plausible that with increasing sampling interval, short run dynamics disappear. Furthermore, for moderate k it is expected that aggregated flow variables are well approximated by a vector MA(1) process. The

reason for this is that according to *(iv)*, the first order autocorrelation is $O(k^{-1})$, while *(v)* implies that higher order autocorrelations are $o(k^{-1})$.

Next, assume that y_t is a vector of integrated variables such that y_t is difference stationary.

THEOREM 3.2 *Let Δy_t be generated by an n dimensional linear process $\Delta y_t = \varepsilon_t + C_1\varepsilon_{t-1} + C_2\varepsilon_{t-2} + \cdots$, where it is assumed that $E(\varepsilon_t\varepsilon_t') = \Omega$, $\sum_{j=1}^{\infty} j||C_j|| < \infty$ and the matrix $\bar{C} = \sum_{j=0}^{\infty} C_j$ has full rank. As $k \to \infty$, the processes for the aggregated vectors Y_N and $\bar{Y}_N$ are characterized by:*

stock variables:

$$(i) \quad \lim_{k\to\infty} \frac{1}{k} E(Y_N - Y_{N-1})(Y_N - Y_{N-1})' = 2\pi f_{\Delta y}(0)$$

$$(ii) \quad \lim_{k\to\infty} \frac{1}{k} E(Y_N - Y_{N-1})(Y_{N+j} - Y_{N+j-1})' = 0 \quad \textit{for } j \geq 1,$$

flow variables:

$$(iii) \quad \lim_{k\to\infty} \frac{1}{k^2} E(\bar{Y}_N - \bar{Y}_{N-1})(\bar{Y}_N' - \bar{Y}_{N-1})' = \frac{4\pi}{3} f_{\Delta y}(0)$$

$$(iv) \quad \lim_{k\to\infty} \frac{1}{k^2} E(\bar{Y}_N - \bar{Y}_{N-1})(\bar{Y}_{N+1} - \bar{Y}_N)' = \frac{\pi}{3} f_{\Delta y}(0)$$

$$(v) \quad \lim_{k\to\infty} \frac{1}{k^2} E(\bar{Y}_N - \bar{Y}_{N-1})(\bar{Y}_{N+j} - Y_{N+j-1})' = 0 \quad \textit{for } j \geq 2,$$

where $f_{\Delta y}(\omega)$ denotes the spectral density matrix of Δy_t at frequency ω.

PROOF: The difference

$$Y_N - Y_{N-1} = y_{kN} - y_{kN-k} = \sum_{i=1}^{k} \Delta y_{(k-1)N+i}$$

is a partial sums process with asymptotic covariance matrix

$$\begin{aligned} \lim_{k\to\infty} k^{-1} E(Y_N - Y_{N-1})(Y_N - Y_{N-1})' &= \Omega + \Gamma + \Gamma' \\ &= 2\pi f_{\Delta y}(0) \ . \end{aligned}$$

(ii) Define the partial sum $S_1 = \sum_{i=1}^{k} u_i$ and $S_2 = \sum_{i=k+1}^{2k} u_i$, where u_t is stationary with covariance function Γ_j. The covariance between S_1 and S_2 is given by

$$E(S_1 S_2') = \Gamma_1 + 2\Gamma_2 + \cdots + k\Gamma_k + (k-1)\Gamma_{k+1} + \cdots + \Gamma_{2k-1} \ .$$

For $\sum_{j=1}^{\infty} j||\Gamma_j|| < \infty$ we have

$$\begin{aligned} ||E(S_1 S_2')|| &< ||\sum_{j=1}^{\infty} j\Gamma_j|| \\ &\leq \sum_{j=1}^{\infty} j||\Gamma_j|| < \infty \end{aligned}$$

and, thus, by letting $S_1 = Y_N - Y_{N-1}$ and $S_2 = Y_{N+1} - Y_N$ it follows that $E(Y_N - Y_{N-1})(Y_{N+1} - Y_N)$ is $O(1)$. A similar result is obtained for higher order autocovariances.

(iii) Let

$$\begin{aligned} k(\bar{Y}_N - \bar{Y}_{N-1}) &= y_{kN} - y_{kN-k} + y_{kN-1} - y_{kN-k-1} + \cdots + y_{kN-k+1} - y_{kN-2k+1} \\ &= S_k(L)\Delta y_{kN} + S_k(L)\Delta y_{kN-1} + \cdots + S_k(L)\Delta y_{kN-k+1} \\ &= S_k(L)^2 \Delta y_{kN}, \end{aligned}$$

where

$$S_k(L) = 1 + L + L^2 + \cdots + L^{k-1}$$

and

$$\begin{aligned} S_k(L)^2 &= 1 + 2L + 3L^2 + \cdots + kL^{k-1} + (k-1)L^k + \cdots + L^{2k-2} \\ &= w_0 + w_1 L + w_2 L + \cdots + w_{2k-2} L^{2k-2} \end{aligned}$$

is a symmetric filter with triangular weights.

The covariance matrix is given by

$$\begin{aligned} k \cdot E(\bar{Y}_N - \bar{Y}_{N-1})(\bar{Y}_N - \bar{Y}_{N-1})' &= E\left(\sum_{i=0}^{2k-2} w_i \Delta y_{kN-i}\right)\left(\sum_{i=0}^{2k-2} w_i \Delta y_{kN-i}'\right) \\ &= \sum_{p=-2k+2}^{2k-2} \sum_{i=1}^{2k-1-|p|} w_i w_{i+|p|} \Gamma_p \ , \end{aligned}$$

where $\Gamma_p = E(\Delta y_t \Delta y_{t-p}')$.

Consider the odd values $p = \pm 1, \pm 3, \pm 5, \ldots$. We have

$$\sum_{i=1}^{2k-1-|p|} w_i w_{i+|p|} = 2 \sum_{i=1}^{k-(|p|+1)/2} i(i+p)$$

and as $k \to \infty$

$$\begin{aligned}\lim_{k\to\infty} \sum_{i=1}^{k-|(p+1)/2|} 2(i^2 - ip) &= 2(\sum_{i=1}^{\infty} i^2) - 2p(\sum_{i=1}^{\infty} i) \\ &= \frac{2}{3}k^3 + O(k^2).\end{aligned}$$

For even values $p = 0, \pm 2, \pm 4, \ldots$ we have

$$\sum_{i=1}^{2k-1-|p|} w_i w_{i+|p|} = (k - |p|/2)^2 + 2 \sum_{i=1}^{k-|p|/2-1} i(i+p)$$

and, thus,

$$\lim_{k\to\infty} \sum_{i=1}^{k-|(p+1)/2|} 2(i^2 - ip) = \frac{2}{3}k^3 + O(k^2) \ .$$

Using these results yields

$$\begin{aligned}E(\bar{Y}_N - \bar{Y}_{N-1})(\bar{Y}_N - \bar{Y}_{N-1})' &= \frac{2}{3}k^2[\Gamma_0 + \sum_{j=1}^{\infty}(\Gamma_j + \Gamma_j')] + o(k^2) \\ &= \frac{4\pi}{3}k^2 f_{\Delta y}(0) + o(k^2).\end{aligned}$$

(iv) The first order autocovariance matrix is given by

$$k \cdot E(\bar{Y}_N - \bar{Y}_{N-1})(\bar{Y}_{N+1} - \bar{Y}_N)' = \sum_{p=-2k+2}^{2k-2} \sum_{i=1}^{2k-1-|p|} w_{i+k} w_{i+k+|p|} \Gamma_p$$

where $\Gamma_p = E(\Delta y_t \Delta y_{t-p}')$.

For an odd value of p we have

$$\begin{aligned}\lim_{k\to\infty} \sum_{i=1}^{2k-1-|p|} w_{i+k} w_{i+k+|p|} &= \sum_{i=1}^{\infty}(k-i)(i+p) + O(k^2) \\ &= k\left(\sum_{i=1}^{\infty} i\right) - \left(\sum_{i=1}^{\infty} i^2\right) + k^2 p - p\left(\sum_{i=1}^{\infty} i\right) + O(k^2) \\ &= \frac{1}{6}k^3 + O(k^2) \ .\end{aligned}$$

It follows that

$$\begin{aligned}E(\bar{Y}_N - \bar{Y}_{N-1})(\bar{Y}_{N+1} - \bar{Y}_N)' &= \frac{1}{6}k^2\left(\Gamma_0 + \sum_{j=1}^{\infty}\Gamma_j + \Gamma_j'\right) + o(k^2) \\ &= \frac{\pi}{3}k^2 f_{\Delta y}(0) + o(k^2).\end{aligned}$$

(v) To simplify the proof we assume that the degree of the polynomial $C(L) = I + C_1 L + \cdots + C_q L^q$ is finite, that is, Δy_t has a vector MA(q) representation with $q < k$. Since $k \to \infty$ the proof is valid for $q \to \infty$, as well. Of course, the assumption $q < k$ imposes the restriction that q does not grow at a faster rate than k.

The second order autocovariance matrix is given by

$$\begin{aligned} k \cdot E(\bar{Y}_N - \bar{Y}_{N-1})(\bar{Y}_{N+2} - \bar{Y}_{N+1})' &= E\left(\sum_{i=0}^{2k-2} w_i \Delta y_{kN-i}\right)\left(\sum_{i=0}^{2k-2} w_i \Delta y'_{kN+2k-i}\right) \\ &= \sum_{p=1}^{k}\sum_{i=1}^{|p|} w_i w_{2k-i-|p|+1}(\Gamma_p + \Gamma'_p) \ . \end{aligned}$$

There exists a constant $c < \infty$ such that for all p

$$\sum_{i=1}^{|p|} w_i w_{2k-i-|p|+1} = \sum_{i=1}^{p} i(p-i+1) < cp^3 \ .$$

Thus, it follows that

$$\begin{aligned} \sum_{p=1}^{k}\sum_{i=1}^{|p|} w_i w_{2k-i-|p|+1}|\Gamma_p + \Gamma'_p| &< \sum_{p=1}^{k} 2cp^3|\Gamma_p| \\ &< 2ck^2 \sum_{p=1}^{k} p|\Gamma_p|, \end{aligned}$$

where we have used $p < k$. From $\sum_{p=1}^{k} p|\Gamma_p| < \infty$ it finally follows that

$$\lim_{k\to\infty} \frac{1}{k^2} E(\bar{Y}_N - \bar{Y}_{N-1})(\bar{Y}_{N+2} - Y_{N+1})' = 0 \ .$$

Similarly it can be shown that the higher order autocorrelations converge to zero as well. ■

From Theorem 3.2 it follows that as k tends to infinity, the vector of aggregated flow variables has the vector MA(1) representation:

$$k^{-1}(\bar{Y}_N - \bar{Y}_{N-1}) = U_N + (2 - \sqrt{3})U_{N-1}, \tag{3.2}$$

where

$$E(U_N U'_N) = \frac{2\pi}{1 + (2-\sqrt{3})^2} f_{\Delta y}(0) \ .$$

Note that for the special case where $n = 1$ (a single time series), our results correspond to the result of Working (1960) who shows that the first order autocorrelation of the increments from an aggregated random walk is 0.25.

The asymptotic results of Theorems 3.1 and 3.2 imply that for difference stationary stock variables as well as for stationary and difference stationary flow variables the contemporaneous relationship of the limiting process reflects the causal relationship at frequency zero in the sense of Geweke (1986) and Granger and Lin (1995).

It is interesting to consider the case that y_t is cointegrated. To discuss the effects of aggregating a set of cointegrated variables it is useful to define the matrix $Q = [\beta, \beta_\perp]'$, where β is an $n \times r$ matrix of cointegration vectors such that $z_t = \beta' y_t$ is stationary. The matrix $\beta_\perp$ is a $n \times (n-r)$ orthogonal complement of β. The linear combinations $w_t = \beta_\perp' y_t$ are assumed to be $I(1)$. From Theorem 3.2 (iii) it follows that $\bar{Z}_N - \bar{Z}_{N-1}$ is $O_p(1)$ whereas $\bar{W}_N - \bar{W}_{N-1}$ is $O_p(k^{1/2})$. Hence, as the aggregation interval k tends to infinity, the variance of the "nonstationary linear combinations" dominates the variance of the "error correction terms". Consequently,

$$\begin{aligned} \frac{1}{k^{3/2}}(\bar{Y}_N - \bar{Y}_{N-1}) &= \frac{1}{k^{3/2}} Q^{-1} \begin{bmatrix} \bar{Z}_N - \bar{Z}_{N-1} \\ \bar{W}_N - \bar{W}_{N-1} \end{bmatrix} \\ &= \frac{1}{k^{3/2}} Q_2^{-1}(\bar{W}_N - \bar{W}_{N-1}) + O_p(k^{-1}) \end{aligned}$$

where Q_2^{-1} is the lower $n \times (n-r)$ block of Q^{-1}. This implies that the differences of $\bar{Y}_N$ possess a singular distribution as k tends to infinity. It is important to note that the limiting processes of the aggregated variables have a singular spectral density matrix for all frequencies $0 \leq \omega \leq \pi$, while the spectral density matrix of y_t is singular at $\omega = 0$ only. In other words, the limiting behavior of the aggregated time series is dominated by the stochastic trends and, thus, the standardized variance of the error correction terms tends to zero. Since this does not seem to be a relevant feature of observed time series, the aggregation of cointegrated variables is excluded from the subsequent considerations.

3.3 Contemporaneous causality

Using the causality definition given in Section 4.2, I am able to consider the relationship between Granger causality at the natural frequency ($k = 1$) and contemporaneous causality in the limiting process ($k \to \infty$). As mentioned above, this comparison is of interest as it is unlikely that any given vector of time series will be observed with a time scale which allows an unambiguous ordering of *cause* and *effect* in time. Thus, it is helpful to know in which situation contemporaneous causality corresponds to Granger causality and whether it is possible that "spurious causality" results from the aggregation process.

Spurious causality between two variables x_t and y_t occurs when the variables do not possess a causal relationship at the natural frequency but for the aggregated variables we find $\rho(X_N, Y_N|Z_N) \neq 0$. In this case, the causality in the aggregated process is obviously due to the aggregation procedure. For the practical application of a concept of causality it is therefore important to have conditions that rules out such kind of spurious causality. The following theorem gives sufficient conditions for non-causality between two aggregated variables.

THEOREM 3.3 *Let $\xi_t = [x_t, y_t, z_t']'$, where z_t is a $(n-2) \times 1$ vector. Assume that either:*

(i) ξ_t *is a vector of stationary flow variables, or*

(ii) ξ_t *is a vector of difference stationary flow variables, or*

(iii) ξ_t *is a vector of difference stationary stock variables.*

If there is no Granger causality between x_t and y_t and

$$(a) \quad x_t \not\to z_{j,t} \quad or \quad (b) \quad y_t \not\to z_{j,t} \quad for\ all\ j = 1, \ldots, n-2,$$

where $z_{j,t}$ is the j'th element of z_t, then, as $k \to \infty$, we have for the partial correlations of the aggregated variables that:

$$\begin{aligned} for\ case\ (i) \qquad & \rho(\bar{X}_N, \bar{Y}_N|\bar{Z}_N) = 0 \\ for\ case\ (ii) \qquad & \rho(\Delta\bar{X}_N, \Delta\bar{Y}_N|\Delta\bar{Z}_N) = 0 \\ for\ case\ (iii) \qquad & \rho(\Delta X_N, \Delta Y_N|\Delta Z_N) = 0\ . \end{aligned}$$

PROOF: For convenience, we confine ourselves to a trivariate VAR(p) process. The proof can easily be generalized to systems with $n > 3$.

First consider a VAR process obeying the conditions:

$$\begin{aligned} & x_t \not\to y_t \\ & y_t \not\to x_t \\ (a) \quad & x_t \not\to z_t, \end{aligned}$$

that is, there is no causality between $y_{1,t}$ and $y_{2,t}$ and condition (a) is satisfied. From the Theorems 3.1 and 3.2 we know that the limiting processes for the cases (i) – (iii) are white noise with a covariance matrix proportional to the spectral density matrix of the original process. Thus, the limiting process for case (i), for example, has a representation of the form:

$$\left(\sum_{j=0}^{\infty} C_j\right)^{-1} \begin{bmatrix} \bar{X}_N \\ \bar{Y}_N \\ \bar{Z}_N \end{bmatrix} = \begin{bmatrix} U_{1,N} \\ U_{2,N} \\ U_{3,N} \end{bmatrix}$$

where $E(U_N U_N') = \Omega$. A similar representation exists for the cases (ii) and (iii). We therefore confine ourselves to case (i). The proof for case (ii) and (iii) is straightforward.

Since we assume that the MA representation is invertible there exists an autoregressive representation with autoregressive polynomial

$$I_n - A_1 L - A_2 L^2 - \cdots = \left(\sum_{j=0}^{\infty} C_j L^j\right)^{-1}$$

and thus the limiting process can be written as

$$\begin{bmatrix} \bar{X}_N \\ \bar{Y}_N \\ \bar{Z}_N \end{bmatrix} = \bar{A} \begin{bmatrix} \bar{X}_N \\ \bar{Y}_N \\ \bar{Z}_N \end{bmatrix} + \begin{bmatrix} U_{1,N} \\ U_{2,N} \\ U_{3,N} \end{bmatrix},$$

where

$$\bar{A} = \sum_{j=1}^{\infty} A_j = \begin{bmatrix} \bar{a}_{11} & 0 & \bar{a}_{13} \\ 0 & \bar{a}_{22} & \bar{a}_{23} \\ 0 & \bar{a}_{32} & \bar{a}_{33} \end{bmatrix}.$$

The zero restrictions in the matrix $\bar{A}$ result from the assumptions on the causal relationship between the variables. Accordingly, we find

$$(1 - \bar{a}_{11})\bar{X}_N = \bar{a}_{13}\bar{Z}_N + U_{1,N}$$

Since $\rho(\bar{Y}_N, \bar{Z}_N|\bar{Z}_N) = 0$ and $\rho(\bar{Y}_N, U_{3,N}|\bar{Z}_N) = 0$ it immediately follows that $\rho(\bar{X}_N, \bar{Y}_N|\bar{Z}_N) = 0$.

Second, consider the condition (b) $y_t \not\to z_t$ instead of (a) $x_t \not\to z_t$. In this case the limiting process can be represented as

$$\begin{bmatrix} \bar{X}_N \\ \bar{Y}_N \\ \bar{Z}_N \end{bmatrix} = \begin{bmatrix} \bar{a}_{11} & 0 & 0 \\ 0 & \bar{a}_{22} & \bar{a}_{23} \\ \bar{a}_{31} & \bar{a}_{32} & \bar{a}_{33} \end{bmatrix} \begin{bmatrix} \bar{X}_N \\ \bar{Y}_N \\ \bar{Z}_N \end{bmatrix} + \begin{bmatrix} U_{1,N} \\ U_{2,N} \\ U_{3,N} \end{bmatrix}$$

This gives:

$$\begin{aligned} (1 - \bar{a}_{22})\bar{Y}_N &= \bar{a}_{23}Z_N + U_{2,N} \\ (1 - \bar{a}_{22})\bar{X}_N Y_N &= \bar{a}_{23}X_N Z_N + X_N U_{2,N} \end{aligned}$$

Since $\rho(\bar{X}_N, \bar{Z}_N|\bar{Z}_N) = 0$ and $\rho(\bar{X}_N, U_{2,N}|\bar{Z}_N) = 0$ it follows that $\rho(\bar{X}_N, \bar{Y}_N|\bar{Z}_N) = 0$.

To generalize the proof to the case $k > 3$, we let $\bar{Z}_N$ be a $(n-2)$-dimensional vector. The reasoning of the proof applies to this case in a straightforward manner. ■

Theorem 3.3 gives sufficient conditions for ruling out spurious contemporaneous causality. If conditions (a) or (b) are violated it may be the case that there is contemporaneous causality between aggregated variables, although there is no Granger causality at the original time scale. This is the case of *spurious* contemporaneous causality.

Necessary *and* sufficient conditions for ruling out spurious contemporaneous causality in aggregated time series can be derived from the relationship between the original process at $k = 1$ and the limiting process for $k \to \infty$. Unfortunately, such conditions are nonlinear and depend on the precise parameter values of the process describing y_t. Since these parameter values are usually unknown and

cannot be estimated without observing the process at its natural frequency, the practical value of the necessary and sufficient conditions is quite limited.

If $x_t \to y_t$ and $y_t \to x_t$ we say that there is feedback causality between x_t and y_t. An important consequence of Theorem 3.3 can be derived for the case that there is no feedback causality among the variables.

COROLLARY 3.1 *For the cases (i) – (iii) of Theorem 3.3 and under the assumption that there is no feedback causality among the variables it follows that, as $k \to \infty$, there is no spurious causality among the aggregated variables of the system.*

Whenever there is no feedback Granger causality, the variables of the system can be arranged such that one of the conditions (a) and (b) in Theorem 3.3 is satisfied. This rules out the case of spurious contemporary causality.

Another (less trivial) consequence of Theorem 3.3 can be derived for a trivariate system. Following Dufour and Renault (1998, Definition 2.2) and Lütkepohl and Burda (1997) we say that x_t *does not cause* y_t *at horizon* h if

$$E(y_{t+h}|\mathcal{I}_t) = E(y_{t+h}|\mathcal{I}_t^+), \tag{3.3}$$

where $\mathcal{I}_t$ and $\mathcal{I}_t^+$ are the same information sets as in (3.1). Obviously, the usual definition of Granger causality given in (3.1) is a special case with $h = 1$. If x_t does not cause y_t *at any horizon* we write $x_t \not\to_\infty y_t$. For a trivariate system the following result holds.

COROLLARY 3.2 *Let $\xi_t = [x_t, y_t, z_t]'$ be a stationary or the differences of a difference stationary 3×1 vector with an invertible MA representation. If $y_t \not\to_\infty z_t$ and (i) – (iii) of Theorem 3.3 hold, then as $k \to \infty$, there is no spurious causality among the aggregated counterparts.*

PROOF: Consider the AR representation of the system

$$\xi_t = A_1\xi_{t-1} + A_2\xi_{t-2} + \cdots + \varepsilon_t \ .$$

Causality at horizon h can be deduced from the conditional expectation

$$E(\xi_t|\xi_{t-h}, \xi_{t-h-1}, \ldots) = \pi_1^{(h)}\xi_{t-h} + \pi_2^{(h)}\xi_{t-h-1} + \pi_3^{(h)}\xi_{t-h-2} + \cdots$$

where the matrices $\pi_j^{(h)}$ are given in Dufour and Renault (1998). Non-causality between x_t and y_t at horizon h implies

$$e_2'\pi_j^{(h)}e_1 = 0 \quad \text{for } j = 1, 2, \ldots,$$

where e_i is the i'th column of the 3×3 identity matrix. The matrix $\pi_1^{(h)}$ is identical to the matrix B_h in the moving average representation

$$\xi_t = \varepsilon_t + B_1\varepsilon_{t-1} + B_2\varepsilon_{t-2} + \cdots$$

Thus, noncausality at all horizons implies

$$e_i'B_he_j = 0 \quad \text{for } h = 1, 2, \ldots \tag{3.4}$$

Assuming stationary flow variables it follows from Theorem 3.1 that the limiting process can be represented as $(I_3 - \bar{A})\bar{Y}_N = U_N$. Non-causality of the form $y_t \not\to z_t$ at $h = 1$ implies that the $(3, 1)$ elements of the matrices A_k, $k = 1, 2, \ldots$ are zero. Accordingly, the limiting distribution can be represented as

$$(1 - \bar{a}_{11})\bar{X}_N = \bar{a}_{12}\bar{Y}_N + \bar{a}_{13}\bar{Z}_N + U_{1,N} \tag{3.5}$$

$$(1 - \bar{a}_{22})\bar{Y}_N = \bar{a}_{21}\bar{X}_N + \bar{a}_{23}\bar{Z}_N + U_{2,N} \tag{3.6}$$

$$(1 - \bar{a}_{33})\bar{Z}_N = \bar{a}_{31}\bar{X}_N + U_{3,N} \tag{3.7}$$

where $\bar{a}_{ij}$ denotes the (i, j) element of the matrix $\bar{A} = \sum A_k$. From the MA representation we get $(I - \bar{A})^{-1} = (I - \bar{B})$, where $\bar{B} = \sum B_k$. From (3.4) it follows that for $y_t \not\to_\infty z_t$ we have $\bar{b}_{32} = 0$. Thus, the representation $\bar{\xi}_N = (I_3 - \bar{A})^{-1}U_N = (I_3 - \bar{B})U_N$ gives rise to the equation

$$\bar{Z}_N = -\bar{b}_{31}U_{1,N} + (1 - \bar{b}_{33})U_{3,N} \ .$$

Solving this equation for $U_{3,N}$ and inserting in (3.7) gives

$$\bar{X}_N = c_1\bar{Z}_N + c_2U_{1,N} \ , \tag{3.8}$$

where c_1 and c_2 are functions of $\bar{a}_{31}, \bar{a}_{33}, \bar{b}_{31}$ and $\bar{b}_{33}$. Comparing (3.8) with (3.5) shows that $\bar{a}_{12}$ must be zero and, thus, implies the same restriction for the limiting process as the assumption that y_t does not cause x_t. From Theorem 3.3 it follows

that in this case there is no spurious contemporaneous causality between $\bar{Y}_N$ and $\bar{Z}_N$.

The proofs for the cases (i) and (iii) are essentially the same. ■

This result is intuitively plausible because the assumption of no causality at any horizon rules out indirect causal effects via the remaining variable z_t. Accordingly, for $x_t \rightarrow y_t$ we must rule out that x_t causes z_t at longer lag horizons h, since otherwise y_t may be used to predict z_{t+h} via x_{t+j}, where $0 < j < h$. Unfortunately, I was not able to generalize this result to higher dimensional systems with $n > 3$.

In order to illustrate the results in this section, it is useful to construct some examples based on a trivariate VAR(1) model for y_t, where all innovation terms represent mutually uncorrelated white noise processes.

Example A: Assume that $\xi_t = [x_t, y_t, z_t]$ is stationary and has a causal structure given by $x_t \rightarrow y_t$ and $y_t \rightarrow z_t$. In particular, assume a VAR(1) representation of the form

$$\begin{aligned} x_t &= \varepsilon_{1,t} \\ y_t &= a x_{t-1} + \varepsilon_{2,t} \\ z_t &= b y_{t-1} + \varepsilon_{3,t} \ . \end{aligned}$$

Since $x_t \not\rightarrow z_t$ and $z_t \not\rightarrow y_t$, it follows from Theorem 3.3 that $\rho(\bar{X}_N, \bar{Z}_N | \bar{Y}_N) = 0$, and that there is no contemporaneous causality between $\bar{X}_N$ and $\bar{Z}_N$. Furthermore, there is no feedback causality among the variable, so that the result immediately follows from the Corollary 3.1.

Example B: Assume that a vector of flow variables is generated by a stationary process given by:

$$\begin{aligned} x_t &= a y_{t-1} + b z_{t-1} + \varepsilon_{1,t} \\ y_t &= \varepsilon_{2,t} \\ z_t &= \varepsilon_{3,t} \ . \end{aligned}$$

Applying Granger's concept of causality, there is no causality between y_t and z_t. Further, a simple calculation shows that for the limiting process, $\rho(\bar{Y}_N, \bar{Z}_N | \bar{X}_N) =$

$-ab/(a^2+a^2+1)$. Thus, a necessary and sufficient condition for the aggregated variables $\bar{Y}_N$ and $\bar{Z}_N$ to have no contemporaneous causal relationship is that either a or b, or both parameters are equal to zero. This result is also an immediate consequence from Theorem 3.3, which states that there is no contemporaneous causality if either y_t or z_t is not Granger causal for x_t.

Example C: To illustrate the problems with aggregated stock variables which are discussed above, consider the stationary process given by:

$$\begin{aligned} x_t &= \varepsilon_{1,t} \\ y_t &= a x_{t-1} + \varepsilon_{2,t} \\ z_t &= b y_{t-1} + \varepsilon_{3,t}. \end{aligned}$$

In this system, $x_t \to y_t$ and $y_t \to z_t$. For $k \geq 3$ the aggregated process becomes white noise with:

$$\begin{aligned} X_N &= U_{1,N} \\ Y_N &= U_{2,N} \\ Z_N &= abX_N + U_{3,N}\ . \end{aligned}$$

For $ab \neq 0$ there exists *spurious* contemporaneous causality between X_N and Z_N, as there is no Granger causality between x_t and z_t. Stated another way, the indirect causal relationship between x_t and z_t via y_t becomes a direct causal link under aggregation. See also Lütkepohl and Burda (1997).

3.4 Monte Carlo experiments

In this section, the asymptotic implications of Theorem 3.3 and Corollary 3.1 are examined via a simple Monte Carlo experiment. In particular, the following VAR(1) model is considered:

$$\begin{bmatrix} \Delta^d x_t \\ \Delta^d y_t \\ \Delta^d z_t \end{bmatrix} = \begin{bmatrix} a & 0 & 0 \\ b & a & 0 \\ 0 & b & a \end{bmatrix} \begin{bmatrix} \Delta^d x_{t-1} \\ \Delta^d y_{t-1} \\ \Delta^d z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}, \tag{3.9}$$

where $d \in \{0, 1\}$ and $\varepsilon_{i,t}$ is an i.i.d. vector of standard normal random variables. For $b \neq 0$, the Granger causal structure of this system is: $x_t \rightarrow y_t$ and $y_t \rightarrow z_t$. From Theorem 3.3 and Corollary 3.1 it follows that as $k \rightarrow \infty$, the limiting process has a partial correlation structure such that $E(u_{1,N}u_{3,N}|u_{2,N}) = 0$ and all other partial correlations are nonzero, where the $u_{j,N}$ $(j = 1, 2, 3)$ denote the innovations from an estimated VAR[4] model using data generated according to (3.9) and aggregated appropriately. Swanson and Granger *henceforth: SG* (1997) propose tests for assessing whether the above partial correlation restriction holds which are based on Fisher's z-statistics or alternatively on t-statistics from regressions involving the residuals. Here we use the Fisher's z-statistic.

Empirical sizes corresponding to 5% nominal size tests and for various parameterizations of the VAR are reported in Tables 4.1. Note also that results are reported for stationary stock variables, which are not treated in Theorem 3.3. In all experiments, b is set equal to 0.5, as results were found not to be sensitive to the choice of b. The parameter a is set equal to $\{0.0, 0.2, 0.4, 0.6, 0.8\}$. Not surprisingly, the magnitude of the parameter a is crucial when k is small, as a determines the roots of the autoregressive polynomial in our model. Thus, our asymptotic results may be a poor guide to finite sample behavior for small k and $|a|$ close to unity.[1] All entries in the Table 4.1 are based on 10,000 Monte Carlo replications, and all estimations use 100 observations of appropriately aggregated data.

Tables 4.1a-c contain results for cases (i) – (iii) in 3.3. Interestingly, the empirical sizes approach the nominal size quite quickly when k increases, for small and moderately sized values of a. For $a = 0.8$, however, the SG test is upwards biased, even for relatively large values of k.

Table 4.1d reports results for stationary stock variables, for which Theorem 3.3 does not apply. For $a = 0$, however, note that $A^n = 0$ for $n \geq 3$, and thus the covariance matrix of the limiting process is $\Omega + A\Omega A' + A^2\Omega(A^2)'$ (see Theorem

[1]Recall also that the aggregated processes which we construct are VARMA processes, in general. Thus, lower order VAR approximations may not yield good estimates of the errors of the process.

Table 4.1: Empirical size of the SG test procedure

a) Stationary flow variables (case (i))					
k	a=0	a=0.2	a=0.4	a=0.6	a=0.8
2	0.10	0.10	0.11	0.11	0.11
5	0.07	0.08	0.16	0.33	0.61
10	0.07	0.07	0.10	0.35	0.95
20	0.07	0.07	0.08	0.19	0.98
50	0.07	0.07	0.07	0.09	0.80
100	0.07	0.07	0.07	0.07	0.47
b) Difference stationary flow variables (case (ii))					
k	a=0	a=0.2	a=0.4	a=0.6	a=0.8
2	0.11	0.11	0.11	0.11	0.11
5	0.08	0.10	0.17	0.34	0.61
10	0.07	0.07	0.10	0.31	0.93
20	0.07	0.07	0.07	0.14	0.95
50	0.07	0.07	0.07	0.07	0.53
100	0.07	0.07	0.07	0.07	0.17
c) Stationary stock variables (case (iii))					
k	a=0	a=0.2	a=0.4	a=0.6	a=0.8
2	0.10	0.11	0.11	0.11	0.11
5	0.07	0.09	0.16	0.33	0.61
10	0.07	0.07	0.10	0.35	0.95
20	0.07	0.07	0.08	0.19	0.98
50	0.07	0.07	0.07	0.09	0.80
100	0.07	0.07	0.07	0.07	0.47
d) Difference stationary flow variables					
k	a=0	a=0.2	a=0.4	a=0.6	a=0.8
2	0.07	0.07	0.07	0.07	0.11
5	0.07	0.07	0.07	0.09	0.26
10	0.07	0.07	0.07	0.10	0.52
20	0.07	0.07	0.07	0.10	0.66
50	0.07	0.07	0.07	0.10	0.66
100	0.07	0.07	0.07	0.10	0.66

Notes: Entries correspond to the frequency of times that the correct contemporaneous causal structure is uncovered, based on empirical procedure given in SG (1997). Results are based on estimations using 100 observations of data generated according to (3.9), and aggregated according to the aggregation interval, k. All entries are based on 10,000 Monte Carlo replications.

3.1 (i)), which is a diagonal matrix in this special case. Therefore, all partial correlations are zero, and *spurious* contemporaneous correlation should not arise for $a = 0$. This is the reason why the empirical size of the SG test is still close to the nominal size when stationary stock variables, as long as a is close to zero. In contrast, for a substantially different from zero, partial correlations need not die out as k increases. Indeed, from Table 4.1 it is seen that the empirical sizes are far from the nominal size for large values of a(i.e. when $a = 0.8$ and $k = 100$ empirical size is 0.66).

3.5 Aggregation of SVAR models

The asymptotic results of Section 4.3 can be used to analyse the (asymptotic) effects of the aggregation procedure on the estimated structural model. Assume that at some ("natural") sampling frequency the process can be represented as a stationary VAR[p] model with

$$B\varepsilon_t = Ru_t \ .$$

The vector u_t represents the structural shocks, where it is assumed that $\Omega = E(u_t u_t')$ is a diagonal matrix. We therefore have $E(\varepsilon_t \varepsilon_t') = \Sigma = B^{-1}R\Omega R'B'^{-1}$. From Theorem 3.1 it follows that for $k \to \infty$ the covariance matrix of the aggregated process is given by

$$E(y_N y_N') = (I - A_1 - \cdots - A_p)^{-1}\Sigma(I - A_1' - \cdots - A_p')^{-1}$$

and, thus, the corresponding structural model is

$$(I - A_1 - \cdots - A_p)^{-1}B\varepsilon_N = Ru_N \ .$$

In general, the aggregation of the data therefore may have a substantial effect on the structural model and may even change the sign of the structural parameter.

To assess the effect of temporal aggregation in practice, a simple example is considered. Let U_t and π_t denote the unemployment rate and the yearly inflation rate of the US, which is observed on a monthly basis from 1960 (i) to 1998 (xii).

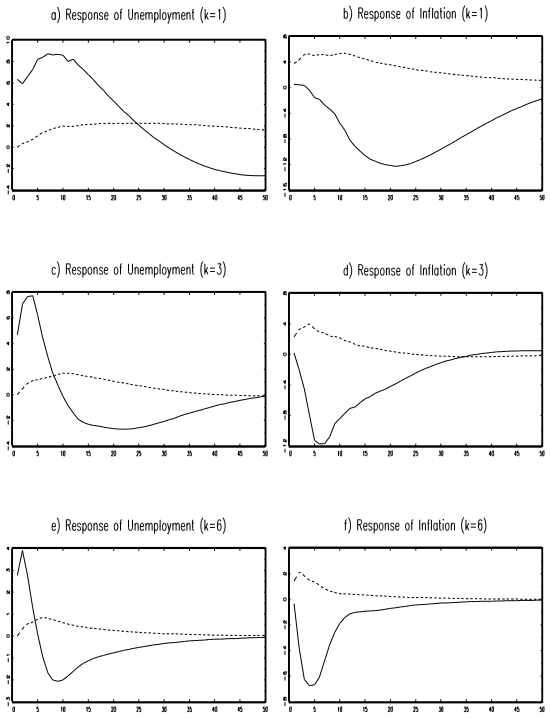


Figure 3.1: Impulse responses at different aggregation intervals.

Neoclassical specification

solid line: supply shock	dashed line: demand shock

Two different specifications are considered. One is the neoclassical version of the Phillips curve (cf. Dolado et al. 1999). This model employs a triangular identification given by

$$\begin{aligned} \varepsilon_t^U &= u_t^s \\ \varepsilon_t^\pi &= r_{12} u_t^s + u_t^d \ , \end{aligned}$$

where $\varepsilon_t^U, \varepsilon_t^\pi$ are the residuals from an estimated VAR[15] model with a constant mean, which is suggested by the AIC criterion. The supply shock is denoted by u_t^s and u_t^d denotes the demand shock. The notion behind this specification is that in a (neo)classical framework, monetary shocks should not affect real variables.

The estimated impulse response functions for this model are depicted in Figure 4.1. Both shocks have a positive short run effect on the unemployment rate so that these shocks represent an unfavourable change in supply or demand. The supply shock has a negative effect on the inflation rate at 1–3 years.[2] Since we are interested in the effect of temporal aggregation we do not give a detailed discussion of the economic implications (see Dolado et al. (1999) for a detailed discussion of alternative specifications).

If the data is aggregated by computing the averages[3] of k time periods, we find that for a quarterly sample frequency ($k = 3$) the estimated model yields qualitatively similar results. There is again a peak at 1 year in the impulse response of the unemployment rate with respect to a supply shock. Similarly the minimum of the impulse response of inflation following a demand shock is a little less than 2 years for $k = 1$ and $k = 3$. Therefore, the general shape of the impulse response functions does not change a lot when aggregating monthly data to a quarterly interval. Similar conclusions can be drawn be aggregating the

[2] It is not the intention to draw economic conclusions from this example. Therefore, no confidence intervals are presented. Furthermore, temporal aggregation reduces the number of observations so that it is difficult to compare the confidence intervals for different k.

[3] The reader may notice that the unemployment rate is not a flow variable but a stock variable. However, since in the theoretical work I did not consider models that mix stock and flow data, I assume that both variables are flow data. In fact, official agencies also report the average unemployment rate within a month (quarter).

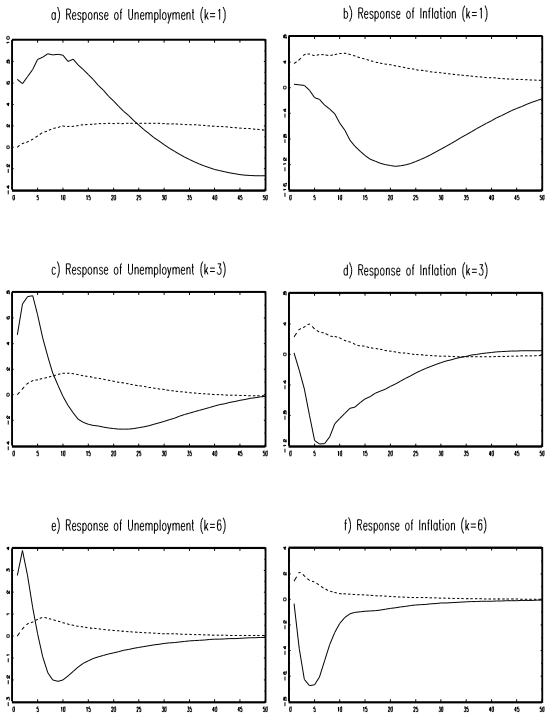


Figure 3.2: Impulse responses at different aggregation intervals.

Monetarist specification

solid line: supply shock	dashed line: demand shock

data further to a bi-anual sample interval. Again the maximum and minimum of the impulse response functions occur at roughly the same time but the size of the impulse responses seems to be affected to a greater extent. For example, the minimum in the impulse responses of the inflation rate with respect to supply shocks is -12 for $k = 3$ but -7 for $k = 6$. The maximum response of the unemployment rate following a supply shock is roughly 8 for $k = 3$ and only 4 for $k = 6$. Therefore, it seems that the aggregation of the data "smoothes" the unemployment rate.

In order to check whether the conclusion is a result of the particular neoclassical specification, we repeat the experiment by using a monetarist identification scheme. In this specification it is assumed that the monetary (demand) shock may affect the unemployment rate at a short horizon but this effect dies out with an increasing lag horizon. That is, we impose the long run restriction as in Blanchard and Quah (1989). The estimated impulse responses are depicted in Figure 4.2. It is seen that this identification scheme yields quite different impulse responses. Again we do not discuss the economic implications of this estimate but focus on the effect of temporal aggregation.

It turns out that the general shape of the impulse response functions is compatible for the different aggregation intervals. In particular the relative maxima and minima of the impulse response functions occur at roughly the same time. However, the magnitude of the effects becomes less important for higher aggregation intervals. For example, the long run effect of supply shocks is 0.30 for $k = 1$ but 0.15 for $k = 3$ and 0.10 for $k = 6$. Similarly, the maximum of the impulse response of inflation with respect to a supply shock is 0.36 at a monthly basis but 0.75 for quarterly data and 1.0 for the bi-anual aggregation interval.

Our experiment suggests that temporal aggregation may have an important effect on the magnitude of the impulse response but seems to have little effect on the general shape of the impulse response functions. Of course, it remains to show that this conclusions applies to other applications as well.

Chapter 4

Inference in nonlinear models

Since the introduction of the concept of cointegration by Granger (1981) the analysis of cointegrated models was intensively studied in a linear context, whereas the work on the extension to nonlinear cointegration is still comparatively limited. Useful reviews of recent work in the analysis of nonlinear cointegration are provided by Granger and Teräsvirta (1993), Granger (1995), and Granger, Inoue and Morin (1997).

In many cases, economic theory suggests a nonlinear relationship as for the production function or the Phillips curve, for example. However, theory does not always provide a precise specification of the functional form so that it is desirable to have nonparametric tools for estimation and inference.

Unfortunately, the usual nonparametric techniques based on kernel functions are not applicable in a nonstationary framework. The reason is that the nonparametric estimate of $E(y|x_0)$ for some given value x_0 is based on neighbourhood values of x_0, where the kernel function employs a weight function decreasing in $|x - x_0|$. If x_t is a random walk, then the process drifts away from the initial value and, thus, the *relative* number of observations in a neighbourhood $[x_0 - \epsilon, x_0 + \epsilon]$ tends to zero. This can be illustrated by considering the number of axis crossings, that is, the number of times x_t crosses the zero level. The number of axis crossings can be seen as a measure of the frequency the process returns to a zero initial value. For a stationary Gaussian AR(1) process, Kedem (1980) shows that

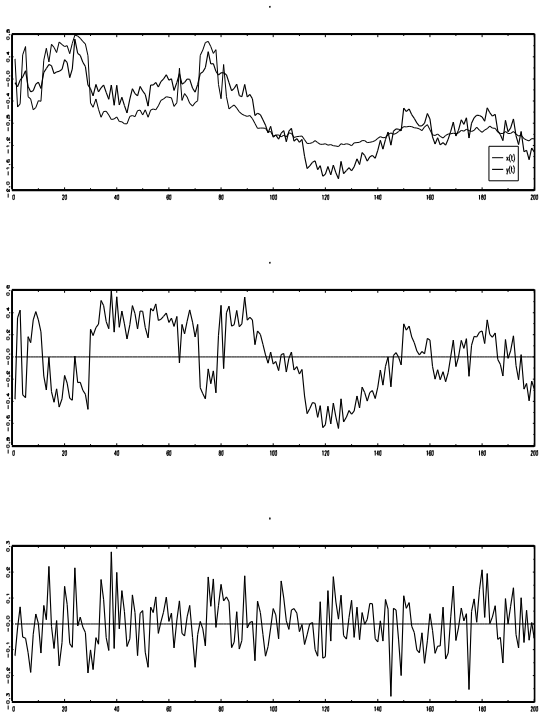
the probability of axis crossings is

$$\text{Prob}(x_t < 0 | x_{t-1} > 0) = \frac{1}{2} - \frac{1}{\pi} \sin^{-1} \rho \ ,$$

where ρ denotes the first order autocorrelation of x_t. Note that if ρ approaches one, then this probability tends to zero. Indeed, for a random walk process Burridge and Guerre (1996) show that the probability of an axis crossing tends to zero with $T^{-1/2}$. Similarly, it can be shown that for a random walk the relative number of observations in the interval $[x_0 - \epsilon, x_0 + \epsilon]$ tends to zero with $T^{-1/2}$ and asymptotically does not depend on the value x_0 (cf. Burridge and Guerre 1996). Therefore, the asymptotic theory for kernel estimates is not applicable in the case of nonstationary variables.

Alternatively, Granger and Hallman (1991a, 1991b) and Breitung and Gouriéroux (1997) suggest using the rank transformation. It is well known that the ranks of the observations are invariant under monotonic transformations of the data. Thus, statistics based on the ranks of the observations do not depend on (monotonous) nonlinear functions of the data. To motivate the test procedure I first show in Section 5.1 that linear cointegration tests may have a poor power against nonlinear cointegration relationships. In Section 5.2 this idea is adopted to test for nonlinear cointegration. Furthermore, the rank transformation is used to decide whether the cointegration relationship is linear or nonlinear (Section 5.3).

An alternative strand of literature was initiated by Bierens (1997a,b) and Vogelsang (1998a,b). These papers allow for unknown nonlinearity in the short run dynamics of the process. Accordingly, such tests are well suited to test for cointegrated models with a nonlinear adjustment process. In Section 5.4 a new test procedure is suggested. The small sample properties of the test are considered in Section 5.5. Furthermore, the test procedures are applied to investigate the term structure of interest rates and the relationship between German dual-class shares.


x(t)
y(t)

4.1 Inconsistency of linear cointegration tests

To illustrate the problems of applying linear cointegration tests to a nonlinear relationship, it is helpful to consider a simple example. Let x_t be a nonlinear random walk given by $x_t^3 = \sum_{j=1}^{t} v_j$, where $\{v_t\}_1^T$ is a white noise sequence with $v_t \sim N(0, \sigma_v^2)$. Furthermore, y_t is given by

$$y_t = x_t^3 + u_t \ , \tag{4.1}$$

where u_t is white noise with $u_t \sim N(0, \sigma_u^2)$. Figure 5.1 presents a realization of the sequences x_t and y_t, where $\sigma_v^2 = \sigma_u^2 = 0.01$. The sample size is $T = 200$. Apparently, there is a fairly strong comovement between both series suggesting a stable long run relationship. However, applying an augmented Dickey-Fuller test (with four lagged differences and a constant), the value of the t-statistic is -2.77, which is insignificant with respect to the 0.05 significance level. In fact, the residuals of the linear cointegration regression (see Figure 5.2) do not look stationary. On the other hand, if the cointegration regression is based on the underlying nonlinear relationship, that is, y_t is regressed on x_t^3, the resulting residual series resembles white noise (see Figure 5.3). Applying the rank tests suggested in Section 5.3 yields significant test statistics with respect to all reasonable significance levels. This example illustrates that ignoring the nonlinear nature of the cointegration relationship may lead to the misleading conclusion that there exists no long run relationship between the series.

On the other hand, one may argue that there is no problem with a test that fails to reject in the presence a nonlinear alternative because we are interested in detecting a *linear* cointegrating relationship. In many applications, however, it is not clear whether the variables must be transformed (e.g. to logarithms) to achieve a linear cointegrating relationship (e.g. Franses and McAleer 1998) and thus the robustness of the test against such monotonic transformation is a desirable property of a cointegration test.

For a theoretical analysis of a nonlinear cointegration relationship, different concepts are used. Granger and Hallman (1991a) and Granger (1995) consider time series which are long memory in mean but have a nonlinear relationship

which is short memory in mean. Corradi (1995) considers "non-strong mixing" processes (processes with a long memory) and strong mixing processes (short memory). Here we I adopt the definition of an integrated process due to Phillips (1987a). The degree of integration is defined as follows:

DEFINITION 4.1 *(i) A time series z_t is $I(0)$ if, as $T \to \infty$,*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[aT]} z_t \;\Rightarrow\; \bar{\sigma}_z W(a) \;,$$

where $\bar{\sigma}_z^2 = \lim_{T\to\infty} E(T\bar{z}^2)$, $\bar{z} = T^{-1}\sum_{t=1}^{T} z_t$, and $W(a)$ represents a standard Brownian motion. (ii) If $(1-B)^d z_t \sim I(0)$, then z_t is integrated of order d, denoted as $z_t \sim I(d)$.

Different sets of sufficient conditions for $z_t \sim I(0)$ are given in Phillips (1987a), Gallant and White (1988), and Phillips and Solo (1992), for example. If we assume that y_t is stationary with an invertible MA representation then the usual definition used in Section 1.3 results. However, since we assume that y_t is a nonlinear process, it may not possess a MA representation and, therefore, the assumption of a stationary process is overly restrictive within a nonlinear framework (e.g. Corradi 1995).

In this section I consider the nonlinear relationship between two real valued time series $\{x_t\}_1^T$ and $\{y_t\}_1^T$

$$y_t = f(x_t) + u_t \;, \tag{4.2}$$

where $y_t \sim I(1)$ and $f(x_t) \sim I(1)$. Under the null hypothesis, u_t is assumed to be $I(1)$, whereas under the alternative of nonlinear cointegration, $u_t \sim I(0)$.

As demonstrated by Granger and Hallman (1991b), the Dickey-Fuller test may perform poorly when applied to a nonlinear transformation of a random walk. To investigate the effects of a nonlinear cointegration relationship on the power of a residual based cointegration test, it is convenient to consider a variant of the Dickey-Fuller test due to Sargan and Bhargava (1983) and Phillips and Ouliaris[1]

[1]The Sargan-Bhargava statistic equals the inverse of the variance ratio statistic suggested by Phillips and Ouliaris.

(1990). The statistic is given by

$$S_T^2 = \frac{1}{\omega_{11\cdot 2} T^2} \sum_{t=1}^{T} (y_t - \hat{\beta} x_t)^2, \tag{4.3}$$

where the $\hat{\beta}$ is the least-squares estimator from a regression of y_t on x_t. The parameter $\omega_{11\cdot 2}$ is defined in Phillips and Ouliaris (1990, below eq. 12) and can be neglected in what follows.

For the function $z = f(x)$ the following assumption is made:

Assumption 4.1 *(i) The function $f(x)$ is monotonically increasing. (ii) There exists a function $h(a)$ such that $f^{-1}(az) = h(a) f^{-1}(z)$, where $f^{-1}(z)$ indicates the inverse function.*

In the following theorem, it is stated that a test against linear cointegration may be inconsistent for some class of nonlinear functions. Further results can be obtained by using the framework of Park and Phillips (1999). However, since the latter approach requires some specialized concepts that are of no interest here, I will confine myself to a simple class of nonlinear functions, which includes the function $y_t = x_t^a$ as a special case.

THEOREM 4.1 *Let $z_t \sim I(1)$ and y_t is generated as in (4.2), where $u_t \sim I(0)$. Under Assumption 4.1, a test based on the statistic S_T^2 given in (4.3) is consistent if and only if $f(x)$ is a linear function.*

PROOF: Using

$$\frac{1}{T^2} \sum_{t=1}^{T} \widehat{u}_t^2 \;=\; \frac{1}{T^2} \sum_{t=1}^{T} y_t^2 - \frac{\left(T^{-2} \sum_{t=1}^{T} y_t x_t \right)^2}{T^{-2} \sum_{t=1}^{T} x_t^2}$$

the test is seen to be consistent if the difference on the right hand side converges to zero as $T \to \infty$. From the continuous mapping theorem it follows that

$$T^{-2} \sum_{t=1}^{T} y_t^2 = T^{-2} \sum_{t=1}^{T} (z_t + u_t)^2 \;\Rightarrow\; \bar{\sigma}_z^2 \int_0^1 W(a)^2 da. \tag{4.4}$$

Furthermore,

$$\begin{aligned} T^{-2}\sum_{t=1}^{T} x_t^2 &= T^{-2}\sum_{t=1}^{T} f^{-1}(z_t)^2 \\ &= h(\bar{\sigma}_z T^{-1/2})^{-2} T^{-2} \sum_{t=1}^{T} f^{-1}(\bar{\sigma}_z T^{-1/2} z_t)^2 \\ &\Rightarrow h(\bar{\sigma}_z T^{-1/2})^{-2} \int_0^1 f^{-1}[W(a)]^2 da \end{aligned}$$

and

$$\begin{aligned} T^{-2}\sum_{t=1}^{T} y_t x_t &= T^{-2}\sum_{t=1}^{T} z_t f^{-1}(z_t) + o_p(1) \\ &= h(\bar{\sigma}_z T^{-1/2})^{-1} T^{-2} \sum_{t=1}^{T} z_t f^{-1}(\sigma_z T^{-1/2} z_t) \\ &\Rightarrow \bar{\sigma}_z h(\bar{\sigma}_z T^{-1/2})^{-1} \int_0^1 W(a) f^{-1}[W(a)] da \ . \end{aligned}$$

It follows that

$$\frac{1}{T^2}\sum_{t=1}^{T} \hat{u}_t^2 \ \Rightarrow \ \bar{\sigma}_z^2 \int_0^1 W(a)^2 da - \bar{\sigma}_z^2 \frac{\left\{\int_0^1 W(a) f^{-1}[W(a)] da\right\}^2}{\int_0^1 f^{-1}[W(a)]^2 da} \ . \tag{4.5}$$

Since $x = f^{-1}(z)$ is an affine mapping it is seen that the right hand side of (4.5) is zero iff $f(x) = bx$ with some constant b. ■

This Theorem shows that residual based cointegration tests are inconsistent for the class of functions given by Assumption 4.1. With some straightforward modifications the result extends to other unit root tests like the original Dickey-Fuller test.

An example may help to illustrate the result. Let y_t be generated as in the previous example given in eq. (4.1). Using (4.5) and $f^{-1}(z) = x^{1/3}$ we have for $\beta \neq 0$:

$$S_T^2 \ \Rightarrow \ \int_0^1 W(a)^2 da - \frac{[\int W(a)^{4/3} da]^2}{\int W(a)^{2/3} da} \ .$$

Thus, under the alternative of nonlinear cointegration, the test statistic is $O_p(1)$. Accordingly, a test based on S_T^2 is inconsistent against a nonlinear cointegration relationship as given in (4.1).

4.2 Rank tests for unit roots

Before considering tests for cointegration based on ranks, it is useful to review the theory on rank tests for a unit root. Assume that the time series x_t is generated by the autoregressive process:

$$x_t = \alpha x_{t-1} + u_t \ ,$$

where u_t is $I(0)$ as defined in Definition 4.1.

The test statistic is based on the ranks of y_t defined as

$$R_T(y_t) = \text{Rank[of } y_t \text{ among } y_1, \ldots, y_T] \ .$$

An asymptotic theory for the ranks is developed by Breitung and Gouriéroux (1997) and is given in a slightly more general form in the following theorem:

THEOREM 4.2 *Let $f(y_t) \sim I(1)$, where $f(\cdot)$ is a strictly monotonous function. Then, as $T \to \infty$, the limiting distribution of the sequence of ranks can be represented as*

$$T^{-1} R_T(x_{[aT]}) \ \Rightarrow \ a\mathcal{A}_1 + (1-a)\mathcal{A}_2 \ ,$$

where $\mathcal{A}_1$ and $\mathcal{A}_2$ are two independent random variables with an arcsine distribution.

PROOF: Let $\mathbb{1}(a)$ be an indicator function that is one if the argument is true and zero otherwise. Then,

$$\begin{aligned}
T^{-1} R_T(x_{[aT]}) &= T^{-1} \sum_{t=1}^{T} \mathbb{1}(x_t < x_{[aT]}) \\
&= \sum_t \mathbb{1}\left(\frac{1}{\sqrt{T}} z_{[\frac{t}{T}T]} < \frac{1}{\sqrt{T}} z_{[aT]} \right) \left[\frac{t}{T} - \frac{t-1}{T} \right] \\
&\Rightarrow \int_0^1 \mathbb{1}[W(u) < W(a)] \mathrm{d}u \\
&= \int_0^a \mathbb{1}[W(u) < W(a)] \mathrm{d}u + \int_a^1 \mathbb{1}[W(u) < W(a)] \mathrm{d}u \ .
\end{aligned}$$

Since the increments of the Brownian motion are independent, the two parts of the integral are independent as well.

Using $\stackrel{d}{=}$ to indicate equality in distribution we have

$$\begin{aligned}\int_0^a \mathbb{1}[W(u) < W(a)]\mathrm{d}u &= \int_0^a \mathbb{1}[W(a) - W(u) > 0]\mathrm{d}u \\ &\stackrel{d}{=} \int_0^a \mathbb{1}[W(a-u) > 0]\mathrm{d}u \\ &\stackrel{d}{=} a\int_0^1 \mathbb{1}[W(u) > 0]\mathrm{d}u \\ &\stackrel{d}{=} a\mathcal{A}_1 \ ,\end{aligned}$$

where $\mathcal{A}_1 = \int_0^1 \mathbb{1}[W(u) > 0]\mathrm{d}u$ is a random variable with an arcsine distribution (cf. Breitung and Gouriéroux, 1997). Similarly, we find

$$\int_a^1 \mathbb{1}[W(u) < W(a)]\mathrm{d}u = (1-a)\mathcal{A}_2 \ ,$$

where $\mathcal{A}_2$ is another random variable with an arcsine distribution independent of $\mathcal{A}_1$. ■

The arcsine distribution has a quite different shape than the normal distribution and some properties are discussed in Breitung and Gouriéroux (1997). Theorem 4.2 generalizes Theorem 1 of Breitung and Gouriéroux (1997) to serially correlated time series. It is important to notice that short run dynamics in Δy_t do not change the distribution of the ranks. Thus, the ranks do not depend on the long run variance of the process.

These results can be used to construct a rank test for cointegration. Consider the null hypothesis that $f(x_t)$ and $g(y_t)$ are independent random walk sequences. This null hypothesis is tested against the alternative of a cointegration relationship of the form:

$$u_t = g(y_t) - f(x_t) \ , \tag{4.6}$$

where $f(x_t) \sim I(1)$, $g(y_t) \sim I(1)$ and $u_t \sim I(0)$. The functions $g(y)$ and $f(x)$ are monotonically increasing. If it is not known whether these functions are monotonically increasing or decreasing, a two-sided test is available. A similar framework is considered in Granger and Hallman (1991a).

The rank statistic is constructed by replacing $f(x_t)$ and $g(y_t)$ by $R_T(x_t)$ and $R_T(y_t)$, respectively. Since it is assumed that $f(x_t)$ and $g(y_t)$ are two random

walk series, it follows that $R_T(x_t) = R_T[f(x_t)]$ and $R_T(y_t) = R_T[g(y_t)]$ behave like ranked random walks for which the limiting distribution is given in Theorem 4.2. The advantage of a statistic based on the sequence of ranks is that the functions $f(\cdot)$ and $g(\cdot)$ need not be known.

We consider two "distance measures" between the sequences $R_T(x_t)$ and $R_T(y_t)$:

$$\kappa_T = T^{-1} \sup_t |d_t| \tag{4.7}$$

$$\xi_T = T^{-3} \sum_{t=1}^{T} d_t^2 \ , \tag{4.8}$$

where $d_t = R_T(y_t) - R_T(x_t)$. It should be noted that d_t is $O_p(T)$ and, thus, the normalization factors are different from other applications of these measures. The statistic κ_T is a Kolmogorov-Smirnov type of statistic considered by Lo (1991) and ξ_T is a Cramer-von-Mises type of statistic used by Sargan and Bhargava (1983). The null hypothesis of no (nonlinear) cointegration between x_t and y_t is rejected if the test statistics are too small.

It is interesting to note that the statistic ξ_T allows for different interpretations. Let $\tilde{b}_T$ denote the least-squares estimate from a regression of $R_T(y_t)$ on $R_T(x_t)$. Using $\sum R_T(x_t)^2 = \sum R_T(y_t)^2 = T^3/3 + O(T^2)$ we have

$$\begin{aligned} \xi_T &= \frac{1}{T^3} \sum_{t=1}^{T} [R_T(y_t)^2 - 2R_T(y_t)R_T(x_t) + R_T(x_t)^2] \\ &= \frac{2 - 2\tilde{b}_T}{T^3} \sum_{t=1}^{T} R_T(x_t)^2 \\ &= \frac{2}{3}(1 - \tilde{b}_T) + o_p(1) \ . \end{aligned}$$

If y_t and x_t are not cointegrated, then $\tilde{b}_T$ has a nondegenerate limiting distribution (see Phillips (1987a) for the linear case). On the other hand, if y_t and x_t are cointegrated, then $\tilde{b}_T$ converges to one in probability and therefore ξ_T converges to zero.

Second, consider a Cramer-von Mises type of statistic based on the residuals of a cointegration regression on the ranks:

$$\widetilde{\xi}_T = \frac{1}{T^3} \sum_{t=1}^{T} [R_T(y_t) - \tilde{b}_T R_T(x_t)]^2$$

$$
\begin{aligned}
&= \frac{1}{T^3}\sum_{t=1}^{T}[R_T(y_t)^2 - 2\tilde{b}_T R_T(y_t)R_T(x_t) + \tilde{b}_T^2 R_T(x_t)^2] \\
&= \frac{1-\tilde{b}_T^2}{T^3}\sum_{t=1}^{T} R_T(x_t)^2 \\
&= \frac{1}{3}(1-\tilde{b}_T^2) + o_p(1) \ .
\end{aligned}
$$

Hence, a two-step approach similar to the one suggested by Engle and Granger (1987) can be seen as a two-sided version of a test based on ξ_T.

Third, the statistic ξ_T is related to the rank correlation coefficient, which is known as "Spearman's rho". Spearman's rho is defined as

$$r_s = 1 - \frac{6}{T^3 - T}\sum_{t=1}^{T} d_t^2 \tag{4.9}$$

(e.g. Kendall and Gibbons 1990, p.8). The statistic r_s can therefore be seen as a mapping of ξ_T into the interval $[-1, 1]$. If x_t and y_t are cointegrated, Spearman's rho converges in probability to one as $T \to \infty$.

Theorem 4.2 implies that, if $f(x_t)$ and $g(y_t)$ are independent random walk sequences, we have

$$T^{-1}d_{[aT]} \ \Rightarrow \ a(\mathcal{A}_1 - \mathcal{A}_3) + (1-a)(\mathcal{A}_2 - \mathcal{A}_4) \ ,$$

where $\mathcal{A}_1, \ldots, \mathcal{A}_4$ are independent random variables with an arcsine distribution. Since Theorem 5.2 allows for heteroscedastic and serially correlated increments, the asymptotic distributions of the differences d_t is not affected by weak forms of heteroskedasticity and short run serial correlation.

Power. Under the alternative of a cointegration relationship as given in (4.6) we have

$$
\begin{aligned}
T^{-1}d_{[aT]} &= T^{-1}\left\{R_T[T^{-1/2}g(y_t)] - R_T[T^{-1/2}f(x_t)]\right\} \\
&= T^{-1}\left\{R_T[T^{-1/2}f(x_t) + o_p(1)] - R_T[T^{-1/2}f(x_t)]\right\} \\
&\Rightarrow \ 0 \ .
\end{aligned}
$$

Hence, κ_T and ξ_T converge to zero as $T \to \infty$, i.e., both rank tests are consistent.

Apart from this general statement it is quite difficult to obtain analytical results for the (local) power of the test. Nevertheless, some interesting properties of the rank test can be observed when the parametric analog of the ranked differences is considered. Let the normalized difference of the series be defined as

$$\delta_t = \frac{y_t}{\bar{\sigma}_y} - \frac{x_t}{\bar{\sigma}_x} \ ,$$

where y_t and x_t are $I(1)$, $E(y_t) = E(x_t) = 0$ for all t, and $\bar{\sigma}_x^2, \bar{\sigma}_y^2$ denote the respective long run variances (cf Definition 4.1). Accordingly, a parametric analog of the statistic ξ_T is constructed as

$$D_T = \frac{1}{T^2} \sum_{t=1}^{T} \delta_t^2 \tag{4.10}$$

and under the null hypothesis of two uncorrelated random walk sequences, the statistic is asymptotically distributed as $\int W_1(a)^2 \mathrm{d}r + \int W_2(a)^2 \mathrm{d}r$, where $W_1(a)$ and W_2 are independent standard Brownian motions. Under the alternative hypothesis assume that $u_t = y_t - \beta x_t$ is stationary and $\beta > 0$. In this case we have $\bar{\sigma}_y = \beta \bar{\sigma}_x$ so that

$$D_T = \frac{1}{T^2} \sum_{t=1}^{T} \left(\frac{\beta x_t + u_t}{\beta \bar{\sigma}_x} - \frac{x_t}{\bar{\sigma}_x} \right)^2 = \frac{1}{\beta^2 \bar{\sigma}_x^2 T^2} \sum_{t=1}^{T} u^2.$$

It is seen that for large T the power of the test depends on the "signal-to-noise ratio" $\beta^2 \bar{\sigma}_x^2 / \sigma_u^2$, where σ_u^2 is the variance of u_t. As a consequence, the power of the test is a monotonically increasing function of the parameter β. In contrast, β does not affect the power of a Dickey-Fuller cointegration test. As a result, we expect that a test based on D_T or its ranked counterpart ξ_T has more (less) power than the Dickey-Fuller type cointegration test if β is large (small).

Extensions. So far I have assumed that $f(x_t)$ and $g(y_t)$ are independent $I(1)$ series. Of course, this assumption is quite restrictive and in many applications it is reasonable to assume that the series are correlated. We therefore relax this assumption and instead assume that $f(x_t)$ and $g(y_t)$ converge to two correlated Brownian motions $W_1(a)$ and $W_2(a)$ with correlation coefficient

$$\rho = E[W_1(1) W_2(1)] \ .$$

Since $f(x_t)$ and $g(y_t)$ are not observed it is not possible to estimate ρ directly. Rather I will investigate the relationship between ρ and the expected correlation coefficient of the rank differences

$$\rho_T^R = \frac{\sum_{t=2}^{T} \Delta R_T(x_t) \Delta R_T(y_t)}{\sqrt{\left(\sum_{t=2}^{T} \Delta R_T(x_t)^2\right)\left(\sum_{t=2}^{T} \Delta R_T(y_t)^2\right)}} \ . \tag{4.11}$$

If there exists an (asymptotic) one-to-one relationship between ρ and $E(\rho_T^R)$, then it is possible to derive the limiting distributions of the test statistics. Unfortunately, the relationship between ρ and $E(\rho_T^R)$ is very complicated and an analytical evaluation appears intractable. Therefore, Monte Carlo, simulations are employed to approximate the functional relationship between the two parameters.

Figure 5.4 presents the estimated relation between ρ and $E(\rho_T^R)$ using 5000 Monte Carlo replications with $T = 100$. It is seen that the correlation between the ranked differences ρ_T^R tends to underestimate ρ in absolute value. However, the difference is small for moderate values of ρ, ρ_T^R can be used as a first guess of ρ. This suggests that for small values of ρ, the test statistic can be corrected in a similar manner as in the linear case:

$$\xi_T^* = \xi / \widehat{\sigma}_{\Delta d}^2 \quad \text{and} \quad \kappa_T^* = \kappa_T / \widehat{\sigma}_{\Delta d} \ , \tag{4.12}$$

where

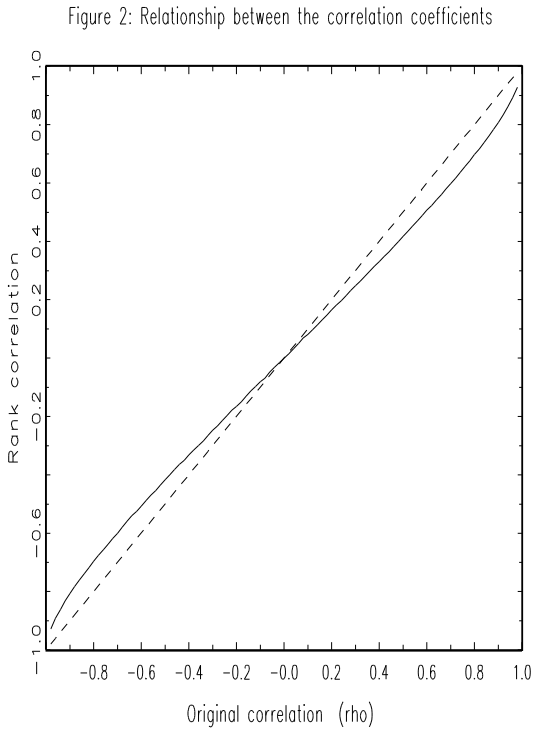
$$\widehat{\sigma}_{\Delta d}^2 = T^{-2} \sum_{t=2}^{T} (d_t - d_{t-1})^2 \ .$$

Critical values for the corrected statistic ξ_T^* and κ_T^* are presented in Table A.5.1 in the appendix to this chapter. If the absolute value of ρ is close to one, this correction performs poorly, however, and a more precise correction is required. In general, the critical values of the statistics ξ_T^* and κ_T^* depend on the correlation coefficient ρ or - by using the relationship between ρ and $E(\rho_T^R)$, on $E(\rho_T^R)$. Let $c_\xi^\alpha(\rho)$ denote the critical value of ξ_T^* with respect to the significance level of α. Using the relationship between ρ and $E(\rho_T^R)$, the critical value may alternatively be expressed as

$$c_\xi^\alpha(\rho) = c_\xi^\alpha(0) \lambda_\xi^\alpha(E\rho_T^R) \ ,$$

Figure 5.4:

The relationship between ρ and $E(\rho_T^R)$



where $\lambda_\xi^\alpha(\cdot)$ is an (unknown) function and $c_\xi^\alpha(0)$ is the critical value of ξ_T^* as presented in Table A.5.1. Accordingly, a test with a correct size in the case of a substantial value of the correlation coefficient is obtained as

$$\xi_T^{**} = \xi_T^*/\lambda_\xi^\alpha(E\rho_T^R) \quad \text{and} \quad \kappa_T^{**} = \kappa_T^*/\lambda_\kappa^\alpha(E\rho_T^R) \ . \tag{4.13}$$

Unfortunately, the determination of the function $\lambda_\xi^\alpha(E\rho_T^R)$ seems intractable so that an approximation is obtained using Monte Carlo simulations. Specifically I use 5000 Monte Carlo replications with $T = 100$ in the range $\rho = [-0.98, -0.96, \ldots, 0.96, 0.98]$. For the resulting 99 values of $\lambda_\xi^\alpha = c_\xi^\alpha(\rho)/c_\xi^\alpha(0)$ with $\alpha = 0.05$ the following regression equations were fitted[2]

$$\begin{aligned} \lambda_\kappa^{0.05} &\simeq 1 - 0.174(\rho_T^R)^2 \qquad (R^2 = 0.985) \\ \lambda_\xi^{0.05} &\simeq 1 - 0.462\rho_T^R \ , \qquad (R^2 = 0.929) \end{aligned}$$

where the uncentered R^2 is given in parentheses. The statistics ξ_T^{**} and κ_T^{**} have the same limiting distributions as for the case assuming uncorrelated series (see Table A.5.1 for critical values).

Furthermore, it is possible to generalize the test in order to test for cointegration among the $k+1$ variables $y_t, x_{1t}, \ldots, x_{kt}$, where it is assumed that $g(y_t)$ and $f_j(x_{jt})$ $(j = 1, \ldots, k)$ are monotonic functions. Let $R_T(\mathbf{x}_t) = [R_T(x_{1t}), \ldots, R_T(x_{kt})]'$ be a $k \times 1$ vector and $\widetilde{\mathbf{b}}_T$ is the least-squares estimate from a regression of $R_T(y_t)$ on $R_T(\mathbf{x}_t)$. Using the residuals

$$\tilde{u}_t = R_T(y_t) - \widetilde{\mathbf{b}}_T' R_T(\mathbf{x}_t).$$

a multivariate rank statistic is obtained from the normalized sum of squares of the residuals:

$$\Xi_T[k] = T^{-3} \sum_{t=1}^{T} (\tilde{u}_t^R)^2 \ ,$$

[2]To select an appropriate regression equation, a fourth order polynomial was estimated and the dominant regressor is selected for the approximation. Somewhat surprisingly, the dominant regressor for $\lambda_\kappa^{0.05}(\rho_T^R)$ is the squared correlation coefficient, whereas $\lambda_\xi^{0.05}(\rho_T^R)$ is well approximated by using a linear function.

Using

$$T^{-3}\sum_{t=1}^{T} R_T(y_t)R_T(x_{jt}) = \frac{1}{2}\left[\frac{2}{3} - \xi_{jt}\right] + o_p(1) \ ,$$

where ξ_{jt} is the bivariate rank statistic for y_t and x_{jt} defined as $\xi_{jt} = T^{-3}\sum_{t=1}^{T}[R_T(y_t) - R_T(x_{jt})]^2$, it is not difficult to show that the multivariate test statistic can be represented as

$$\Xi_T[k] = \frac{1}{3} - \frac{1}{4}\delta_T'\Psi_T\delta_T \ ,$$

where

$$\delta_T = \begin{bmatrix} \xi_{1T} - 2/3 \\ \xi_{2T} - 2/3 \\ \vdots \\ \xi_{kT} - 2/3 \end{bmatrix}$$

and

$$\Psi_T = T^{-3}\sum_{t=1}^{T} R_T(\mathbf{x}_t)R_T(\mathbf{x}_t)' \ .$$

To account for a possible correlation between the series, a modified statistic can be constructed:

$$\Xi_T^*[k] = \Xi_T[k]/\widehat{\sigma}_{\Delta u}^2 \ , \tag{4.14}$$

where

$$\widehat{\sigma}_{\Delta u}^2 = T^{-2}\sum_{t=2}^{T}(\tilde{u}_t - \tilde{u}_{t-1})^2 \ .$$

Critical values for the test statistic $\Xi_T^*[k]$ are presented in Table A.5.1.

4.3 A rank test for neglected nonlinearity

Whenever the rank test for cointegration indicates a stable long run relationship, it is interesting to know whether the cointegration relationship is linear or nonlinear. For a convenient representation of such null and alternative hypotheses I follow Granger (1995) and write the nonlinear relationship as

$$y_t = \gamma_0 + \gamma_1 x_t + f^*(x_t) + u_t \ , \tag{4.15}$$

where $\gamma_0 + \gamma_1 x_t$ is the linear part of the relationship. Under the null hypothesis it is assumed that $f^*(x_t) = 0$ for all t. If $f^*(x_t)$ is unknown, it may be approximated by Fourier series (e.g. Gallant 1981) or a neural network (Lee, White and Granger 1993). Here we suggest to use the multiple of the rank transformation $\theta R_T(x_t)$ instead of $f^*(x_t)$.

It is interesting to note that the rank transformation is to some extent related to the neural network approach suggested by Lee, White and Granger (1993). If x_t is a $k \times 1$ vector of "input variables" and α is a corresponding vector of coefficients, the neural network approach approximates $f^*(x_t)$ by $\sum_{j=1}^{q} \beta_j \psi(x_t' \alpha_j)$, where $\psi(\cdot)$ has the properties of a cumulated distribution function. A function often used in practice is the logistic $\psi(x) = x/(1-x)$. In our context, x_t is a scalar variable, so that the neural network term simplifies to $\beta\psi(\alpha x_t)$. Using $T^{-1}R_T(x_t) = \widehat{F}_T(x_t)$, where $\widehat{F}_T(x_t)$ is the empirical distribution function, the rank transformation can be motivated by letting $\psi(\alpha x_t)$ be the empirical distribution function with the attractive property that the parameter α can be dropped due to the invariance of the rank transformation.

If it is assumed that x_t is exogenous and u_t is white noise with $u_t \sim N(0, \sigma^2)$, a score test statistic is obtained as the $T \cdot R^2$ of the least-squares regression

$$\tilde{u}_t = c_0 + c_1 x_t + c_2 R_T(x_t) + e_t \ , \tag{4.16}$$

where $\tilde{u}_t = y_t - \tilde{\gamma}_0 - \tilde{\gamma}_1 x_t$ and $\tilde{\gamma}_0$ and $\tilde{\gamma}_1$ are the least-squares estimates from a regression of y_t on a constant and x_t.

A problem with applying the usual asymptotic theory to derive the limiting null distribution of the test statistic is that the regression (4.16) involves the nonstationary variables x_t and $R_T(x_t)$. However under some (fairly restrictive) assumptions, the following theorem shows that under the null hypothesis $c_2 = 0$ the score statistic is asymptotically χ^2 distributed.

THEOREM 4.3 *Let* $x_t = \sum_{j=1}^{t} v_j$ *and*

$$y_t = \gamma_0 + \gamma_1 x_t + u_t \ ,$$

where it is assumed that v_t *is* $I(0)$ *and* u_t *is white noise with* $E(u_t) = 0$ *and*

$E(u_t^2) = \sigma_u^2$. As $T \to \infty$, the score statistic for $H_0 : c_2 = 0$ in the regression (4.16) has an asymptotic χ^2 distribution with one degree of freedom.

PROOF: It is convenient to introduce the matrix notation:

$$X_1 = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_T \end{bmatrix} \quad \text{and } X_2 = \begin{bmatrix} R_T(x_1) \\ \vdots \\ R_T(x_T) \end{bmatrix} ,$$

$y = [y_1, \ldots, y_T]'$ and $\tilde{u} = [\tilde{u}_1, \ldots, \tilde{u}_T]'$. With this notation, the score statistic can be written as

$$T \cdot R^2 = \frac{1}{\tilde{\sigma}^2}(\hat{\beta}_2)^2[X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2] ,$$

where $\hat{\beta}_2$ is the least-squares estimator of β_2 in the regression $y = X_1\beta_1 + X_2\beta_2 + u$ and $\tilde{\sigma}^2 = \tilde{u}'\tilde{u}/T$. As shown by Park and Phillips (1988), the least-squares estimator in a regression with strictly exogenous $I(1)$ regressors is conditionally normally distributed, so that conditional on $X = [X_1, X_2]$, $\tilde{\beta}_2$ is asymptotically distributed as $N(0, V_2)$, where

$$V_2 = \sigma_u^2[X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2]^{-1}.$$

Since $\tilde{\sigma}_u^2$ converges weakly to σ_u^2 it follows that $T \cdot R^2$ has an asymptotic χ^2 distribution with one degree of freedom. ■

Unfortunately, the assumptions of Theorem 4.3 are too restrictive to provide a useful result for practical situations. In many situations, the errors u_t are found to be serially correlated and the regressor x_t may be endogenous. However, using standard techniques for cointegration regressions (Saikonnen 1991, Stock and Watson 1993) the test can be modified to accommodate serially correlated errors and endogenous regressors. For this purpose assume that

$$\begin{aligned} u_t &= E(u_t|\Delta x_t, \Delta x_{t\pm1}, \Delta x_{t\pm2}, \ldots) + v_t \\ &= \sum_{j=-\infty}^{\infty} \pi_j \Delta x_{t-j} + v_t \end{aligned}$$

and v_t admits the autoregressive representation

$$v_t = \sum_{j=1}^{\infty} \alpha_j v_{t-j} + \varepsilon_t \ ,$$

where the roots of the lag polynomial $\alpha(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \cdots$ are bounded away rom the unit circle.

Under the null hypothesis of linear cointegration we thus have the representation

$$y_t = \gamma_0^* + \sum_{j=1}^{\infty} \alpha_j y_{t-j} + \gamma_1^* x_t + \sum_{j=-\infty}^{\infty} \pi_j^* \Delta x_{t-j} + \varepsilon_t \tag{4.17}$$

(see Stock and Watson 1993 and Inder 1995).

A test for non-linear cointegration can be obtained by truncating the infinite sums appropriately and forming $T \cdot R^2$ for the regression of the residuals $\tilde{\varepsilon}_t$ on the regressors of (4.17) and $R_T(x_t)$. Along the lines of Theorem 4.3 it can be shown that the resulting score statistic is asymptotically χ^2 distributed under the null hypothesis of a linear cointegration relationship.

4.4 Nonlinear short run dynamics

In recent papers by Bierens (1997a,b) and Vogelsang (1998a,b) it was observed that it is possible to construct test statistics that asymptotically do not depend on parameters involved in the short run dynamics of the process. Accordingly, it is not necessary to estimate the nuisance parameters such as the coefficients for the lagged differences in a Dickey-Fuller regression or the "long run variance" (2π times the spectral density at frequency zero) by using a kernel estimate as in Phillips and Perron (1988). Such an approach is called "model free" in Bierens (1997a) and "nonparametric" in Bierens (1997b). Albeit both terms may be somewhat misleading, I follow Bierens (1997b) and use the term "nonparametric". In fact, it is difficult to think of any test, which is "less parametric".

The idea behind this approach is the following. Under suitable conditions on the sequence $\varepsilon_1, \varepsilon_2, \ldots$ the functional central limit theorem (FCLT) implies

$$T^{-1/2} \sum_{t=1}^{[aT]} \varepsilon_t \ \Rightarrow \ \bar{\sigma} W(a) \ . \tag{4.18}$$

In what follows I consider tests, which by construction do not depend on the parameter $\bar{\sigma}^2$, asymptotically.

To test the hypothesis that y_t is $I(0)$ against the alternative $y_t \sim I(1)$, Tanaka (1990) and Kwiatkowski et al. (1992) suggest an LM type test statistic given by

$$\varrho_T = \frac{T^{-2} \sum_{t=1}^{T} Y_t^2}{T^{-1} \sum_{t=1}^{T} y_t^2} , \tag{4.19}$$

where $Y_t = y_1 + \cdots + y_t$ denotes the partial sum process and it is assumed that y_t is white noise. If y_t is serially correlated, the denominator is replaced by the estimated long run variance (cf Kwiatkowski et al. 1992). Note that ϱ_T is the (normalized) variance ratio of the partial sums and the original series. This test statistic is therefore referred to as the "variance ratio statistic".

In contrast to Kwiatkowski et al. (1992), the variance ratio statistic is employed to test the null hypothesis that y_t is $I(1)$ against the alternative $y_t \sim I(0)$. Thus, our test flips the null and alternative hypothesis of the test suggested by Kwiatkowski et al. (1992). The following theorem presents the limiting null distribution of such a test procedure.

THEOREM 4.4 *Let $y_t \sim I(1)$. Then, as $T \to \infty$, we have*

$$T^{-1}\varrho_T \;\Rightarrow\; \frac{\int_0^1 [\int_0^a W(s)ds]^2 da}{\int_0^1 W(a)^2 da} .$$

PROOF: From Definition (4.1) it follows that

$$\begin{aligned} T^{-1/2} y_{[aT]} &\Rightarrow \bar{\sigma} W(a) \\ T^{-3/2} Y_{[aT]} &\Rightarrow \bar{\sigma} \int_0^a W(s)ds \end{aligned}$$

Thus, we get

$$\begin{aligned} T^{-1}\varrho_T &= \frac{T^{-4} \sum_{t=1}^{T} Y_t^2}{T^{-2} \sum_{t=1}^{T} y_t^2} \\ &\Rightarrow \frac{\int_0^1 [\int_0^a W(s)ds]^2 da}{\int_0^1 W(a)^2 da} \end{aligned}$$

■

It is important to notice that the null distribution does not depend on nuisance parameters. This is due to the fact that the parameter $\bar{\sigma}^2$ cancels in the variance ratio. The following theorem shows that the test is consistent against stationary alternatives and the usual class of local alternatives (e.g. Phillips 1987b).

THEOREM 4.5 *Let y_t be stationary with the Wold representation $y_t = \sum_{j=0}^{\infty} \gamma_j \varepsilon_{t-j}$, where $\gamma_0 = 1$, $\sum_{j=0}^{\infty} \gamma_j^2 < \infty$, and ε_t is white noise with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma_\varepsilon^2$. Under this alternative we have as $T \to \infty$*

$$\varrho_T \Rightarrow \frac{\bar{\sigma}^2 \int_0^1 W(a)^2 dr}{\sigma_y^2},$$

where $\bar{\sigma}^2 = (\sum_{j=0}^{\infty} \gamma_j)^2 \sigma_\varepsilon^2$ and $\sigma_y^2 = \sum_{j=0}^{\infty} \gamma_j^2 \sigma_\varepsilon^2$.

Under the local alternative $\phi_T = 1 - c/T$ in $y_t = \phi_T y_{t-1} + \varepsilon_t$ the limiting distribution is given by

$$T^{-1}\varrho_T \;\Rightarrow\; \frac{\int_0^1 [\int_0^a J_c(s)ds]^2 da}{\int_0^1 J_c(a)^2 da} \;,$$

where $J_c(a)$ represents an Ornstein-Uhlenbeck process defined as $J_c(a) = \int_0^a e^{(a-s)c} dW(s)$.

PROOF: Under a stationary alternative we have

$$T^{-2} \sum Y_t^2 \;\Rightarrow\; \bar{\sigma}^2 \int_0^1 W(a)^2 da \;.$$

Using these results and $T^{-1} \sum_{t=1}^{T} y_t^2 \xrightarrow{p} \sigma_y^2$, the limiting distribution against a stationary alternative follows immediately.

Under the sequence of local alternatives we have (cf Phillips 1987b)

$$\begin{aligned} T^{-1/2} y_{[aT]} &\Rightarrow J_c(a) \\ T^{-3/2} Y_{[aT]} &\Rightarrow \int_0^a J_c(s)ds \;. \end{aligned}$$

Therefore, the limiting distribution results from replacing the Brownian motion $W(a)$ in Theorem 4.4 by $J_c(a)$. ■

Testing the cointegration rank. The variance ratio statistic for a nonparametric unit root test can be generalized straightforwardly to test hypotheses on the cointegration rank in the spirit of Johansen (1988, 1991). To this end I make the following assumption.

Assumption 4.2 *There exists an invertible matrix* $Q = [\gamma, \beta]'$*, where* γ *and* β *are linearly independent* $n \times q$ *and* $n \times (n-q)$ *matrices, respectively, with* $0 \leq q < n$ *such that*

$$Qy_t = \begin{bmatrix} \gamma' y_t \\ \beta' y_t \end{bmatrix} \equiv \begin{bmatrix} \xi_t \\ u_t \end{bmatrix} = z_t$$
$$T^{-1/2} \xi_{[aT]} \;\Rightarrow\; W_q(a)$$
$$T^{-2} \sum_{t=1}^{T} u_t u_t' = o_p(1) \ ,$$

where $W_q(a)$ *is a* q*–dimensional Brownian motion with unit covariance matrix.*

To allow for some general nonlinear processes generating u_t, it is not assumed that the "error correction term" u_t is stationary. Instead I assume that the trend component ξ_t is "variance dominating" in the sense that the variance of ξ_t diverges with a faster rate than u_t.

The dimension of the stochastic trend component ξ_t is related to the cointegration rank of a linear system by $q = n - r$, where r is the rank of the matrix Π in the VECM representation

$$\Delta y_t = \Pi y_{t-1} + v_t \ , \tag{4.20}$$

and v_t is a stationary error vector. In a linear system, the hypothesis on the number of stochastic trends is equivalent to a hypothesis on the cointegration rank as in Johansen (1988). However, since we do not assume that the process is linear, the representation of the form (4.20) may not exist.

Our test statistic is based on the eigenvalues λ_j of the problem

$$|\lambda_j B_T - A_T| = 0 \ , \tag{4.21}$$

where

$$A_T = \sum_{t=1}^{T} y_t y_t' \ , \quad B_T = \sum_{t=1}^{T} Y_t Y_t'$$

and $Y_t = \sum_{j=1}^{t} y_j$ denotes the n-dimensional partial sum with respect to y_t. The eigenvalues of (4.21) are identical to the eigenvalues of the matrix $R_T = A_T B_T^{-1}$. For $n = 1$ the eigenvalue is identical to the statistic $1/\varrho_T$ and, thus, the test can be seen as a generalization of the variance ratio statistic to multivariate processes.

The eigenvalues of (4.21) are given by

$$\lambda_j = \frac{\eta_j' A_T \eta_j}{\eta_j' B_T \eta_j} \ , \tag{4.22}$$

where η_j is the eigenvector associated with the eigenvalue λ_j. If the vector η_j falls inside the space spanned by the columns of γ, then $\eta_j' A_T \eta_j$ is $O_p(T^2)$ and $\eta_j' B_T \eta_j$ is $O_p(T^4)$ so that the eigenvalue is $O_p(T^{-2})$. On the other hand, if the eigenvector η_j falls into the space spanned by the columns of β, it follows that $T^2\lambda_j$ tends to infinity, as $T \to \infty$. Therefore, the test statistic

$$\Lambda_q = T^2 \sum_{j=1}^{q} \lambda_j \tag{4.23}$$

has a nondegenerate limiting distribution, where $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ denote the eigenvalues of the matrix R_T. In contrast, if the number of stochastic trends is smaller than q, then Λ_q diverges to infinity. The following theorem presents the limiting null distribution for the test statistic Λ_q.

THEOREM 4.6 *Assume that y_t admits a decomposition as in Assumption 4.2 with $0 < q \leq n$. Then, as $T \to \infty$*

$$\Lambda_q \ \Rightarrow \ tr \left\{ \int_0^1 W_q(a) W_q(a)' da \left[\int_0^1 V_q(a) V_q(a)' da \right]^{-1} \right\} \ ,$$

where $W_q(a)$ is a q-dimensional standard Brownian motion and $V_q(a) = \int_0^a W_q(s) ds$.

PROOF: Let $Z_t = \sum_{j=1}^{t} z_j$ denote the partial sum with respect to $z_t = Q y_t = [\xi_t', u_t']'$. Then, the eigenvalues of problem (4.21) also solves the problem

$$|\lambda_j D_T - C_T| = 0 \ ,$$

where

$$C_T = \sum_{t=1}^{T} z_t z_t' \ , \quad D_T = \sum_{t=1}^{T} Z_t Z_t' \ .$$

Partition the corresponding eigenvectors $\tilde{\eta}_j = [\tilde{\eta}'_{1j}, \tilde{\eta}'_{2j}]'$ such that $\tilde{\eta}'_j z_t = \tilde{\eta}'_{1j}\xi_{1t} + \tilde{\eta}'_{2j}u_{2t}$, and Z_t is partitioned accordingly. We normalize the eigenvectors as

$$\tilde{\eta}_1 = [\tilde{\eta}_{11}, \ldots, \tilde{\eta}_{1q}] = \begin{bmatrix} I_q \\ \Phi_T \end{bmatrix}$$

so that $\tilde{\eta}'_{1j} z_t = \xi_{jt} + \Phi'_T u_t$, where ξ_{jt} denotes the j-th component of the vector ξ_t. It follows that

$$\begin{aligned} \lambda_j &= \frac{\tilde{\eta}'_j C_T \tilde{\eta}_j}{\tilde{\eta}'_j D_T \tilde{\eta}_j} \\ &= \frac{\sum_{t=1}^{T} \xi_{jt}^2 + o_p(T^2)}{\sum_{t=1}^{T} Z_{jt}^2 + o_p(T^4)} \\ &= \frac{\sum_{t=1}^{T} \xi_{jt}^2}{\sum_{t=1}^{T} Z_{jt}^2} + o_p(1), \end{aligned}$$

where $Z_{jt} = \sum_{s=1}^{t} \xi_{js}$. As $T \to \infty$ we therefore have

$$T^2 \sum_{j=1}^{q} \lambda_j \;\Rightarrow\; tr\left\{ \int_0^1 W_q(a) W_q(a)' da \left[\int_0^1 V_q(a) V_q(a)' da \right]^{-1} \right\}$$

■

From this theorem it follows that the asymptotic distribution of the q smallest eigenvalues of the problem (4.21) does not depend on nuisance parameters and, thus, we do not need to select the lag order of the VAR process as in Johansen's approach or the truncation lag as for the test of Quintos (1998).

Including Deterministic Terms. To accommodate processes with a nonzero mean assume that the mean function $E(y_t) = Cd_t$ is a linear function of deterministic variables like a constant, time trend or dummy variables stacked in the $k \times 1$ vector d_t and C is a matrix (or vector) of unknown coefficients. In this case it is natural to remove the mean of the time series by using the residuals from the

regression $y_t = \widehat{C}d_t + \widehat{u}_t$, where $\widehat{C}$ denotes the least-squares estimator of C. The partial sums are then constructed by using the residuals $\widehat{Y}_t = \widehat{u}_1 + \cdots + \widehat{u}_t$. This procedure is referred to as "OLS-detrending".[3]

In cointegrated systems it is often the case that the deterministic terms are constrained under the cointegration hypothesis. In particular, it is assumed that y_t has a linear time trend, whereas the cointegrating relations $\beta' y_t$ have a constant mean. This specification of the mean function is used if the model is estimated with Johansen's ML procedure including unrestricted constants. This specification implies that the linear combinations $u_t = \beta' y_t$ are adjusted for a mean, whereas the vector of permanent components $\xi_t = \gamma' y_t$ is adjusted for a time trend. Thus, in order to impose these restrictions on the deterministic terms, estimates for the matrices β and γ are needed. A possible way to estimate these matrices is to use the principle component estimator. An attractive property of this estimator is that in this case γ is estimated to be the orthogonal complement of the cointegration matrix. It follows, that this matrix is estimated with the same convergence rate as the cointegration matrix (cf Harris 1997).

Let $\widehat{\beta}$ and $\widehat{\gamma}$ denote the estimates from a principal component procedure. Then, the adjusted vector of time series results as

$$x_t^* = \begin{bmatrix} \widehat{\gamma}' y_t - \widehat{a}_0 - \widehat{a}_1 t \\ \widehat{\beta}' y_t - \widehat{b} \end{bmatrix},$$

where $\widehat{a}_0$ and $\widehat{a}_1$ are the least-squares estimates from a regression of $\widehat{\xi}_t = \widehat{\gamma}' y_t$ on a constant and a time trend and $\widehat{b}$ denotes the mean of $\widehat{u}_t = \widehat{\beta}' y_t$. Then, the statistic is computed by using x_t^* instead of y_t and the critical values for a test with time trend are applied. Note that

$$x_t^* = \begin{bmatrix} \gamma' y_t - E(\gamma' y_t) + o_p(T^{1/2}) \\ \beta' y_t - E(\beta' y_t) + O_p(1) \end{bmatrix}$$

and, thus, the differences between the estimated and true nonstationary components are asymptotically negligible but the transitory components are measured

[3] Alternatively, the parameters in C may be estimated by applying quasi differences to the nonstationary components as in Xiao and Phillips (1999).

with an nonvanishing error. However, since the transitory components are asymptotically irrelevant under the null hypothesis, this does not affects the null distribution of the test.

4.5 Small sample properties

Rank tests. To investigate the small sample properties of the rank tests for nonlinear cointegration I follow Gonzalo (1994) and generate two time series according to the model equations

$$\begin{aligned} y_t &= \beta z_t + u_t \ , \qquad u_t = \alpha u_{t-1} + \varepsilon_t \\ z_t &= z_{t-1} + v_t \ , \end{aligned} \tag{4.24}$$

where

$$\begin{bmatrix} \varepsilon_t \\ v_t \end{bmatrix} \sim \text{iid } N\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right).$$

The variable x_t is obtained from the random walk z_t by using the inverse function $x_t = f^{-1}(z_t)$.

Under the null hypothesis $H_0 : \ \alpha = 1$ there is no cointegration relationship between the series. If in addition $\beta = 0$ and $\rho = 0$, then x_t and y_t are two independent random walks with serially uncorrelated increments. For this specification, Table 5.1 reports the rejection frequencies of different cointegration tests. The rank tests κ_T and ξ_T are computed as in (4.7) and (4.8) and "CDF" indicates the Dickey-Fuller t-test applied to the residuals of a linear regression of y_t on x_t and a constant. The results for the linear process is given in the left half of Table 5.1 indicated by $f(x) = x$. It might be surprising to see that for $\beta = 1$ the rank test is much more powerful than the CDF test if α is close to one. As was argued in Section 5.3, this is due to the fact that the rank statistics depend on the parameter β, whereas the Dickey-Fuller type test for cointegration does not depend on β. In fact, the simulation results indicate that the power of the rank tests are very sensitive to the value of β. For $\beta = 0.5$ the Dickey-Fuller tests perform better, whereas for $\beta = 1$ the rank tests clearly outperform the Dickey-Fuller type tests.

Table 5.1: Size and power ($\theta = 0$)

$f(x) =$	x			x^3	$\log(x)$	$\tan(x_t)$
α	κ_T	ξ_T	CDF	CDF	CDF	CDF
Size ($\beta = 0$)						
1.00	0.049	0.049	0.050	0.098	0.051	0.077
Power ($\beta = 0.5$)						
0.98	0.232	0.248	0.080	0.310	0.156	0.186
0.95	0.288	0.332	0.225	0.491	0.229	0.295
0.90	0.387	0.484	0.698	0.730	0.323	0.474
0.80	0.551	0.708	0.999	0.866	0.410	0.616
Power ($\beta = 1$)						
0.98	0.594	0.616	0.080	0.310	0.156	0.186
0.95	0.733	0.792	0.225	0.491	0.229	0.295
0.90	0.861	0.930	0.698	0.730	0.323	0.474
0.80	0.953	0.993	0.999	0.866	0.410	0.616

Note: Rejection frequencies resulting from 10,000 replications of the process given in (4.24). The sample size is $T = 200$. The nominal size is 0.05. The test statistic κ_T and ξ_T are defined in (4.7) and (4.8). CDF indicates a Dickey-Fuller t-test on the residuals of a cointegrating regression including a constant term.

Table 5.2: Testing correlated random walks

ρ	κ_T^*	ξ_T^*	κ_T^{**}	ξ_T^{**}	$\Xi_T^*[1]$
-0.900	0.130	0.007	0.054	0.033	0.003
-0.600	0.073	0.016	0.048	0.039	0.012
-0.400	0.058	0.023	0.048	0.042	0.020
-0.200	0.052	0.030	0.049	0.042	0.029
0.000	0.048	0.041	0.047	0.040	0.042
0.200	0.050	0.053	0.047	0.040	0.056
0.400	0.052	0.070	0.043	0.037	0.074
0.600	0.062	0.096	0.039	0.036	0.107
0.900	0.105	0.234	0.038	0.053	0.255

Note: Rejection frequencies resulting from two random walks with $\text{corr}(\Delta x_t, \Delta y_t)=\rho$. The sample size is $T = 200$. The statistics κ_T^* and ξ_T^* are defined in (4.12) and κ_T^{**} and ξ_T^{**} are given in (4.13). The statistic $\Xi_T^*[1]$ is the two-sided test statistic given in (4.14).

Table 5.3: Power against nonlinear cointegration relationships

$f(x) =$	x^3		$\log(x)$		$\tan(x)$	
regressor:	$R_T(x_t)$	$f(x_t)$	$R_T(x_t)$	$f(x_t)$	$R_T(x_t)$	$f(x_t)$
$\beta = 0.01$	0.267	0.252	0.246	0.216	0.237	0.226
$\beta = 0.05$	0.473	0.485	0.701	0.676	0.549	0.548
$\beta = 0.1$	0.714	0.746	0.957	0.955	0.834	0.855
$\beta = 0.5$	0.974	0.988	1.000	1.000	0.999	1.000

Note: Simulated power from a score tests using $R_T(x_t)$ and $f(x_t)$ as additional regressors. The sample size is $T = 200$.

It should also be noted that the rank test does not require to estimate the cointegration parameter β. Accordingly, this test has the same power as for the case of a known cointegration relationship. Furthermore, the rank tests impose the *one-sided* hypothesis that $f(x_t)$ is an *increasing* function.

Since the rank tests are invariant to a monotonic transformation of the variables, the power function is the same as for the linear case. Comparing the power of the CDF test with the rank counterparts, it turns out that the power of the CDF test may drop dramatically for nonlinear alternatives (see also Granger and Hallman (1991b)), while the rank test performs as well as in the linear case. In particular, for the case $f(x) = \log(x)$ the parametric CDF test performs quite poorly.

To study the ability of the modified statistics suggested in Section 5.3 to account for correlated random walks, I simulate correlated data by varying the correlation coefficient in the range $\rho = -0.9, \ldots, 0.9$. The statistics κ_T^* and ξ_T^* use a correction that is similar to the correction in the linear case. As argued in Section 5.3, this test statistic should perform well if the correlation is moderate. For more substantial correlation coefficients, the improved statistics κ_T^{**} and ξ_T^{**} defined in (4.13) should be used.

Table 5.2 presents the empirical sizes for testing the null hypothesis of no cointegration with $\beta = 0$ and $\alpha = 1$. It turns out that the statistic κ_T^* performs well in the range $\rho \in [-0.4, 0.4]$, whereas the statistic ξ_T^* should only be used for a small correlation in the range $\rho \in [-0.2, 0.2]$. In contrast, the statistics κ_T^{**} and ξ_T^{**} turn out to be very robust against a correlation among x_t and y_t.

Next, consider the small sample properties of the rank test for nonlinear cointegration suggested in Section 5.4. It is assumed that y_t and x_t are cointegrated so that $y_t - \beta f(x_t)$ is stationary. By setting $\alpha = 0.5$ I generate serially correlated errors and by letting $\rho = 0.5$, the variable x_t is correlated with the errors u_t, that is, x_t is endogenous. The rank test for nonlinear cointegration is obtained by regressing y_t on $x_t, y_{t-1}, \Delta x_{t+1}, \Delta x_t, \Delta x_{t-1}$ and a constant. The score statistic is computed as $T \cdot R^2$ from a regression of the residuals on the same set of regressors and the ranks $R_T(x_t)$.

To study the power of the tests I consider three different nonlinear functions. As a benchmark I perform the tests using $f(x_t)$ instead of the ranks $R_T(x_t)$. Of course, using the true functional form, which is usually unknown in practice, the test is expected to have better power than the test based on the ranks. Surprisingly, the results of the Monte Carlo simulations (see Table 5.3) suggest that the rank test may even be (slightly) more powerful than the parametric test, whenever the nonlinear term enters the equation with a small weight ($\beta = 0.01$). However, the gain in power is quite small and falls in the range of the simulation error. In any case, the rank test performs very well and seems to imply no important loss of power in comparison to the parametric version of the test.

Nonlinear short run dynamics. Next, I consider the small sample properties of nonparametric tests of Section 5.5 that were suggested to allow for nonlinear short run dynamics. Four nonlinear processes are considered, where the assumptions of the Dickey-Fuller test are violated. The first process it the bilinear process:

$$\text{(bilin):} \quad \Delta y_t = \phi \varepsilon_{t-1} \Delta y_{t-1} + \varepsilon_t \ . \tag{4.25}$$

Note that the correlation between ε_{t-1} and Δy_{t-1} implies a linear time trend in y_t. The second process is a variable coefficient model of the form

$$\text{(VCM):} \quad \Delta y_t = \alpha_t \Delta y_{t-1} + \varepsilon_t \ \text{ with } \alpha_t = \cos(2\pi t/T) \ . \tag{4.26}$$

Kim (1999) shows that in such an "evolutionary model" the augmented Dickey-Fuller test has a limiting distribution that is different from the distribution derived in Dickey and Fuller (1979).

Third, I consider a threshold autoregressive process given by

$$\text{(TAR):} \quad y_t = \begin{cases} y_{t-1} + \varepsilon_t & \text{for } |\varepsilon_t| < 2\sigma_\varepsilon \\ 0.8 y_{t-1} + \varepsilon_t & \text{for } |\varepsilon_t| \geq 2\sigma_\varepsilon \end{cases} \ . \tag{4.27}$$

The fourth process is a stochastic unit root process as considered in Granger and Swanson (1997):

$$\text{(STUR):} \quad y_t = \alpha_t y_{t-1} + \varepsilon_t \ \text{ with } \alpha_t = 0.4 + 0.6\alpha_{t-1} + \eta_t \ , \tag{4.28}$$

where $E(\varepsilon_t^2) = 1$ and $E(\eta_t^2) = 0.01^2$. Note that $E(\alpha_t) = 1$.

The empirical sizes computed from 10,000 realizations with $T = 200$ are presented in Table 5.4. All tests allow for a linear trend. For the TAR and STUR model, the size bias of the ADF(1) test is very small and the variance ratio statistic ϱ_T performs only marginally better. However, for the bilinear model and the VCM model, the size bias of the ADF(1) test is more substantial. In these cases the nonparametric statistic ϱ performs much better.

To investigate the properties of the nonparametric cointegration test I generate data according to the "canonical" process (Toda 1994) with MA(1) errors

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} - \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix} \ , \tag{4.29}$$

where $E(\varepsilon_{1t}^2) = E(\varepsilon_{2t}^2) = 1$ and $E(\varepsilon_{1t}\varepsilon_{2t}) = \theta$. To test the hypothesis $r = 1$, I let $\phi_1 = 0$ and $\phi_2 = -0.2$. Under the alternative I set $\phi_1 \in \{-0.05, -0.1, -0.2\}$. Furthermore, I let $\theta = 0$ and $\theta = 0.8$ to investigate the impact of the error correlation. The sample size is $T = 200$ and 10,000 samples are generated to compute the rejection frequencies of the tests.

For Johansen's LR trace test, the process is approximated by a VAR[p] process, where p is 4 and 12, respectively. The respective tests are denoted by LR(4) and LR(12). Unrestricted constants are included in each equation. The nonparametric test statistic is denoted by Λ_q and the critical values are taken from Table A.5.3 in the appendix to this chapter. First, consider the results for testing $H_0 : q = r = 1$ presented in Table 5.5. From the empirical sizes it turns out that for $\theta = 0$, a VAR[4] model is not sufficient to approximate the infinite VAR process, whereas a VAR[12] approximation yields an accurate size. The nonparametric statistic Λ_q possesses a negligible size bias, only. The power of Λ_q is substantially smaller than the power of LR(4) but clearly higher than the power of LR(12). Similar results apply for the tests letting $\theta = 0.8$. However, the LR(12) statistic now possesses a moderate size bias, whereas Λ_q is nearly unbiased. Moreover, the power of Λ_q is closer to the (favorable) LR(4) statistic than in the case of $\theta = 0$.

We now turn to the test of $H_0 : r = 0$. Under the null hypothesis the difference of the variables are generated by a multivariate MA process. In this case, all three test statistic are substantially biased, where the size bias does not depend on the parameter θ. Although the size bias differs for the three tests, the differences are moderate and some general conclusions with respect to the relative power of the tests can be drawn. For $\theta = 0$ and ϕ_1 close to unity, the nonparametric test Λ_q is slightly more powerful than the LR(4) test, whereas for $\phi_1 = 0.8$ the power of LR(4) is slightly higher. Finally, the power of LR(12) is much smaller than the power of the other two tests. For $\theta = 0.8$ a different picture emerges. The relative power of Λ_q drops substantially and for ϕ_1 close to one, the power is even lower than the power of the LR(12) test. The results for a model with a linear time trend are qualitatively similar and are not presented here.

Table 5.4: Empirical sizes for some nonlinear processes

process	ADF(1)	ϱ
bilin	0.091	0.046
VCM	0.138	0.072
TAR	0.055	0.053
STUR	0.053	0.051

Note: The entries of the table display the empirical sizes computed from 10,000 replications of model (4.25) – (4.28). "bilin" is a bilinear process, "VCM" is a process with a structural changes in the short run dynamics, "TAR" is a threshold unit root process, and "STUR" is a stochastic unit root process. The sample size is $T = 200$ and the nominal size of the test is 0.05.

Table 5.5: Testing hypotheses on the cointegration rank

$H_0 : r = 1$, $\phi_2 = -0.2$				
test statistic	$\phi_1 = 0$	$\phi_1 = -0.05$	$\phi_1 = -0.10$	$\phi_1 = -0.20$
$\theta = 0$				
Λ_T	0.059	0.346	0.604	0.853
LR(4)	0.072	0.428	0.894	0.999
LR(12)	0.048	0.180	0.389	0.636
$\theta = 0.8$				
Λ_T	0.043	0.295	0.566	0.853
LR(4)	0.057	0.310	0.793	0.999
LR(12)	0.063	0.190	0.382	0.636
$H_0 : r = 0$, $\phi_2 = 0$				
test statistic	$\phi_1 = 0$	$\phi_1 = -0.05$	$\phi_1 = -0.10$	$\phi_1 = -0.20$
$\theta = 0$				
Λ_T	0.107	0.300	0.582	0.900
LR(4)	0.083	0.241	0.558	0.962
LR(12)	0.094	0.166	0.290	0.506
$\theta = 0.8$				
Λ_T	0.107	0.240	0.508	0.854
LR(4)	0.083	0.511	0.949	1.000
LR(12)	0.094	0.352	0.581	0.768

Note: The entries of the table report the rejection frequencies based on 10,000 replications of model (4.29), where $E(y_t)$ is constant.

4.6 Empirical applications

The term structure of interest rates. The rank tests are applied to test for a possible nonlinear cointegration between interest rates with different time to maturity. Recent empirical work suggests that interest rates with a different time to maturity are nonlinearly related.[4] The data set consists of monthly yields of government bonds with different time to maturity as published by the *Deutsche Bundesbank*. The sample runs from 1967(i) through 1995(xii) yielding 348 monthly observations for each variable.

The nonlinear relationship between yields for different times to maturity can be motivated as follows. Let r_t denote the yield of a one-period bond and R_t represents the yield of a two-period bond at time t. The expectation theory of the term structure implies that

$$R_t = \phi_t + 0.5 r_t + 0.5 E_t(r_{t+1}) \ , \tag{4.30}$$

where E_t denotes the conditional expectation with respect to the relevant information set available at period t and ϕ_t represents the risk premium. Letting $r_{t+1} = E_t(r_{t+1}) + 2\nu_t$ and subtracting r_t from both sides of (4.30) gives

$$R_t - r_t = 0.5(r_{t+1} - r_t) + \phi_t + \nu_t \ .$$

Assuming that r_t is $I(1)$ and $\phi_t + \nu_t$ is stationary implies that R_t and r_t are (linearly) cointegrated (e.g. Wolters 1995). However, if the risk premium depends on r_t such that $\phi_t = f^*(r_t) + \eta_t$ with η_t stationary, then the yields are nonlinearly cointegrated:

$$R_t - f(r_t) = u_t \ \sim \ I(0) \ ,$$

where $f(r_t) = r_t + f^*(r_t)$ and $u_t = 0.5(r_{t+1} - r_t) + \eta_t + \nu_t$. Note that u_t is correlated with r_t and, therefore, r_t is endogenous. Furthermore, if the sampling interval is shorter than the time to maturity, then the errors are serially correlated even if ν_t and η_t are white noise.

[4] See, e.g., Campbell and Galbraith (1993), Pfann et al. (1996), and the reference therein.

To test whether interest rates possess a (nonlinear) cointegration relationship various tests for unit roots are applied first. Neither the conventional Dickey-Fuller t-test nor the ranked counterpart discussed in Section 5.2 reject the null hypothesis that the interest rates are $I(1)$ (not presented).

The results presented in Table 5.6 highlight the cointegration properties of yields with different time to maturity. The parametric as well as the rank tests indicate a cointegration relationship between the short term bonds, while the evidence for cointegration between short run and long run bonds (e.g. R1 and R10) is much weaker. Furthermore, the rank test for nonlinear cointegration ("nonlin") does not reveal any evidence against a linear cointegration relationship. All values of the test statistic suggested in Section 5.4 are much smaller than the critical value of 3.84. This result suggests that a linear version of the expectation theory for the term structure yields an appropriate description of the long run relationship between interest rates with different time to maturity.

Table 5.6: The cointegration relationship with R1

Var.	CDF	$Q_T[1]^*$	κ_T^{**}	ξ_T^{**}	ρ_T^R	nonlin
R2	–4.138*	0.010*	0.407	0.017*	0.893	0.003
R3	–3.591*	0.014*	0.405	0.022	0.808	0.012
R4	–3.373*	0.016*	0.404	0.023	0.730	0.034
R5	–3.213	0.018*	0.413	0.025	0.678	0.107
R10	–2.702	0.025	0.498	0.033	0.582	0.489

Note: "CDF" denotes the Dickey-Fuller t-test applied to the residuals of the cointegration regression. $\Xi_T[1]^*$ is the two-sided test statistic given in (4.14) and κ_T^{**} and ξ_T^{**} are defined in (4.13). "nonlin" indicates the test for nonlinearity using the ranks as additional regressors, and ρ_T^R is the correlation coefficient between the differences of the rank sequences. "*" indicates significance at the 0.05 significance level.

The relationship between dual-class shares. In an early paper on cointegration Granger (1986, p. 218) states that "[i]f x_t, y_t are a pair of prices from a jointly efficient, speculative market, they cannot be cointegrated." The reason is that whenever two variables are cointegrated, there exists an error correction representation so that at least one variable can be forecasted by using the lagged error correction term.

Recently, however, this statement was called into question in a number of studies. For example, Kasa (1992) finds evidence for common stochastic trends (and thus of cointegration) in international stock markets. Kehr (1997) shows that stock prices traded at different regional markets in Germany are cointegrated and Krämer (1999) and Dittmann (1998) find (fractional) cointegration between different classes of stocks of the same or very similar German companies.

These conflicting views of the efficient market hypothesis can be resolved by assuming a *nonlinear* error correction mechanism (ECM). The idea is that small deviations from the long run relationship are not predictable as claimed by Granger (1986). If the deviations become large, however, an effective adjustment process prevents stock prices of fundamentally related assets from drifting too far away. The economic reason behind such a nonlinear adjustment process is that transaction costs make it unprofitable to exploit small deviations from the fundamental relationship. When undervaluation (or overvaluation) becomes more substantial, agents will buy (or sell) the respective assets until the fundamental relationship is re-established. This reasoning naturally leads to a nonlinear version of the error correction model (e.g. Escribano and Mira 1998).

We apply a variety of cointegration tests to time series data of pairs of German voting and non-voting shares issued by the same firm (dual-class firm). Daily stock price data adjusted for stock splits, dividends and other corporate events are from the *"Deutsche Finanzdatenbank"* (DFDB) in Karlsruhe. We use logarithms of stock prices in our statistical analysis. According to the criteria liquidity and availability of long time series, a sample of 6 dual-class firms is chosen for examination. Among those four are contained in the index of the 30 largest German blue-chip stocks (DAX), the remaining two are in the German mid-cap index

MDAX. In Table 5.7 the sample is further described.

German corporate law requires that holders of non-voting shares must be compensated for the lack of corporate control by a dividend advantage. This usually takes the form of a minimum preferred dividend (stated as percentage of par value) which will be carried forward in particular years of dividend omissions (cumulative preferred dividend). Both the cumulative (past) preferred dividends and the current preferred dividends have to be paid out before the common shareholders can receive anything. In addition some firms commit themselves to pay the non-voting shareholders a certain (non-cumulative) amount in excess of the common stock dividend.

To test for (linear) cointegration I first compute the augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) test applied to the residuals of a OLS cointegrating regression (cf. Phillips and Ouliaris 1990). For the ADF test ten lagged differences and a constant are included in the regression. The truncation lag for the Phillips-Perron test on the residuals is set to 20. Using these tests, only a weak evidence for cointegration is found for RWE, and RHM (see Table 5.8), whereas in the other cases, the null hypothesis of no cointegration is rejected at a significance level of 0.05. Applying Johansen's (1988) likelihood-ratio test procedure based on a VAR[10] model with a constant restricted to the cointegration relationship gives a slightly different picture. This test finds a cointegration relationship for RWE, MAN and BMW. In the remaining cases, the LR statistics are insignificant.

An important problem is, however, that these test procedures are based on the assumption of a linear process and thus may be problematical in applications using financial time series data. Applying the nonparametric test statistic suggested in Section 5.5, the evidence for cointegration decrease substantially. Only for BMW the test clearly points to a cointegration relationship. For RHM a cointegration relationship is found at the 0.05 significance level but not on the 0.01 significance level. The diminished evidence may be due to an improved robustness against nonlinear short run dynamics. However, it may also be due to a lack of power compared to the parametric counterparts.

Interestingly, the rank tests for cointegration unambiguously reject the hypoth-

esis of no cointegration. This may be the result of a structural instability of the cointegration relationships, that may be represented by a nonlinear cointegration relationship. In fact, as shown by Breitung and Wulff (1999), there is a strong evidence for structural changes in RWE, MAN, BMW, and BOSS. Furthermore, tests for nonlinearity in the error correction mechanism suggest that the stock prices of MAN, RHM and BOSS indeed react nonlinearly to a deviation to the long run relationship in the previous period.

Table 5.7: Details of the dual-class shares

Firm	Abbrev.	No. obs.	Sample range	Index	Div. adv.
RWE	RWE	5894	1/2/74 – 7/31/97	DAX	5 / –
MAN	MAN	5588	4/24/75 – 7/31/97	DAX	4 / –
BMW	BMW	1983	8/25/89 – 7/31/97	DAX	2 / –
Volkswagen	VW	2706	10/6/86 – 7/31/97	DAX	4 / 2
Rheinmetall	RHM	3182	10/31/84 – 7/31/97	MDAX	6 / 2
Boss	BOSS	2052	5/22/89 – 7/31/97	MDAX	3 / 3

Note: Dividend advantage ("Div. adv.") is expressed in percent of par value, whereby the first figure indicates the minimum dividend and the excess dividend is given after the slash.

Table 5.8: Cointegration tests for voting and non-voting stocks

Test	RWE	MAN	BMW	VW	RHM	BOSS
ADF	–3.197	–5.638**	–6.641**	–3.551*	–2.824	–3.812*
PP	–3.744*	–8.707**	–11.19**	–4.139**	–3.517*	–5.701**
LR	21.42*	36.21**	47.14**	15.87	17.78	18.52
Λ	94.15	236.6	501.38**	83.53	340.91*	302.6
RDF	–22.47**	–15.43**	–10.74**	–6.859**	–7.219**	–9.380**
Rdiff	0.0005**	0.0011**	0.0024**	0.0058*	0.0047**	0.0030

Note: "ADF" denotes the augmented Dickey-Fuller test including 10 lagged differences and a constant. "PP" is the unit root test of Phillips and Perron (1988) applied to the residuals of the cointegration regression. "LR" is Johansen's trace statistic for the hypothesis that the cointegration rank is zero. Λ is the nonparametric test for cointegration. "RDF" is rank tests for cointegration as suggested by Granger and Hallman (1991) and "Rdiff" is the statistic ξ_T as defined in (4.8). * and ** indicate significance with respect to the 0.05 and 0.01 significance level, respectively.

4.7 Appendix: Critical values

Table A.5.1: Critical values for the rank statistics

T	0.10	0.05	0.01
κ	0.644	0.552	0.422
ξ	0.057	0.042	0.024
κ^*	0.394	0.364	0.316
ξ^*	0.023	0.019	0.013
$\Xi^*[1]$	0.025	0.020	0.014
$\Xi^*[2]$	0.020	0.016	0.012
$\Xi^*[3]$	0.016	0.014	0.010
$\Xi^*[4]$	0.014	0.012	0.009
$\Xi^*[5]$	0.012	0.010	0.008
$\Xi^*[6]$	0.010	0.009	0.008

Note: Critical values computed from 10,000 realizations of independent random walk sequences with $T = 500$.

Table A.5.2: Critical values for $T^{-1}\varrho_T$ statistic (ϱ_T)

T	0.1	0.05	0.01
mean adjusted			
100	0.01435	0.01004	0.00551
250	0.01433	0.01003	0.00561
500	0.01473	0.01046	0.00536
trend adjusted			
100	0.00436	0.00342	0.00214
250	0.00442	0.00344	0.00223
500	0.00450	0.00355	0.00225

Note: The hypothesis of a unit root process is rejected if the test statistic falls below the respective critical values reported in this table.

Table A.5.3: Critical values for Λ_q

$q_0 = n - r_0$	0.1	0.05	0.01
mean adjusted			
1	67.89	95.60	185.0
2	261.0	329.9	505.8
3	627.8	741.1	1024
4	1200	1360	1702
5	2025	2255	2761
6	3177	3460	4045
7	4650	5049	5905
8	6565	7061	8032
trend adjusted			
1	222.4	281.1	443.6
2	596.2	713.3	976.1
3	1158	1330	1689
4	1972	2184	2699
5	3107	3429	4120
6	4572	4954	5780
7	6484	6984	8012
8	8830	9388	10714

Note: The hypothesis $r = r_0$ is rejected if the test statistic exceeds the respective critical value. The simulation are based on a sample size of $T = 500$.

Chapter 5

Conclusions and outlook

The previous chapters give an account of recent research in multiple time series analysis that intends to bridge the gap between reduced form VAR models used in Johansen's (1995) work and structural models based on sets of simultaneous equations advocated by the influential Cowles commission. A variant of this methodology was developed by Sims (1986) and others who suggest to impose restrictions on the covariance matrix of the errors in order to identify macroeconomic shocks. The latter approach became recently very popular and plays a dominating role during the last few years. It was argued that the structural framework is able to account for the cointegration properties of the data so that the long run relationship of the variables can be given a structural interpretation. This is an important advantage over the (reduced form) cointegration analysis that can only indicate some stationary linear combination of the variables. Whether the observed long run relationship is a "money demand schedule" or a "money supply schedule" or a mixture of both cannot be decided without imposing additional assumption needed to identify the structure.

It is sometimes argued (see e.g. Möller 1993) that it does not make sense to attach the usual structural interpretation to the long run solution of a system because the variables in a long run relationship cannot be qualified as dependent and predetermined variables. This is indeed correct as long as the short run dynamics of the system are left unspecified. However, if the dynamic structure

(that is the long and short run parameters) is identified appropriately, the error correction representation is a particular reformulation of the structural model and structural inference applies in the usual way (Hsiao 1997).

If the errors of the system are identified as economic shocks, it is attractive to impose cointegration relationships as it allows to qualify shocks to be transitory or permanent. For example, it is often assumed that demand shocks affect real quantities only in the short run, whereas supply shocks may have a permanent effect on variables like output and unemployment. As demonstrated in Chapter 3, such assumptions can easily be imposed in a structural cointegrated VAR model. Using impulse response analysis the dynamic effects of structural stocks on the variables of interest can be studied.

To discuss the alternative approaches to draw structural inference from linear dynamic models it was assumed that sufficient identifying restrictions can be derived from economic theory so that the remaining problem is how to impose these assumptions on the cointegrated VAR system. In many applications, however, economic theory gives only a vague motivation of some structural properties but the initial enthusiasm that fully specified models can be derived from some basic assumptions on optimizing agents under limited resources is unwarranted. Usually, economic theory is not a "sharp theory" in the sense that it attach a prior probability of one to a single structure and a probability of zero to all other structures (see also Leamer 1983). Rather, theoretical reasoning suggests that some specifications are more plausible than others. Empirical work should therefore start with specifications that correspond best to the preferred theoretical model and try out other specifications that are similarly plausible.

Such a "robustness analysis" is often encountered in practice (for example Bernanke and Mihow 1997, Cochrane 1998, Rudebush 1998). Although it is clearly desirable to have robust results for the range of all possible specifications it is questionable to claim that robustness of the results is a necessary condition for a credible empirical finding (e.g. Cooley and Dwyer 1998). First, it does not seem sensible to claim for a "robust" judgement in the sense that it is always be compatible with the addresses to the jury delivered from both the prosecutor

and the defense counsel. Of course the judge is in a comfortable position if the prosecutor and the defense counsel demand for the same judgement but it is unrealistic to claim that this should always be the case. Second, a result that is robust against a change in the assumptions may simply demonstrate that the assumptions are uninformative. Why should we confine ourselves to use uninformative assumptions?

In many cases theoretical considerations give rise to a particular structural assumption but is silent about other aspects of the model that have to be specified as well. For example, (keynesian) money demand theory states that money demand depends on output and interest rates so that an increase in output has a positive effect on money demand, whereas a raise in interest rates leads to a decrease in the desired monetary base. For a dynamic analysis of the relationship between money, output and interest rates, we need further structural assumptions on the effect of interest rates on output (such as an IS-schedule) and an equation with interest rates as the dependent variable (e.g. a money supply function). Hence, the structural system employs quite different theoretical ingredients that may interact in a complex way. Indeed, as argued forcefully by Sims (1980), the theory of rational expectation implies that it does not make sense to specify structural elements of the system in isolation. On the other hand, theoretical models that try to model the all aspects of economic behaviour jointly are difficult to handle and yield structural models with nonlinear cross-equation constraints (see, e.g., Wickens 1982) that are not considered in this thesis. Future work should therefore allow for more complex specifications. Examples for current research in that field are Johansen and Swensen (1999) and Kozicki and Tinsley (1988).

Bibliography

Amemiya, T. and R.Y. Wu (1972), The Effect of Aggregation on Prediction in the Autoregressive Model, Journal of the American Statistical Association, 339, 628-632.

Amisano, G. and C. Giannini (1997), Topics in Structural VAR Econometrics, 2nd ed., Berlin: Springer.

Balke, N.S. and T.B. Fomby (1997), Threshold Cointegration, International Economic Review, 38, 627–645.

Benkwitz, A., Lütkepohl, H., Neumann, M.H. (1999), Problems Related to Confidence Intervals for Impulse Responses of Autoregressive Processes. Econometric Reviews, forthcoming.

Bernanke, B.S. und H. Mihov (1997), What does the Bundesbank Target?, European Economic Review, 41, 1025–1053.

Bewley, R. and D. Orden (1994), Alternative Methods for Estimating Long-Run Responses With Applications to Australian Import Demand, Econometric Reviews, 13, 179–204.

Bewley, R. and M. Yang (1995), Tests for Cointegration Based on Canonical Correlation Analysis, Journal of the American Statistical Association, 90, 990–996.

Bierens, H.J. (1997a), Testing the Unit Root With Drift Hypothesis Against Nonlinear Trend Stationarity, With an Application to the US price Level and Interest Rate, Journal of Econometrics, 81, 29–64.

Bierens, H.J. (1997b), Cointegration Analysis, in: C. Heij, H. Schumacher, B. Hanzon, and C. Praagman (eds), System Dynamics in Economic and Financial Models, New York: John Wiley.

Blanchard, O.J (1989), A Traditional Interpretation of Macroeconomic Fluctuations, American Economic Review, 79, 1146–1164.

Blanchard, O.J. and D. Quah (1989), The Dynamic Effects of Aggregate Supply and Demand Disturbances, American Economic Review, 79, 655–673.

Bossaerts, P. (1988), Common Nonstationary Components of Asset Prices, Journal of Economic Dynamics and Control 12, 347–364.

Boswijk, H.P. (1995), Identifiability in Cointegrated Systems, Paper presented at the 7th World Congress of the Econometric Society, Tokyo 1995.

Boswijk, H.P. (1996), Testing Identifiability of Cointegration Vectors, Journal of Business & Economic Statistics, 14, 153–160.

Boswijk, H.P. and J.-P. Urbain (1997), Lagrange-Multiplier Tests for Weak Exogeneity: A synthesis, Econometric Review, 16, 21–38.

Box, G.E.P. and G.M. Jenkins (1970), Time Series Analysis, Forecasting and Control, San Francisco: Holden Day.

Breitung, J. (1995a), A Simultaneous Equations Approach to Cointegrated Systems, SFB 373 Discussion Paper No. 46/1995, Humboldt University Berlin

Breitung, J. (1995b), Modified Stationarity Tests With Improved Power in Small Samples, Statistical Papers, 36, 77–95.

Breitung, J. (1998), Canonical Correlation Statistics for Testing the Cointegration Rank in a Reversed Order, SFB 373 Discussion Paper No. 105/1998.

Breitung, J. and C. Gouriéroux (1997), Rank tests for Unit Roots, Journal of Econometrics, 81, 7-28.

Breitung, J. and M. Heinemann (1998), Short Run Comovement, Persistent Shocks and the Business Cycle, Jahrbücher für Nationalökonomie und Statistik, 436–448.

Breitung, J. and C. Wulff (1999), Nonlinear Error Correction and the Efficient Market Hypothesis: The Case of German Dual-Class Shares, SFB 373 Discussion Paper No. 67/1999.

Breitung, J., F. Haslinger and M. Heinemann (1993), Ist die Makroökonomik eine wissenschaftliche Illusion?, Ökonomie und Gesellschaft, 10, 150–203.

Brewer, K.R.W. (1973), Some Consequences of Temporal Aggregation and Systematic Sampling for ARMA and ARMAX Models, Journal of Econometrics, 1, 133-154.

Burridge, P. and E. Guerre (1996), The Limit Distribution of Level Crossings of a Random Walk and a Simple Unit Root Test, Econometric Theory, 12, 705–723.

Caporale, G.M. and N. Pittis (1999), Efficient Estimation of Cointegrating Vectors and Testing for Causality in Vector Autoregressions, Journal of Economic Surveys, 13, 1–35.

Campbell, B. and J.W. Galbraith (1993), Inference in Expectations Models of the Term Structure: A non-parametric approach, Empirical Economics 18, 623–638.

Choi, I. (1994), Residual-Based Tests for the Null of Stationarity With Applications to U.S. Macroeconomic Time Series, Econometric Theory, 10, 720–746.

Choi, I. and B.C. Ahn (1995), Testing for Cointegration in Systems of Equations, Econometric Theory, 11, 952–983.

Choi, I. and B.C. Ahn (1999), Testing the Null of Stationarity for Multiple Time Series, Journal of Econometrics, 88, 41–77.

Christiano and Eichenbaum (1987), Temporal Aggregation and Structural Inference in Macroeconomics, Carnegie-Rochester Conference Series on Public Policy, 26, 63–130.

Christiano, L. and M. Eichenbaum (1992), Current Real-Business-Cycle Theories and Aggregate Labour-Market Fluctuations, American Economic Review, 82, 430–450.

Clarida, R. and M. Gertler (1996), How the Bundesbank Conducts Monetary Policy, NBER Working Paper No. 5581.

Clemens, M.P. and G.E. Mizon (1991), Empirical Analysis of Macroeconomic Time Series: VAR and structural models, European Economic Review, 35, 887–932.

Cochrane, J.H. (1998), What do the VARs mean? Measuring the Output Effects of Monetary Policy, Journal of Monetary Economics, 41, 277–300.

Cogley, T. and J.M. Nason (1995), Output Dynamics in Real-Business-Cycle Models, American Economic Review, 85, 492–511.

Cooley, T.F. and M. Dwyer (1998), Business Cycle Analysis Without Much Theory: A Look at Structural VARs, Journal of Econometrics, 83, 57–88.

Cooley, T.F. and S.F. LeRoy (1985), Atheoretical Macroeconometrics: A Critique, Journal of Monetary Economics, 16, 283–308.

Corradi, V. (1995), Nonlinear Transformations of Integrated Time Series: A reconsideration, Journal of Time Series Analysis, 16, 537–549.

Crowder, W.J. (1995), The Dynamic Effects of Aggregate Demand and Supply Disturbances: Another look, Eonomics Letters, 49, 231–237.

Cuthbertson, K. and M.P. Taylor (1989), Anticipated and Unanticipated Variables in the Demand for M1 in the U.K., Manchester School, 57, 319–339.

Davidson, J. (1998), Structural Relations, Cointegration and Identification: Some simple results and their application, Journal of Econometrics, 87, 87-113.

Davidson, J. and S. Hall (1991), Cointegration in Recursive Systems, Economic Journal, 101, 239–251.

Dickey, D.A. and W.A. Fuller (1979), Distribution of the Estimates for Autoregressive Time Series With a Unit Root, Journal of the American Statistical Association 74, 427–431.

Dickey, D.A. and S.G. Pantula (1987), Determining the Order of Differencing in Autoregressive Processes, Journal of Business and Economic Statistics, 5, 455–461.

Dittmann, I. (1998), Fractional Cointegration of Voting and Non-Voting Shares, Discussion Paper, Universität Dortmund, Nr. 98-02.

Dolado, J.J. (1992), A Note on Weak Exogeneity in VAR Cointegrated Models, Econometric Letters, 38, 139–143.

Dolado, J.J., J.D. Lopez-Salido, and J.L. Vega (1999), Spanish Unemployment and Inflation Persistence: Are there Phillips Trade-offs?, Spanish Economic Review, forthcoming.

Dhrymes, P. and D.D. Thomakos (1998), Structural VAR, MARMA, and Open Economy Models, International Journal of Forecasting, 14, 187–198.

Dufour, J.-M. and E. Renault (1998), Short Run and Long Run Causality in Time Series: Theory, Econometrica, 66, 1099–1126.

Dwyer, M. (1998), Impulse Response Priors for Discriminating Structural Vector Autoregressions, Woking Paper, Los Angeles.

Elliot, G., T.J. Rothenberg and J.H. Stock (1996), Efficient Tests for an Autoregressive Unit Root, Econometrica, 813–836.

Engle, R.F. (1969), Biases from Time-Aggregation of Distributed Lag Models, Ph.D. Dissertation, Cornell University.

Engle, R.F. and C.W.J. Granger (1987), Co-integration and Error Correction: Representation, Estimation, and Testing, Econometrica, 55, 251-276.

Engle, R.F. and B.S. Yoo (1991), Cointegrated Economic Time Series: An Overview with New Results, in: Fomby, T.B. and G.F. Rhodes (eds.), Advances in Econometrics: Co-Integration, Spurious Regression, and Unit Roots, Greenwich: JAI Press, 237–266.

Engle, R.F., D.F. Hendry, and J.-F. Richard (1983), Exogeneity, Econometrica, 51, 277–304.

Escribano, A. and S. Mira (1998), Nonlinear Cointegration and Error Correction, Working paper, Department of Statistics and Econometrics, Universidad Carlos III de Madrid.

Faust, J. (1998), The Robustness of Identified VAR Conclusions about Money, Carnegie-Rochester Conference Series on Public Policy, 49, 207–244.

Faust, J. and E.M. Leeper (1997), When do Long-Run Identifying Restrictions Give Reliable Results?, Journal of Business and Economic Statistics, 15, 345–352.

Fisher, F.M. (1966), The Identification Problem in Econometrics, New York: McGraw-Hill.

Fountis, N.G. and D.A. Dickey (1989), Testing for a Unit Root Nonstationarity in Multivariate Autoregressive Time Series, 17, 419–428.

Franses, P.H. and McAlleer, M. (1998), Testing for Unit Roots and Nonlinear Transformations, Journal of Time Series Analysis, 19, 147–164.

Fung, B. S. and M. Kasumovich (1998), Monetary Shocks in the G-6 Countries: Is there a Puzzle?, Journal of Monetary Economics, 42, 575–592.

Gali, J. (1992), How Well Does the IS-LM Model Fit Postwar U.S. Data?, Quarterly Journal of Economics, 709–738.

Gallant, A.R. (1981), Unbiased Determination of Production Technologies, Journal of Econometrics, 30, 149–169.

Gallant, A.R. and H. White (1988), A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models, Oxford: Basil Blackwell.

Garratt, A., K. Lee, M.H. Pesaran, and Y. Shin (1999), A Structural Cointegrating VAR Approach to Macroeconometric Modelling, mimeo.

Geweke, J. (1986), Measurement of Linear Dependence and Feedback Between Multiple Time Series, Journal of American Statistical Association, 77, 304–313.

Gonzalo, J. (1994), Five Alternative Methods of Estimation Long-run Equilibrium Relationships, Journal of Econometrics, 60, 203–233.

Gordon, S. and D. Boccanfuso (1998), What Can We Learn from Structural Vector Autoregression Models, Université Laval, Canada, mimeo.

Gordon, D.B. and E.M. Leeper (1994), The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification, Journal of Political Economy, 102, 1228–1247.

Granger, C.W.J. (1969), Investigating Causal Relations by Econometric Models and Cross Spectral Methods, Econometrica, 37, 428-438.

Granger, C.W.J. (1980), Testing for Causality, A Personal Viewpoint, Journal of Economic Dynamics and Control, 2, 329-352.

Granger, C.W.J. (1981), Some Properties of Time Series Data and their Use in Econometric Model Specification, Journal of Econometrics, 16, 121–130.

Granger, C.W.J. (1986), Developments in the Study of Co-integrated Economic Variables, Oxford Bulletin of Economics and Statistics, 48, 213–228.

Granger, C.W.J. (1988), Some Recent Developments in a Concept of Causality, Journal of Econometrics, 39, 199-211.

Granger, C.W.J. (1995), Modelling Nonlinear Relationships Between Extended-memory Variables, Econometrica, 63, 265–279.

Granger, C.W.J. and J. Hallman (1991a), Long Memory Processes With Attractors, Oxford Bulletin of Economics and Statistics, 53, 11–26.

Granger, C.W.J. and J. Hallman (1991b), Nonlinear Transformations of Integrated Time Series, Journal of Time Series Analysis, 12, 207–224.

Granger, C.W.J. and N.R. Swanson (1997), An Introduction to Stochastic Unit-Root Processes, Journal of Econometrics, 80, 35–62.

Granger, C.W.J., and J.-L. Lin (1995), Causality in the Long Run, Econometric Theory, 11, 530-536.

Granger, C.W.J. and T. Teräsvirta (1993), Modelling Nonlinear Economic Relationships, Oxford: Oxford University Press.

Granger, C.W.J., T. Inoue and N. Morin (1997), Nonlinear Stochastic Trends, Journal of Econometrics, 81, 65–92.

Granger, C.W.J. and P. Newbold (1977), Forecasting Economic Time Series, New York: Academic Press.

Grenander, U. and M. Rosenblatt (1957), Statistical Analysis of Stationary Time Series, New York: John Wiley.

Haavelmo, T. (1944), The Probability Approach in Econometrics, Econometrica, 12, 1–118, Supplement.

Haldrup, N. (1994), The Asymptotics of Single-Equation Cointegration Regressions with I(1) and I(2) Variables, Journal of Econometrics, 63, 153–181.

Hamilton, J.D. (1994), Time Series Analysis, Princeton: Princeton University Press.

Harris, D. (1997), Principal Components Analysis of Cointegrated Time Series, Econometric Theory, 13, 529–557.

Harris, D. and B. Inder (1994), A Test of the Null Hypothesis of Cointegration, in: C.P. Hargreaves (ed.), Nonstationary Time Series Analysis and Cointegration, Oxford: Oxford University Press, 133–152.

Harbo, I., S. Johansen, B. Nielsen, A. Rahbek (1998), Asymptotic Inference on Cointegrating Rank in Partial Systems, Journal of Business & Economic Statistics, 16, 388–399.

Hartley, P.R. and C.E. Walsh (1992), A Generalized Method of Moments Approach to Estimating a Structural Vector Autoregression, Journal of Macroeconomics, 14, 199–232.

Haug, A. (1996), Tests for Cointegration: A Monte Carlo Comparison, Journal of Econometrics, 71, 89–115.

Heckman, J.L. (1999), Causal Parameters and Policy Analysis in Econometrics: A twentieth century retrospective, NBER Working Paper No. 7333.

Hendry, D.F. and G.E. Mizon (1983), Evaluating Dynamic Econometric Models by Encompassing the VAR, in: P.C.B. Phillips (ed.) Models, Methods and Applications of Econometrics: The XXth anniverary of CORE, Cambridge (MA): MIT Press.

Hoffman, D.L. and R.H. Rasche (1996), Aggregate Money Demand Functions, Boston: Kluwer.

Holly, A. (1988), Specification Tests: an overview, in: T.F. Bewley, Fifth World Congress, Vol 1, p. 59–97.

Hood, W.C. and T.C. Koopmans (1953), Studies in Econometric Method, New York: John Wiley.

Horvath, M.T.K. and M.W. Watson (1995), Testing for Cointegration When Some of the Cointegrating Vectors are Prespecified, Econometric Theory, 11, 984–1014.

Hsiao, C. (1983), Identification, in: Z. Griliches and M.D. Intriligator (eds), Handbook of Econometrics, Vol. 1., Amsterdam: North-Holland.

Hsiao, C. (1997), Cointegration and Dynamic Simultaneous Equation Models, Econometrica, 65, 647–670.

Hubrich, K., H. Lütkepohl, and P. Saikkonen (1998), A Review of System Cointegration Tests, Humboldt-University Berlin, SFB 373 Working Paper, No. 101/1998.

Inder, B. (1995), Finite Sample Arguments for Appropriate Estimation of Cointegrating Vectors, Discussion Paper, Monash University, Australia.

Jöreskog, K.G. (1969), A General Approach to Confirmatory Factor Analysis, Psychometrika, 34, 183–202.

Jöreskog, K.G. and D. Sörbom (1993), LISREL 8: User's Reference Guide, Scientific Software International, 1525 East 53rd Street, Chicago, Illinois 60615.

Johansen, S. (1988), Statistical Analysis of Cointegration Vectors, Journal of Economic Dynamics and Control, 12, 231–254.

Johansen, S. (1991), Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, Econometrica, 59, 1551–1580.

Johansen, S. (1992a), Determination of Cointegration Rank in the Presence of a Linear Time Trend, Oxford Bulletin of Economics and Statistics, 54, 383–397.

Johansen, S. (1992b), Testing Weak Exogeneity and the Order of Cointegration in UK Money Demand Data, Journal of Policy Modeling, 14, 313–334.

Johansen, S. (1994), The Role of the Constant and Linear Terms in Cointegration Analysis of Nonstationary Variables, Econometric Reviews, 13, 205–229.

Johansen, S. (1995a), Likelihood-based Inference in Cointegrated Vector Autoregressive Models, Oxford: University Press.

Johansen, S. (1995b), Identifying Restrictions of Linear Equations With Applications to Simultaneous Equations and Cointegration, Journal of Econometrics, 69, 111-132.

Johansen, S. (1995c), A Statistical Analysis of Cointegration for I(2) Variables, Econometric Theory, 11, 25–59.

Johansen, S. and K. Juselius (1990), Maximum Likelihood Estimation and Inference on Cointegration – With applications to the demand for money, Oxford Bulletin of Economics and Statistics, 52, 169–210.

Johansen, S. and K. Juselius (1992), Structural Tests in a Multivariate Cointegration Analysis of the PPP and the UIP for UK, Journal of Econometrics, 53, 211-244.

Johansen, S. and K. Juselius (1994), Identification of the Long-Run and the Short-Run Structure: An application to the ISLM model, Journal of Econometrics, 63, 7–36.

Judge, G.G., R.C. Hill, W.E. Griffiths, H. Lütkepohl, and T.-C. Lee (1988), Introduction to the Theory and Practice of Econometrics, 2. ed., New York: Wiley.

Juselius, K. (1993), VAR Modelling and Haavelmo's Probability Approach to Macroeconomic Modelling, Empirical Economics, 18, 595–622.

Juselius, K. (1998), A Structured VAR for Denmark Under Changing Monetary Regimes, Journal of Business & Economic Statistics, 16, 400–411.

Kasa, K. (1992), Common Stochastic Trends in International Stock Markets, Journal of Monetary Economics, 29, 95–124.

Kedem, B. (1980), Binary Time Series, New York: Marcel Dekker.

Kehr, L.-H. (1997), Preisfindung bei verteilter Börsenstruktur, Wiesbaden: DUV.

Kendall, M. and J.D. Gibbons (1990), Rank Correlation Methods, London: Edward Arnold.

King, R.G., C.I. Plosser, J.H. Stock and M.W. Watson (1991), Stochastic Trends and Economic Fluctuations, The American Economic Review, 81, 819–840.

King, R. and M. Watson (1997), Testing Long-Run Neutrality, Federal Reserve Bank of Richmond, Economic Quarterly, 83, 69–101.

Krämer, W. (1997), Kointegration von Aktienkursen, Universität Dortmund, mimeo.

Kydland, F.E. and E.C. Prescott (1982), Time to Build and Aggregate Fluctuations, Econometrica, 50, 1345–1370.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin (1992), Testing the Null Hypothesis of Stationary Against the Alternative of a Unit Root: How sure are we that economic time series have a unit root?, Journal of Econometrics 54, 159–178.

Leamer, E.E. (1983), Let's Take the Con Out of Econometrics, American Economic Review, 73, 31–43.

Lee T.-H., White H., Granger C.W.J. (1993), Testing for Neglected Nonlinearity in Time Series Models: A comparison of neural network methods and alternative tests, Journal of Econometrics, 56, 269–290.

Lehfeldt, R.A. (1915), review of H.L. Moore: Economic Cycles: Their Law and Cause, Economic Journal, 25, 409–411.

Leybourne, S.J. and B.P.M. Mc Cabe (1994a), A Simple Test for Cointegration, Oxford Bulletin of Economics and Statistics, 56, 97–103.

Leybourne, S.J. and B.P.M. Mc Cabe (1994b), A Consistent Test for a Unit Root, Journal of Business & Economic Statistics, 12, 157–166.

Levtchenkova, S., A.R. Pagan and J.C. Robertson (1998), Shocking Stories, Journal of Economic Surveys, 12, 507–532.

Lewis, R. and G.C. Reinsel (1985), Prediction of Multivariate Time Series by Autoregressive Model Fitting, Journal of Multivariate Analysis, 16, 393–411.

Lippi, M. and L. Reichlin (1993), The Dynamic Effects of Aggregated Demand and Supply Disturbances: Comment, American Economic Review, 83, 644–658.

Lo, A.W. (1991), Long-Term Memory in Stock Market Prices, Econometrica, 59, 1279–1314.

Lucas, R.E. (1976), Econometric Policy Evolution: A Critique, in: K. Brunner and A. Meltzer (eds.), The Phillips Curve and Labor Markets, Vol.1 of Carnegie-Rochester Conferences on Public Policy, Amsterdam: North Holland.

Lütkepohl, H. (1987), Forecasting Aggregated Vector ARMA Processes, Berlin, Springer.

Lütkepohl, H. (1991), Introduction to Multiple Time Series Analysis, Berlin: Springer.

Lütkepohl, H. (1993), Testing for Causation Between Two Variables in Higher Dimensional VAR Models, in: H. Schneeweiß and K.F. Zimmermann (eds.), Studies in Applied Econometrics, 73–91, Heidelberg: Physica.

Lütkepohl, H. (1999a), Vector Autoregressions, SFB 373 Discussion Paper No. 4/1999.

Lütkepohl, H. (1999b), Forecasting Cointegrated VARMA Processes, SFB 373 Discussion Paper No. 68/1999.

Lütkepohl, H. und J. Breitung (1997), Impulse Response Analysis of Vector Autoregressive Processes, in: C. Heij, H. Schumacher, B. Hanzon und C. Praagman (eds.), System Dynamics in Economic and Financial Models, New York: John Wiley.

Lütkepohl, H. and M.M. Burda (1997), Modified Wald Tests Under Nonregular Conditions, Journal of Econometrics, 78, 315–332.

Lütkepohl, H. and P. Saikkonen (1999), Testing for the Cointegrating Rank of a VAR Process with a Time Trend, Journal of Econometrics, forthcoming.

Lütkepohl H. and J. Wolters (1998), A Money Demand System for German M3, Empirical Economics, 23, 371-386.

Marcellino, M. (1999), Some Consequences of Temporal Aggregation in Empirical Analysis, Journal of Economic and Business Statistics, 17, 129–136.

McCallum, B.T. (1999), Analysis of the Monetary Transmission Mechanism: Methodological issues, NBER Working Paper No. 7395.

Mills, T.C. (1998), Recent Developments in Modelling Nonstationary Vector Autoregressions, Journal of Economic Surveys, 12, 279–313.

Monfort, A. and R. Rabemananjara (1990), From a VAR Model to a Structural Model, with an Application to the Wage-Price Spiral, Journal of Applied Econometrics, 5, 203–227.

Moore, H.L. (1914), Economic Cycles – Their Law and Cause, New York: Macmillan.

Mosconi, R. (1998), Malcolm: The Theory and Practice of Cointegration Analysis in RATS, Milano: Cafoscarina.

Nautz, D. and J. Wolters (1999), The Response of Long-Term Interest Rates to News about Monetary Policy Actions: Empirical Evidence for the U.S. and Germany, Weltwirtschaftliches Archiv, 135, 397–412

Pagan, A. (1995), Three Econometric Methodologies: An update, in: L. Oxley, D.A.R. George, C.J. Roberts und S. Sayer (eds.), Surveys in Econometrics, Oxford: Basil Blackwell.

Pagan, A.R. and J.C. Robertson (1998), Structural Models of the Liquidity Effect, Review of Economics and Statistics, 80, 203–217.

Park, J.Y. and P.C.B. Phillips (1988), Statistical Inference in Regressions With Integrated Processes: Part 1, Econometric Theory, 4, 468–497.

Park, J.Y. and P.C.B. Phillips (1999), Asymptotics for Nonlinear Transformations of Integrated Time Series, Econometric Theory, 15, 269–298.

Perron, P. and J.Y. Campbell (1993), A Note on Johansen's Cointegration Procedure when Trends are Present, Empirical Economics, 18, 777-789.

Pesaran, M.H. and Y. Shin (1995), Long-Run Structural Modelling, Paper presented at the 7th World Meeting of the Econometric Society, Tokyo, 1995.

Pesaran, M.H., Y. Shin and R. Smith (1999), Structural Analysis of Vector Error Correction Models with Exogenous I(1) Variables, Journal of Econometrics, forthcoming.

Pfann, G., P. Schotman, and R. Tschernig (1996), Nonlinear Interest Rate Dynamics and Implications for the Term Structure, Journal of Econometrics 74, 149–176.

Phillips, P.C.B. (1987a), Time Series Regression With a Unit Root, Econometrica 55, 277–301.

Phillips, P.C.B. (1987b), Towards a Unified Asymptotic Theory of Autoregression, Biometrika, 74, 535–48.

Phillips, P.C.B. (1991), Optimal Inference in Co-integrated Systems, Econometrica, 59, 282–306.

Phillips, P.C.B. (1995), Fully Modified Least Squares and Vector Autoregressions, Econometrica, 1023–1078.

Phillips, P.C.B. (1998), Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VARs, Journal of Econometrics, 83, 21–56.

Phillips, P.C.B. and S. Durlauf (1986), Multiple Time Series Regression With Integrated Processes, Review of Economic Studies, 53, 473–495.

Phillips, P.C.B. and B.E. Hansen (1990), Statistical Inference in Instrumental Variable Regression with I(1) Processes, Review of Economic Studies, 57, 99–125.

Phillips, P.C.B. and S. Ouliaris (1990), Asymptotic Properties of Residual Based Tests for Cointegration, Econometrica, 58, 165–193.

Phillips, P.C.B. and P. Perron (1988), Testing for a Unit Root in Time Series Regression, Biometrika 75, 335–346.

Phillips, P.C.B. and V. Solo (1992), Asymptotics for Linear Processes, Annals of Statistics, 20, 971–1001.

Quenouille, M.H. (1957), The Analysis of Multiple Time Series, London: Griffin.

Quintos, C.E. (1998), Fully Modified Vector Autoregressive Inference in Partially Nonstationary Models, Journal of the American Statistical Association, 93, 783–795.

Rahbek, A. and R. Mosconi (1999), Cointegration Rank Inference with Stationary Regressors in VAR Models, Econometrics Journal, 2.

Rasche, R.H. (1994), Money Demand and the Term Structure: Some New Ideas on an Old Problem, Michigan State University.

Renault, E. and A. Szafarz (1991), True Versus Spurious Instantaneous Causality, Working paper, GREMAQ and IDEI, Toulouse, France.

Roberts, J.M. (1993), The Sources of Business Cycles: A Monetarist Interpretation, International Economic Review, 34, 923–934.

Rothenberg, T.J. (1971), Identification in Parametric Models, Econometrica, 39, 577–592.

Rudebusch, G. (1998), Do Measures of Monetary Policy in a VAR Make Sense?, International Economic Review, 39, 907–948.

Saikkonen, P. (1991), Asymptotically Efficient Estimation of Cointegration Regressions, Econometric Theory, 7, 1–21.

Sargan, J.D. and A. Bhargava (1983), Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk, Econometrica, 51, 153–174.

Schmidt, P. (1976), Econometrics, New York: Marcel Decker.

Shapiro, M. and M.W. Watson (1988), Sources of Business Cycle Fluctuations, NBER Macroeconomcs Annual, 3, 111–156.

Shin, Y. (1994) A Residual-Based Test of the Null of Cointegration Against the Alternative of No Cointegration, Econometric Theory, 10, 91–115.

Sims, C.A. (1980), Macroeconomics and Reality, Econometrica, 48, 1–47.

Sims, C.A. (1981), An Autoregressive Index Model for the U.S. 1948-1975, in: J. Kmenta and J.B. Ramsey (eds.), Large-Scale Macro-Econometric Models, Amsterdam: North-Holland, 283-327.

Sims, C.A. (1986), Are Forecasting Models Usable for Policy Analysis?, Quarterly Review, Federal Reserve Bank of Minneapolis, Winter.

Sims, C.A. and T. Zha (1994), Error Bands for Impulse Responses, mimeo.

Sims, C.A. and T. Zha (1998), Bayesian Methods for Dynamic Multivariate Models, International Economic Review, 39, 949–968.

Sims, C.A., J.H. Stock, M.W. Watson (1990), Inference in Linear Time Series Models With Some Unit Roots, Econometrica, 58, 113–144.

Spanos, A. (1990), The Simultaneous-Equations Model Revisited: Statistical adequacy and identification, Journal of Econometrics, 44, 87-105.

Snell, A. (1998), Testing for r Versus $r-1$ Cointegrating Vectors, Journal of Econometrics, 88, 151–191.

Stock, J.H. (1994a), Deciding Between I(1) and I(0), Journal of Econometrics, 105–131.

Stock, J.H. (1994b), Unit Roots, Structural Breaks and Trends, in: R.F Engle and D.L. McFadden (eds.), Handbook of Econometrics, Vol. 4, Amsterdam: North Holland.

Stock, J.H. and M.W. Watson (1988), Testing for Common Trends, Journal of the American Statistical Association, 83, 1097-1107.

Stock, J.H. and M.W. Watson (1993), A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems, Econometrica, 61, 783–820.

Stram, D.O. and W.W.S. Wei (1986), Temporal Aggregation in the ARIMA Process, Journal of Time Series Analysis, 7, 279-292.

Swanson, N.R. und C.W.J. Granger (1997), Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions, Journal of the American Statistical Association, 92, 357–367.

Tanaka, K. (1990), Testing for a Moving Average Root, Econometric Theory 6, 433–444.

Theil, H. (1971), Principles of Econometrics, New York: John Wiley.

Tiao, G.C. (1972), Asymptotic Behaviour of Temporal Aggregates of Time Series, Biometrika, 59, 525–531.

Tiao, G.C. (1999), The ET Interview: Professor George C. Tiao, Econometric Theory, 15, 389–424.

Toda, H.Y. (1994), Finite Sample Properties of Likelihood Ratio Tests for Cointegrating Ranks When Linear Trends are Present, The Review of Economics and Statistics, , 66–79.

Toda, H.Y. and Phillips, P.C.B. (1993), Vector Autoregressions and Causality, Econometrica, 61, 1367–1394.

Toda, H.Y. and P.C.B. Phillips (1994), Vector Autoregression and Causality: A Theoretical Overview and a Simulation Study, Econometric Reviews, 13, 259–285.

Tsay, R.S. (1993), Testing for Noninvertible Models With Applications, Journal of Business & Economic Statistics, 11, 225–233.

Uhlig, H. (1997), What are the Effects of Monetary Policy? Results from an Agnostic Identification Procedure, CentER, Tilburg University, Juni 1997.

Vlaar, P.J.G. (1999) On the Asymptotic Distribution of Impulse Response Functions With Long Run Restrictions, Amsterdam, mimeo.

Vogelsang, T.J. (1998a), Trend Function Hypothesis Testing in the Presence of Serial Correlation, Econometrica, 66, 123–148.

Vogelsang, T.J. (1998b), Testing for the Shift in Mean Without Having to Estimate Serial-Correlation Parameters, Journal of Business & Economic Statistics, 16, 73–80.

Waggoner, D.F. and T. Zha (1997), Normalization, Probability Distribution, and Impulse Responses, Federal Reserve Bank of Atlanta, Working Paper 97-11.

Wallis, K.F. (1977), Multiple Time Series Analysis and the Final Form of Econometric Models, Econometrica, 45, 1481-1491.

Warne, A. (1990), Vector Autoregressions and Common Trends in Macro and Financial Economics, PhD thesis, Stockholm School of Economics.

Watson, M.W. (1994), Vector Autoregressions and Cointegration, in: Engle, R.F. and D.L. McFadden (eds.), Handbook of Econometrics, Vol. IV, New York: Elsevier.

8

Weber, C.E. (1995), Cyclical Output, Cyclical Unemployment, and Okun's Coefficient: A new approach, Journal of Applied Econometrics, 10, 433–445.

Wei, W.W.S. (1978), The Effect of Temporal Aggregation on Parameter Estimation in Distributed Lag Models, Journal of Econometrics, 8, 237-246.

Wei, W.W.S. (1981), Effect of Systematic Sampling on ARIMA Models, Communications in Statistical-Theoretical Mathematics A, 10, 2389-2398.

Weiss, A.A. (1984), Systematic Sampling and Temporal Aggregation in Time Series Models, Journal of Econometrics, 26, 271-281.

Wickens, M.R. (1982), The Efficient Estimation of Econometric Models with Rational Expectations, Review of Economic Studies, 49, 55–67.

Wickens, M.R. (1996), Interpreting Cointegrating Vectors and Common Stochastic Trends, Journal of Econometrics, 74, 255–271.

Wold, H. (1960), A Generalization of Causal Chain Models, Econometrica, 28, 443–463.

Wolters, J. (1995), On the Term Structure of Interest Rates – Empirical Results for Germany, Statistical Papers, 36, 193–214.

Working, H. (1960), Note on the Correlation of First Differences of Averages in a Random Chain, Econometrica, 26, 916–918.

Wright, P.G. (1915), review of H.L. Moore: Economic Cycles – Their Law and Cause, Quarterly Journal of Economics, 29, 631–641.

Yang, M. (1998), System Estimators of Cointegrating Matrix in Absence of Normalising Information, Journal of Econometrics, 2, 317-337.

Xiao, Z. and P.C.B. Phillips (1999), Efficient Detrending in Cointegrating Regression, Econometric Theory, 15, 519-548.

Zellner, A. (1971), An Introduction to Bayesian Inference in Econometrics, New York: John Wiley.

Zellner, A. (1992), Statistics, Science and Public Policy, Journal of the American Statistical Association, 87, 1–6.

Zellner, A. and C. Montmarquette (1971), A Study of Some Aspects of Temporal Aggregation Problems in Econometric Analyses, Review of Economics and Statistics, 335-342.

Zellner, A. and F. Palm (1974), Time Series Analysis and Simultaneous Equations Econometric Models, Journal of Econometrics, 2, 17–54.

Zellner, A. and H. Theil, H. (1962), Three–Stage Least-Squares: Simultaneous Estimation of Simultaneous Equations, Econometrica, 30, 54–78.