AN ECONOMETRIC ANALYSIS OF I(2) VARIABLES

Niels Haldrup

Aarhus University

Abstract. This paper provides a selective survey of the recent literature dealing with I(2) variables in economic time series, that is, processes that require to be differenced twice in order to become stationary. With reference to particular economic models intuition is provided of why I(2)-and polynomial cointegration are features likely to occur in economics. The properties of I(2) series are discussed and I review topics such as: Testing for double unit roots, representations of I(2) cointegrated systems, and hypothesis testing in single equations as well as in systems of equations. Different data sets are used to illustrate the various econometric and statistical techniques.

Keywords. I(2) processes; Unit root tests; VAR models; Error correction models.

1. Introduction

The past decade has witnessed a veritable explosion of research on the analysis and implications of non-stationary time series with unit roots. By now there seems to be general consensus in the literature of how such time series can be characterized, represented, tested, and analyzed, both within univariate and multivariate settings. Moreover, many economic theories provide good arguments why time series may have such non-stationary features. There exist numerous recent contributions in the literature that survey this literature. A list of references which is far from being exhaustive includes Diebold and Nerlove (1990), Dolado and Jenkinson (1990), Muscatelli and Hurn (1992), Banerjee *et al.* (1993), Hamilton (1994), Stock (1994), and Watson (1994).

Notwithstanding, an extended class of non-stationary time series models has been attached only a minor interest in these surveys, that is, processes with double unit roots. Perhaps this is not too surprising since processes with a single unit root seems best at describing the behaviour of most economic time series. However, some time series like prices, wages, money balances, stock-variables etc., appear to be more smooth and more slowly changing than what is normally observed for variables integrated of order one, I(1). Such time series are potentially integrated of order 2 such that double differencing is needed to make the series stationary. If the series are log-transformed, the differenced series (i.e. the growth rates) will therefore be I(1). Since both levels and growth rates of economic time series are important for many economic theories the complicated interaction between I(2) and I(1) series is thus of great importance in the econometric analysis of such

0950–0804/98/05 0595–56 JOURNAL OF ECONOMIC SURVEYS Vol. 12, No. 5

models. The purpose of this survey is to demonstrate why I(2) processes are a relevant class of models to consider in economics and I shall review different econometric tools that are presently available in the literature in the study of I(2) models.

In section 2 some introductory concepts are presented in order to discuss the cointegration possibilities in I(2) systems. The subsequent section presents various economic models justifying cointegration amongst doubly integrated time series and polynomial cointegration, that is, models where both the levels and the first differences of I(2) variables are needed in the equilibrium relations. In particular, I address models of intertemporal economic behaviour and stock-flow models of the inventory accelerator type. Following this motivation a selective survey of the econometric literature is given which includes: The basic properties of I(2) processes compared to I(1) processes and other classes of non-stationary time series models (section 4), testing for I(2)-ness (section 5), and single equation analysis in multivariate I(2) systems (section 6). In particular, section 6 addresses spurious regression problems in I(2) models and based on this analysis the natural extension is made to residual based testing for cointegration when I(2) and I(1) variables are present. The general properties of single equation cointegration models with stationary errors are characterized as well. In section 7 cointegration in systems is motivated from the triangular representation. Next, section 8 is dedicated to a review of the different ways cointegrated I(2) VAR systems can be represented. In practice most empirical analyses of I(2) systems are conducted within the context of the so-called Johansen maximum likelihood procedure so it is natural to emphasize this technique in particular. This is done in section 9. The next section briefly discusses other representation, estimation and testing procedures in multivariate systems and finally section 11 concludes. Throughout the paper the various techniques are demonstrated by empirical examples.

2. Introductory definitions and the cointegration possibilities amongst I(2) variables

Before discussing the different ways that I(2) variables may appear in economic models it will be useful to define formally what is meant by I(2) series and I discuss briefly, at an introductory level, how cointegration can occur amongst such series.

Following the initial benchmark description of Box and Jenkins (1970), a univariate variable x_t that needs to be differenced d times in order to have a stationary and invertible ARMA representation is said to be integrated of order d, ($x_t \sim \mathrm{I}(d)$). That is, an I(2) variable is one for which $\Delta^2 x_t$, can be given an ARMA representation with no unit roots in the AR or MA parts of the process. I use standard notation and hence let Δ and L signify difference and lag operators: $\Delta x_t = (1 - L)x_t = x_t - x_{t-1}$. Initially it is assumed that x_t is free of deterministic components. The two series displayed in the first graph of figure 1 are examples of I(2) series.

Following Engle and Granger (1987), we can consider a ($p \times 1$) vector time series x_t for which each component is integrated[1] of order d. More compactly we then write $\mathrm{x}_t \sim \mathsf{I}(d)$ and by construction this also implies that $\Delta \mathrm{x}_t \sim \mathsf{I}(d-1)$. Now the possibility exists that the levels of the series cointegrate, i.e. such that one can find a p-vector β satisfying that $\beta' \mathrm{x}_t \sim \mathsf{I}(d-b)$, $d \geqslant b > 0$, and where β is the so-called cointegration vector. In Engle and Granger's terminology we write $\mathrm{x}_t \sim \mathsf{CI}(d, b)$. Generally several linearly independent cointegration vectors may exist and in this case β becomes a matrix with rank less than p. How exactly the cointegration rank is given for systems where $d > 1$ is non-trivial (I return to this in sections 7 and 8) so to clarify the arguments in what follows it is assumed that β is just a vector.

In order to focus particularly on $\mathsf{I}(2)$ systems we define the vector series $\mathrm{x}_{2t} \sim \mathsf{I}(2)$ and $\mathrm{x}_{1t} \sim \mathsf{I}(1)$, so that really we have a combination of $\mathsf{I}(2)$ and $\mathsf{I}(1)$ variables. The processes have dimension m_2 and m_1, respectively. The cointegration possibilities that may occur for $\mathsf{I}(2)$ systems can now be described as the following three cases:

$$\beta_1' \mathrm{x}_{2t} \sim \mathsf{I}(0) \tag{1}$$

$$\beta_2' \mathrm{x}_{2t} \sim \mathsf{I}(1) \tag{2}$$

$$\beta_3' \mathrm{x}_{2t} + \beta_4' \mathrm{x}_{1t} + \beta_5' \Delta \mathrm{x}_{2t} \sim \mathsf{I}(0) \tag{3}$$

In the above expressions $\beta_i, i = 1, \ldots, 5$, are parameter vectors of matching dimensions. In (1) linear combinations exist between the $\mathsf{I}(2)$ variables such that $\mathrm{x}_{2t} \sim \mathsf{CI}(2, 2)$; hence, in this situation no $\mathsf{I}(1)$ variables are needed to induce stationarity amongst the $\mathsf{I}(2)$ series. It is interesting to observe that for this particular case the *timing* of the variables will be of importance. For instance, consider two cointegrated $\mathsf{I}(2)$ variables, $x_{21,t}$ and $x_{22,t}$, such that e.g. $x_{21,t} - x_{22,t}$ is stationary $\mathsf{I}(0)$. Instead, if $x_{22,t}$ is lagged by one period and we focus on the relation $x_{21,t} - x_{22,t-1} = (x_{21,t} - x_{22,t}) + \Delta x_{22,t}$ this is seen to be $\mathsf{I}(0) + \mathsf{I}(1)$ which in sum is $\mathsf{I}(1)$. The phenomenon that the timing of the variables will matter for the cointegration property to hold is something which does not apply to $\mathsf{I}(1)$ systems.

The second type of cointegration that may arise is when variables are $\mathsf{CI}(2, 1)$. This is the situation described in (2). If the relation $\beta_2' \mathrm{x}_{2t}$, does not cointegrate with other $\mathsf{I}(1)$ variables in the system, then the relation can be described alternatively as an $\mathsf{I}(0)$ relation in the *differenced* series since $\beta_2' \Delta \mathrm{x}_{2t}$ is stationary. If this is the only cointegration possibility present, then we can just model the series as an $\mathsf{I}(1)$ system for the growth rates (given that the original variables are in logs) and the standard $\mathsf{I}(1)$ analysis can be adopted.

The final cointegration possibility (3) is perhaps the most interesting because this is the one creating new insights compared to the $\mathsf{I}(1)$ model. In this case[2] $\beta_3' \mathrm{x}_{2t}$ is integrated of order one, ($\mathrm{x}_{2t} \sim \mathsf{CI}(2, 1)$), but with the further implication that there exist $\mathsf{I}(1)$ variables in the system cointegrating with $\beta_3' \mathrm{x}_{2t}$. This could be the variables x_{1t}, but as indicated in (3) it may also be necessary to include the differenced $\mathsf{I}(2)$ variables, $\Delta \mathrm{x}_{2t}$, which obviously will be integrated of order one as well. Sometimes the situation where first differences of the $\mathsf{I}(2)$ variables are included in cointegration relations is referred to as *polynomial* cointegration,

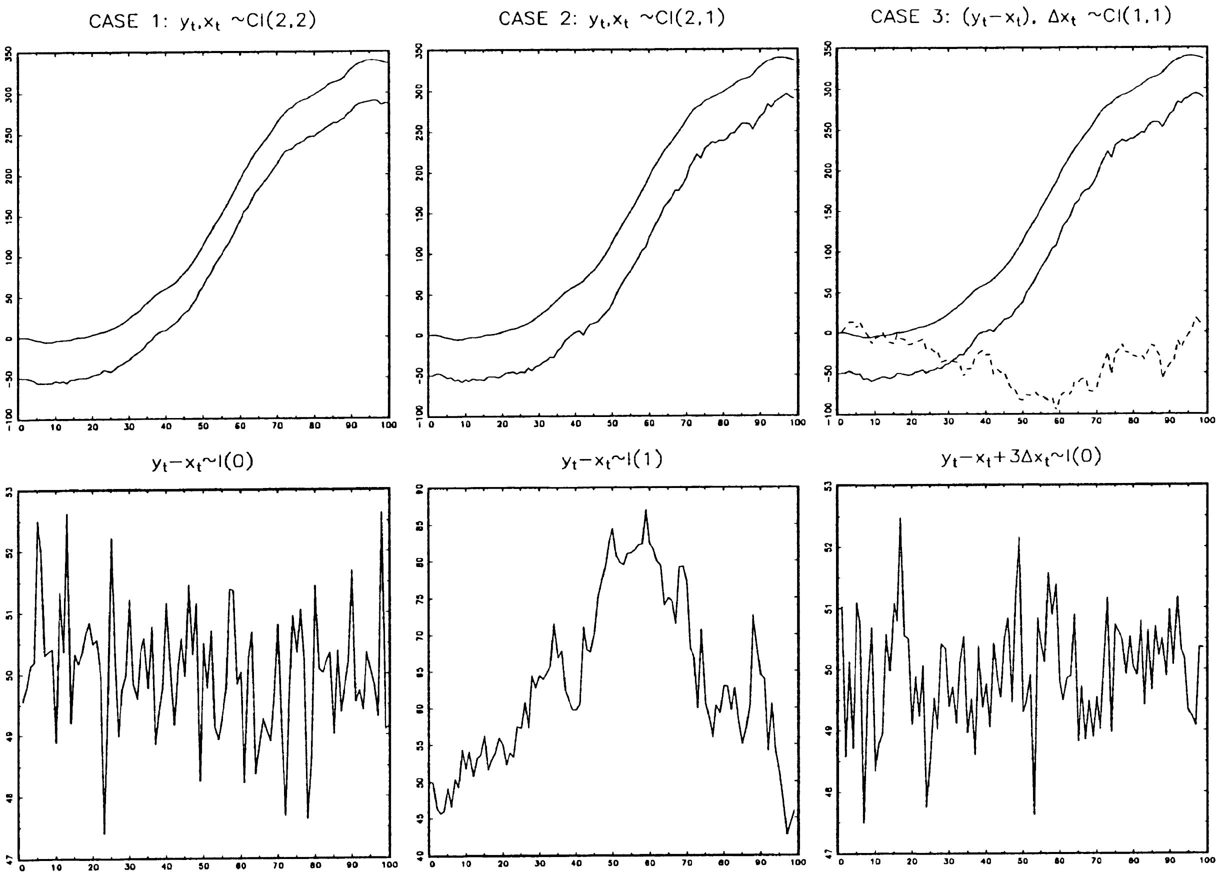

Figure 1. Examples of cointegrated I(2) series for $T=100$ observations. Case 1: $y_t, x_t \sim \mathrm{CI}(2,2)$. Case 2: $y_t, x_t \sim \mathrm{CI}(2,1)$. Case 3: $y_t, x_t \sim \mathrm{CI}(2,1)$, and (y_t-x_t), $\Delta x_t \sim \mathrm{CI}(1,1)$: *polynomial cointegration*. Note: Δx_t in case 3 has been scaled to ease comparison with y_t-x_t in the second panel of case 2.

because the cointegration relation can be expressed in terms of a polynomial in the lag operator.

In figure 1 I have simulated different bivariate examples of two cointegrated I(2) series corresponding to the 3 cases outlined above.

3. Some motivation for economic models with I(2) variables

Recently cointegration amongst I(2) series and polynomial cointegration has been analyzed in a growing number of empirical studies. As I shall demonstrate in section 4 a dominating empirical regularity is that many nominal variables, price indices, and stock variables have characteristics that mimic those of I(2) processes, and hence models involving such variables may potentially be (polynomially) cointegrated according to the possibilities sketched in the previous section. Examples include models of money demand relations, see *inter alia.* King *et al.* (1991), Hallman *et al.* (1991), Johansen (1992a), Juselius and Hargreaves (1992), Stock and Watson (1993), Juselius (1994, 1996, 1998), Rahbek *et al.* (1998) and Haldrup (1994), and models of the Purchasing Power Parity (PPP), see, *inter alia*, Johansen and Juselius (1992), Juselius (1992, 1995a,b). Banerjee *et al.* (1998) presents an interesting I(2) analysis for inflation and the markup. In many of these studies the purpose has been to test empirically whether unit elasticity and homogeneity restrictions could be satisfied by the data, for instance whether money demand should be specified in real or in nominal terms. Consider also the PPP; if p_{1t}, p_{2t} are the logarithms of domestic and foreign prices, respectively, and $e_{12,t}$ is the logarithm of the domestic exchange rate in domestic currency per unit of the foreign currency, then the theory predicts that $p_{2t} - p_{1t} - e_{12t} = 0$ is an equilibrium relation. When the price series are I(2) and the exchange rate is I(1), as is frequently found in empirical studies, the theory suggests cointegration amongst the variables in a particular way.

Polynomial cointegration is not only a statistical property that can sometimes be found in the data but can be justified in many cases on theoretical grounds. In the following I shall describe some models where agent's intertemporal behaviour will lead to equilibrium relations with polynomial cointegration.

3.1. *Polynomial cointegration in models of intertemporal behaviour*

Cagan's model of hyperinflation. Cagan's model of hyperinflation, see Cagan (1956), has been investigated in numerous empirical studies, see e.g. Sargent (1977), Goodfriend (1982), Taylor (1991), and Engsted (1993, 1994). The model is derived under the assumption of rational expectations and instantaneous market clearing in the money market and dictates that

$$m_t - p_t = \alpha - \delta(\mathsf{E}_t p_{t+1} - p_t) + u_t \tag{4}$$

where m_t and p_t are the natural logarithms of the stock of money and the price level, respectively, and u_t is a variable capturing velocity and demand shocks. E_t is the mathematical expectations operator with respect the information available at time t, and α and δ are parameters to be estimated. Cagan's model predicts that if real balances tend to rise it indicates expected reductions in future inflation.

There are different ways of writing the cointegration implications that follow from this model. First, it is easy to see that (4) can be rewritten as

$$\Delta p_{t+1} - \alpha\delta^{-1} + \delta^{-1}(m_t - p_t) - \delta^{-1}u_t = \varepsilon_{t+1} \tag{5}$$

where $\varepsilon_{t+1} = p_{t+1} - \mathsf{E}_t p_{t+1}$ is the rational expectations forecast error at time $t+1$ and hence is a martingale difference. If the properties of the time series are such that the demand/velocity shock variable u_t is stationary $\mathsf{I}(0)$ whilst real balances and inflation variables are $\mathsf{I}(1)$ then $((m_t - p_t), \Delta p_{t+1})' \sim \mathsf{CI}(1, 1)$. However, the theory also predicts then, that $(m_t, p_t) \sim \mathsf{CI}(2, 1)$ with a cointegration vector given by $(1, -1)$. Hence $(m_t, p_t, \Delta p_t)$ constitutes an $\mathsf{I}(2)$ cointegrated system with polynomial cointegration.

Linear Quadratic Adjustment Cost models. The costs of adjusting economic policy instruments provides another important justification for considering polynomial cointegration. A class of intertemporal models having this feature is the Linear Quadratic Adjustment Cost (LQAC) model which has been extensively used in studies of e.g. labour and money demand, and models of price adjustment, see *inter alia* Sargent (1978), Kennan (1979), Dolado *et al.* (1991), Pesaran (1991), Price (1992), Gregory *et al.* (1993), West (1995), and Engsted and Haldrup (1994, 1997). In this model economic agents are assumed to choose some decision variable, y_t, in order to minimize the conditional expectation of an intertemporal loss function which in its most simple form is given as

$$L_t = \sum_{i=0}^{\infty} \delta^i [y_{t+i} - y^*_{t-i})^2 + (y_{t+i} - y_{t+i-1})^2]. \tag{6}$$

The variable y_t^* is the desired level of the decision variable and costs are incurred by discrepancies between this level and the actual level.[3] However, since adjustments are penalized through the presence of the term $(y_{t+i} - y_{t+i-1})$ in the loss function there is no instantaneous reaction towards the equilibrium. The parameter θ measures the relative cost of the two cost terms and δ is the subjective discount rate of the agents.

The first order condition of the problem which minimizes the conditional expectation of the loss function (6) yields an Euler equation which can be reparametrized as

$$\Delta^2 y_{t+1} = (\delta^{-1} - 1)\Delta y_t + \frac{\theta}{\delta}(y_t - y_t^*) + u_{t+1} \tag{7}$$

where $u_{t+1} = y_{t+1} - \mathsf{E}_t y_{t+1}$ is again an $\mathsf{I}(0)$ rational expectations prediction error. Equation (7) is deliberately formulated such that it facilitates analysis when y_t is an $\mathsf{I}(2)$ variable. In this situation, unless we are in the degenerate case with no discounting, $\delta = 1$, y_t and its target y_t^* will cointegrate into an $\mathsf{I}(1)$ relation which further cointegrates with Δy_t; hence y_t and the variables determining y_t^* are polynomially cointegrated. An important feature of this model is that due to adjustment costs the gap between y_t and y_t^* will not be closed completely.

The above examples of intertemporal models show that if certain variables can be considered $\mathsf{I}(2)$, then theory predicts variables to be polynomially cointegrated. It should be emphasized, however, that cointegration testing *per se* is *not* sufficient for rejection or non-rejection of this sort of economic theories. Theoretical models (and in particular intertemporal models based on Euler equations) provide many more (overidentifying) restrictions that need to be tested against the data; cointegration only delivers a necessary condition to be fullfilled for such models to be valid.[4]

3.2. *Multicointegration as defined by Granger and Lee (1989, 1990)*

It is normally the case that if one considers just two $\mathsf{I}(1)$ series, then only a single cointegration relation can exist between the variables. However, in some situations it may happen that more than one cointegration vector exists amongst the series because the cumulated equilibrium errors cointegrate with the basic economic variables. This property refers to *multicointegration*. As we shall see shortly it may be advisable in this situation to formulate the system in terms of an $\mathsf{I}(2)$ system which will facilitate the analysis although the original economic variables are really $\mathsf{I}(1)$.

The notion of multicointegration was initially introduced and defined by Granger and Lee (1989, 1990) and is also considered in Salmon (1988). They consider the case with a system of two $\mathsf{I}(1)$ variables, production, y_t, and sales, x_t, of some commodity and assume that these variables are cointegrated, $(y_t, x_t)' \sim \mathsf{CI}(1, 1)$, with a cointegrating vector implying that $y_t - x_t$ is stationary. The variable $y_t - x_t$ has a nice interpretation as this is just the change of inventories, ΔI_t, and, assuming no depreciation, the inventory stock is then given as $I_t = \Delta^{-1}(y_t - x_t) = \sum_{i=1}^{t} \Delta I_i$, i.e. the cumulated equilibrium errors of the equilibrium relation between y_t and x_t. Naturally the inventory stock is $\mathsf{I}(1)$ by construction and hence it can happen that this variable cointegrates with y_t and/or x_t such that

$$I_t - \gamma_{11} y_t - \gamma_{12} x_t \sim \mathsf{I}(0). \tag{8}$$

This is the condition for multicointegration amongst just two variables.

In the literature polynomial cointegration, and multicointegration are frequently used as synonyms, simply because multicointegrated series can be rewritten in terms of polynomial cointegration amongst $\mathsf{I}(2)$ variables including their differences. Note that (8) can be written in the alternative way $\Delta^{-1} y_t - \Delta^{-1} x_t - \gamma_{11} y_t - \gamma_{12} x_t$ where e.g. $\Delta^{-1} y_t = \sum_{i=0}^{t} y_i \equiv Y_t$ and capital letters indicate cumulated series which then are $\mathsf{I}(2)$. In terms of the new series we can rewrite (8) as

$$Y_t - X_t - \gamma_{11} \Delta Y_t - \gamma_{12} \Delta X_t \sim \mathsf{I}(0), \tag{9}$$

and hence this indicates polynomial cointegration in the variables[5] (Y_t, X_t).

There are numerous examples of economic models that imply multi-cointegration[6] and in particular models which describe the interaction amongst stock and flow variables. Granger and Lee's inventory model is one such example, but one could also consider, for instance, a model of stock-adjustment in housing construction where y_t is housing units started, x_t is housing units completed, and $I_t = \sum_{i=0}^{t}(y_i - x_i)$ is the stock of housing units under construction, see Lee (1992, 1996) and Engsted and Haldrup (1998). Similarly, according to the life-cycle hypothesis of income another example is when y_t is income, x_t is consumption, and I_t measures wealth (cumulated savings), see e.g. Lee (1996), and Hendry and von-Ungern Sternberg (1981). In these papers many examples are given supporting that multicointegration is an empirically relevant concept in stock-flow models.

In the above examples it could seem as if the first level cointegration vector is always trivially given as $(1, -1)$. This is not the case, however. Sometimes it is an empirical problem to measure for instance durable goods and the corresponding service flows, see e.g. Campbell (1987). Similarly, in the housing example some housing starts may never be completed and hence the relationship between y_t and x_t is not given beforehand as the difference between the variables, but rather as the quasi-difference $y_t - \kappa x_t$ where κ is a parameter which needs to be estimated.

An interesting property of the multicointegration concept is that the implied decision rules that correspond to generalized error correction models can be derived from LQAC type of models with *proportional*, *integral*, *and derivative* (*PID*) control mechanisms, see e.g. Phillips (1954), Holt *et al.* (1960), and Hendry and von-Ungern Sternberg (1981). Granger and Lee (1990) provide this formal connection, see also Engsted and Haldrup (1998).

4. A comparison of some properties of univariate processes

In this section I describe some of the univariate features characterizing I(2) processes. To facilitate comparison with other time series models, possibly with maintained deterministic components, we consider the general data generating process (DGP)

$$x_t = \gamma' \mathrm{c}_t + x_t^0, \qquad t = 1, 2, \ldots T. \tag{10}$$

The deterministic part of the time series is given by $\gamma' \mathrm{c}_t$ whereas x_t^0 contains the stochastic part. For instance, c_t can be a polynomial trend vector $\mathrm{c}_t = (1, t, t^2, \ldots, t^k)'$ such that $\gamma' \mathrm{c}_t = \sum_{i=0}^{k} \gamma_i t^i$ is a general trend polynomial. In practice $\mathrm{c}_t = (1, t, t^2)'$ is most common as the specification of the trend part, however, other (even non-polynomial) characterizations of the deterministic part may be considered.

Asymptotic behaviour of the processes. It is of interest to compare some of the asymptotic properties of I(0), I(1), and I(2) series, viewed as possible DGP's, and their relation to various sorts of deterministic components. To ease the comparison the following (stochastic) processes are addressed:

$$x_{0t}^0 = \varepsilon_t, \qquad \Delta x_{1t}^0 = \varepsilon_t, \qquad \Delta^2 x_{2t}^0 = \varepsilon_t \tag{11}$$

Notice that subscripts $i = 0, 1, 2$ indicate the order of integration of the series.

The assumptions I shall make about the errors driving the processes are quite general and certainly need not be just martingale difference sequences. In fact, the errors are only needed to be $\mathsf{I}(0)$ with a dependency that is not too long in memory. The regularity conditions of the errors are rather familiar in the literature and are discussed in detail by e.g. Phillips (1987).

For later reference we need to define $\sigma^2 = \lim_{T\to\infty} \mathsf{E}(T^{-1}(\sum_{i=1}^{T} \varepsilon_i)^2)$ which is frequently denoted the 'long run variance'. This can be written in the alternative way: $\sigma^2 = \sigma_\varepsilon^2 + 2\lambda$, where $\sigma_\varepsilon^2 = \mathsf{E}(\varepsilon_1^2)$ and $\lambda = \sum_{j=2}^{\infty} \mathsf{E}(\varepsilon_1 \varepsilon_j)$.

One of the basic properties of $\mathsf{I}(0)$ series is that certain functions of the sample values will converge to constants as the number of observations tends to infinity. For instance laws of large numbers guarantee that the sample mean converges in probability to its true mean for a wide range of processes, including stationary processes, see e.g. White (1984). However, one of the basic features of $\mathsf{I}(1)$ and $\mathsf{I}(2)$ processes is that this sort of convergence theorems will fail to hold. Instead, sample moments, suitably normalized, will converge (weakly) to random variables rather than to constants.

I(1) Processes. Initially, consider the $\mathsf{I}(1)$ series x_{1t}^0 and assume for convenience a zero initial condition, $x_{1,0}^0 = 0$. In levels the series can thus be written as the stochastic trend

$$x_{1t}^0 = \sum_{j=1}^{t} \varepsilon_j = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t. \tag{12}$$

This demonstrates an important feature of $\mathsf{I}(1)$ series, namely, that a shock to the series occurring in the past will persist and hence will have an everlasting influence on the levels of the series.

Under weak regularity conditions the limiting behaviour of x_{1t}^0, after appropriate scaling, can be shown to be a Brownian motion. More specifically it holds that

$$T^{-1/2} x_{1,[Tr]}^0 = T^{-1/2} \sum_{j=1}^{[Tr]} \varepsilon_j + o_p(1) \Rightarrow B(r) \equiv \sigma W(r) \tag{13}$$

where '$\Rightarrow$' signifies weak convergence and $[\cdot]$ selects the integer value of its argument. $B(r)$ is a Brownian motion defined on the unit interval $r \in [0, 1]$ with long run variance σ^2. $W(r)$ is a standard Brownian motion.

In calculating the limiting behaviour of the sample mean and sample variance of x_{1t}^0 it can be shown that $\bar{x}_1^0 = O_p(T^{1/2})$ and $\widehat{\mathrm{Var}}(x_{1t}^0) = O_p(T)$. Hence the process will be without bound asymptotically and the empirical variance will increase with the sample size.

I(2) Processes. Turning to I(2) processes it is easy to see that by assuming $x^0_{2,-1} = x^0_{2,0} = 0$, the series can be written in levels as

$$x^0_{2t} = \sum_{k=1}^{t}\sum_{j=1}^{k} \varepsilon_j = t\varepsilon_1 + (t-1)\varepsilon_2 + \cdots + 3\varepsilon_{t-2} + 2\varepsilon_{t-1} + \varepsilon_t. \tag{14}$$

This is also a stochastic trend but it is better to denote it an I(2) trend to discriminate it from usual I(1) trends. It is obvious that since an I(2) trend is a double sum of errors such series will be very smooth by nature. Notice that Δx^0_{2t} is just the I(1) trend given in (12). As seen the influence from a shock that occurred in the past will have an amplifying impact on the levels of the series as the time goes. This may seem as an odd property and researchers have argued that this feature makes I(2) processes an inconvenient class of models to consider in reality, see e.g. Granger (1997, p. 173). However, one way of looking at the amplified persistency phenomenon is to consider a shock to the growth rate of an already persistent time series, e.g. inflation; of course, this will have an accumulating effect on the levels of the series since this is just how the price level, for instance, is defined.

It is important to make a brief side remark on explosive processes, i.e. processes with roots outside the unit circle. Such processes have characteristics that mimic I(2) processes, at least for sample sizes of moderate length. A simple example of an explosive process is $x^0_{\exp,t} = \alpha x^0_{\exp,t-1} + \varepsilon_t$, with $|\alpha| > 1$. With a zero initial condition the series can be given the levels representation

$$x^0_{exp,t} = \sum_{j=1}^{j} \alpha^{t-j}\varepsilon_j = \alpha^{t-1}\varepsilon_1 + \alpha^{t-2}\varepsilon_2 + \cdots + \alpha\varepsilon_{t-1} + \varepsilon_t. \tag{15}$$

Since $|\alpha| > 1$ a shock in the past will have an increasing effect on the series as t increases, just like I(2) processes. However, whilst the impact of shocks increases exponentially in (15) the rate is the slower linear rate for I(2) series as shown in (14). Asymptotically it is easy to discriminate I(2) series from explosive processes, but in finite samples it can be difficult, especially if α is not too large.

Returning to I(2) series, I proceed by describing their asymptotic behaviour. By suitable scaling it can be shown that the limiting process of an I(2) variable is an integrated Brownian motion. More precisely we have that

$$T^{-3/2}x^0_{2t} = T^{-3/2}\sum_{s=1}^{[Tr]}\sum_{j=1}^{[Ts]} \varepsilon_j + o_p(1) \Rightarrow \int_0^r B(s)\,\mathrm{d}s \equiv \sigma\int_0^r W(s)\,\mathrm{d}s. \tag{16}$$

The easier notation $\overline{W}(r) = \int_0^r W(s)\mathrm{d}s$ is used to indicate an integrated Brownian motion. For I(2) processes the limits of the sample mean and variance imply that $\bar{x}^0_1 = O_p(T^{3/2})$ and $\widehat{\mathrm{Var}}(x^0_{1t}) = O_p(T^3)$. Hence the scaling of these quantities needs to be increased in comparing I(1) and I(2) processes.

Deterministic Components. Deterministics are important because they appear in most time series in one form or another, for instance as a trend polynomial $\gamma' \mathrm{c}_t$ as in (10). If not taking proper account of such components it may flaw and possibly invalidate the statistical analysis. Hence some insight of their behaviour is necessary. In Eq. (10), consider a single component of c_t, which I denote c_{it}. Since the Brownian motions previously discussed are defined on the unit interval $r \in [0, 1[$ we may want a scaling of the deterministics which facilitates comparison. Obviously, we can scale c_{it} and define $f_{Ti}(r)$ such that

$$f_{Ti}(r) = \frac{c_{it}}{T^i} = \frac{[Tr]^i}{T^i} + o(1) \longrightarrow r^i \text{ for } T \longrightarrow \infty$$

where $i = 0, 1, 2, \ldots$ In particular it is seen that a linear and a quadratic trend require normalizations T and T^2, respectively, for $f_{Ti}(r)$ to be bounded.

A time series x_t like in Eq. (10) is generally a mixture of deterministic and stochastic components; the latter possibly with components of different integration orders. With respect to the asymptotic behaviour of x_t it is therefore the component of the largest order which dominates the series in the limit. In finite samples the relative influence of the components will naturally depend on the parameters as well as the signal to noise ratio. If we let ' > ' signify 'dominance in variation', the following ranking can be made:

$$t^2 > \mathsf{I}(2) > t > \mathsf{I}(1) > \mathsf{I}(0). \tag{17}$$

This ordering becomes especially important with respect to hypothesis testing.

5. Determining the order of differencing

In the univariate time series literature following the Box and Jenkins tradition it is commonplace to difference the series an appropriate number of times until, by visual inspection, the autocorrelations of the transformed series exhibit fast decay. The idea about unit root testing is to use a more formal statistical testing procedure to identify the number of differences needed to render the time series stationary. However, in doing so it is quite important to be aware of the alternative hypothesis under consideration, i.e. whether I(2) is tested against I(1), I(0), or explosive processes, possibly with maintained deterministic components. A large number of tests for double unit roots exist in the literature, but their properties differ depending upon the actual alternative that power is wanted against.

In the following some of these tests are described. The discussion is separated into two types of tests: parametric tests which assume that the series can be given a finite order AR representation, and the semi-parametric equivalents of the tests which permit a more general class of processes. To simplify the discussion I assume that I(2) is the largest order of integration to consider; to my knowledge there are no examples of economic time series which have been found to be I(3) or of higher order.

5.1. *Parametric approaches to testing for I(2)*

Initially I shall assume that the time series x_t is free of deterministic components and can be given the finite order $\mathsf{AR}(p)$ representation

$$A(L)x_t = u_t \tag{18}$$

where u_t is a sequence of iid errors.

A joint test for double unit roots. $A(L)$ can be factorized as $A(L) = A(1)L + A^*(1)\Delta L + A^{**}(L)\Delta^2$ where $A^{**}(L)$ is a polynomial of order $p-2$ with $A(0) = 1$. By defining $A(1) = 1 - \alpha_1$ and $A^*(1) = 1 - \alpha_2$ the regression equivalent of (18) can be written in the generalized augmented Dickey-Fuller form

$$\Delta^2 x_t = (\hat{\alpha}_1 - 1)x_{t-1} + (\hat{\alpha}_2 - 1)\Delta x_{t-1} + \sum_{j=1}^{p-2} \hat{\varphi}_j \Delta^2 x_{t-j} + \hat{u}_t. \tag{19}$$

Hasza (1977) and Hasza and Fuller (1979) were the first to suggest using this auxiliary regression as the basis to test for double unit roots. They used the fact that when x_t is $\mathsf{I}(2)$, $\alpha_1 = \alpha_2 = 1$, and hence it is natural to test this joint hypothesis by a standard F-test, for instance. Observe that since the F-test is two-sided, the alternative hypothesis is quite general as it covers all the situations where x_t is either explosive, $\mathsf{I}(0)$, or $\mathsf{I}(1)$. In the paper by Hasza and Fuller the non-standard limiting distribution of the F-test statistic is derived; by using the notation previously given, the distribution reads

$$F_{\alpha_1, \alpha_2} \Rightarrow \frac{1}{2} \int_0^1 dWG(r)' \left(\int_0^1 G(r)G(r)'dr \right)^{-1} \int_0^1 G(r)dW \tag{20}$$

where $G(r) = (W(r), \overline{W}(r))'$.

Empirical critical values of this distribution are reported by Hasza and Fuller. As it is the case for the Dickey-Fuller test for integration of order 1, it is important that the correct augmentation is used in the regression (19): A too low order implies that the errors will not be iid and hence gives rise to size distortion of the test. On the other hand, including too many lags of the lagged second differences will give power loss. This trade-off is a difficult compromise in practical situations, especially when the errors contain moving average components.

An interesting property which follows from estimation of the parameters in (19) is that, given the true process is $\mathsf{I}(2)$, the parameters α_1 and α_2 converge to their true value 1 very rapidly. In fact, it can be shown that

$$T^2(\hat{\alpha}_1 - 1) = \mathrm{O}_\mathrm{p}(1) \tag{21}$$

$$T(\hat{\alpha}_2 - 1) = \mathrm{O}_\mathrm{p}(1). \tag{22}$$

So, the parameter α_1 is estimated at a more rapid rate than the usual super-consistent rate $O_p(T^{-1})$ applying for $\mathsf{I}(1)$ systems.

Despite the fast convergence rates of the least squares estimates the use of single t-tests of the hypotheses H_0: $\alpha_1 = 1$, and H_0: $\alpha_2 = 1$ based on the regression (19) is of little use in practice when testing for $\mathsf{I}(2)$. For instance, one may consider a sequential testing procedure where one first tests H_0: $\alpha_1 = 1$ and, given acceptance, proceeds to testing for a second unit root, H_0: $\alpha_2 = 1$. But this procedure is invalid as the first test will depend upon whether the second hypothesis is actually true or false. Observe that both when a single and double unit root exist in the process it will hold that $\alpha_1 = 1$. However, in the case of an I(2) process ($\alpha_2 = 1$) the distribution of $t_{\alpha 1}$ is given by a complicated functional of the stochastic integrals $W(r)$, and $\overline{W}(r)$, whereas when x_t is I(1)($|\alpha_2| < 1$), $t_{\alpha 1}$ will follow the usual Dickey-Fuller distribution,[7] see Fuller (1976), Dickey and Fuller (1979) and Phillips (1987):

$$t_{\alpha_1} \Rightarrow \left(\int_0^1 W(r)\mathrm{d}W\right)\left(\int_0^1 W^2(r)\mathrm{d}W\right)^{-1/2}. \tag{23}$$

Hence the t_{α_1} test of H_0: $\alpha_1 = 1$ is not invariant to the actual value of α_2. Dickey and Pantula (1987) conducted a simulation study to examine the empirical relevance of the size distortions that result from using the Dickey-Fuller test in the first step of the sequential procedure sketched above. They found that for a sample of length 50 the Dickey-Fuller test has an actual size against the stationary alternative which exceeds the nominal 5% level when, in fact, two unit roots are present. In other words, it is more likely that the sequential procedure gives the misleading inference that the series is stationary when the process is I(2) compared to when it is I(1); intuition says that we should strongly reject stationarity when two unit roots are present rather than just a single unit root.

A consistent sequential testing procedure for the number of unit roots. As an alternative strategy to the one outlined above Dickey and Pantula (1987) suggest to reverse the sequence of testing by starting with the highest possible integration order and then testing down the integration order of the model.[8] By using this sequence of testing the hypotheses a consistent α-level of the procedure is obtained. The test procedure is conditional[9] in the sense that first I(2)-ness is tested given at least the series is I(1) and, provided rejection, I(1)-ness is subsequently tested against the stationary I(0) alternative. In the first step of the procedure a unit root is present both under the null and the alternative hypothesis so the appropriate auxiliary regression to focus on is (19) with the restriction $\alpha_1 = 1$ imposed. Hence the auxiliary regression reads

$$\Delta^2 x_t = (\hat{\alpha}_2 - 1)\Delta x_{t-1} + \sum_{j=1}^{p-2} \hat{\varphi}_j \Delta^2 x_{t-j} + \hat{u}_t \tag{24}$$

where the t-ratio associated with the regressor Δx_{t-1} is used to test H_0: $\alpha_2 = 1$. Since this is nothing else than a standard I(1) problem for the first differenced data, the test statistic follows the usual Dickey-Fuller distribution[10] (23) which is tabulated in e.g. Fuller (1976) and McKinnon (1991). When the null hypothesis is rejected, I(1) is tested against I(0) using standard procedures, for instance a Dickey-Fuller test on the levels of the series. Of course this is equivalent to testing H_0: $\alpha_1 = 1$ in (19).

Dickey and Pantula's procedure seems to be the most dominant procedure in applied work and obviously has the advantage that in each step the null distributions is given by the well-known Dickey-Fuller distribution and with a size which is controllable by choice of the significance level. In terms of power against the I(1) alternative, Dickey and Pantula demonstrate that the sequential procedure is preferable compared to the joint test of Hasza and Fuller, because the imposition of a unit root in the first step will be correct, both under the null and the alternative hypothesis. They also argued that some power is gained as the joint test is two-sided in nature, whereas the alternative is one-sided by using the sequential t-test procedure.

Notwithstanding, it is important to note that if power is wanted against the *explosive* alternative the Dickey-Pantula testing strategy is inappropriate. As indicated in section (4) explosive processes are a relevant class of models to consider as their properties (at least in finite samples) mimic those of I(2) series. If the true underlying process is explosive, the prior imposition of a unit root in the first stage of Dickey and Pantula's procedure may give the misleading inference that the series is really I(2) because, in principle, an explosive process can be differenced infinitely many times without becoming stationary. In addition, wrongly differencing an explosive process may produce a noninvertible error term which may cause serious problems with respect to estimation and inference. A Monte Carlo study reported in Haldrup (1994a) shows that when $\alpha_1 > 1$ the power following from the Dickey-Pantula procedure tends rapidly to zero as α_1 and the sample size T increase.

If power against the explosive alternative is wanted the Hasza-Fuller I(2)-test should be used. When this test rejects the series can be either I(1), I(0) or explosive. Subsequently the usual Dickey-Fuller t-test for the I(1)-null can be performed on the levels of the series but with a two-sided alternative: the lower tail of the Dickey-Fuller distribution indicates rejection of I(1) in favour of the stationary alternative; the right tail of the distribution favours the explosive alternative. In most applications only the one-sided stationary alternative is considered but in the present case the upper tail is naturally of interest as well.

A symmetrized joint test for double unit roots. A different class of (non-sequential) tests is the so-called symmetric tests which, in an I(2) setting, were initially suggested by Sen (1986) and Sen and Dickey (1987). The test is a symmetric version of Hasza and Fuller's (1979) joint F-test. The motivation arises from the interesting property that if the difference equation defining the time series is given by

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + u_t \tag{25}$$

where u_t is white noise with variance σ_u^2, then the series with the same difference equation equal to

$$x_t = \phi_1 x_{t+1} + \phi_2 x_{t+2} + \cdots + \phi_p x_{t+p} + v_t \tag{26}$$

will also have white noise errors v_t with the same error variance as for u_t, see

Fuller (1976). The basic idea is thus to jointly estimate (25) and (26), in a symmetric fashion, and use it to test for double unit roots.

The symmetrized version of the Hasza-Fuller regression model (19) is given by the pair of regression equations

$$
\begin{aligned}
\Delta^2 x_t &= (\hat{\alpha}_1 - 1)x_{t-1} + (\hat{\alpha}_2 - 1)\Delta x_{t-1} + \sum_{j=1}^{p-2} \hat{\varphi}_j \Delta^2 x_{t-j} + \hat{u}_t,\ t = p+1, \ldots, n \\
\Delta^2 x_t &= (\hat{\alpha}_1 - 1)x_{t-1} - (\hat{\alpha}_2 - 1)\Delta x_t + \sum_{j=1}^{p-2} \hat{\varphi}_j \Delta^2 x_{t+j} + \hat{v}_t,\ t = 3, \ldots, n-p+2.
\end{aligned}
\tag{27}
$$

In table 1 the regressand and the p columns of the independent variables are completely characterized. The above system corresponds to a SURE regression model with allowance made for the constraint that parameters are the same and hence this will provide more efficient estimates of the parameters of interest, α_1 and α_2.

From this 'extended' Hasza-Fuller regression the F-test of the hypothesis H_0: $\alpha_1 = \alpha_2 = 1$ can be constructed. Notice that using just the first $n-p-1$ rows of Table 1 (i.e. the first panel) corresponds to the usual Hasza-Fuller regression. Due to the symmetry of the regression model, the limiting distribution of the F-test statistic is somewhat simplified as cross terms appear to cancel. The distribution is still non-standard, of course. More specifically it reads

$$
F^{Sym}_{\alpha_1, \alpha_2} \Longrightarrow \left(\int_0^1 \overline{W}^2(r)\mathrm{d}r\right)^{-1} \left(\int_0^1 \overline{W}(r)\mathrm{d}W\right)^2 + \frac{1}{4}\left(\int_0^1 W^2(r)\mathrm{d}r\right)^{-1}. \tag{28}
$$

The empirical distributions are reported in Sen and Dickey's article.

The purpose of the symmetrized F-test is to gain power and simulations reported by Sen and Dickey support this intuition. They find that for a sample size of either 50 or 100 a wide range of parameters under the stationary alternative

Table 1. Regressand and regressors in the Sen and Dickey (1987) symmetric test.

Dependent variable	Independent variables				
$\Delta^2 x_{p+1}$	x_p	Δx_p	$\Delta^2 x_p$	$\cdots$	$\Delta^2 x_3$
$\Delta^2 x_{p+2}$	x_{p+1}	Δx_{p+1}	$\Delta^2 x_{p+1}$	$\cdots$	$\Delta^2 x_4$
$\Delta^2 x_{p+3}$	x_{p+2}	Δx_{p+2}	$\Delta^2 x_{p+2}$	$\cdots$	$\Delta^2 x_5$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
$\Delta^2 x_n$	x_{n-1}	Δx_{n-1}	$\Delta^2 x_{n-1}$	$\cdots$	$\Delta^2 x_{n-p+2}$
$\Delta^2 x_{n-p+2}$	x_{n-p+1}	$-\Delta x_{n-p+2}$	$\Delta^2 x_{n-p+3}$	$\cdots$	$\Delta^2 x_n$
$\Delta^2 x_{n-p+1}$	x_{n-p}	$-\Delta x_{n-p+1}$	$\Delta^2 x_{n-p+2}$	$\cdots$	$\Delta^2 x_{n-1}$
$\Delta^2 x_{n-p}$	x_{n-p-1}	$-\Delta x_{n-p}$	$\Delta^2 x_{n-p+1}$	$\cdots$	$\Delta^2 x_{n-2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
$\Delta^2 x_3$	x_2	$-\Delta x_3$	$\Delta^2 x_4$	$\cdots$	$\Delta^2 x_{p+1}$

leads to considerable power increases compared to the Hasza-Fuller non-symmetric test. For explosive alternatives (for a sample of 50 observations) power is found to be larger as well, but their study is more limited in this respect.

Obtaining power against deterministic trend alternatives. So far little has been said about the appropriate treatment of deterministic components. Notwithstanding, it is of crucial importance that deterministics be adequately dealt with since otherwise very misleading inference may result. A few examples will demonstrate this.

The difficulties are well-known from the I(1) analysis. Consider a trend stationary process $x_t = \gamma_0 + \gamma_1 t + x_t^0$ where $x_t^0 \sim \mathrm{I}(0)$ and $\gamma_1 \neq 0$. In testing for unit roots one may consider a Dickey-Fuller test; assume the analyst wrongly conducts the regression

$$\Delta x_t = \hat{\mu} + (\hat{\alpha} - 1)x_{t-1} + \sum_{j=1}^{p-1} \hat{\varphi}_j \Delta x_{t-j} + \hat{u}_t \tag{29}$$

without a trend. In this case it can be shown that although $|\alpha| < 1$, $\mathrm{plim}_{T\to\infty}\hat{\alpha} = 1$ (unit root) and $\mathrm{plim}_{T\to\infty}\hat{\mu} = \gamma_1$ (drift). The explanation is that the regression (29) does not accomodate a trend stationary process under the alternative. Hence the only way the trending feature of the process can be captured is to bias $\hat{\alpha}$ towards unity and letting $\hat{\mu}$ become an estimate of the trend slope, γ_1.

The problem carries over to the I(2) analysis. For instance, if the data generating mechanism is given by either of the processes

$$\begin{aligned} x_t &= \gamma_0 + \gamma_1 t + \gamma_2 t^2 + x_t^0, && \text{with } \gamma_2 \neq 0,\ x_t^0 \sim \mathrm{I}(0), \\ x_t &= \gamma_0 + \gamma_1 t + x_t^0, && \text{with } \gamma_1 \neq 0,\ x_t^0 \sim \mathrm{I}(1), \end{aligned} \tag{30}$$

then in both cases x_t is dominated by a deterministic trend. Consider the Hasza Fuller regression (19), possibly with an intercept added, but no linear or quadratic trend. In this situation it can be shown that for both processes sketched in (30) $\mathrm{plim}_{T\to\infty}\hat{\alpha}_1 = \mathrm{plim}_{T\to\infty}\hat{\alpha}_2 = 1$ and hence the test has zero asymptotic power and therefore erroneously indicates the series to be I(2). One way of getting power against the processes in (30) is to include also a quadratic trend in the auxiliary regression since in both cases the series have a quadratic trend when written in levels.

Essentially detrending can be undertaken in two different, though equivalent, ways. In accordance with the first route, the following Hasza-Fuller regressions may be conducted:

$$\Delta^2 x_t = \hat{m}_0 + (\hat{\alpha}_1 - 1)x_{t-1} + (\hat{\alpha}_2 - 1)\Delta x_{t-1} + \sum_{j=1}^{p-2} \hat{\varphi}_j \Delta^2 x_{t-j} + \hat{u}_t \tag{31}$$

$$\Delta^2 x_t = \tilde{m}_0 + \tilde{m}_t t + (\tilde{\alpha}_1 - 1)x_{t-1} + (\tilde{\alpha}_2 - 1)\Delta x_{t-1} + \sum_{j=1}^{p-2} \tilde{\varphi}_j \Delta^2 x_{t-j} + \tilde{u}_t \tag{32}$$

$$\Delta^2 x_t = \bar{m}_0 + \bar{m}_1 t + \bar{m}_2 t^2 + (\bar{\alpha}_1 - 1) x_{t-1} + (\bar{\alpha}_2 - 1) \Delta x_{t-1} + \sum_{j=1}^{p-2} \bar{\varphi}_j \Delta^2 x_{t-j} + \bar{u}_t. \tag{33}$$

In (33) the test of H_0: $\alpha_1 = \alpha_2 = 1$ has power against alternatives where the true series is $\mathsf{I}(0)$ + quadratic trend or $\mathsf{I}(1)$ + linear trend. Notice that in both cases the series in levels will have a quadratic trend and this is exactly why this regressor should be included in the regression. Turning to (32), the test based on this regression has power against $\mathsf{I}(0)$ + linear trend and $\mathsf{I}(1)$ + drift processes because in these cases the dominant feature of the series in levels is the linear trend. Finally (31) should be used when the series is suspected to have a non-zero mean.

It is important to realize that the reason for including a linear and quadratic trend in (33) is not because a cubic trend or a trend of order $O(T^4)$ is believed to characterize the series when two unit roots are present, but because the appropriate alternative should be accommodated to ensure a consistent test.

The above reflections apply equally to the symmetric test of Sen and Dickey and of course also to the sequential procedure of Dickey and Pantula. However, when using the sequential procedure the first step deterministic regressors are of order one less compared to the Hasza-Fuller procedure, since the assumption of one unit root increases the order of all deterministic regressors.

The second route to pursue in detrending time series is to consider

$$\begin{aligned} x_t &= \hat{\gamma}_0 + \hat{x}_t^0 \\ x_t &= \tilde{\gamma}_0 + \tilde{\gamma}_1 t + \tilde{x}_t^0 \\ x_t &= \bar{\gamma}_0 + \bar{\gamma}_1 t + \bar{\gamma}_2 t^2 + \bar{x}_t^0 \end{aligned} \tag{34}$$

and proceed with the same analysis as in the case with no deterministics, but by using as new series the detrended processes $\hat{x}_t^0$, $\tilde{x}_t^0$, and $\bar{x}_t^0$. The considerations of which deterministic regressors to use follows the discussion already given. This latter detrending procedure appears to be especially attractive when conducting semiparametric tests as I shall demonstrate later.

Both of the detrending procedures are going to deliver the same results. The way that detrending affects the asymptotic distributions is rather straightforward since in place of the Brownian motion expressions $W(r)$ and $\overline{W}(r)$ used in (20), (23) and (28) one substitutes by appropriately detrended Brownian motions, i.e. the limiting processes of $\hat{x}_t^0$, $\tilde{x}_t^0$ and $\bar{x}_t^0$. These are given by

$$W^i(r) = W(r) - f_i(r)' \left(\int_0^1 f_i(s) f_i(s)' \mathrm{d}s \right)^{-1} \int_0^1 f_i(s) W(s) \mathrm{d}s, \quad i = 0, 1, 2 \tag{35}$$

where, respectively, $f_0(r) = 1$, $f_1(r) = (1, r)'$, and $f_2(r) = (1, r, r^2)'$.

With respect to the repeated Brownian motions $\overline{W}(r) = \int_0^r W(s)\mathrm{d}s$ this is replaced by integrated equivalents of the above detrended Brownian motions.

The finite sample critical values relevant for the different tests are reported in

Table 2. Some references to critical values used in tests for double unit roots.

Detrending Test	None	Constant	Trend	Quadratic trend
Hasza-Fuller, F_{α_1,α_2}	Hasza and Fuller (1979),			Haldrup (1994a)
Dickey-Pantula*, t_{α_1}	Fuller (1976)			
Sen-Dickey, $F^{Sym}_{\alpha_1\alpha,2}$	n.a.	Sen and Dickey (1987)		Shin and Kim (1997)

* The Dickey-Pantula test is equivalent to the Dickey-Fuller test on the differenced data.

various articles. Table 2 provides a review of the appropriate references where the relevant fractiles can be found.

5.2. *Examples of parametric tests for I(2)*

Throughout the paper I will apply the various techniques presented to actual time series. The first data set is the money and price series (in logs) for the German hyperinflation episode 1920:1–1923:6. The period is the same as in other analyses of this data set, see e.g. Engsted (1993, 1994), and Casella (1989).

The next data set is used to analyze money demand in Denmark and has previously been examined by Juselius (1998). The sample is 1974:1–1993:4 and extends the sample period considered by Johansen and Juselius (1990). The series, except the bond rate i_{bt}, are in logs. Potentially, consumer prices, p_{ct}, the implicit GDP deflator, p_t, and nominal money, m_t will have I(2) features. The remaining series, real money, $m_t - p_t$, real GDP, y_t, and the effective bond rate, i_{bt} are more likely to be I(1), but naturally these are empirical questions to be tested.

The final data set is for housing construction in the US and covers the period 1968:1–1994:12. The series are new privately owned housing units started and completed, respectively. This data set is included to demonstrate the idea of multicointegration as we shall see later. The data set has previously been analyzed by Lee (1996) and Engsted and Haldrup (1998).

In table 3 (1st. panel) the money and price series during the German hyperinflation episode have been tested for the presence of double unit roots. All three tests described in the previous section have been conducted.[11] To account for the different deterministic alternatives the cases with constant, trend, and quadratic trend were considered. A quadratic trend was not considered in the Dickey-Pantula procedure since this implies the presence of at least a cubic trend in the levels of the series and power against this alternative does not seem reasonable. As seen, all tests favour the I(2) null.

In the second panel results for the price and money series from the Danish money demand data set are displayed. Both p_{ct} and m_t are accepted to be I(2). The GDP deflator series p_t is on the limit of being I(2) when the Dickey-Pantula procedure is considered. However, when models with trend and quadratic trends are addressed, both the joint F type tests accept the presence of two unit roots.

Table 3. Parametric tests for double unit roots.

		lags	constant	trend	quadratic trend
Germany 1920:1–1923:6 (T = 42)					
	HF-$F_{\alpha_1\alpha_2}$	4	6.72*	6.59	12.37
m_t	SD-$F^{sym}_{\alpha_1\alpha}$	4	6.94	7.78	6.45
	DP-t	4	−2.92*	−3.62*	—
	HF-$F_{\alpha_1\alpha_2}$	4	2.51	4.23	12.33
p_t	SD-$F^{sym}_{\alpha_1\alpha_2}$	4	0.94	1.18	1.81
	DP-t	4	−1.59	−2.92	—
Denmark 1974:1–1993:4 (T = 80)					
	HF-$F_{\alpha_1\alpha_2}$	7	5.14	4.97	6.49
p_{ct}	SD-$F^{sym}_{\alpha_1\alpha_2}$	7	0.46	7.53	9.95
	DP-t	7	−0.54	−3.06	—
	HF-$F_{\alpha_1\alpha_2}$	3	9.23**	8.81	12.91
p_t	SD-$F^{sym}_{\alpha_1\alpha_2}$	3	1.89	11.27	14.06
	DP-t	5	−2.77**	−3.56**	—
	HF-$F_{\alpha_1\alpha_2}$	3	3.41	5.83	6.02
m_t	SD-$F^{sym}_{\alpha_1\alpha_2}$	3	2.20	6.56	8.78
	DP-t	3	−2.41	−2.42	—

Note: The three different tests denote respectively the Hasza-Fuller F-test, the Sen-Dickey symmetric F-test, and the Dickey-Pantula t-test. 'lags' refers to the number of lags in the auxiliary regression. The deterministic components indicate whether these were included. '*', '**', and '***' refer to significance on 10,5, and 1% levels.

5.3. *Semiparametric approaches to testing for I(2)*

A major problem about the 'parametric' approach to test for double unit roots, as well as in testing for a single unit root, is that a finite order AR process is assumed to adequately describe the time series. This may be valid in many cases, at least as an approximation, but clearly the presence of moving average errors may yield poor finite order AR descriptions of the data and may invalidate the tests in the sense that their size may be hard to control without destroying the power. In testing for a single unit root Phillips (1987) and Phillips and Perron (1988) suggest semiparametric tests, known as the Phillips-Perron Z-tests, which are modifications of the standard Dickey-Fuller tests. The idea behind these tests is that when the distribution of the Dickey-Fuller t-ratio depends upon nuisance parameters under the null, the unknown population parameters can be consistently estimated, and the t-ratio subsequently be adjusted in a particular way to yield a distribution free of nuisance parameters. The problem about the semiparametric tests is that although in principle these are valid for MA errors they appear to have rather large size distortions when the root of the MA errors tends to minus one; Schwert (1989) reports these findings in a large simulation study. This criticism applies to any unit root test, since it may always be difficult for a finite stretch of data to identify a unit root when AR and MA roots are close to cancellation. However, size distortions appear to be particularly serious for semiparametric tests.

As an alternative to the Hasza-Fuller test described in the previous section, I (Haldrup (1994a)) develop a semi-parametric equivalent to the Hasza-Fuller joint F-test. The test is a straightforward generalization of the Phillips-Perron Z-tests developed for the I(1) case. The basic regression model to consider is

$$\Delta^2 x_t = (\hat{\alpha}_1 - 1)x_{t-1} + (\hat{\alpha}_2 - 1)\Delta x_{t-1} + \hat{\varepsilon}_t \tag{36}$$

which is just (19) but with no lags of $\Delta^2 x_t$ included in the regression. The error term captures all the short run dynamics which is reflected by the presence of nuisance parameters in the asymptotic distribution of the F-test of H_0: $\alpha_1 = \alpha_2 = 1$. The way the nuisance parameters enter the asymptotic distribution follows from the asymptotic expression

$$\begin{aligned} F_{\alpha_1, \alpha_2} \Rightarrow \frac{\sigma^2}{2\sigma_\varepsilon^2} D^{-1} \Bigg\{ & \left(\int_0^1 \overline{W}(r)\mathrm{d}W \right)^2 \int_0^1 W^2(r)\mathrm{d}r \\ & - 2\left(\int_0^1 W(r)\mathrm{d}W + \lambda' \right)\left(\int_0^1 \overline{W}(r)W(r)\mathrm{d}r \right)\left(\int_0^1 \overline{W}(r)\mathrm{d}W \right) \\ & + \left(\int_0^1 W(r)\mathrm{d}W' + \lambda' \right)^2 \left(\int_0^1 \overline{W}^2(r)\mathrm{d}r \right) \Bigg\} \end{aligned} \tag{37}$$

where $D = \int_0^1 \overline{W}^2(r)\mathrm{d}r \int_0^1 W^2(r)\mathrm{d}r - (\int_0^1 \overline{W}(r)W(r)\mathrm{d}r)^2$ and $\sigma^2, \sigma_\varepsilon^2$ have previously been defined. We also let $\lambda' = (\sigma^2 - \sigma_\varepsilon^2)/2\sigma^2$.

When errors are independent $\sigma^2 = \sigma_\varepsilon^2$ and the limiting distribution can be shown to simplify to (20). To adjust the F-ratio given above such that the subsequent test has the distribution (20), consistent estimates of σ^2 and σ_ε^2 are needed.

A consistent estimator of the long run variance σ^2 is given by

$$\hat{\sigma}^2 = T^{-1} \sum_1^T \hat{\varepsilon}_t^2 + 2T^{-1} \sum_{\tau=1}^{l} \omega_{\tau l} \sum_{t=\tau+l}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-\tau} \tag{38}$$

where $\hat{\varepsilon}_t$ are regression residuals from the regression (36), see Newey and West (1987). $\omega_{\tau l}$ is a weight function which corresponds to the lag window used for spectral smoothing in the frequency domain. Any lag window can be used, e.g. Parzen, Tukey, etc., but in the sequel I use the Bartlett window $\omega_{\tau l} = 1 - \tau/(l+1)$. Choosing the width l of the lag window is difficult in practice. My recommendation is to use different values but ensuring that l increases with the available sample size. Alternatively, an automatic kernal lag selection can be used, see e.g. Andrews (1991). A consistent estimate of the regression standard error $\hat{\sigma}_\varepsilon{}^2$ can be found as

$$\hat{\sigma}_\varepsilon^2 = T^{-1}\sum_{1}^{T}\hat{\varepsilon}_t^2. \tag{39}$$

From these estimates it follows trivially that λ and λ' can be estimated by $\hat{\lambda} = (\hat{\sigma}^2 - \hat{\sigma}_\varepsilon^2)/2$ and $\hat{\lambda}' = \hat{\lambda}/\hat{\sigma}^2$. Phillips (1987) described very carefully the moment and mixing conditions that are needed for the estimates $\hat{\sigma}^2$ and $\hat{\sigma}_\varepsilon^2$, and hence for the construction of the semiparametric statistics, to be valid.

In order to construct the semiparametric equivalent of Hasza and Fuller's test we need to define the following quantities:

$$\begin{aligned}
m_{xx} &= T^{-4}\sum_{1}^{T}x_t^2 & m_{x\Delta x} &= T^{-3}\sum_{1}^{T}x_{t-1}\Delta x_t \\
m_{x\Delta^2 x} &= T^{-2}\sum_{1}^{T}x_{t-1}\Delta^2 x_t & m_{\Delta x\Delta^2 x} &= T^{-1}\sum_{1}^{T}\Delta x_{t-1}\Delta^2 x_t \\
m_{\Delta x\Delta x} &= T^{-2}\sum_{1}^{T}\Delta x_t \Delta x_t & M &= m_{xx}m_{\Delta x\Delta x} - m_{x\Delta x}^2.
\end{aligned} \tag{40}$$

Let $F_{\alpha_1\alpha_2}$ be the F-statistic from the regression (36). The test is now given by

$$Z(F_{\alpha_1\alpha_2}) = F_{\alpha_1\alpha_2}\frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}^2} - \frac{1}{2}M^{-1}\left[2\hat{\lambda}'(m_{\Delta x\Delta^2 x}m_{xx} - m_{x\Delta x}m_{x\Delta^2 x}) - \left(\frac{\hat{\lambda}}{\hat{\sigma}}\right)^2 m_{xx}\right] \tag{41}$$

which follows the Hasza-Fuller distribution (20).

If power is desired against deterministic alternatives the series should be detrended prior to the construction of the $Z(F_{\alpha_1\alpha_2})$ test. This can be done by the regressions (34) and subsequently use the regression residuals as the new (detrended) series.

The above test is found to suffer from the size distortions, which also characterize the Phillips-Perron tests, by the presence of a large negative MA root. However, in cases where the alternative is explosive there is evidence that the semiparametric test has favourable properties compared to the sequential procedure.

The paper by Shin and Kim (1997) is jointly motivated by the symmetric Sen and Dickey $F_{\alpha_1\alpha_2}^{Sym}$-test and the semiparametric $Z(F_{\alpha_1\alpha_2})$-test. A semiparametric version of the Sen-Dickey test is suggested and not surprisingly this test is found to have better (size-corrected) power compared to previously existing tests. More importantly, however, it is found that some of the size distortions of the $Z(F_{\alpha_1\alpha_2})$-test can be considerably reduced by symmetrizing the regression. The test is constructed in a manner like the test of Sen and Dickey but, again, no augmentation through lagged second differences is used. Hence the regressions estimated jointly are

$$\begin{aligned}
\Delta^2 x_t &= (\hat{\alpha}_1 - 1)x_{t-1} + (\hat{\alpha}_2 - 1)\Delta x_{t-1} + \hat{\varepsilon}_t, \quad t = 3, \ldots, n \\
\Delta^2 x_t &= (\hat{\alpha}_1 - 1)x_{t-1} - (\hat{\alpha}_2 - 1)\Delta x_t + \hat{\eta}_t, \quad t = n, \ldots, 3
\end{aligned} \tag{42}$$

where the dependent and independent variables in the joint estimation are given by

$$Y = (\Delta^2 x_3, \Delta^2 x_4, \ldots, \Delta^2 x_n, \Delta^2 x_n, \ldots, \Delta^2 x_4, \Delta^2 x_3)', \text{ and}$$

$$X = [(x_2, x_3, \ldots, x_{n-1}, x_{n-1}, \ldots, x_3, x_2)' \vdots \\ (\Delta x_2, \Delta x_3, \ldots, \Delta x_{n-1}, -\Delta x_n, \ldots, -\Delta x_4, -\Delta x_3)'].$$

Note that this corresponds to setting $p = 2$ in table 1. Based on this regression the F-test of the joint double unit root hypothesis follows the distribution

$$F^{Sym}_{\alpha_1, \alpha_2} \Longrightarrow \frac{\hat{\sigma}^2}{\sigma_\varepsilon^2}\left(\int_0^1 \overline{W}^2(r)\mathrm{d}r\right)^{-1}\left(\int_0^1 \overline{W}(r)\mathrm{d}W\right)^2 + \frac{\sigma_\varepsilon^2}{4\sigma^2}\left(\int_0^1 W^2(r)\mathrm{d}r\right)^{-1} \tag{43}$$

which depends upon nuisance parameters σ_ε^2 and σ^2. These can be estimated exactly as in (38) and (39), that is, by using the first half of the data points.[12] Given these estimates the semiparametric test is defined as

$$Z(F^{Sym}_{\alpha_1\alpha_2}) = \frac{1}{2\hat{\sigma}_\varepsilon^2}\left(\frac{\hat{\sigma}_\varepsilon}{\hat{\sigma}}(\hat{\alpha}_1 - 1), \frac{\hat{\sigma}}{\hat{\sigma}_\varepsilon}(\hat{\alpha}_2 - 1)\right)(X'X)\left(\frac{\hat{\sigma}_\varepsilon}{\hat{\sigma}}(\hat{\alpha}_1 - 1), \frac{\hat{\sigma}}{\hat{\sigma}_\varepsilon}(\hat{\alpha}_2 - 1)\right)' \tag{44}$$

where $(\hat{\alpha}_1 - 1, \hat{\alpha}_2 - 1)' = (X'X)^{-1}X'Y$. The distribution of this test is given as in (43) with $\sigma_\varepsilon^2 = \sigma^2$, i.e. the Sen and Dickey distribution (28). The extension of this test to account for deterministic components follows the same procedure as for the $Z(F_{\alpha_1\alpha_2})$-test.

In a rather comprehensive Monte Carlo study Shin and Kim (1997) find their test to have rather good size properties compared to other semiparametric tests, especially in the situations where the other tests seem to perform rather poorly. They provide a reasonable intuitive argument why this is so. Note simply, that the non-symmetrized $Z(F_{\alpha_1\alpha_2})$-test is adjusting $F_{\alpha_1\alpha_2}$ both by scale (through $\sigma^2/\sigma_\varepsilon^2$) and by location (through λ'); compare the asymptotic formula (37). However, the $Z(F^{Sym}_{\alpha_1\alpha_2})$-test only requires a scale adjustment as some cross terms appear to cancel due to symmetry. Hence, less adjustments of the original statistic is needed in the symmetric test. Concerning the size adjusted powers Shin and Kim's simulation study also turned out to be rather favourable to their own test as powers were significantly higher compared to the $Z(F_{\alpha_1\alpha_2})$-test and the test based on the Dickey-Pantula procedure.

It is obvious that a semiparametric analogue of the Dickey-Pantula procedure can be easily adopted by use of Phillips-Perron tests.

5.4. *Examples of semi-parametric tests for I(2)*

I consider the same data series as in section 5.2, but now by focusing on the semi-parametric versions of the tests.[13] In calculating the long run variance used in the construction of the tests, the Bartlett kernal estimator was used. The truncation parameter was chosen to reflect both the frequency of the data and the number of observations. However, different values were tried and this did not have a significant impact on the test results.

The overall evidence from the semi-parametric tests reported in table 4 is quite different from the parametric tests displayed in table 3. Only in few cases can the I(2) null not be rejected. It seems likely that these results are due to the poor power that semi-parametric tests seem to have more generally. Note however that the Shin-Kim symmetric F test generally seems to be less significant than the other tests, just as their Monte Carlo results indicate.

6. Single equation analysis in multivariate I(2) systems

Determining the integration order of economic time series is not necessarily interesting *per se*, but it provides a valuable input to be used in the formulation of multivariate models. In the following sections various aspects of multivariate regression models will be analyzed. First I focus on single equation analysis and in section 8 the systems approach to multivariate analysis is addressed.

The set up of the analysis can be described as follows. Consider a model where y_t, is a scalar I(2) variable related to the time series x_t which consists of m_1 I(1) variables, x_{1t}, and m_2 I(2) variables, x_{2t}, as well as m_0 separate deterministic

Table 4. Semi-parametric tests for double unit roots.

		lags	constant	trend	quadratic trend
Germany 1921:1–1923:6 ($T=42$)					
m_t	HF-$Z(F_{\alpha_1\alpha_1})$	8	2.41	1.12	5.71
	SK-$Z(F^{Sym}_{\alpha_1\alpha_1})$	8	0.79	3.47	10.66
	DP-$Z(t)$	8	−0.99	−2.45	—
p_t	HF-$Z(F_{\alpha_1\alpha_1})$	8	11.58***	6.52	22.80*
	SK-$Z(F^{Sym}_{\alpha_1\alpha_1})$	8	2.62	5.59	38.45***
	DP-$Z(t)$	8	−3.71***	−5.27***	—
Denmark 1974:1–1993:4 ($T=80$)					
p_{ct}	HF-$Z(F_{\alpha_1\alpha_1})$	4	27.56***	12.66***	35.62***
	SK-$Z(F^{Sym}_{\alpha_1\alpha_1})$	4	1.95	7.47	37.81***
	DP-$Z(t)$	4	−4.99***	−7.89***	—
p_t	HF-$Z(F_{\alpha_1\alpha_1})$	4	80.83***	57.40***	82.90***
	SK-$Z(F^{Sym}_{\alpha_1\alpha_1})$	4	5.91	23.24***	74.25***
	DP-$Z(t)$	4	−10.51***	−12.55***	
m_t	HF-$Z(F_{\alpha_1\alpha_1})$	4	108.98***	107.19***	115.27***
	SK-$Z(F^{Sym}_{\alpha_1\alpha_1})$	4	28.14***	67.17***	75.24***
	DP-$Z(t)$	4	−14.51***	−15.04***	—

Notes: The three different tests denote respectively the semiparametric versions of the Hasza-Fuller F-test, the Shin-Kim symmetric F-test, and the Dickey-Pantula t-test. The latter test is just the Phillips-Perron test on the differenced data. 'lags' refers to the truncation parameter used in the Bartlett kernal estimate of σ^2. The deterministic components indicate whether these were included. '*', '**', and '***' refer to significance on 10,5, and 1% levels.

components of different orders. We let $m = m_0 + m_1 + m_2$ and specify the separate series as in (10):

$$\begin{aligned} y_t &= \gamma_0' \mathrm{c}_t + y_t^0, & & \\ \mathrm{x}_{1t} &= \gamma_1' \mathrm{c}_t + \mathrm{x}_{1t}^0, & \Delta \mathrm{x}_{1t}^0 &= \varepsilon_{1t} \\ \mathrm{x}_{2t} &= \gamma_2' \mathrm{c}_t + \mathrm{x}_{2t}^0, & \Delta^2 \mathrm{x}_{2t}^0 &= \varepsilon_{2t} \\ \mathrm{x}_t &= (\mathrm{c}_t', \mathrm{x}_{1t}', \mathrm{x}_{2t}')'. & & \end{aligned} \tag{45}$$

For the subsequent analysis to be correct x_{1t} and x_{2t} are required to be individually non-cointegrated vector series, and furthermore x_{1t} and $\Delta \mathrm{x}_{2t}$ are assumed not to cointegrate.[14] The stochastic part of y_t denoted y_t^0 is integrated of order two and is related to x_{1t}^0 and $\mathrm{x}^0{}_{2t}$ as follows

$$y_t^0 - \beta_1' \mathrm{x}_{1t}^0 - \beta_2' \mathrm{x}_{2t}^0 = u_t. \tag{46}$$

In terms of the original variables y_t and x_t we therefore have[15]

$$y_t = \beta_0' \mathrm{c}_t + \beta_1' \mathrm{x}_{1t} + \beta_2' \mathrm{x}_{2t} + u_t = \beta' \mathrm{x}_t + u_t \tag{47}$$

with $\beta_0' = (\gamma_0' - \beta_1'\gamma_1' - \beta_2'\gamma_2')$. With respect to the errors u_t these can be $\mathsf{I}(0)$, $\mathsf{I}(1)$ or $\mathsf{I}(2)$. We let $\Delta^d u_t = v_t, d = 0, 1, 2$, so depending upon the integration properties of u_t different situations with distinct cointegration features may arise.

Under rather weak conditions, known as the multivariate invariance principle, it can be shown that in defining the $p = 1 + m_1 + m_2$ dimensional error process $\mathrm{w}_t = (v_t, \varepsilon_{1t}', \varepsilon_{2t}')'$, the scaled vector partial sum process

$$T^{-1/2} \sum_{1}^{[Tr]} \mathrm{w}_t \Longrightarrow \mathrm{B}(r) \quad \text{as } T \longrightarrow \infty, \tag{48}$$

where $\mathrm{B}(r) = (B_0(r), \mathrm{B}_1(r)', \mathrm{B}_2(r)')'$ is a vector Brownian motion (partitioned conformably with w_t) with long-run covariance matrix $\Omega = \lim_{T\to\infty} T^{-1} \mathsf{E}((\sum_1^{\mathrm{T}} \mathrm{w}_t)(\sum_1^{T} \mathrm{w}_t)')$. Alternatively we write $\Omega = \Sigma + \Lambda + \Lambda'$ where $\Sigma = \mathsf{E}(\mathrm{w}_1 \mathrm{w}_1')$ and $\Lambda = \sum_{k=2}^{\infty} \mathsf{E}(\mathrm{w}_1 \mathrm{w}_\mathrm{k}')$. The covariance matrix can be partitioned conformably with w_t, and, in so doing, the sub-covariance matrices associated with x_{1t} and x_{2t} must be of full rank. This condition implies that the right hand side variables in (47) do not cointegrate. This is an important assumption, but I shall slack this in section (8).

Given the above assumptions we consider in the following the regression model

$$y_t = \hat\beta_0' \mathrm{c}_t + \hat\beta_1' \mathrm{x}_{1t} + \hat\beta_2' \mathrm{x}_{2t} + \hat u_t = \hat\beta' \mathrm{x}_t + \hat u_t. \tag{49}$$

The properties of statistics based on this model depend upon the integration order of u_t. When u_t is $\mathsf{I}(0)$, y_t, x_{1t}, and x_{2t} are cointegrated and hence constitute a long-run equilibrium relationship. As noted previously, this definition of cointegration is slightly different from Engle and Granger's (1987) definition. When $u_t \sim \mathsf{I}(1)$, y_t, $\mathrm{x}_{2t} \sim \mathsf{CI}(2, 1)$, but do not cointegrate with x_{1t}. Finally, the case where $u_t \sim \mathsf{I}(2)$ refers to non-cointegration amongst the variables. Sometimes the cases where u_t is either $\mathsf{I}(1)$ or $\mathsf{I}(2)$ may lead to spurious regressions as we shall see below, and therefore these situations should be analyzed with care. However, spurious regressions are typically also the basis for residual based tests for (non) cointegration and hence are of separate interest to examine.

6.1. *Spurious regression*

In a much cited paper Granger and Newbold (1974) demonstrated how regressions that involved statistically independent random walks would lead to the incorrect inference that the series were related if standard statistical tools were used for analysis. Through simulations they also showed that 'spurious' regressions were accompanied with low values of the Durbin-Watson (*DW*) statistic, i.e. residuals were highly positively correlated, and the coefficient of determination, R^2, was found to be higher than expected and hence essentially indicating that a large proportion of the variation in data could be explained by artificial regressors. Phillips (1986) provided an analytical explanation of Granger and Newbold's findings: In regressing independent random walks on each other the t-ratio of a zero coefficient null diverges of order $O_p(T^{1/2})$, the regression coefficients of stochatic regressors are non-degenerate, the DW-statistic tends in probability to zero as $T \longrightarrow \infty$, and finally the R^2 statistic has a non-degenerate distribution.

In Haldrup (1994b) I extend Phillips' analytical results to the case where the regression model is (49) under the assumptions outlined above. These results, together with Phillips' findings, are reported in schematic form in Table 5. The regression errors in my own (Haldrup (1994b)) analysis are assumed to be integrated of either order one or two. In the table the $F(\hat{\beta})$-statistic refers to the general F-test of hypotheses H_0: $R\beta = r$, where R is a $q \times m$ restriction matrix and

Table 5. Orders in probability of regression coefficients and diagnostics. Index d refers to the integration order of u_t.

		I(1)-model of Phillips (1986)*	I(2) model based on the regression (49)	
		$d=1$	$d=1$	$d=2$
constant:	$\hat{\beta}^0_{0t}$	$O_p(T^{1/2})$	$O_p(T^{1/2})$	$O_p(T^{3/2})$
trend:	$\hat{\beta}^1_{0t}$	$O_p(T^{-1/2})$	$O_p(T^{-1/2})$	$O_p(T^{1/2})$
quadratic trend:	$\hat{\beta}^2_{0t}$	—	$O_p(T^{-3/2})$	$O_p(T^{-1/2})$
x_{1t}:	$\hat{\beta}_{1t}$	$O_p(1)$	$O_p(1)$	$O_p(T)$
x_{2t}:	$\hat{\beta}_{2t}$	—	$O_p(T^{-1})$	$O_p(1)$
Diagnostics:				
$F(\hat{\beta})$		$O_p(T)$	$O_p(T)$	$O_p(T)$
DW		$O_p(T^{-1})$	$O_p(T^{-1})$	$O_p(T^{-1})$
R^2		$O_p(1)$	$1 + O_p(T^{-1/2})$	$O_p(1)$

*Note: The Phillips model is based on the regression $y_t = \hat{\beta}^0_0 + \hat{\beta}^1_0 t + \hat{\beta}'_1 x_{1t} + \hat{u}_t$, where y_t, x_{1t}, u_t are I(1) series.

r is a known q-dimensional vector.[16] Concerning the deterministic regressors it is assumed that $c_t = (1, t, t^2)'$ and with β_0^0, β_0^1 and β_0^2 being the associated coefficients.

When u_t is non-stationary, either I(1) or I(2), it is seen that many of the spurious regression results applying to the I(1) model of Granger and Newbold (1974) and Phillips (1986) carry over to the I(2) model; apart from R^2 all spurious regression results for $d = 1$ appear to be the same across the two different models. Generally, if the order of a particular regressor is less than $O_p(T^{d-1/2})$, it is seen that the least squares estimate is inconsistent and diverges at a rate which depends upon the order of the regressor. Consistency is only achieved when the order of the regressor exceeds $O_p(T^{d-1/2})$. F-tests diverge at the same rate $O_p(T)$ (t-ratios $= O_p(T^{1/2})$) regardless of the order d, and similarly the DW statistic tends to zero at the same rate.

The reason why $R^2 \longrightarrow 1$ in the I(2) model when $T \longrightarrow \infty$, even though $d = 1$, follows from the fact that in this case $y_t, x_{2t} \sim \mathrm{CI}(2, 1)$, and thus, because cointegration to I(1) level occurs, the residual variation will be of a lower order in probability than the dependent I(2) variable. Note however, that when $d = 2$ the distribution of R^2 is non-degenerate, but still bounded on $[0, 1]$ of course.

6.2. *Residual based test for (non) cointegration*

Spurious regression occurs when variables are not fully cointegrated and hence this can naturally be used as a basis for testing the hypothesis of no cointegration. Following the initial benchmark description of Engle and Granger's (1987) 2-step procedure, the idea is to test the integration order of residuals from a cointegration regression. A common class of tests adopted in the single equation I(1) analysis is the augmented Dickey-Fuller class of tests including their associated semiparametric analogues. This procedure can be straightforwardly extended to the I(2) case such that regression residuals from a regression like (49) can be tested for the order of integration.

In principle, any of the unit root tests previously discussed in section 5 may be considered in a cointegration framework. Hence the relevant hypotheses are

H_2: There is no cointegration amongst y_t, x_{1t}, x_{2t}, such that u_t is I(2)

H_1: The I(2) variables y_t and x_{2t} cointegrate to I(1) level, but no further cointegration occurs whereby u_t is I(1).

Tests for both types of hypotheses may be constructed. However, in most practical situations the I(2) variables cointegrate at least to an I(1) relation such that the hypothesis H_1 is the most relevant. This is the assumption made by Haldrup (1994b) who derives the properties of the augmented Dickey-Fuller class of tests when the hypothesis H_1 is tested using regression residuals $\hat{u}_t$ based on the regression (49). Since the single equation model depends upon both I(1) and I(2) variables, the asymptotic distributions and hence the critical values need to be modified.[17]

It can be shown that the limiting behaviour of $\hat{u}_t$ after appropriate scaling is a

Brownian motion process which depends upon a number of parameters. Hence the asymptotic distribution of the Dickey-Fuller t-ratio based on the regression

$$\Delta\hat{u}_t = (\hat{\alpha} - 1)\hat{u}_{t-1} + \sum_{j=1}^{q} \hat{\varphi}_j \Delta\hat{u}_{t-i} + \hat{\eta}_{qt} \tag{50}$$

is qualitatively similar to the distributions known from univariate unit root testing and cointegration testing in the I(1) case. Quantitatively, however, the distributions will depend upon m_1 and m_2, i.e. the number of I(1) and I(2) regressors in the regression. In addition, the distributions are affected by the deterministic components included in the first step auxiliary regression (49). Critical values for different values of m_1 and m_2 are reported in Haldrup (1994b) for the case with a constant and in Engsted *et al.* (1997) for the trend and quadratic trend cases. Interestingly, it appears that asymptotic critical values are rather similar to the asymptotic critical values reported in Phillips and Ouliaris (1990) for a given value of the sum $m_1 + m_2$. Hence there is indication that it is the total number of stochastic regressors in the auxiliary regression which matters for the asymptotic distributions. In finite samples, however, one has to discriminate between the number of I(1) and I(2) regressors.

Polynomial cointegration and multicointegration. The analysis above can be straightforwardly extended to the case of testing the null of non-polynomial cointegration by including in x_{1t} first differences of the I(2) variables x_{2t}. This will not affect the validity of the analysis.

In section (3.2) Granger and Lee's (1989,1990) notion of multicointegration was introduced, i.e. the idea that cumulated cointegration errors cointegrate with the original variables. Assume for a moment, that y_t is I(1) and cointegrates with the I(1) vector x_{1t}, such that the *first layer* of cointegration occurs when $y_t - \beta_1' x_{1t} = z_t$ is I(0). Multicointegration implies that $\sum_{j=1}^{t} z_j - \beta_1^{*'} x_{1t} - \beta_1^{**} y_t = u_t$ is also I(0); we denote this the *second layer* of cointegration. When β_1 is unknown there are essentially two ways of testing for multicointegration.

Granger and Lee suggest a two step procedure where the first layer of cointegration is tested by using standard procedures for I(1) models. Given cointegration this gives the estimate $\hat{\beta}_1$ and the generated series $\sum_{j=1}^{t} \hat{z}_j$ of regression residuals; note that this series is I(1) by construction. Subsequently $\sum_{j=1}^{t} \hat{z}_j$ is regressed on the remaining variables to see whether multicointegration occurs. In Engsted *et al.* (1997) we demonstrate that there are a number of important statistical problems about this procedure, one fundamental problem being that the null of no multicointegration in the second step of the procedure is invalidated when using standard residual based cointegration test procedures.

Instead, we suggest a single step procedure with favourable statistical properties, by exploiting the fact that a multicointegrated model can be written as an I(2) model. The idea is thus to use cumulated data series in the cointegration regression such that the model is formulated directly in terms of I(2) and I(1) variables. Note simply, that when u_t is stationary, the multicointegration relation can be written as

$$\sum_{j=1}^{t} y_j = \beta_1' \sum_{j=1}^{t} x_{1j} + \beta_1^{*\prime} x_{1t} + \beta_1^{**} y_t + u_t \qquad (51)$$

where $\sum_{j=1}^{t} y_j$ and $\sum_{j=1}^{t} x_{1j}$ are the generated I(2) series. In practical situations deterministics may be included in (49) since the cumulated series may also generate a trend, for instance. As seen the residual based testing procedure described above naturally encompasses the single step procedure of testing for multicointegration.[18]

6.3. *Examples of residual based testing for I(2) cointegration and multicointegration*

By applying the ML procedure for I(2) models, which I return to in section 8, Juselius (1998) analyzed the Danish money demand data set previously described. In the specification of a (nominal) money demand relation she considers the variables: nominal money, m_t, real income, y_t, the GDP deflator, p_t, and the deposit and bond rates, i_{dt}, and i_{bt}. The parametric tests for double unit roots in section 5.2 indicated that m_t and p_t are I(2). Unit root tests (not reported) also indicated that y_t, i_{bt}, i_{dt}, are I(1). Also the real money stock, $m_t - p_t$, was found to be I(1), hence implying that $m_t, p_t \sim \mathrm{CI}(2,1)$ with a cointegrating vector given by $(1, -1)$. The same cointegrating vector is naturally given for $\Delta m_t, \Delta p_t \sim \mathrm{CI}(1,1)$. These results suggest that we may consider both a real specification of the money demand relation (in terms of I(1) variables) as well as a nominal specification in terms of mixed I(2) and I(1) variables. In both situations Δp_t or Δm_t are likely variables to appear amongst the cointegrating variables.[19]

We start by considering a nominal specification. The variables included in the first model are $[m_t, p_t, y_t]$. According to the residual based procedure outlined above m_t is regressed on p_t and y_t, and an augmented Dickey-Fuller test is subsequently applied to the residuals. A constant was also included in the cointegration regression. The ADF-t value with lags 2, 3, and 4 yields -3.01. The 10% critical value tabulated in Haldrup (1994b), table 1, is -3.55 for $m_1 = m_2 = 1$ (corresponding to the number of I(1) and I(2) regressors) so we cannot reject the null. Next we estimate a model with the bond rate included as well so the information set now consists of $[m_t, p_t, y_t, i_{bt}]$. In this case the ADF-t-value with lags 2 and 4 yields -5.80 which is strongly significant (the 1% critical value for $m_1 = 2$ and $m_2 = 1$ is -4.81). Hence a nominal specification with these 4 variables suggests cointegration. Observe that inflation Δp_t needs not to be included to ensure that the variables cointegrate.

The above results suggest that we may also consider a real specification which includes the data set $[m_t - p_t, y_t, i_{bt}]$. Note that this information set consists only of I(1) variables so the standard Engle-Granger two-step procedure can be conducted in this case. The ADF-t-value with lags 2 and 4 included yields -5.17 which is also strongly significant. Hence the data satisfies a specification where money and prices satisfy long run homogeneity.

To demonstrate testing for multicointegration I consider the US housing data set, 1968:1–1994:12 ($T = 324$), previously analyzed by Lee (1996) and Engsted and Haldrup (1998). Lee argued that if y_t is housing units completed and x_t is housing units started, then, given that these series are I(1), it is likely that $y_t, x_t \sim \mathrm{CI}(1, 1)$. However, it is also likely that the stock of new housing units under construction, $Q_t = \sum_{j=1}^{t}(x_j - y_j)$, cointegrates with e.g. y_t such that $Q_t - ky_t$ is I(0). When this occurs the data series are multicointegrated. Actually Lee (1996) modified the Q_t variable as follows $Q_t = \sum_{j=1}^{t}(0.98x_j - y_j)$ since he was judging that 2% of all new housing starts are never completed. Using the US data set, he found rather strong evidence for the presence of multicointegration. However, rather than imposing the parameter 0.98 in the definition of the Q_t variable this can be estimated from the data by considering a regression of the form:

$$\sum_{j=1}^{t} y_j = \beta_1 \sum_{j=1}^{t} x_j + \beta_2 y_t + \gamma_0 + \gamma_1 t + u_t \tag{52}$$

where the trend is included to account for the fact that the cumulated series may produce a linear trend. Hence the idea of testing for multicointegration amounts to testing whether u_t is stationary. An augmented Dickey-Fuller test of $\hat{u}_t$ from the above regression yields a value of -4.82 which for $m_1 = m_2 = 1$ is significant at a 1% level. Note that the critical value should account for the presence of a linear trend in the cointegration regressions so the critical values of Engsted *et al.* (1997, table 1), should be used. Observe that the estimate $(1 - \hat{\beta}_1)$ provides an estimate of the unknown percentage of housing units that are never completed. In the above case this estimate is 5.3%. As we shall see in section 6.4 this estimate is rather precise as the rate of consistency is $O_p(T^2)$.

6.4. *Properties of single equation cointegration regressions with I(1) and I(2) variables*

When the errors u_t in (47) appear to be stationary the time series $y_t, \mathrm{x}_{1t}, \mathrm{x}_{2t}$, are I(2) cointegrated[20] and in this case it is of interest to examine the properties of statistics based on the regression model (49) where, once again, c_t is assumed to be given by the vector $(1, t, t^2)'$. Haldrup (1994b) shows that from this regression $R^2 = 1 + O_p(T^{-3})$ and therefore an indication is given of a very good fit. Next we focus on the properties of least squares regression estimates. The following distribution result holds:[21]

$$\mathrm{D}_T \mathrm{G}(\hat{\beta} - \beta) \Rightarrow \left(\int_0^1 \mathrm{B}_*(r)\mathrm{B}_*'(r)\mathrm{d}r \right)^{-1} \left(\int_0^1 \mathrm{B}_*(r)\mathrm{d}B_0 + (0', \Delta_{10}', 0')' \right) \tag{53}$$

where $D_t = \mathrm{diag}(T^{1/2}, T^{3/2}, T^{5/2}, T, T^2)$, $\quad B_*(r) = (1, r, r^2, B_1(r)', \overline{B_2(r)}')'$, and

$$G = \begin{pmatrix} I_3 & \gamma_1 & \gamma_2 \\ 0 & I_{m_1} & 0 \\ 0 & 0 & I_{m_2} \end{pmatrix}. \tag{54}$$

The term Δ_{10} is the $(1, 1)$ vector element of $\Delta = \Sigma + \Lambda$.

In particular, it follows from (53) that

$$\begin{aligned} T^2(\hat{\beta}_2 - \beta_2) &= O_p(1) \\ T(\hat{\beta}_1 - \beta_1) &= O_p(1). \end{aligned} \tag{55}$$

Hence, in cointegration regressions with I(2) variables the least squares estimate $\hat{\beta}_2$, corresponding to the I(2) regressors, will tend to its true value β_2 at the rapid *super-super* consistent rate, $O_p(T^{-2})$. This rate is much faster than the usual *super* consistent rate, $O_p(T^{-1})$, found in I(1) cointegrating regressions. However, with respect to the I(1) regressors the associated estimate $\hat{\beta}_1$ is seen to be super consistent. In general, we cannot say what the orders of the deterministic regressors will be since this depends upon which (non-zero) deterministics are present in the time series.[22] The largest orders possible of the elements in $\hat{\beta}_0$ are given by $O_p(T^{-1/2})$, $O_p(T^{-3/2})$, and $O_p(T^{-5/2})$, respectively; a situation occurring, for instance, when the regressors have no deterministic components, $\gamma_1, \gamma_2 = 0$.

Generally, the distributions of the least squares estimates can be expressed in terms of the non-standard distributions given in (53) so standard gaussian inference cannot be undertaken in general, but only in special situations. One such special case is when x_{1t} and x_{2t} can be considered strictly exogenous regressors such that the single Brownian motions are uncorrelated. In this situation the distribution in (53) simplifies to a mixed gaussian distribution (i.e. a gaussian distribution conditional on information) of the form, see Park and Phillips (1989) and Haldrup (1994b),

$$D_T G(\hat{\beta} - \beta) \Rightarrow \int_V N(0, V)\, dP(V) \tag{56}$$

with $V = (\int_0^1 B_*(r) B_*'(r) dr)^{-1}$ being the mixing variate matrix. Hence hypothesis testing can be undertaken using standard procedures, for instance *t*-tests will follow the standard normal distribution.[23]

Strict exogeneity is a very restrictive property of the data and this has led Choi *et al.* (1997) to suggest a generalization of Park's (1992) canonical cointegration regression procedure which is based on transformation of the regressors and the regressand in the model (49) such that the resulting distribution of the least squares estimator becomes mixed gaussian as in (56). I will not go into a technical discussion of this procedure but refer to Choi *et al.* (1997).

An estimation method which is somewhat related to that of Choi *et al.* (1997) is the *residual-based fully modified ordinary least-squares* (RBFM-OLS) procedure suggested by Chang and Phillips (1995). However, the way the dependency of nuisance parameters is accounted for is different; mainly because no prior assumptions are made about the cointegration dimension and the mixture of I(0), I(1), and I(2) variables in the model.

6.5. *Examples of estimation of cointegration relations from single equations*

In section 6.3 I found for the Danish money demand data set that $[m_t, p_t, y_t, i_{bt}]$ is I(2) cointegrated. I also estimated a relation where the model included also a linear trend and this was found to constitute an equally significant relationship. Least squares estimation gave the following regression results:

$$m_t = 1.18p_t + 1.03y_t - 9.80i_{bt} + constant$$
$$R^2 = 0.994, DW = 1.54$$

$$m_t = 1.10p_t + 1.05y_t - 9.42i_{bt} + constant + trend$$
$$R^2 = 0.994, DW = 1.53$$

As seen, despite $O_p(T^2)$-consistency, the estimate of the coefficient associated with p_t seems to depend much upon the presence of a linear trend in the model. It could therefore be of interest to produce a confidence interval for this coefficient. However, simple OLS will not produce the correct standard errors in order to do standard inference, so instead the Choi *et al.* (1997) CCR estimator for I(2) systems was used to reestimate the relations.[24] This gave the following results where the numbers in parenthesis indicate standard errors:

$$\underset{}{m_t} = \underset{(0.03)}{1.17}p_t + \underset{(0.12)}{1.04}y_t - \underset{(0.76)}{10.79}i_{bt} + constant$$

$$\underset{}{m_t} = \underset{(0.08)}{1.19}p_t + \underset{(0.12)}{1.05}y_t - \underset{(0.85)}{10.74}i_{bt} + constant + trend$$

It can be seen that in both cases a standard confidence interval for the p_t coefficient does not include 1 which perhaps is a bit surprising given the fact that the real specification in section 6.3 was found to exhibit cointegration. Note however, that the coefficient of y_t cannot be rejected to equal unity.

7. Cointegrated models in systems: some motivation

In order to clarify the different cointegration possibilities that can occur in systems of variables it may be useful to introduce the so-called triangular I(2) representation.

For I(1) processes Phillips (1991) suggested a parametrization of cointegrated time series in terms of a so-called *triangular system*. He used this parametrization as a convenient framework for conducting optimal inference when the number of unit roots, and hence the number of cointegrating relations, are known a priori. Within the context of triangular systems, generalizations to I(2) processes have been introduced by, *inter alia*, Stock and Watson (1993) and Kitamura (1995), and are discussed in e.g. Chang and Phillips (1995), and Boswijk (1997). I will not go into a detailed discussion of all the different forms triangular representations may take; here I follow Stock and Watson's exposition for $d = 2$ (maximal order of integration) as it nicely demonstrates the main ideas and motivates how cointegration can occur.

In the triangularization of Stock and Watson a p-vector time series x_t can be partitioned (and possibly rearranged) as $x_t = (x'_{0t}, x'_{1t}, x'_{2t})$ where the single components are of dimension r, s, and $p - r - s$, which (rather importantly) are assumed to be known *a priori*. As we shall see these numbers will refer to the so-called integration indices to be defined in section (8.2). For ease of exposition I assume that the series are free of deterministic components. The triangular I(2) representation is now given by:

$$\begin{aligned} x_{0t} &= A_1 x_{1t} + A_2 x_{2t} + A_3 \Delta x_{2t} + u_{0t}, \\ \Delta x_{1t} &= A_4 \Delta x_{2t} + u_{1t}, \\ \Delta^2 x_{2t} &= u_{2t}, \end{aligned} \tag{57}$$

where $u_t = (u_{0t}, u_{1t}, u_{2t})$ is a general stationary process. This model is rather general and encompasses as a special case the vector autoregressive model.[25]

This representation shows that the single time series can be arranged such that $x_{2t} = \sum_{j=1}^{t} \sum_{i=1}^{j} u_{2i}$ with a total of $p - r - s$ (non-cointegrating) I(2) trends. In a similar fashion, s I(1) trends are given by $x_{1t} - A_4 x_{2t} = \sum_{i=1}^{t} u_{1i}$ such that x_{1t} and x_{2t} jointly determine the common I(1) and I(2) stochastic trends of the system. Lastly, the stationary components are given by the first expression in (57), that is, the linear combinations $x_{0t} - A_1 x_{1t} - A_2 x_{2t}$ are generally integrated of order 1 but cointegrates (polynomially) with Δx_{2t}. These are the stationary relations.

Note that if $A_3 = 0$ then no differences are needed in defining the stationary relations (no polynomial cointegration) since $x_t \sim \mathrm{CI}(2, 2)$ in this case. Other special cases are encompassed in (57). For instance, all elements in x_t need not be I(2) since rows of A_4 can have zero elements. In fact, full blocks may be absent in (57).

In the following these different possibilities will be analyzed within the framework of a Gaussian vector autoregressive (VAR) model since this model appears to be especially well suited with respect to estimation and hypothesis testing.

8. System representations of cointegrated I(2) VAR models

It will be instructive to define the class of models we are now dealing with. Cointegrated I(1) and I(2) models can be characterized as restricted sub-models of the general p dimensional kth order VAR model, $A(L)x_t = \phi c_t + \varepsilon_t$. We will prefer to write it in the form[26]

$$x_t = \sum_{i=1}^{k} \Pi_i x_{t-i} + \Phi c_t + \varepsilon_t, \tag{58}$$

where it is assumed that ε_t is a sequence of *i.i.d.* zero mean errors with covariance matrix Ω. In most cases we also assume errors to be Gaussian so more compactly we have $\varepsilon_t \sim N_p(0, \Omega)$. The vector c_t, contains the possible deterministic components of the process, i.e. a constant and trend. Frequently the finite order Gaussian VAR model (58) is found to describe economic time series rather well.

The Gaussian VAR model (58) is the basis for a huge empirical literature analyzing cointegration in I(1) systems by using the so-called Johansen cointegration procedure, see Johansen (1988,1991,1995b). It is probably one of the most frequently adopted techniques in cointegration analysis and giving references to applied papers seems superfluous. Given the popularity of this technique it seems natural to extend its use to the case where variables are I(2), but before presenting results on this model, it will be instructive for comparative reasons to briefly report some important results when variables are I(1).

8.1. *Johansen's representation theorem for I(1) systems*

To simplify the analysis I will first abstract from the possible presence of deterministic components in (58). Johansen (1991) has characterized the Granger representation theorem, see Engle and Granger (1987), in terms of the Gaussian VAR described above. It is useful to reparametrize (58) as the *vector error correction model* (VECM)

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \varepsilon_t, \tag{59}$$

with $\Pi = -I + \sum_{i=1}^{k} \Pi_i$, $\Gamma_i = \sum_{j=i+1}^{k} \Pi_j$. We also define the matrix $\Gamma = I - \sum_{i=1}^{k-1} \Gamma_i$.

If the x_t vector is cointegrated it means that $A(1) = -\Pi$ has reduced rank $r < p$, so that $\Pi = \alpha\beta'$ where α and β are both full rank matrices of dimension $p \times r$.

Johansen shows that the cointegrated VAR model can also be given the alternative *common stochastic trends representation*, see also Stock and Watson (1988),

$$x_t = C \sum_{i=1}^{t} \varepsilon_i + C(L)\varepsilon_t \tag{60}$$

where $C = \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp'$ and $\beta_\perp$ (and similarly with $\alpha_\perp$) is defined to be the orthogonal complement matrix of β with dimension $p \times (p-r)$ such that $\beta_\perp'\beta = 0$ and $\text{rank}(\beta_\perp, \beta) = p$. $C(L)$ is such that $C(L)\varepsilon_t$ corresponds to a p-dimensional I(0) component. It is now easy to see that although x_t is p-dimensional, the vector series is driven by just $p - r$ common stochastic I(1) trends $\alpha_\perp' \sum_{i=1}^{t} \varepsilon_i$. In terms of observable variables it is also possible to calculate the I(1) directions as $\beta_\perp' x_t$ which are just particular linear combinations of the stochastic trends. It is necessary though, that the matrix $\alpha_\perp'\Gamma\beta_\perp$ be invertible, i.e. has full rank $p - r$, (as it can be seen from the definition of C); otherwise the system is not an I(1) system, but a system integrated of order higher than one as we shall see.

It can be easily verified from (60) that r linear combinations of the series appear stationary. These are the I(0) directions and are given by

$$\beta' \mathrm{x}_t = \beta' \mathrm{C}(L)\varepsilon_t \tag{61}$$

and hence constitute the *r cointegrating relations* of the I(1) system.[27] The single cointegration vectors are not individually identified but the space they span is.

Note that when $r=p$ this corresponds to the Π matrix being of full rank and hence x_t is really a vector I(0) series. On the other hand, $r=0$ corresponds to $\Pi=0$ so that the number of variables is just the number of common stochastic trends.

8.2. *Johansen's representation theorem for I(2) systems*

In the I(1) model II is of reduced rank and $\alpha_\perp' \Gamma \beta_\perp$ is of full rank. For the system to be I(2) it is additionally required that $\alpha_\perp' \Gamma \beta_\perp$ be of reduced rank $s<p-r$. To see this, it will be useful to follow Johansen (1992b, 1995a, 1997) and reparametrize the model (59) as

$$\Delta^2 \mathrm{x}_t = \Pi \mathrm{x}_{t-1} - \Gamma \Delta \mathrm{x}_{t-1} + \sum_{i=1}^{k-2} \Psi_i \Delta^2 \mathrm{x}_{t-i} + \varepsilon_t \tag{62}$$

which explicitly includes Γ as a parameter and where $\Psi_i = -\sum_{j=i+1}^{k-1} \Gamma_j$. Notice that this equation corresponds to a multivariate version of the univariate Hasza-Fuller regression previously presented in section 5.1, so when $\Pi=\Gamma=0$, for instance, the system consists of p non-cointegrating double unit root processes. With $\alpha_\perp' \Gamma \beta_\perp$ being of rank s it is possible to define parameter matrices ξ and η, such that the joint pair of reduced rank conditions of the I(2) model can be written:[28]

$$\Pi = \alpha\beta', \quad \text{with } \alpha, \beta \text{ being } p\times r, \quad r<p \tag{63}$$

$$\alpha_\perp' \Gamma \beta_\perp = \xi\eta', \quad \text{with } \xi, \eta \text{ being } (p-r)\times s, \quad s<(p-r). \tag{64}$$

These reduced rank restrictions naturally have testable implications as we shall see in section 9. In order to characterize the stochastic trends driving the system and the cointegration relations, it is necessary to define parameters describing the I(0), I(1), and I(2) directions of the variables. As demonstrated below, matrices $(\beta, \beta_1, \beta_2)$ can be found such that they individually provide a basis of the I(0), I(1), and I(2) relations, respectively, of the p-dimensional system. The associated dimension of each sub-system is given by r, s, and $p-r-s$; Paruolo (1996) denote these numbers the *integration indices* of the VAR.

We use the notation $\bar{\mathrm{a}} = \mathrm{a}(\mathrm{a}'\mathrm{a})^{-1}$ such that $\mathrm{a}'\bar{\mathrm{a}} = \mathrm{I}$, and $P_a = \bar{\mathrm{a}}\mathrm{a}'$ defines the projection onto the space spanned by the columns of a. Then it can be shown that $\beta_1 = \overline{\beta}_\perp \eta$, $\beta_2 = \beta_\perp \eta_\perp$, and similarly $\alpha_1 = \bar{\alpha}_\perp \xi$, and $\alpha_2 = \alpha_\perp \xi_\perp$. It is easy to see from these definitions that $\beta_\perp = (\beta_1, \beta_2)$ and $(\beta, \beta_1)_\perp = \beta_2$ such that β, β_1 and β_2 are mutually orthogonal and thus jointly describe a basis for the p-dimensional space. The $\alpha' s$ have a similar property.

The *common stochastic trends representation* (the I(2) equivalent of (60)) is now given by[29]

$$\mathrm{x}_t = \mathrm{C}_2 \sum_{j=1}^{t} \sum_{i=1}^{j} \varepsilon_i + \mathrm{C}_1 \sum_{i=1}^{t} \varepsilon_i + \mathrm{C}^*(L)\varepsilon_t \tag{65}$$

where

$$\mathrm{C}_2 = \beta_2(\alpha_2'\Theta\beta_2)^{-1}\alpha_2'. \tag{66}$$

The matrix C_1 is rather complicated but it can be shown that it satisfies the restriction

$$\beta_1'\mathrm{C}_1 = \bar{\alpha}_1'\Gamma\mathrm{C}_2 \tag{67}$$

which will show useful later. $\mathrm{C}^*(L)$ is a matrix polynomial with all roots strictly outside the unit circle. All these formulae may seem difficult but are useful in visualizing the cointegrating combinations of the variables.

Note first that in general x_t will have $p-r-s$ common stochastic $\mathrm{I}(2)$ trends given by $\alpha_2' \sum_{j=1}^{t} \sum_{i=1}^{j} \varepsilon_i$. A different way of looking at the $\mathrm{I}(2)$ trends is to consider $\beta_2'\mathrm{x}_t$. Since $\beta_2'\mathrm{C}_2 \neq 0$ it follows that $\beta_2'\mathrm{x}_t$ by construction is a $p-r-s$ dimensional $\mathrm{I}(2)$ trend; indeed, these are just linear combinations of the stochastic trends $\alpha_2' \sum_{j=1}^{t} \sum_{t=1}^{j} \varepsilon_i$.

Consider now the combinations $\beta_{2\perp}'\mathrm{x}_t = (\beta, \beta_1)'\mathrm{x}_t$. Since in this case $(\beta, \beta_1)'\mathrm{C}_2 = 0$ it follows from (65) that in general

$$(\beta, \beta_1)'\mathrm{x}_t = (\beta, \beta_1)'\mathrm{C}_1 \sum_{i=1}^{t} \varepsilon_i + (\beta, \beta_1)'\mathrm{C}^*(L)\varepsilon_t \sim \mathrm{I}(1) \tag{68}$$

so (β, β_1) reduces the integration order from 2 to 1. However, this is not the end of the story because even in the $\mathrm{I}(2)$ model it makes sense to talk about β as the vectors defining the $\mathrm{I}(0)$ relations. It appears that the combinations $\beta'\mathrm{x}_t$ can cointegrate to $\mathrm{I}(0)$ level and/or will have the property that they potentially cointegrate with $\beta_2'\Delta\mathrm{x}_t$ such that *polynomial cointegration* results, see Johansen (1995a). Note that $\beta_2'\Delta\mathrm{x}_t$ by construction is $\mathrm{I}(1)$ and non-cointegrated. These r relations read

$$\beta'\mathrm{x}_t - \delta\beta_2'\Delta\mathrm{x}_t \sim \mathrm{I}(0) \tag{69}$$

with $\delta = \bar{\alpha}'\Gamma\overline{\beta}_2$ of dimension $r \times (p-r-s)$ and hence define the $\mathrm{I}(0)$ directions.

Not all of the r stationary relations given in (69) need include the differenced $\mathrm{I}(2)$ component $\beta_2'\Delta\mathrm{x}_t$. Consider the orthogonal complement matrix of δ denoted $\delta_\perp$ which is of dimension $r \times (r-(p-r-s))$ such that $\delta_\perp'\delta = 0$. Then it follows that

$$\delta_\perp'\beta'\mathrm{x}_t \sim \mathrm{I}(0) \tag{70}$$

and hence these are not polynomially cointegrating relations. On the other hand, the $p-r-s$ relations

$$\delta'\beta'\mathrm{x}_t - \delta'\delta\beta_2'\Delta\mathrm{x}_t \sim \mathrm{I}(0) \tag{71}$$

are all polynomially cointegrating. Note that the number of polynomial cointegrating relations equals the number of I(2) trends which requires, of course, that $r \geqslant p - r - s$.

Table 6, which is motivated by Paruolo (1996), reviews the cointegration possibilities of I(2) and I(1) VAR models. As seen, the cointegration parameters of interest are given by β, β_1 and δ.

A different ay of writing the VAR is in ECM form by directly incorporating the cointegration parameters:

$$\Delta^2 x_t = \alpha(\beta' x_{t-1} - \delta\beta_2' \Delta x_{t-1}) - (\zeta_1, \zeta_2)(\beta, \beta_1)' \Delta x_{t-1} + \sum_{i=1}^{k-2} \Psi_i \Delta x_{t-i} + \varepsilon_i, \quad (72)$$

see Paruolo and Rahbek (1996). ζ_i are appropriately defined adjustment parameters. This way of writing the VAR is perhaps more intuitive since all the cointegration possibilities can be explicitly defined through error correction terms.

As for cointegrated I(1) models it must be emphasized that only the space spanned by the cointegrating vectors is identified; the single cointegration relations are unidentified. However, for I(2) systems this problem is more complex since β, β_1 and δ cannot be determined independently from each other. In section 9.4 I briefly discuss the identification problem.

Example: The Cagan model.
As a simple illustration consider calculating the integration indices of the Cagan (1956) hyper-inflation model discussed in section 3.1. It was shown that when (m_t, p_t) are I(2) series the theory predicts that the series are polynomially cointegrated. The polynomial cointegration relation is given by $m_t - p_t + \delta\Delta m_t$. As seen, in this model $p = 2$, $r = 1$, and $s = 0$: There is one I(2) trend and a single I(0) relation. Granger and Lee's multicointegration model which consists of just two series, i.e. production and sales of some commodity, can be shown to yield the same integration indices.

Table 6. Integration indices and the associated processes in VAR models. Note that the cointegration parameters are given by β, β_1 and δ.

	Dimension	Basis	Associated processes
I(1)-model			
I(0)-relations	r	β	$\beta' x_t$
I(1)-relations	$p - r$	$\beta_\perp$	$\beta_\perp' x_t$
I(2)-model			
I(0)-relations	r	β	$\beta' x_t - \delta\beta_2' \Delta x_t \left\{ \begin{array}{ll} r - (p - r - s): & \delta_\perp' \beta' x_t \\ p - r - s: & \delta' \beta' x_t - \delta' \delta \beta_2' \Delta x_t \end{array} \right\}$
I(1)-relations	s	β_1	$\beta_1' x_t$
I(2)-relations	$p - r - s$	β_2	$\beta_2' x_t$

Source: Partially taken from Paruolo (1996).

8.3. *Deterministic components*

In the presentation given above I have abstracted from the possible presence of deterministic components in the processes. Without going too much into the details I briefly outline how deterministic trends may potentially enter the models. First we consider the case where the VAR model (62) is augmented with an intercept, μ, that is,

$$\Delta^2 x_t = \mu + \Pi x_{t-i} - \Gamma \Delta x_{t-1} + \sum_{i=1}^{k-2} \Psi_i \Delta^2 x_{t-i} + \varepsilon_t. \tag{73}$$

We want to understand the meaning of μ in this model. Paruolo (1996) considers the following factorization of the intercept:

$$\mu = \alpha\mu_0 + \alpha_1\mu_1 + \alpha_2\mu_2, \quad \text{with } \mu_d = \bar{\alpha}_d'\mu, \quad \text{for } d = 0, 1, 2. \tag{74}$$

Hence μ is projected onto the spaces spanned by α, α_1, and α_2, with the respective dimensions r, s, and $p - r - s$. The μ_d-terms will indicate the extent to which the intercept will influence the associated I(d) space. Generally, the I(2) space will have a quadratic trend, the I(1) space will have a linear trend, and the I(0) space will have no trend. To see this, consider the stochatic trends representation of the I(2)-VAR given in (65) where ε_i is replaced by $\varepsilon_i + \mu$:

$$x_t = \tau_2 t^2 + (\tau_2 + C_1\mu)t + C_2 \sum_{j=1}^{t}\sum_{i=1}^{j} \varepsilon_i + C_1 \sum_{i=1}^{t} \varepsilon_i + C^*(L)\varepsilon_t \tag{75}$$

$$\tau_2 = \tfrac{1}{2} C_2 \mu. \tag{76}$$

It follows that in general x_t will have quadratic trends but more can be said. Note simply from the decomposition (74), that $\tau_2 = \frac{1}{2}C_2\mu = \frac{1}{2}\beta_2(\alpha_2'\Theta\beta_2)^{-1}\alpha_2'\alpha_2\mu_2$ so it is the presence of $\mu_2 \neq 0$ that causes the quadratic trend. By focusing explicitly on the I(2) directions $\beta_2' x_t$ these are seen to have quadratic trends,[30] unless $\mu_2 = 0$.

Next we address the I(1) directions $\beta_1' x_t$. Naturally there will be no quadratic trend in these directions since $\beta_1'\tau_2 = 0$. The C_1 matrix is rather complicated, however, unless μ_1 and μ_2 are both equal to zero, $\beta_1' x_t$ will have a linear trend.

Finally, by using the fact that $\beta' C_1 = \bar{\alpha}'\Gamma C_2$ it can be shown that the stationary polynomially cointegrating relations $\beta' x_t - \delta\beta_2'\Delta x_t$ will be free of deterministic trends.

Rahbek *et al.* (1998) choose a different parametrization of the model corresponding to the factor representation $x_t = x_t^0 + \gamma_0 + \gamma_1 t$, where x_t^0 is the stochastic part of the process defined in (65), that is,

$$x_t = \gamma_0 + \gamma_1 t + C_2 \sum_{j=1}^{t}\sum_{i=1}^{j} \varepsilon_i + C_1 \sum_{i=1}^{t} \varepsilon_i + C^*(L)\varepsilon_t. \tag{77}$$

As in Paruolo (1996), who assumes $\mu_2 = 0$, it is seen that this model is also unable to produce quadratic trends. However, by choosing the formulation (77) the parameters γ_0 and γ_1 can vary freely and consequently even the polynomially cointegrating I(0) relations may contain linear trends and hence will be trend stationary. This appears to be especially useful with respect to hypothesis testing about the cointegration rank when deterministics are considered in the processes.

9. Estimation and hypothesis testing using the Johansen procedure

In empirical applications dealing with I(2) variables in systems, the Johansen I(2) procedure seems to be dominating. We will here give some of the main results for this procedure. Also there is a growing number of empirical contributions. Understanding the complex nature of the models and the hypotheses to be defined in the various spaces of different integration order is difficult. We start by presenting a natural hierarchical ordering of the testable hypotheses which is essential in understanding the subsequent issues related to estimation and hypothesis testing.

9.1. *A hierarchical ordering of testable sub-models*

As we have seen in section 8.2 the I(1) and I(2) models result from particular restrictions of the general VAR model which may be formulated either as in (59) or as in (62). In table 7 the various sub-models that can be formulated in the I(1) model are displayed, and in table 8 the sub-models of the I(2) VAR are reported separately given a fixed value of r, that is, the cointegration rank defined from the I(1) model is fixed. The structure of the two tables are basically the same but the focus in each case is on the I(1) and I(2) reduced rank conditions, respectively. Notice in particular, that the least restricted model is H_p, the model H_0 is a VAR in first differences (no I(0) relations), and finally the model H_{00} is a VAR in second differences (all p relations are I(2)). All the intermediate cases describe situations with cointegration.

Observe also that a superscript '0', for instance in H_r^0, is used to indicate the submodel of H_r where α and β have full rank, r. No superscript indicates that the

Table 7. Hypotheses in the I(1) model

	Restriction	Parameter space	Comments
H_p	None	$(\Pi, \Gamma_1, \Gamma_2, \ldots, \Gamma_{k-1}, \Omega)$	VAR in levels
H_r	$\Pi = \alpha\beta', \alpha, \beta\ p \times r, r = 0, 1, \ldots, p$	$(\alpha, \beta, \Gamma_1, \Gamma_2, \ldots, \Gamma_{k-1}, \Omega)$	Reduced rank
H_0	$\Pi = 0$	$(\Gamma_1, \Gamma_2, \ldots, \Gamma_{k-1}, \Omega)$	VAR in first differences
H_r^0	$\Pi = \alpha\beta'$, $\text{rank}(\alpha\beta') = r$	$(\alpha, \beta, \Gamma_1, \Gamma_2, \ldots, \Gamma_{k-1}, \Omega)$	α, β of full rank r

The relation between the various hypotheses:

$$H_r = \cup_{i=0}^{r} H_i^0$$
$$H_0 \subset H_1 \subset \ldots \subset H_p$$

rank of the matrices is less than or equal to r: $H_r = \cup_{i=0}^{r} H_i^0$. The sub-models $H_{r,s}$ are indexed in a similar way.

Table 9 is taken from Johansen (1995a) and shows how the various I(1) and I(2) sub-models are related. In particular, it should be noticed that by moving from H_{00} to the right, down to H_{10} to the right, and so on, defines models which become less and less restricted as one proceeds.

9.2. *Determining the integration indices*

The ordering of hypotheses defined in section (9.1) is useful in determining the integration indices, that is, the dimension of each of the I(d) spaces, $d = 0, 1, 2$. This is the focus of the present section whilst the subsequent section addresses hypotheses about the parameters and the cointegration spaces after the integration indices have been fixed.

Table 8. Hypotheses in the I(2) model. All hypotheses are conditional on rank$(\alpha\beta') = r$.

	Restriction	Parameter space	Comments
$H_{r,p-r}$	No further	$(\alpha, \beta, \Gamma, \Psi_1, \Psi_2, \ldots, \Psi_{k-2}, \Omega)$	Cointegrated I(1) VAR model. $H_{r,p-r} = H_r^0$
$H_{r,s}$	$\alpha_\perp' \Gamma \beta_\perp = \xi\eta'$, ξ, η $(p-r) \times s$, $s = 0, 1, \ldots, p-r$	$(\alpha, \beta, \Gamma, \Psi_1, \Psi_2, \ldots, \Psi_{k-2}, \Omega)$ subject to restriction	Reduced rank of I(2) model
$H_{r,0}$	$\Gamma = 0$	$(\alpha, \beta, \Psi_1, \Psi_2, \ldots, \Psi_{k-2}, \Omega)$	No I(1) trends
$H_{r,s}^0$	$\alpha_\perp' \Gamma \beta_\perp = \xi\eta'$, rank$(\xi\eta') = s$	$(\alpha, \beta, \Gamma, \Psi_1, \Psi_2, \ldots, \Psi_{k-2}, \Omega)$ subject to restriction	ξ, η of full rank s

The relation between the various hypotheses:

$$H_{r,s} = \cup_{i=0}^{s} H_{r,i}^0$$
$$H_{r,0} \subset H_{r,1} \subset \ldots \subset H_{r,p-r} = H_r^0 \subset H_r$$

Table 9. The relations between the various I(1) and I(2) models.

$p-r$	r													
p	0	H_{00}	$\subset$	H_{01}	$\subset$	$\ldots$	$\subset$	$H_{0,p-1}$	$\subset$	$H_{0,p}$	$= H_0^0$	$\subset$	H_0
													$\cap$
$p-1$	1			H_{10}	$\subset$	$\ldots$	$\subset$	$H_{1,p-2}$	$\subset$	$H_{1,p-1}$	$= H_1^0$	$\subset$	H_1
													$\cap$
											$\vdots$		$\vdots$
													$\cap$
1	$p-1$							$H_{p-1,0}$	$\subset$	$H_{p-1,1}$	$= H_{p-1}^0$	$\subset$	H_p
$p-r-s$		p		$p-1$		$\ldots$		1		0			

Source: Johansen (1995a)

It is well known from the I(1) analysis, see Johansen (1988, 1991), that likelihood analysis of the models H_r can be conducted as a combination of regression and reduced rank regression;[31] regression is conducted to filter the data such that nuisance parameters are eliminated whereas reduced rank regression explicitly aims for determining the cointegration rank. The I(2) analysis is discussed in Johansen (1995a, 1997) and follows a similar train of thought, but now two reduced rank conditions need to be examined, i.e. the restrictions (63) and (64) which define the I(2) model $H_{r,s}$. However, the analysis is further complicated by the fact that the second reduced rank condition associated with the matrix $\alpha_\perp'\Gamma\beta_\perp = \xi\eta'$ depends upon the first reduced rank condition of $\Pi = \alpha\beta'$. Joint estimation of the indices r and s which define these matrices is therefore rather complicated. Instead, Johansen has suggested a two step procedure; the basic structure of the reduced rank problems is the following:

Step 1 consists of solving the reduced rank problem associated with $\Pi = \alpha\beta'$ and calculating for *each* value of $r = 0, \ldots, p-1$, the estimates $\hat{\alpha}_r, \hat{\beta}_r, \hat{\alpha}_{r\perp}$ and $\hat{\beta}_{r\perp}$. Subscript r indicates for which value of r the calculated estimates.

Step 2 considers subsequently the reduced rank regression problem associated with $\alpha_\perp'\Gamma\beta_\perp = \xi\eta'$ which is solved for $s = 0, 1, \ldots, p-r-1$, by replacing the unknown matrices $\alpha_\perp$ and $\beta_\perp$ by the estimates obtained in step 1. Note that estimates are available for each value of r.

From this triangular array of sub-models corresponding to the hypotheses $H_{r,s}^0$ it remains to be determined which values of r and s to choose. To do so we first need to be specific about the reduced rank regression problems of steps 1 and 2. I will not go into technical details about this, but just note that the reduced rank problem of step 1 amounts to solving an eigenvalue problem after the influence of the nuisance parameters Γ, Ψ_i, $i = 1, 2, k-2$, has been eliminated by prefiltering.[32] The eigenvalue problem delivers the eigenvalues $1 > \hat{\lambda}_1 > \hat{\lambda}_2 > \ldots > \hat{\lambda}_p > 0$ with the associated eigenvectors $(v_1, v_2, \ldots, v_p)$. The Maximum Likelihood (ML) estimator of β for a given value of r is then given by $(\hat{\beta}_r = (v_1, v_2, \ldots, v_r)$, that is, the eigenvectors corresponding to the r largest eigenvalues. The estimates $\hat{\alpha}_r$ can also be found. The maximized value of the likelihood function can be shown to be proportional to

$$L_{\max}^{-2/T} \propto \prod_{i=1}^{r} (1 - \hat{\lambda}_i) \tag{78}$$

and hence the likelihood ratio (LR) test, $Q(H_r|H_p)$ of the I(1) model H_r with $\text{rank}\Pi \leqslant r$ against the unrestricted VAR, H_p, is given by

$$Q_r = -2\ln Q(H_r \mid H_p) = -T \sum_{i=r+1}^{p} \ln(1 - \hat{\lambda}_i), \quad r = 0, \ldots, p-1. \tag{79}$$

By the presence of I(1) (but no I(2)) trends the distribution of this statistic is the Johansen trace statistic well known from the I(1) analysis. Unfortunately the distribution will be different by the presence of I(2) trends and since the number of these is unknown we cannot use the Q_r statistic to determine r in practice. Instead we have to proceed to step 2 of the procedure after Q_r is calculated for each possible value of $r = 0, \ldots, p-1$.

In step 2 the variables are transformed using the estimates $\hat{\alpha}_r, \hat{\beta}_r, \hat{\alpha}_{r\perp}$, and $\hat{\beta}_{r\perp}$ which replace the unknown population parameters in the subsequent calculations. The reduced rank regression problem now consists of regressing $\hat{\alpha}'_{r\perp}\Delta^2 \mathrm{x}_t$ on $\hat{\beta}'_{r\perp}\Delta \mathrm{x}_{t-1}$ after appropriate prefiltering has been made of the series.[33] Again this problem can be formulated as a particular eigenvalue problem yielding the eigenvalues $1 > \hat{\rho}_1 > \hat{\rho}_2 > \ldots > \hat{\rho}_{p-r} > 0$ and eigenvectors $(w_1, w_2, \ldots, w_{p-r})$. The ML estimate of η is obtained as $\hat{\eta} = (w_1, w_2, \ldots, w_s)$. A solution can be found for $\hat{\xi}$ as well. For given values of r, α, and β the contribution to the likelihood function is proportional to

$$L_{\max}^{-2/T} \propto \prod_{i=1}^{s} (1 - \hat{\rho}_i). \tag{80}$$

Therefore, the LR test of the models $H_{r,s}$ conditional on the model H_r^0, is given by

$$Q_{r,s} = -2 \ln Q(H_{r,s} \mid H_r^0) = -T \sum_{i=s+1}^{p-r} \ln(1 - \hat{\rho}_i), \quad s = 0, \ldots, p-r-1. \tag{81}$$

This test assumes that r is known and conditional on the value of r the distribution is again given by the I(1) Johansen trace statistic. But since r is unknown in practice it is more relevant to test $H_{r,s}$ (corresponding to the models with $r \leqslant p$, and $s \leqslant p - r$) against H_p (the unrestricted VAR) by focusing on the statistic

$$S_{r,s} = -2 \ln Q(H_{r,s} \mid H_p) = Q_{r,s} + Q_r. \tag{82}$$

This is valid since the hypotheses are nested. Johansen (1995a) shows that the distribution of $S_{r,s}$ is given by the sum of the trace of functionals of standard Brownian motions and integrated standard Brownian motion processes. In fact, this corresponds to multivariate versions of the Hasza-Fuller distribution discussed in section 5.1 within a univariate context. Generally $S_{r,s}$ will depend upon the integration indices, s and $p - r - s$, i.e. the number of I(1) and I(2) trends in the system. The distributions for the case with no deterministics in the model are given in Johansen (1995a, table 3).

The $S_{r,s}$ statistic can be used to determine the integration indices. The model $H_{r,s}$ is rejected if $H_{i,j}$ is rejected for all $i < r$ and $j \leqslant s$. The idea is therefore to use the hierarchical ordering given in table 9 from left to right and from top to bottom, i.e. following the sequence starting with H_{00} to the right, down to H_{10} to the right *etc*. The integration indices $\hat{r}, \hat{s}$ are determined as the first pair of r, s which is not being rejected.[34]

It can be shown that this order of testing hypotheses which become less and less restricted is a consistent procedure and with a type-1 error which can be controlled by choice of the significance level.[35]

Deterministic components. As previously mentioned, Paruolo (1994, 1996) focuses on the VAR model discussed above augmented with an intercept μ. Basically the presence of an unrestricted intercept is accommodated for by correcting the regressors and the regressand for an intercept prior to the reduced rank regression problems. In Paruolo (1996, table 13), the asymptotic critical values for the corresponding $S_{r,s}$ test with μ being unrestricted are reported. If restrictions are wanted on the intercept such that there is no quadratic trends in the data, $\mu_2 = 0$, or no linear trend, $\mu_1 = \mu_2 = 0$, then the trending behaviour of the different models has to be decided jointly with the determination of the integration indices. Due to the complexity of this analysis it will not be pursued here, but the interested reader is referred to Paruolo (1996) for the details. Note, however, that when the restriction $\mu_2 = 0$ is imposed yet another set of critical values needs to be used to validate the analysis, Paruolo (1996, table 6). If $\mu_1 = \mu_2 = 0$ is imposed in estimation neither linear nor quadratic trends are allowed for and Paruolo's table 5 should be used to determine the integration indices in this case.[36]

In practical situations the factor representation of Rahbek *et al.* (1998) given in (77) seems to be the most appropriate as the benchmark for tests about the cointegration rank. In their VAR model the presence of quadratic trends are excluded but linear trends are allowed, even in the cointegration space. The common stochastic trends representation (77) can be reparametrized such that the model reads

$$\Delta^2 x_t = \alpha\beta^{*\prime} x^*_{t-1} - \Gamma\Delta x_{t-1} + \mu_0 + \sum_{i=1}^{k-2} \Psi_i \Delta^2 x_{t-i} + \varepsilon_t \tag{83}$$

where $\beta^{*\prime} = (\beta', \beta_0')$ and $x^*_{t-1} = (x'_{t-1}, t)$ and with a particular restriction on the intercept term μ_0. How exactly the parameters μ_0, β_0 of (83) are related with the parameters γ_0, γ_1 of the model (77) can be seen from their paper. The procedure adopted to determine the integration indices for the case with no deterministics can be straightforwardly extended to cover the above situation. Variables just need to be appropriately redefined (as seen in (83)) by taking into account the restrictions on μ_0. LR statistics in each step of the analysis define the LR test of $H^*_{r,s}$ against H^*_p (a '*' signifies that linear trends are allowed in the various directions) and is given by

$$S^*_{r,s} = Q^*_{r,s} + Q^*_r. \tag{84}$$

Rahbek *et al.* (1998) report the distribution of this statistic as well as the relevant asymptotic critical values (their tables 1 and 4).

Interestingly, the distributions are similar with respect to the actual deterministic components present (given, of course, that there are no quadratic trends). Hence, after the cointegration indices have been fixed, it is rather simple to test whether the trend-stationarity of the cointegration relations implied by (83) is actually satisfied by the data. Absence of the trends means that $\beta_0 = 0$ and hence the idea is to compare the likelihood values for the model (83) and the restricted model which is obtained by replacing x^*_t by x_t. The likelihood ratio test is given by

$$Q_{\beta_0=0} = T\sum_{i=1}^{r} \ln((1-\hat{\lambda}_i^*)/(1-\hat{\lambda}_i)) \tag{85}$$

where $\hat{\lambda}_i^*$ and $\hat{\lambda}_i$ correspond to the first step eigenvalues for the restricted and unrestricted models, respectively. The test is asymptotic χ^2 with r degrees of freedom.

9.3. *Examples. Estimation of the integration indices and cointegration relations in the various data sets*[37]

German hyperinflation, 1920:1–1923:6.
The Cagan model predicts that for the pair ($p = 2$) of time series, m_t and p_t, $s = 0$, and $r = 1$, that is, one I(0) relation and $p - r - s = 1$ I(2) trend. For the present data set a VAR(3) was estimated with a constant restricted not to produce quadratic trends. This corresponds to the model (73) with $\mu_2 = 0$. The $S_{r,s}$-trace statistics (82) which are associated with the hierarchical hypotheses in table 9 are displayed in table 10.

According to the hierarchical principle the first hypothesis tested is H_{00}. The test value is 45.09 and is clearly significant.[38] The next test yields 14.38 which is insignificant. The conclusion is therefore that $r = 0$ and $s = 1$, that is, m_t and p_t have a common I(2) trend but there is no polynomial cointegration with Δm_t or Δp_t.

US housing data set, 1968:1–1994:12.
A VAR with 12 lags was estimated for the cumulated housing starts and housing completion series. The choice of 12 lags was found to whiten the errors reasonably well. Note that when time series are potentially multicointegrated in the sense of Granger and Lee, the appropriate way of defining the VAR, according to Engsted and Johansen (1997), is to consider the VAR for the integrated time series, see also Engsted and Haldrup (1998). Since the cumulated

Table 10. Joint tests of the cointegration indices. German hyperinflation data set, 1920:1–1923:6.

$p-r$	r	$S_{r,s}$		Q_r
2	0	45.09***	**14.38**	13.03
		36.12	*22.60*	*15.41*
1	1		5.55	2.94
			12.93	*3.84*
$p-r-s$		2	1	0

Note: Numbers in italics are 95 per cent quantiles, (Paruolo (1996), table 6.). r and $p-r-s$ are the number of I(0) and I(2) components.

series will have a trend by construction I decided to allow for a trend also amongst the $\mathsf{I}(0)$ relations as in (52), c.f. the Rahbek *et al.* (1998) specification. Table 11 summarizes the results.

As seen the first hypothesis which cannot be rejected at the 90% level suggests the indices $(r, s) = (1, 0)$. Hence the system has one $\mathsf{I}(2)$ trend and a single multicointegrating relation. This relation corresponds to Eq. (69) and is given by

$$\sum_{j=1}^{t} y_j - 0.972 \sum_{j=1}^{t} x_j - 3.908 y_t - 4.022 x_t + const + trend \sim \mathsf{I}(0)$$

so the estimate of the percentage of housing construction starts never being completed is 2.8%. The value Lee (1996) imposed a priori was 2%.

Danish money demand data set, 1974:1–1993:4.

The present information set is of course much larger than in the previous two examples which naturally complicates the structural analysis of the equilibrium relations. The analysis presented here is for the purpose of illustration and a more detailed analysis can be found in Juselius (1998) who also consider the possibility of structural shifts in the data. For ease of exposition we abstract from such generalizations. A $\mathsf{VAR}(4)$ model was fitted to the data set consisting of the variables $\mathrm{x}_t = (m_t, p_t, y_t, i_{bt}, i_{dt})'$. The analysis allowed for seasonal dummy variables and an intercept restricted such that quadratic trends $(\mu_2 = 0)$ were excluded. Table 12 shows the test results.

In this case it is seen that the hierarchical test procedure suggests one $\mathsf{I}(2)$ trend and 4 cointegrating relations,[39] $(r, s) = (4, 0)$. Observe that this implies a total number of two unit roots in the entire VAR model.

It is easy to see that the variables constituting the $\mathsf{I}(2)$ trend is some combination of the m_t and p_t variables. The $\mathsf{I}(2)$ directions are given by $\beta_2' \mathrm{x}_t$ and in the present case this is estimated as $\hat{\beta}_2' \mathrm{x}_t = (1, 1.07, -0.02, 0.01, 0.01)\mathrm{x}_t$. In fact, the $\mathsf{I}(2)$ trend is close to an average of the m_t and p_t series.

Since we assume there is a total of 4 stationary relations it may show useful to separate these into the $r - (p - r - s) = 4 - (5 - 4 - 0) = 3$ relations $\delta_\perp' \beta' \mathrm{x}_t$

Table 11. Joint tests of the cointegration indices. US Housing data set, 1968:1–1994:12.

$p-r$	r		$S_{r,s}$	Q_r
2	0	76.90***	32.72*	22.55*
		47.60	*34.36*	*25.43*
1	1		**13.90**	4.20
			19.87	*12.49*
$p-r-s$		2	1	0

Note: Numbers in italics are 95 per cent quantiles, (Rahbek *et al.* (1998), tables 1 and 4). r and $p-r-s$ are the number of $\mathsf{I}(0)$ and $\mathsf{I}(2)$ components.

Table 12. Joint tests of the cointegration indices. Danish money data set 1974:1–1993:4.

$p-r$	r	$S_{r,s}$					Q_r
5	0	252.64***	206.70***	163.58***	133.88***	114.65***	100.12
		171.89	*142.57*	*117.63*	*97.97*	*81.93*	*68.52*
4	1		180.20***	137.12***	95.32***	76.70***	61.59***
			116.31	*91.41*	*72.99*	*57.95*	*47.21*
3	2			111.43***	69.02***	41.45**	32.31**
				70.87	*51.35*	*38.82*	*29.68*
2	3				57.29***	29.47***	15.36*
					36.12	*22.60*	*15.41*
1	4					**10.55**	5.63
						12.93	*3.84*
$p-r-s$		5	4	3	2	1	0

Note: Numbers in italics are 95 per cent quantiles, (Paruolo(1996), table 6). r and $p-r-s$ are the number of I(0) and I(2) components.

which are I(0) without the differences, and the remaining single relation $\delta'\beta'x_t - \delta'\delta\beta_2'\Delta x_t$ which is polynomially cointegrating, see (70)–(71) and table 6. By appropriate choice of normalization it can be found that

$$\delta_\perp'\beta' x_t = \begin{pmatrix} 1 & -1.22 & -0.18 & -4.49 & 48.18 \\ -0.95 & 1 & 0.36 & -27.08 & 27.80 \\ -0.51 & 0.51 & 1 & -6.17 & 6.46 \end{pmatrix} (m_t, p_t, y_t, i_{bt}, i_{dt})'$$

$$\begin{aligned}\delta'\beta'x_t - \delta'\delta\beta_2'\Delta x_t = &(1 \quad 1.18 \quad 6.54 \quad -9.94 \quad -270.94)x_t - \\ &(215.99 \quad 231.52 \quad -3.99 \quad 2.57 \quad 1.61)\Delta x_t.\end{aligned}$$

Of course, one should be careful about making structural interpretations based on the above estimates since the single cointegrating relations are unidentified. However, the estimates suggest several overidentfying restrictions. For instance one may consider addressing a real specification of the money variable, at least this seems to hold for the first three relations. Moreover, the second and third relation suggests a specification where the difference between the bond and deposit rates is considered, and finally the polynomially cointegrating relation, due to the magnitude of the parameters, seems to suggest a cointegrating relationship between real money growth, $\Delta(m_t - p_t)$ and the deposit rate, i_{dt}. These hypotheses are of course testable.

9.4. *Hypothesis testing of the model parameters*

When the cointegration ranks have been determined it is of interest to test restrictions on the cointegration vectors and the adjustment coefficients. Since the cointegration and non-stationary relations are defined such that e.g. β_1 and β_2 are not in the space spanned by β, it is clear that any restriction on β will have implications for the other vectors as well. When using the two step procedure this becomes apparent since the analysis of the vectors β_1 and β_2 is considered after the choice of β has been made.

Recently Johansen (1997) considered a particular parametrization of the Gaussian VAR model (62) which allows all model parameters to vary freely. This makes the likelihood analysis simpler as the model can be formulated as a general non-linear regressions problem upon which ML estimates can be obtained. Interestingly, it can be shown that the computationally simpler two step procedure is fully efficient in estimating the I(2) model, see also Paruolo (1997) for the model with a constant. This means that likelihood ratio tests concerning hypotheses on β from the I(1) analysis are valid although I(2) components are present and hence the distributions are asymptotically χ^2 distributed.

However, in discussing hypotheses on the model parameters it is important to realize, that β and α are unidentified as any non-singular matrix ζ can be chosen whereby $\alpha\zeta$ and $\zeta^{-1}\beta'$ will give the same value of $\Pi = \alpha\beta'$. Therefore, in order to interpret the single cointegration vectors, (exactly) identifying restrictions need to be imposed on β. An easy way of doing this is to normalize the single vectors such that the cointegration relations can be solved, for instance by chosing the identification scheme $\beta_c = (\mathrm{I}_r \vdots \mathrm{B}')'$. Of course any exactly identifying restrictions can be chosen; generally the restrictions can be phrased as $\beta_c = \beta(\mathrm{c}'\beta)^{-1}$.

In doing so Johansen (1995a) shows that[40]

$$T(\hat{\beta}_c - \beta) \longrightarrow \int_{\Xi} N(0, \Xi)\, \mathrm{d}P(\Xi), \tag{86}$$

that is, the distribution is mixed Gaussian with Ξ being the mixing variate matrix. An estimate of the covariance matrix can be obtained by

$$(\mathrm{I} - \hat{\beta}_c \mathrm{c}') \left[\sum_{t=1}^{T} \mathrm{R}_t \mathrm{R}_t' \right]^{-1} (\mathrm{I} - \mathrm{c}\hat{\beta}_c') \otimes (\hat{\alpha}_c' \hat{\Omega}^{-1} \hat{\alpha}_c)^{-1} \tag{87}$$

where R_t is the vector of residuals from the regression of x_{t-1} on lagged first differences (and deterministics). This result means that for the normalization chosen it is easy to do asymptotic inference on the single coefficients associated with β_c by using the standard normal distribution for single t tests.[41]

With respect to the other cointegration parameters similar identification problems arise but of course these problems cannot be solved independently as the parameters are linked together in a particular way. For instance, observe that the parameter δ associated with polynomial cointegration can be defined from $\alpha\delta\beta_2'$ in the representation (72) but only the product of the matrices is identified from the outset. However, whenever α and β_2 are identified also δ is identified. But this means that the whole identification problem for δ is solved when $(\beta, \beta_1, \beta_2)$ are identified since the identification of β also identifies[42] α.

The estimator of δ can be shown to be super-consistent, but unfortunately it appears that inference about δ can be less easily carried out as the distribution of the estimator is not mixed Gaussian. Any test involving this parameter is therefore difficult, see Boswijk (1997) and Paruolo (1997) for details.

Finally, also the estimate of β_1 can be shown to be T-consistent and mixed Gaussian. However, the β directions are probably the most interesting relations from an economic perspective.

With respect to general hypotheses of *overidentifying restrictions* on the cointegration relations (apart from restrictions involving the polynomial cointegration parameter δ), these can for instance be of the form

$$\beta = (\mathrm{H}_1\phi_1, \mathrm{H}_2\phi_2) \tag{88}$$

where different restrictions are imposed on the single vectors. ϕ_i are the freely varying parameters of the restricted model. The restrictions mean that also the other cointegration relations are restricted as β_1 and β_2 are defined orthogonal to β. In practice likelihood inference is based on solving the eigenvalue problems for the constrained model. LR tests can now be conducted by comparing the restricted and unrestricted eigenvalues like in the trend test (85). The LR-test statistic is χ^2 distributed with degrees of freedom corresponding to the number of over-identifying restrictions.

Similarly hypothesis testing can be undertaken with respect to the adjustment coefficients. In the I(1) analysis such tests are frequently associated with testing for weak exogeneity. In this analysis, if rows of the α matrix are found to have zero elements, it means that the corresponding variables can be treated as weakly exogenous with respect to the cointegration parameters, such that a partial system can be analyzed without loss of efficiency. Paruolo and Rahbek (1996) have developed the theory for the I(2) model. Generally they find that the weak exogeneity restrictions can take many different forms depending on the parametrization of the model. For instance, in the parametrization (62) the $(p-m)$ dimensional process $\mathrm{b}'\mathrm{x}_t$ is weakly exogenous for the cointegration parameters (β, β_1, δ) when $\mathrm{b}'\Pi = \mathrm{b}'\Gamma = 0$. A necessary condition for this to be the case is that the weakly exogenous process has dimension such that $m \geqslant r+s$. In general the maximum number of weakly exogenous variables can be shown to equal the number of I(2) trends $p-r-s$. In empirical studies this number is frequently found to be rather small; a negative implication of this is that it becomes difficult to formulate reduced partial models without loosing efficiency. The likelihood analysis and testing strategies are discussed in Paruolo and Rahbek's paper which the interested reader can consult for the details.

9.5. *Examples. Testing hypotheses on the cointegration relations in the Danish money demand data set*

In section 9.3 different restrictions of the cointegrated I(2) VAR for the Danish money demand data set were suggested. Here I briefly demonstrate how the hypothesis that a real specification of money demand will suffice in the specification of the model. If such a restriction can be validly imposed the entire analysis simplifies considerably since in place of m_t and p_t the standard I(1) analysis can be carried out for the variables $[(m_t - p_t), y_t, \Delta p_t, i_{bt}, i_{dt}]$.

The hypotheses tested correspond to considering in (88) the same restrictions

$H_i = H$, for all $i = 1, 2, \ldots, 4$. In the present case H is given by

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and ϕ_i are 4×1 matrices of parameters. Note that this set of hypotheses does not identify the model structure; this would require different restrictions on each of the cointegration relations.

The LR test of the above hypothesis yields the test value $LR^* = 2.53$ and is distributed as $\chi^2(4)$ and hence is insignificant. The I(1) analysis for the transformed (real) system is carefully analyzed in Juselius (1998).

10. Other results for I(2) variables in systems

This section addresses very briefly some remaining results applying to I(2) systems.

In sections 8 and 9 I have focused attention on the I(2) analysis relating to the Gaussian VAR model. In a recent paper Chang (1997) generalizes Chang and Phillips (1995) and addresses the problem about efficiently estimating (in Phillips' (1991) sense) VAR models when the model has an unknown mixture of I(0), I(1), and I(2) components, i.e. no prior or pretest information is required with respect to the number of unit roots and the integration indices. Chang shows that his estimator (denoted the *residual-based fully modified* -VAR (RBFM-VAR) estimator, see also Chang and Phillips (1997)) is identical to that of the maximum likelihood estimator under Gaussian assumptions with the precise knowledge about the order of unit roots in individual series and the cointegrating relationships in the model. Simulations show that the estimator is behaving well in large samples, and reasonably well in finite samples.

The triangular system representation was introduced in section 7 and has been the benchmark representation for a number theoretical studies of I(2) systems. For instance, Stock and Watson (1993) used the representation (assuming known integration indices) in order to propose a likelihood-based estimation method by augmenting the single equations by leads and lags of the differenced series. Within this framework they showed that their estimator is asymptotically efficient, even when the errors are not necessarily Gaussian. Kitamura (1995) showed, for particular cointegrated I(2) models, how optimal inference applies when other formulations of triangular systems are used.

Although different representations of multivariate systems are typically derived to facilitate statistical analysis, these may also be of separate interest as they provide different insights into the different features of such systems. Many different representations including common stochastic trends, triangular, and error correction representations can be derived from the so-called Smith-McMillan form of a rational polynomial matrix, see Kailath (1980). This tool has been applied to I(1) models, see Yoo (1986) and Hylleberg and Mizon (1989) as well as to models exhibiting seasonal cointegration, see Engle and Yoo (1991) and Hylleberg *et al.* (1993). In the context of I(2) systems this approach has been considered by, *inter alia*, Engle and Yoo (1991) and Haldrup and Salmon (1998).

11. Final comments

In this paper I have provided a survey of recent developments in the econometric analysis of time series models involving I(2) series. The review has deliberately been selective, but I have tried to address the topics I believe are potentially useful for practitioners. Especially with respect to multivariate analysis there exist other contributions to this literature, for instance Gregoir and Laroque (1993, 1994) have derived a rather comprehensive and self-contained procedure for models with polynomial cointegration which includes all of the elements: representation, estimation and hypothesis testing. In relation to the multivariate analysis I decided to focus especially on the Johansen procedure since this has proven to be the most commonly used technique in empirical applications.

It is important to realize the limitations of the I(2) analysis since in general the I(2) property is found to describe only a limited number of time series like e.g. nominal time series and stock variables. However, as I have demonstrated there are in fact many economic theories and models indicating that relations may exist between I(2) and I(1) variables. When I(2) series are potentially present it is important to consider what to do next since inappropriate treatment of the I(2) components may invalidate the I(1) analysis. Sometimes it may be useful to transform the time series *a priori* such that I(2)-ness is excluded, for instance by defining economic models in real rather than in nominal terms or by including the differenced I(2) variables in the analysis. When such transformations are not obvious or it is desirable to test the restrictions on the data, the I(2) analysis becomes especially important.

Acknowledgements

This review paper was written while the author was visiting the European University Institute in Florence in the spring 1997. I would like to thank the Department of Economics for its hospitality during my stay. A longer and more detailed version of the paper can be obtained from the author upon request. The comments and suggestions of three referees and the editor are acknowledged. I also appreciate several stimulating discussions and comments on I(2) issues with Peter Boswijk, Clive Granger, H.C. Kongsted, Søren Johansen, Katarina Juselius, Paolo Paruolo, Anders Rahbek, and Andreu Sansó i Roselló. Some of the data and computer programs used in the applications were kindly made available to me by Tom Engsted, Katarina Juselius, Tae-Hwy Lee, and Byungchul Yu.

Notes

1. At this stage we assume (for simplicity) that each component of x_t is $\mathsf{I}(d)$. Later we will permit x_t to have components of different orders of integration and with d refering to the components of maximal integration order.
2. Note that since we have a mixture of $\mathsf{I}(1)$ and $\mathsf{I}(2)$ variables cointegrating the cointegration possibility (3) actually goes beyond the original definition of cointegration given in Engle and Granger (1987). When case (3) occurs I will just refer to x_{1t}, x_{2t} as being $\mathsf{I}(2)$ cointegrated and hence indicating that the maximial order of integration of the relevant variables is $\mathsf{I}(2)$.
3. In Engsted and Haldrup (1997) we assume y_t^* is determined from a vector of forcing $\mathsf{I}(1)$ and $\mathsf{I}(2)$ variables.
4. For instance Engsted (1993) and Engsted and Haldrup (1997) show how Cagan's model and the LQAC model can be formally tested by extending the method initially suggested by Campbell and Shiller (1987) for present value relations.
5. In a recent paper Engsted and Johansen (1997) have extended Granger and Lee's notion of multicointegration to general VAR models and formalized the link existing between $\mathsf{I}(2)$- and multicointegrated systems. See also Engsted *et al.* (1997).
6. Despite the equivalence, I will continue in the following to discriminate between $\mathsf{I}(2)$ (and polynomial) cointegration and multicointegration since the latter concept fundamentally is a property which occurs amongst $\mathsf{I}(1)$ variables with an obvious economic interpretation. The fact that the system can be formulated as an $\mathsf{I}(2)$ system is a different technical issue.
7. It is rather easy to see why the Dickey-Fuller distribution of the *t*-ratio applies when $|\alpha_2| < 1$, because when x_t is $\mathsf{I}(1)$ the regression (19) can be equivalently written in terms of Δx_t regressed on x_{t-1} and lags of Δx_t, that is, the usual augmented Dickey-Fuller regression.
8. In Boswijk *et al.* (1997) a similar procedure is used in testing for multiple unit roots in periodic models.
9. Actually Dickey and Pantula's procedure is more general than outlined here as they consider d to be larger than 2. For practical purposes $d \leqslant 2$ so their analysis simplifies considerably in the current presentation.
10. Note, however, that any unit root test for a single unit root can be applied sequentially in this procedure.
11. In the augmentation of the tests lags of the second differenced variables were included to remove autocorrelation. Some of the series seemed to be weakly heteroscedastic but this is not seriously invalidating the tests.
    The Dickey-Pantula test and the Hasza-Fuller test can be easily calculated from standard econometrics software packages. A GAUSS procedure to calculate the Sen-Dickey symmetric F test can be obtained from the author upon request.
12. Shin and Kim (1997) also consider using in place of $\hat{\varepsilon}_t$ the series $\Delta^2 y_t = \varepsilon_t$ which they argue provide improved size of their semiparametric test.
13. GAUSS routines to calculate the semi-parametric tests can be obtained from the author upon request.
14. It is notationally convenient to focus just on $\mathsf{I}(1)$ variables x_{1t} and $\mathsf{I}(2)$ variables x_{2t}. However, all the subsequent results are not qualitatively affected by assuming that part of x_{1t} could in fact be Δx_{2t}.
15. The multivariate version of this regression equation, i.e. when y_t is a vector, has been intensively studied by Park and Phillips (1989) for the case where $u_t \sim \mathsf{I}(0)$.
16. For technical reasons R is assumed to be such that no cross-restrictions are imposed on coefficients corresponding to regressors of different orders of integration.
17. The asymptotic properties of the residual based 'cointegration' augmented Dickey-Fuller test and the Phillips Z-test have been analyzed extensively by Engle and Yoo (1987) and Phillips and Ouliaris (1990).

18. Note that from a cointegration point of view, one of the I(1) regressors y_t or x_{1t} may be excluded from the regression since these series are themselves cointegrated.
19. From an economic perspective Δp_t is perhaps the most reasonable variable to consider since Δp_t is just inflation which may act as the opportunity costs of holding money.
20. Sometimes it may be of interest to consider non-linear least squares estimation of cointegrated models. One such example is the LQAC model with driving I(2) variables, see section 3.1 and in particular Eq. (7). Engsted and Haldrup (1997) discuss different issues concerning the estimation and testing of this particular model.
21. We need to emphasize once more that the analysis relies on the assumption that it is known that only one cointegration relation exists amongst the variables.
22. Some special cases where mirror image distributions of the regressors are obtained can be found in Haldrup (1994b).
23. In Haldrup (1994b) I describe another situation where normality applies asymptotically. That is, when the single series are dominated by deterministic trends and there is full I(0) cointegration and the trends are cotrending. By 'cotrending' of the deterministic components I mean that when weighted by the cointegration vector, some or all of the higher order trends will vanish. In this case, given that the regression model does not include the higher trends, asymptotic Gaussian distributions result. This generalizes a similar result by West (1988) for the I(1) model.
24. A GAUSS procedure to conduct the Choi *et al.* (1997) CCR procedure for I(2) variables is available from the author upon request.
25. If comparisons with the Gaussian VAR model is wanted it may be useful to assume that u_t is i.i.d. $N(0, \Omega)$ as in Boswijk (1997).
26. The specification (58) allows variables to be integrated of different orders. Generally, x_t is said to be I(d), if d is the smallest positive integer for which $\Delta^d x_t$ is a stationary process and $\Delta^{d-1}x_t$ is a nonstationary process.
27. Observe that if some variables in x_t are I(0) these will be elements of $\beta' x_t$, i.e. cointegration vectors of the form $(0, \ldots, 1, \ldots 0)$ can lie in the space spanned by the columns of β.
28. A further condition for the system to be I(2), but not higher order integrated, is that $\Theta = (\Gamma\overline{\beta}\overline{\alpha}'\Gamma + I - \sum_{i=1}^{k-2} \Psi_i)$ is a full rank matrix. This can be seen immediately from the definition of C_2 below.
29. In the stochastic trends representation we have excluded nuisance parameters which follow from the presence of the initial conditions, see Johansen (1995b) for details.
30. Paruolo (1996) discusses the model under the maintained assumption that $\mu_2 = 0$ so no quadratic trends occur in the model. Within this set-up linear trends will only arise in the I(1) and I(2) directions.
31. A detailed discussion of the reduced rank regression technique is beyond the scope of the present survey. The reader is assumed to be familiar with this tool from the I(1) analysis. Since the notation is rather involved for the I(2) model I prefer to describe non-technically this part of the analysis.
32. That is, $\Delta^2 x_t$ and x_{t-1} are regressed on $\Delta x_{t-1}, \Delta^2 x_{t-j}$, (and maybe a constant) for $j = 1, \ldots, k-2$, and regression residuals are used in the subsequent reduced rank regression analysis.
33. The variables $\hat{\alpha}'_{r\perp}\Delta^2 x_t$ and $\hat{\overline{\beta}}'_{r\perp}\Delta x_{t-1}$ are corrected for the influence of $\hat{\beta}'_r\Delta x_{t-1}$ and lagged second differences of the variables. In fact, the basic regression model can be written as (ignoring '$\hat{\ }$' and subscript 'r') $\alpha'_\perp\Delta^2 x_t = -\alpha'_\perp\Gamma\overline{\beta}(\beta'\Delta x_{t-1}) - \xi\eta'(\overline{\beta}'_\perp\Delta x_{t-1}) + \alpha_\perp\Theta z_t + \alpha'_\perp\varepsilon_t$ where $z_t = (\Delta^2 x'_{t-1}, \ldots, \Delta^2 x'_{t-k+2})$. This shows that the regression explicitly contains the parameters ξ and η.
34. Note that although the above procedure is referred to as a two-step procedure, the integration indices are in fact determined jointly.
35. Remember, by first testing r using the statistic Q_r and then proceeding to fitting s by the statistic $Q_{r,s}$ gives rise to an incorrect size in the first step of the procedure given that I(2) trends are present.

36. When trends and integration indeces are determined jointly the sequence of testing still follows the hierarchical ordering shown in table 9. However, now there exist submodels of the hypotheses given in table 9, corresponding to the restrictions made on the deterministic components. For instance the most restricted model has $\mu_1 = \mu_2 = 0$. Under this restriction, if all hypotheses H_{ij} are rejected by using the usual sequence of testing the hypotheses one proceeds to the model with $\mu_2 = 0$. This model is naturally less restricted. Under this maintained model the same sequence of tests is used for the new model. The first model which cannot reject the hypothesis for given restrictions of the deterministic components determines the integration indeces.
37. All calculations concerning the I(2) VAR analysis has been based on the I(2) platform of the CATS in RATS programme. The programme is written by Clara M. Jørgensen and can be downloaded from http://www.estima.com/procs/i2index.htm.
38. Generally one should be rather careful about using the various asymptotic critical values when the sample size is small. In finite samples there may happen to be serious discrepancies from the limiting distributions.
39. Note that, in fact, the presence of more than a single cointegration relation invalidates the single equation analysis previously conducted.
40. Similar results apply to models with deterministics components, see e.g. Paruolo (1996,1997) and Rahbek *et al.* (1998). Note that in certain directions (orthogonal to (β, β_1)) the rate of normalization in (86) is going to be T^2 by the presence of I(2) components, but this does not change the basic result that the scaled estimators will follow a mixed Gaussian distribution.
41. Note, however, that this facility is not yet available in the CATS in RATS I(2) procedure.
42. Remember that once β and η are estimated the identification of β_1 follows from the relationship $\beta_1 = \overline{\beta}_\perp \eta$. Similarly, we have $\beta_2 = \beta_\perp \eta_\perp$. Note also that when β_c is identified by $\beta_c = \beta(c'\beta)^{-1}$ also α can be identified as $\alpha_c = \alpha\beta' c$ since $\alpha_c \beta_c' = \alpha\beta$.

References

Andrews, D. W. K. (1991) Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* 59, 817–858.

Billingsley, P. (1968) *Convergence of probability measures*. New York: Wiley.

Banerjee, A., L. Cockerell, and B. Russell (1998) An I(2) analysis of inflation and the markup. Working paper, University of Oxford.

Boswijk, H. P. (1997) Mixed normality and ancillarity in I(2) systems. Conditionally accepted for *Econometric Theory*.

Boswijk, H. P., P. H. Franses, and N. Haldrup (1997) Multiple unit roots in periodic autoregression. *Journal of Econometrics*, 80, 167–193.

Cagan, P. (1956) The monetary dynamics of hyperinflation. In Milton Friedman (ed.), *Studies in the quantity theory of money*. 25–120. The University of Chicago Press.

Campbell, J. Y. (1987) Does saving anticipate declining labour income? An alternative test of the permanent income hypothesis. *Econometrica*, 55, 1249–1273.

Campbell, J. Y., and R. J. Shiller (1987) Cointegration and tests of present value models. *Journal of Political Economy*, 95, 1052–1088.

Casella, A. (1989) Testing for rational bubbles with exogenous or endogenous fundamentals. The German hyperinflation once more. *Journal of Monetary Economics*, 24, 109–122.

Chang, Y. (1997) Vector autoregressions with unknown mixtures of I(0), I(1), and I(2) components. Working paper, Dept. of Economics, Rice University.

Chang, Y. and P. C. B. Phillips (1995) Time series regression with mixtures of integrated processes. *Econometric Theory*, 11, 1033–1094.

Choi, I. J. Y. Park, and B. Yu (1997) Canonical cointegration regression and testing for cointegration in the presence of I(1) and I(2) variables. *Econometric Theory*, forthcoming.

Dickey, D. A., and W. A. Fuller (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of 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.

Diebold, F. X., and M. Nerlove (1990) Unit roots in economic time series: A selective survey. *Advances in Econometrics*, 8, 3–69.

Dolado, J. J., T. Jenkinson (1990) Cointegration and unit roots. *Journal of Economic Surveys*, 4, 249–273.

Dolado, J. J., J. W. Galbraith, and A. Banerjee (1991) Estimating intertemporal quadratic adjustment cost models with integrated series. *International Economic Review*, 32, 919–36.

Engle, R. F., and C. W. J. Granger (1987) Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55, 2, 251–276.

Engle, R. F., and S. Yoo (1987) Forecasting and testing in cointegrated systems. *Journal of Econometrics*, 35, 143–159.

Engle, R. F. and B. S. Yoo (1991) Cointegrated economic time series: A survey with new results. In C. W. J. Granger, and R. F. Engle (eds.). *Long-run economic relations. Readings in cointegration.* Oxford: Oxford University Press.

Engle, R. F., C. W. J. Granger, S. Hylleberg, and H. S. Lee (1993) Seasonal cointegration: The Japanese consumption function. *Journal of Econometrics*, 55, 275–298.

Engsted, T. (1993) Cointegration and Cagan's model of hyperinflation under rational expectations. *Journal of Money, Credit, and Banking*, 25, 350–360.

Engsted, T. (1994) The classic European hyperinflations revisited: Testing the Cagan model using a cointegrated VAR approach. *Economica*, 61, 331–343.

Engsted, T., and N. Haldrup (1994) The linear quadratic adjustment cost model and the demand for labour. *Journal of Applied Econometrics*, 9, (supplement), S145–S159.

Engsted, T. and N. Haldrup (1997) Estimating the LQAC Model with I(2) Variables. Forthcoming in *Journal of Applied Econometrics*.

Engsted, T. and N. Haldrup (1998) Multicointegration in stock-flows models. Forthcoming in *Oxford Bulletin of Economics and Statistics*.

Engsted, T., J. Gonzalo, and N. Haldrup (1997) Testing for multicointegration. *Economics Letters*, 56, 259–266.

Engsted, T., and S. Johansen (1997) Granger's representation theorem and multicointegration. EUI Working papers, ECO No. 97/15, European University Institute, Florence.

Fuller, W. A. (1976) Introduction to statistical time series, New York, John Wiley & Sons.

Goodfriend, M. S. (1982) An alternative method of estimating the Cagan money demand function in hyperinflation under rational expectations. *Journal of Monetary Economics*, 9, 43–57.

Granger, C. W. J. (1997) On modelling the long run in applied economics. *Economic Journal*, 107, 169–177.

Granger, C. W. J., and T. Lee (1989) Investigation of production, sales and inventory relations using multicointegration and non-symmetric error correction models. *Journal of Applied Econometrics*, 4 (supplement), S145–S159.

Granger, C. W. J., and T. Lee (1990) Multicointegration. In *Advances in Econometrics*., Vol. 8, ed. by G. F. Rhodes and T. B. Fomby, Greenwich, CT: JAI Press, 71–84.

Granger, C. W. J., and P. Newbold (1974) Spurious regressions in econometrics. *Journal of Econometrics*, 26, 1045–1066.

Gregoir, S. and G. Laroque (1993) Multivariate time series: a polynomial error correction theory. *Econometric Theory*, 9, 329–342.

Gregoir, S. and G. Laroque (1994) Polynomial cointegration: Estimation and tests. *Journal of Econometrics*, 63, 183–214.
Gregory, A. W., A. R. Pagan, and G. W. Smith (1993) Estimating linear quadratic models with integrated processes. In P. C. B. Phillips (ed.), *Models, Methods, and Applications in Econometrics*. Oxford: Basil Blackwell. 220–39.
Haldrup, N. (1994a) Semi-parametric tests for double unit roots. *Journal of Business and Economic Statistics*, 12, 109–122.
Haldrup, N. (1994b) The asymptotics of single-equation cointegration regressions with I(1) and I(2) variables. *Journal of Econometrics*, 63, 153–81.
Haldrup, N. and M. Salmon (1998) Representations of I(2) Cointegrated Systems using the Smith-McMillan Form. *Journal of Econometrics*, 84, 303–325.
Hallman, J. J., R. D. Porter, and D. H. Small (1991) Is the price level tied to the M2 monetary aggregate in the long run?, *American Economic Review*, 81, 841–858.
Hamilton, J. (1994) *Time Series Analysis*. Princeton University Press: Princeton, NJ.
Harvey, A. C. (1981) *Time Series Models*. London: Philip Allen.
Hasza, D. P. (1977) Estimation in nonstationary time series. Unpublished Ph.D. thesis. Iowa state University, Dept. of Statistics.
Hasza, D. P. and W. A. Fuller (1979) Estimation of autoregressive processes with unit roots. *The Annals of Statistics*, 7, 1106–1120.
Hendry, D. F. and T. von-Ungern Sternberg (1981) Liquidity and inflation effects of consumer expenditure. In A. S. Deaton (ed.), *Essays in the theory and measurement of consumers' behaviour*. Cambridge University Press.
Holt, C. C., F. Modigliani, J. F. Muth, and H. Simon (1960) *Planning production inventories, and work force*. Prentice Hall, Englewood Cliffs.
Hylleberg, S., and G. E. Mizon (1989) Cointegration and error correction mechanisms. *Economic Journal*, 99, 113–125.
Hylleberg, S., R. F. Engle, C. W. J. Granger, and B. S. Yoo (1990) Seasonal integration and cointegration. *Journal of Econometrics*, 44, 215–238.
Johansen, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 8, 231–254.
Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59, 1551–1580.
Johansen, S. (1992a) Testing weak exogeneity and the order of cointegration in UK money demand data. *Journal of Policy Modeling*, 14, 313–334.
Johansen, S. (1992b) A representation of vector autoregressive processes integrated of order 2. *Econometric Theory*, 8, 188–202.
Johansen, S. (1995a) A statistical analysis of cointegration for I(2) variables. Econometric Theory, 11, 25–59.
Johansen, S. (1995b) *Likelihood-based inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.
Johansen, S. (1997) A likelihood analysis of the I(2) model. *Scandinavian Journal of Statistics*, forthcoming.
Johansen, S., and K. Juselius (1990) Maximum likelihood estimation and inference on cointegration with applications to money demand. *Oxford Bulletin of Economics and Statistics*, 52, 169–210.
Johansen, S., and K. Juselius (1992) Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK. *Journal of Econometrics*, 53, 211–244.
Juselius, K. (1992) Domestic and foreign effects on prices in an open economy: The case of Denmark. *Journal of Policy Modeling*, 14, 401–428.
Juselius, K. (1994) On the duality between long-run relations and common trends in the I(1) versus I(2) model. An application to aggregate money holdings. *Econometric Reviews*, 13, 151–178.
Juselius, K. (1995a) Do purchasing power parity and uncovered interest rate parity hold in

the long run? An example of likelihood inference in a multivariate time-series model. *Journal of Econometrics*, 69, 211–240.

Juselius, K. (1995b) Predictable and unpredictable components of the longrun growth in nominal prices. *Mathematics and Computers in Simulation*, 39, 257–263.

Juselius, K. (1996) An empirical analysis of the changing role of the German Bundesbank after 1983. *Oxford Bulletin of Economics and Statistics*, 58, 791–819.

Juselius, K. (1998) A structured VAR in Denmark under changing monetary regimes. *Journal of Business and Economic Statistics*, 16(4), 400–411.

Juselius, K., and C. Hargreaves (1992) Long-run relations in Australian monetary data. In: C. Hargreaves (ed.), *Macroeconomic modelling of the long run*. Vermont: Edward Elgar.

Kailath, T. (1980) *Linear Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Kennan, J. (1979) The estimation of partial adjustment models with rational expectations. *Econometrica*, 47, 1441–1455.

King, R. G., C. I. Plosser, J. H. Stock, and M. W. Watson (1991) Stochastic trends and economic fluctuations. *American Economic Review*, 81, 819–840.

Kitamura, Y. (1995) Estimation of cointegrated systems with I(2) processes. *Econometric Theory*, 11, 1–24.

Lee, T.-H. (1992) Stock-flow relationships in housing construction. *Oxford Bulletin of Economics and Statistics*, 54, 419–430.

Lee, T.-H. (1996) Stock adjustment for multicointegrated series. *Empirical Economics*, 21, 633–639.

McKinnon, J. G. (1991) Critical values for cointegration tests. In R. F. Engle, and C. W. J. Granger (eds), *Long run Economic Relationships*. Oxford: Oxford University Press, 267–276.

Muscatelli, V. A., and S. Hurn (1992) Cointegration and dynamic time series models. *Journal of Economic Surveys*, 6, 1–43.

Newey, W. and K. D. West (1987) A simple positive semi-definite heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Pantula, S. G. (1989) Testing for unit roots in time series data. *Econometric Theory*, 5, 256–271.

Park, J. Y. (1992) Canonical cointegration regression. *Econometrica*, 60, 119–143.

Park, J. Y., and P. C. B. Phillips (1988) Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, 4, 468–498.

Park, J. Y. and P. C. B. Phillips (1989) Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory*, 5, 95–131.

Paruolo, P. (1994) The role of the drift in I(2) systems. *Journal of the Italian Statistical Society*, 3, 65–96.

Paruolo, P. (1996) On the determination of integration indices in I(2) systems. *Journal of Econometrics*, 72, 313–356.

Paruolo, P. (1997) Asymptotic efficiency of the two stage estimator in I(2) systems. *Econometric Theory*, forthcoming.

Paruolo, P., and A. Rahbek (1996) Weak exogeneity in I(2) systems. Working paper.

Peseran, M. H. (1991) Costly adjustment under rational expectations: A generalization. *Review of Economics and Statistics*, 73, 353–358.

Phillips, A. W. (1954) Stabilization policy in the closed economy. *Economic Journal*, 64, 290–323.

Phillips, P. C. B. (1986) Understanding spurious regressions in econometrics. *Journal of Econometrics*, 33, 311–340.

Phillips, P. C. B. (1991) Optimal inference in cointegrated systems. *Econometrica*, 59, 283–306.

Phillips, P. C. B. (1987) Time series regression with unit root. *Econometrica*, 55, 277–302.

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.
Price, S. (1992) Forward looking price setting in UK manufacturing. *Economic Journal*, 102, 497–505.
Rahbek, A., C. Jørgensen, and H. C. Kongsted (1998) Trend-stationarity in the I(2) cointegration model. Forthcoming in *Journal of Econometrics*.
Salmon, M. (1988) Error correction models, cointegration and the internal model principle. *Journal of Economic Dynamics and Control*, 12, 523–549.
Sargent, T. J. (1977) The demand for money during hyperinflation under rational expectations I. *International Economic Review*, 14, 59–82.
Sargent, T. J. (1978) Estimation of dynamic labor demand schedules under rational expectations. *Journal of Political Economy*, 86, 1009–1044.
Schwert, G. W. (1989) Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics*, 7, 147–159.
Sen, D. L. (1986) Robustness of single unit root test statistics in the presence of multiple unit roots. Unpublished Ph.D. thesis, North Carolina state University, Dept. of statistics.
Sen, D. L. and D. A. Dickey (1987) Symmetric test for second differencing in univariate time series. *Journal of Business and Economic Statistics*, 5, 463–473.
Shin, Y. (1994) A residual based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory*, 10, 91–115.
Shin, D. W., and H. J. Kim (1997) Semiparametric tests for double unit roots based on symmetric estimators. *Journal of Business and Economic Statistics*, forthcoming.
Stock, J. H. (1994) Unit Roots, Structural Breaks and Trends. In Engle, R. F., and D. L. McFadden (eds.), *Handbook of Econometrics*, 4, ch. 46, 2739–2841.
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.
Taylor, M. P. (1991) The hyperinflation model of money demand revisited. *Journal of Money, Credit, and Banking*, 23, 327–56.
Watson, M. W. (1994) Vector Autoregressions and Cointegration. In Engle, R. F., and D. L. McFadden (eds.), *Handbook of Econometrics*, 4, ch. 47, 2843–2915.
West, K. (1988) Asymptotic normality when regressors have a unit root. *Econometrica*, 56, 1397–1417.
White, H. (1984) *Asymptotic Theory for Econometricians*. Orlando: Academic Press.
White, J. S. (1958) The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics*, 29, 1188–1197.
Yoo, S. (1986) Multi-cointegrated time series and a generalized error correction model. UCSD Discussion paper.