Søren Johansen

A LIKELIHOOD ANALYSIS OF THE I(2) MODEL

7
Preprint
August
1994

IMS
Institute of Mathematical Statistics
University of Copenhagen

A likelihood analysis of the I(2) model

Søren Johansen
Institute of Mathematical Statistics
University of Copenhagen

Abstract

The I(2) model is defined as a sub-model of the general vector autoregressive model, by two reduced rank conditions. The model describes stochastic processes with stationary second difference. A parametrization is suggested which makes likelihood inference feasible. The asymptotic distribution of the maximum likelihood estimators is given. It is shown that the asymptotic distribution is either Gaussian, mixed Gaussian or, in some cases, even more complicated.

1 Introduction

The vector autoregressive model is often applied in statistics to describe a stationary time series but such models can also be used to describe the fluctuations of non stationary processes by imposing suitable restrictions on the parameter space.

In econometrics the observed time series like prices, money and income are often best described by non stationary processes and a class of processes that has often been used consists of a random walk plus a stationary process, that is, a process for which the difference is stationary. We call such a process an $I(1)$ process, that is, integrated of order 1. A closer inspection of the price variables (in logs) shows that the changes, that is the inflation rate, is sometimes best described by an $I(1)$ process. If the changes of a process is an $I(1)$ process the process itself is called an $I(2)$ process and it can be decomposed into a cumulated random walk, a random walk, and a stationary process, see (7).

The statistical analysis of $I(1)$ processes is now a well established technique in econometrics, see Johansen [3], Reinsel and Ahn [14] and Phillips [13].

The statistical analysis of $I(2)$ processes is discussed by Johansen [7] in the context of the vector autoregressive model using a modification of likelihood methods, and it is followed by papers by Paruolo [10] and [11]. Regression models have been used by Stock and Watson [16].

See Johansen [8] for an analysis of prices, interest rates and exchange rates in Australia and US, and Juselius [9] for an analysis of purchasing power parity between Denmark and Germany using the notion of $I(2)$ processes.

The purpose of this paper is to analyse the likelihood function for the $I(2)$ model as a sub-model of the general VAR model as defined in Johansen [7]. An algorithm for calculating the maximum likelihood estimator is given in [6], and in this paper we want to find the asymptotic distribution of the maximum likelihood estimator and compare it with the results for the above two step procedure.

It is known that by restricting the coefficient matrix of the levels in a reduced form error correction model to have reduced rank, the process is forced to be non-stationary, and conditions exist for eliminating the possibility of $I(2)$, see [5]. For an $I(2)$ model we need two reduced rank conditions (5) and (6) and the first result, in section 4, of this paper is a parametrisation of this model where the parameters vary independently, which makes the analysis of the likelihood function feasible. We next discuss briefly the asymptotic analysis of the process and product moments derived from it. Section 7 contains the derivatives of the likelihood function with respect to the relevant parameters. Using these results we can then prove the existence and consistency of the maximum likelihood estimator, and find their asymptotic distribution. It turns out that the asymptotic distribution is not mixed Gaussian as one finds in the $I(1)$ analysis.

2 The analysis of I(1) variables

The vector autoregressive model can be rewritten as an error correction model in the form

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \epsilon_t, \tag{1}$$

and it is a well known result that if the roots of the characteristic polynomial

$$A(z) = |(1-z)I - \Pi z - \sum_{i=1}^{k-1} \Gamma_i (1-z) z^i|$$

are outside the unit disk, then the process generated by (1) is stationary. If unit roots are allowed then $A(1) = -\Pi$ is of reduced rank, and can be written as $\Pi = \alpha\beta'$. Under this condition X_t has the representation

$$X_t = C \sum_{i=1}^{t} \epsilon_i + C(L)\epsilon_t, \tag{2}$$

where

$$C = \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp, \quad \Gamma = I - \sum_{i=1}^{k-1} \Gamma_i, \tag{3}$$

provided $\alpha'_\perp \Gamma \beta_\perp$ has full rank. Thus X_t has the form of an $I(1)$ variable, that is, a stationary process plus an random walk. This representation shows the main result in cointegration analysis, see Granger [2], namely that although X_t is a non stationary process it still holds that $\beta' X_t$ is stationary. This result shows that there can be stationary relations between non stationary variables. It accounts for the interest that this type of analysis has gained in econometrics, where the existence of stationary relations between variables allows one to formulate structural economic relations as stationary relations, and the error correction model shows how the process reacts and adjusts to a disequilibrium error $(\beta' X_t)$ through the adjustment coefficients α. The random walk $\alpha'_\perp \sum_{i=1}^t \epsilon_i$ which accounts for the non stationarity of the process is called the common trend.

The $I(1)$ model is defined by equation (1) where the parameters are

$$(\alpha, \beta, \Gamma_1, ..., \Gamma_{k-1}, \Omega),$$

which vary freely.

The statistical analysis of this model by likelihood methods shows that α and β can be estimated by reduced rank regression of ΔX_t on X_{t-1} on $\Delta X_{t-1}, ..., \Delta X_{t-k+1}$, see Johansen [3]. It turns out that the maximum likelihood estimator of β is superconsistent in the sense that $T(\hat{\beta} - \beta)$ is weakly convergent, and that the limit is mixed Gaussian, which means that (asymptotic) inference on β can be conducted in the usual way, using the χ^2 distribution, see section 6.

3 The representation of I(2) variables

For an analysis of $I(2)$ variables it is convenient to rewrite the autoregressive model as

$$\Delta^2 X_t = \Gamma \Delta X_{t-1} + \Pi X_{t-2} + \sum_{i=1}^{k-2} \Psi_i \Delta^2 X_{t-i} + \epsilon_t, \tag{4}$$

where ϵ_t are *i.i.d.* $N_p(0, \Omega)$. We assume as before that the parameters α, β are defined by

$$\Pi = \alpha\beta', \ \alpha, \beta \ (p \times r), \ r < p. \tag{5}$$

If α and β are of full rank we define $\alpha_\perp$ and $\beta_\perp$ of dimension $p \times (p-r)$ and full rank such that $\alpha'\alpha_\perp = \beta'\beta_\perp = 0$. We need the notation $\bar{\alpha} = \alpha(\alpha'\alpha)^{-1}$ and similarly for other matrices of full rank. We let $|A|$ denote the determinant of the square matrix A. In order that (4) should generate $I(2)$ variables we must assume that $\alpha'_\perp \Gamma \beta_\perp$ has reduced rank, so that we can define the parameters φ and η by

$$\bar{\alpha}'_\perp \Gamma \bar{\beta}_\perp = \varphi\eta', \ \varphi, \eta \ (p-r) \times s, \ s < p - r. \tag{6}$$

We then define the parameter functions

$$\beta_1 = \beta_\perp \eta, \ \alpha_1 = \alpha_\perp \varphi, \ \beta_2 = \bar{\beta}_\perp \eta_\perp, \ \alpha_2 = \bar{\alpha}_\perp \varphi_\perp.$$

Note that $(\beta, \beta_1, \beta_2)$ are mutually orthogonal and the same holds for $(\alpha, \alpha_1, \alpha_2)$. Note also that the parameter η depends on the choice of $\beta_\perp$, whereas (β_1, β_2) are independent of this choice. In [5] it is shown that under the above conditions and the condition

$$|\alpha_2'\Theta\beta_2| = \left|\alpha_2'(\Gamma\bar{\beta}\bar{\alpha}'\Gamma + I - \sum_{i=1}^{k-2}\Psi_i)\beta_2\right| \neq 0,$$

it holds that the process X_t is an I(2) process with the representation

$$X_t = C_2 \sum_{j=1}^{t}\sum_{i=1}^{j} \epsilon_i + C_1 \sum_{i=1}^{t} \epsilon_i + C(L)\epsilon_t. \tag{7}$$

The coefficient matrices C_1 and C_2 are complicated functions of the parameters of the model, and we here only quote the result

$$C_2 = \beta_2(\alpha_2'\Theta\beta_2)^{-1}\alpha_2'. \tag{8}$$

It follows from (7) that X_t needs two differences to become stationary as long as $C_2 \neq 0$. From (7) and (8) it is seen that $\beta_2'X_t$ is $I(2)$ and that no linear combination of this process has lower order of integration. Since $(\beta, \beta_1)'\beta_2 = 0$ we have $(\beta, \beta_1)'C_2 = 0$ and hence $(\beta, \beta_1)'X_t$ is $I(1)$ in general, but more can be said. It turns out, see [5], that

$$\beta'X_t + \bar{\alpha}'\Gamma\bar{\beta}_2\beta_2'\Delta X_t,$$

and therefore also the process

$$\beta'X_t + \bar{\alpha}'\Gamma\Delta X_t,$$

are stationary. Thus the representation (7) implies that $\tau = (\beta, \beta_1)$ are $C(2,1)$ in the sense that they reduce the order of the process from 2 to 1. The relation $\beta'X_t$ cointegrates with the $I(1)$ process ΔX_t to stationarity. The process $\alpha_2' \sum_{j=1}^{t}\sum_{i=1}^{j}\epsilon_i$ is called the common $I(2)$ trend.

It is the purpose of this paper is to find the asymptotic distribution of the maximum likelihood estimators for the parameters $(\beta, \beta_1, \beta_2)$ together with a matrix ξ such that $\beta'X_t + \xi'\Delta X_t$ is stationary, see the definitions in Table 1.

4 A reparametrization of the I(2) model

In this section we define the VAR model with 2 lags written in the error correction form, and define the $I(2)$ model as a sub-model which has two reduced rank conditions (5) and (6) on the coefficient matrices. We then define a different parametrization, which has the property that the parameters vary unrestrictedly. This make the analysis of the likelihood easier.

The reparametrization only involves the parameters Π, Γ and Ω so in the next sections we only analyse the VAR(2) model

$$\Delta^2 X_t = \Gamma \Delta X_{t-1} + \Pi X_{t-2} + \epsilon_t, \tag{9}$$

where ϵ_t are *i.i.d.* $N_p(0, \Omega)$.

The restrictions imposed on the model are

$$\Pi = \alpha\beta',$$

where α and β are $p \times r$ matrices of full rank, and

$$\bar{\alpha}'_\perp \Gamma \bar{\beta}_\perp = \varphi\eta', \tag{10}$$

where φ and η are $p \times s$ matrices. Thus the parameters are $\theta^* = (\alpha, \beta, \varphi, \eta, \Gamma, \Omega)$ which vary freely except for the restriction (10).

The purpose of this section is to define another set of parameters that vary unrestrictedly. In order to motivate the new parameters we first analyse the model equations as follows: Multiplying (9) by $\bar{\alpha}'$ and $\bar{\alpha}'_\perp$ we find

$$\bar{\alpha}'\Delta^2 X_t = \bar{\alpha}'\Gamma\Delta X_{t-1} + \beta' X_{t-2} + \bar{\alpha}'\epsilon_t, \tag{11}$$

$$\bar{\alpha}'_\perp \Delta^2 X_t = \bar{\alpha}'_\perp \Gamma \Delta X_{t-1} + \bar{\alpha}'_\perp \epsilon_t. \tag{12}$$

Using the condition $\bar{\alpha}'_\perp \Gamma \bar{\beta}_\perp = \varphi\eta'$, the identity $I = \bar{\beta}_\perp \beta'_\perp + \bar{\beta}\beta'$ and the definition $\beta_1 = \beta_\perp \eta$ the right hands side of (12) equals

$$\bar{\alpha}'_\perp \Gamma \bar{\beta}\beta'\Delta X_{t-1} + \bar{\alpha}'_\perp \Gamma \bar{\beta}_\perp \beta'_\perp \Delta X_{t-1} + \bar{\alpha}'_\perp \epsilon_t$$

$$= \bar{\alpha}'_\perp \Gamma \bar{\beta}\beta'\Delta X_{t-1} + \varphi\eta'\beta'_\perp \Delta X_{t-1} + \bar{\alpha}'_\perp \epsilon_t$$

$$= (\bar{\alpha}'_\perp \Gamma \bar{\beta}, \varphi)(\beta, \beta_1)'\Delta X_{t-1} + \bar{\alpha}'_\perp \epsilon_t$$

$$= \kappa'\tau'\Delta X_{t-1} + \bar{\alpha}'_\perp \epsilon_t. \tag{13}$$

We define $\epsilon_{1t} = \bar{\alpha}'_\perp \epsilon_t$, $\omega = \bar{\alpha}'\Omega\bar{\alpha}_\perp(\bar{\alpha}'_\perp \Omega \bar{\alpha}_\perp)^{-1}$ and $\epsilon_{2t} = \bar{\alpha}'\epsilon_t - \omega\bar{\alpha}'_\perp \epsilon_t$, which is independent of ϵ_{1t}, and find from (11) and (13) the conditional model for $\bar{\alpha}'\Delta^2 X_t$ given $\bar{\alpha}'_\perp \Delta^2 X_t$ and the past

$$\bar{\alpha}'\Delta^2 X_t = \omega\bar{\alpha}'_\perp \Delta^2 X_t + (\bar{\alpha}' - \omega\bar{\alpha}'_\perp)\Gamma\Delta X_{t-1} + \beta' X_{t-2} + \epsilon_{2t}. \tag{14}$$

The new set of parameters is given as $\theta = (\alpha, \tau, \rho, \kappa, \xi, \Omega_1, \Omega_2, \omega)$ and they are defined in Table 1

Table 1
Definition of new parameters θ
in terms of the old parameters θ^*

θ		θ^*	dimension
α	$=$	α	$p \times r$
τ	$=$	(β, β_1)	$p \times (r+s)$
ρ	$=$	$(\tau'\tau)^{-1}\tau'\beta$	$(r+s) \times r$
ξ'	$=$	$(\bar{\alpha}' - \omega\bar{\alpha}'_{\perp})\Gamma$	$r \times p$
κ'	$=$	$(\bar{\alpha}'_{\perp}\Gamma\bar{\beta}, \varphi)$	$(p-r) \times (r+s)$
ω	$=$	$\bar{\alpha}'\Omega\bar{\alpha}_{\perp}(\bar{\alpha}'_{\perp}\Omega\bar{\alpha}_{\perp})^{-1}$	$r \times (p-r)$
Ω_1	$=$	$\bar{\alpha}'_{\perp}\Omega\bar{\alpha}_{\perp}$	$(p-r) \times (p-r)$
Ω_2	$=$	$\bar{\alpha}'\Omega\bar{\alpha} - \bar{\alpha}'\Omega\bar{\alpha}_{\perp}(\bar{\alpha}'_{\perp}\Omega\bar{\alpha}_{\perp})^{-1}\bar{\alpha}'_{\perp}\Omega\bar{\alpha}$	$r \times r.$

The vectors that are $C(2,1)$ are collected in the matrix τ, and the matrix ρ recovers from τ those vectors that span the row space of the Π matrix. In terms of these parameters the model equations (14) and (12) can be written

$$\bar{\alpha}'\Delta^2 X_t = \omega\bar{\alpha}'_{\perp}\Delta^2 X_t + \xi'\Delta X_{t-1} + \rho'\tau' X_{t-2} + \epsilon_{2t}, \tag{15}$$

$$\bar{\alpha}'_{\perp}\Delta^2 X_t = \kappa'\tau'\Delta X_{t-1} + \epsilon_{1t}. \tag{16}$$

By the choice of $\tau = (\beta_1, \beta_2)$ we have that the first r rows are orthogonal to the last s rows, and that we can choose $\rho = (I, 0)'$, but this requires that the first columns of τ are the β vectors, and we want to let τ vary freely, and hence need the parameters ρ in order to pick out the β vectors.

From given values of the new parameters θ we can reconstruct the parameters of the original model θ^* which satisfy the restriction (6). To see this we use the definitions in Table 2.

Table 2
Definition of old parameters θ^*
in terms of new parameters θ

θ^*		θ	dimension
α	$=$	α	$p \times r$
β	$=$	$\tau\rho$	$p \times r$
φ	$=$	$\kappa'\bar{\rho}_{\perp}$	$(p-r) \times s$
η	$=$	$\bar{\beta}'_{\perp}\tau\rho_{\perp}$	$(p-r) \times s$
Ω	$=$	$\alpha\Omega_2\alpha' + (\alpha_{\perp} + \alpha\omega)\Omega_1(\alpha_{\perp} + \alpha\omega)'$	$p \times p$
Γ	$=$	$\alpha_{\perp}\kappa'\tau' + \alpha\omega\kappa'\tau' + \alpha\xi'$	$p \times p.$

With this choice we see that $\Pi = \alpha\beta' = \alpha\rho'\tau'$ has reduced rank $(\leq)r$. Further we find that if β has full rank then

$$\bar{\alpha}'_{\perp}\Gamma\bar{\beta}_{\perp} = \kappa'\tau'\bar{\beta}_{\perp} = \kappa'(\bar{\rho}\rho' + \bar{\rho}_{\perp}\rho'_{\perp})\tau'\bar{\beta}_{\perp} = (\kappa'\bar{\rho}_{\perp})(\rho'_{\perp}\tau'\bar{\beta}_{\perp}) = \varphi\eta',$$

since $\rho'\tau'\bar{\beta}_{\perp} = \beta'\bar{\beta}_{\perp} = 0$. This matrix has reduced rank $(\leq)s$, since $\rho_{\perp}$ is $(r+s) \times s$. Thus any values of the new parameters θ correspond to values of the old parameters θ^* with the required restrictions, hence the new parameters vary unrestrictedly.

Note that only if ρ and τ are chosen such that the product has full rank do we get a β of full rank, and hence a parameter value from θ^*. Thus strictly speaking the parameters in θ should have the restriction that $\tau\rho$ has full rank in order that a parameter point in the old parameter set is produced. We shall, however, let the parameters in θ vary freely, since the extra points that we add form a small set with Lebesgue measure zero, which does not influence the analysis of the likelihood function.

We conclude by pointing out that the parameters we are really interested in are the matrices

$$\begin{aligned} \beta &= \beta &&= \tau\rho \\ \beta_1 &= \beta_{\perp}\eta &&= \bar{\tau}\rho_{\perp} \\ \beta_2 &= \bar{\beta}_{\perp}\eta_{\perp} &&= \tau_{\perp} \end{aligned}$$

as well as the matrix ξ which describe the cointegration properties of the process. The relation $\beta_1 = \bar{\tau}\rho_{\perp}$ is seen from the fact that $\beta = \tau\rho$ by definition, β_2 is orthogonal to $\tau = (\beta, \beta_1)$ hence $\beta_2 = \tau_{\perp}$ and finally the remaining vectors are orthogonal to $\tau_{\perp}$ and $\tau\rho$, and therefore must have the form $\bar{\tau}\rho_{\perp}$. The idea in the following is to work with the parameters τ, ρ and ξ in the analysis of the likelihood function and then derive the distributions of β, β_1 and β_2 at the end.

The basic equation (9) can be written in terms of the new parameters in an error correction form

$$\Delta^2 X_t = \alpha(\rho'\tau' X_{t-2} + \xi'\Delta X_{t-1}) + (\alpha_{\perp} + \alpha\omega)\kappa'\tau'\Delta X_{t-1} + \epsilon_t, \tag{17}$$

which shows that α, $\alpha_{\perp}$, ω and κ are coefficients of the stationary variables $\rho'\tau' X_{t-2} + \xi'\Delta X_{t-1}$ and $\tau'\Delta X_{t-1}$ which represent disequilibrium errors.

5 Asymptotic properties of the process

The asymptotic results needed for the analysis are collected in the next table. The results are found in the literature see Phillips and Durlauf [12], Chan and Wei [1], and Johansen [5]. We assume that ϵ_t is *i.i.d.* with mean zero and variance Ω, even though this assumption can be relaxed.

The basic result is that a random walk is approximated by a Brownian motion, in the sense of weak convergence, that is,

$$T^{-\frac{1}{2}}\sum_{t=1}^{[Tu]}\epsilon_t \xrightarrow{w} W(u),$$

where W is a Brownian motion with variance matrix Ω. From this result follows a number of other results which are relevant for the discussion of the asymptotic properties of the process X_t. We collect these in the Table 3.

Table 3

Asymptotic properties of product moments of the $\epsilon's$

$T^{-2}\Sigma_{t=1}^{T}(\Sigma_{i=1}^{t}\epsilon_i)(\Sigma_{i=1}^{t}\epsilon_i')$	$\xrightarrow{w}$	$\int_0^1 W(u)W(u)'du$
$T^{-3}\Sigma_{t=1}^{T}(\Sigma_{i=1}^{t}\Sigma_{j=1}^{i}\epsilon_j)(\Sigma_{i=1}^{t}\epsilon_i')$	$\xrightarrow{w}$	$\int_0^1(\int_0^u W(s)ds)W(u)'du$
$T^{-4}\Sigma_{t=1}^{T}(\Sigma_{i=1}^{t}\Sigma_{j=1}^{i}\epsilon_j)(\Sigma_{i=1}^{t}\Sigma_{j=1}^{i}\epsilon_j')$	$\xrightarrow{w}$	$\int_0^1(\int_0^u W(s)ds)\int_0^u W(s)'dsdu$
$T^{-1}\Sigma_{t=1}^{T}(\Sigma_{i=1}^{t-1}\epsilon_i)\epsilon_t'$	$\xrightarrow{w}$	$\int_0^1 W(dW)'$
$T^{-2}\Sigma_{t=1}^{T}(\Sigma_{i=1}^{t-1}\Sigma_{j=1}^{i}\epsilon_j)\epsilon_t'$	$\xrightarrow{w}$	$\int_0^1(\int_0^u W(s)ds)(dW)'$.

From these results we can derive properties of the process X_t if it is $I(1)$ or if it is $I(2)$. The $I(1)$ process X_t is given by the representation (2) provided the matrix C given by (3) is well defined. We then find the results in Table 4.

Table 4

Asymptotic properties of the process and the product moments for I(1) processes

$T^{-\frac{1}{2}}\bar{\beta}_\perp' X_{[Tt]}$	$\xrightarrow{w}$	$\bar{\beta}_\perp' CW_t$
$T^{-2}\bar{\beta}_\perp' \sum_{t=1}^{T} X_{t-1}X_{t-1}'\bar{\beta}_\perp$	$\xrightarrow{w}$	$\bar{\beta}_\perp' C\int_0^1 W_tW_t'dtC'\bar{\beta}_\perp$
$T^{-1}\bar{\beta}_\perp' \sum_{t=1}^{T} X_{t-1}\epsilon_t'$	$\xrightarrow{w}$	$\bar{\beta}_\perp' C\int_0^1 W(dW)'$.

If X_t is defined by the equations (4) under conditions (5) and (6), and if C_2 given by (6) is well defined, then the process is $I(2)$ and given by (7) or as the solution of (12) and (14). In this case we define W_1 and W_2 from ϵ_{1t} and ϵ_{2t} and apply the representation (7) to derive results about the process in various directions as well as properties of the product moments M_{ij} defined by

$$M_{ij} = T^{-1}\sum_{t=1}^{T}\Delta^{2-i}X_{t-i}\Delta^{2-j}X_{t-j}',\ i,j=0,1,2,$$

$$M_{i\epsilon_j} = T^{-1}\sum_{t=1}^{T}\Delta^{2-i}X_{t-i}\epsilon_{jt}',\ i=0,1,2,\ j=1,2.$$

Table 5
Asymptotic properties of the process and the product moments for I(2) processes

$T^{-\frac{1}{2}}\bar{\beta}_2'\Delta X_{[Tt]}$	$\xrightarrow{w}$	$\bar{\beta}_2' C_2 W_t$	$=$	H_{0t}
$T^{-\frac{1}{2}}\bar{\beta}_1' X_{[Tt]}$	$\xrightarrow{w}$	$\bar{\beta}_1' C_1 W_t$	$=$	H_{1t}
$T^{-3/2}\bar{\beta}_2' X_{[Tt]}$	$\xrightarrow{w}$	$\bar{\beta}_2' C_2 \int_0^t W_u du$	$=$	H_{2t}
$T^{-1}\bar{\beta}_2' M_{11}\bar{\beta}_2$	$\xrightarrow{w}$	$\int_0^1 H_0 H_0' du$		
$T^{-1}\bar{\beta}_1' M_{22}\bar{\beta}_1$	$\xrightarrow{w}$	$\int_0^1 H_1 H_1' du$		
$T^{-2}\bar{\beta}_1' M_{22}\bar{\beta}_2$	$\xrightarrow{w}$	$\int_0^1 H_1 H_2' du$		
$T^{-3}\bar{\beta}_2' M_{22}\bar{\beta}_2$	$\xrightarrow{w}$	$\int_0^1 H_2 H_2' du$		
$T^{-1}\bar{\beta}_2' M_{12}\bar{\beta}_1$	$\xrightarrow{w}$	$\int_0^1 H_0 H_1' du$		
$T^{-2}\bar{\beta}_2' M_{12}\bar{\beta}_2$	$\xrightarrow{w}$	$\int_0^1 H_0 H_2' du$		
$\bar{\beta}_2' M_{1\epsilon_1}$	$\xrightarrow{w}$	$\int_0^1 H_0 (dW_1)'$		
$\bar{\beta}_1' M_{2\epsilon_2}$	$\xrightarrow{w}$	$\int_0^1 H_1 (dW_2)'$		
$T^{-1}\bar{\beta}_2' M_{2\epsilon_2}$	$\xrightarrow{w}$	$\int_0^1 H_2 (dW_2)'$.		

The process $H = (H_0', H_1', H_2')$ is given by the first three rows in Table 5 and consists of the non-stationary components of the limit process derived from X_t. Thus H_2 is the limit of the process in the directions $\bar{\beta}_2$ and therefore the process is the continuous analogue of an $I(2)$ process, that is, an integral of a Brownian motion. The differences of this process has a limit process that is the Brownian motion H_0, and therefore $H_{2t} = \int_0^t H_{0u} du$. Note the different normalizations of the process and its differences. When multiplying by $\bar{\beta}_1$ the matrix C_2 vanishes and the next term of (7) takes over and defines in the limit the process H_1.

To get an overview of such results one can summarize them in another table: Let Y_{it} be integrated of order $I(i)$, $i = 0, 1, 2$, then

$$S_{ij} = \Sigma_{t=1}^T Y_{it} Y_{jt}' \in O_P(T^{(i+j)\vee 1}),$$

that is, the order of magnitude of these product moments are given by Table 6.

Table 6
The order of magnitude of product moments of integrated variables

i/j	0	1	2
0	T	T	T^2
1	T	T^2	T^3
2	T^2	T^3	T^4

These results will be applied below for the analysis of the likelihood equations.

6 Asymptotic analysis of the I(1) model

This section is included to justify certain steps in the analysis of the $I(2)$ model and is a brief summary of the $I(1)$ analysis as it can be found in the papers by Johansen (1988) or Reinsel and Ahn (1990). To simplify we consider only the model given by

$$\Delta X_t = \alpha\beta' X_{t-1} + \epsilon_{t,} \tag{18}$$

and assume in the calculations below that α and β are chosen so that X_t is $I(1)$. In this case $C = \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha'_\perp$. We apply the following notation for the derivatives of matrix valued functions with respect to matrix arguments.

Let $f(x)$ be defined on the space of matrices of dimension $n \times m$ ($M_{n,m}$) and with values in the space $M_{k,p}$, say, and let $h \in M_{n,m}$. We assume that we have the following Taylor approximation

$$f(x+h) = f(x) + D_x f(h) + \frac{1}{2} D_{xx} f(h,h) + O(|h|^3).$$

Here $D_x f(h)$ is linear in h and denotes the derivative of f wit respect to x in the direction h, and $D_{xx} f(h,k)$ is linear in h and k and denotes the second derivative of f with respect to x in the directions h and k. Finally $|h|$ denotes a norm on the space of matrices $M_{n,m}$. The condition for a point x to be stationary is that the derivative with respect to x in the direction h is zero for all directions, or that $D_x f(h) = 0$ for all $h \in M_{n,m}$.

The likelihood function for the model given by (28) is

$$\log L(\alpha,\beta,\Omega) = -\frac{T}{2}\log|\Omega| - \frac{1}{2} tr\{\Omega^{-1}\Sigma_{t=1}^T \epsilon_t(\alpha,\beta)\epsilon_t(\alpha,\beta)'\}.$$

From this function we find the derivatives with respect to the various parameters. The derivatives and their order of magnitude is given in Table 7.

Table 7

Derivatives and their order of magnitude for the I(1) model

$D_\beta LogL(b)$	$=$	$tr\left\{\Omega^{-1}\sum_{t=1}^T \epsilon_t X'_{t-1} b\alpha'\right\}$	$\in$	$O_P(T)$
$D_\alpha LogL(a)$	$=$	$tr\left\{\Omega^{-1}\sum_{t=1}^T \epsilon_t X'_{t-1}\beta a'\right\}$	$\in$	$O_P(T^{\frac{1}{2}})$
$D_\Omega LogL(h)$	$=$	$\frac{T}{2} tr\left\{\Omega^{-1}h\Omega^{-1}(T^{-1}\sum_{t=1}^T \epsilon_t\epsilon'_t - \Omega)\right\}$	$\in$	$O_P(T^{\frac{1}{2}})$
$D_{\beta\beta} LogL(b,b)$	$=$	$-tr\left\{\Omega^{-1}\alpha b'\sum_{t=1}^T X_{t-1}X'_{t-1} b\alpha'\right\}$	$\in$	$O_P(T^2)$
$D_{\beta\alpha} LogL(b,a)$	$=$	$-tr\left\{\Omega^{-1}a\beta'\sum_{t=1}^T X_{t-1}X'_{t-1} b\alpha'\right\}$		
		$+tr\left\{\Omega^{-1}\sum_{t=1}^T \epsilon_t X'_{t-1} b a'\right\}$	$\in$	$O_P(T)$
$D_{\beta\Omega} LogL(b,h)$	$=$	$-tr\left\{\Omega^{-1}h\Omega^{-1}\sum_{t=1}^T \epsilon_t X'_{t-1} b\alpha'\right\}$	$\in$	$O_P(T)$
$D_{\alpha\alpha} LogL(a,a)$	$=$	$-tr\left\{\Omega^{-1}a\beta'\sum_{t=1}^T X_{t-1}X'_{t-1}\beta a'\right\}$	$\in$	$O_P(T)$
$D_{\alpha\Omega} LogL(a,h)$	$=$	$-tr\left\{\Omega^{-1}h\Omega^{-1}\sum_{t=1}^T \epsilon_t X'_{t-1}\beta a'\right\}$	$\in$	$O_P(T^{\frac{1}{2}})$
$D_{\Omega\Omega} LogL(h,h)$	$=$	$\frac{T}{2} tr\left\{\Omega^{-1}h\Omega^{-1}h\right\} - tr\left\{\Omega^{-1}h\Omega^{-1}h\Omega^{-1}\sum_{t=1}^T \epsilon_t\epsilon'_t\right\}$	$\in$	$O_P(T)$.

It is seen from Table 4 and Table 7 that

$$T^{-1}D_\beta \log L(b) \xrightarrow{w} tr\left\{\Omega^{-1}\int_0^1 (dW)W'C'b\alpha'\right\},$$

and that $T^{-\frac{1}{2}}D_\alpha \log L(a)$ and $T^{-\frac{1}{2}}D_\Omega \log L(h)$ are asymptotically Gaussian. The asymptotic variance of the first is $\Omega^{-1}\otimes aVar(\beta'X_t)a'$ and the second has a limit distribution which can be derived from that of

$$T^{-\frac{1}{2}}\sum_{t=1}^{T}(\epsilon_t\epsilon_t' - \Omega).$$

If the second derivatives are normalized similarly we find that

Table 8
Limits of second derivatives in the I(1) model

$T^{-2}D^2_{\beta\beta}\log L(b,b)$	$\xrightarrow{w}$	$-tr\left\{\Omega^{-1}\alpha b'C\int_0^1 WW'dtC'b\alpha'\right\}$
$T^{-1}D^2_{\alpha\alpha}\log L(a,a)$	$\xrightarrow{P}$	$-tr\left\{\Omega^{-1}aVar(\beta'X_t)a'\right\}$
$T^{-1}D^2_{\Omega\Omega}\log L(h,h)$	$\xrightarrow{P}$	$-\frac{1}{2}tr\left\{\Omega^{-1}h\Omega^{-1}h\right\}$

whereas all mixed derivatives normalized by the corresponding powers of T tend to zero in probability. This has the consequence that inference about any of the parameters α, β and Ω can be conducted as if the others were known. The reason for this is that the information matrix becomes block diagonal in the limit, a result that holds for the $I(2)$ analysis as well. Since inference on Ω and α is usual inference in the sense that the limit distributions Gaussian we focus in the following on the asymptotic properties of the estimator for β.

First, however, we need to discuss the normalization of the estimator, since evidently α and β are not identified since only their product enters the equations. A general way of normalization is to choose a matrix c $(p\times r)$ and define $\beta_c = \beta(c'\beta)^{-1}$, so that$\hat{\beta}_c = \hat{\beta}(c'\hat{\beta})^{-1}$ is the same for any choice of maximum likelihood estimator. A very special choice is found as follows:

$$\hat{\beta} = \beta b_0 + \beta_\perp b_1$$

implies that

$$\tilde{\beta} = \hat{\beta}b_0^{-1} = \beta + \beta_\perp b_1 b_0^{-1},$$

so that

$$\tilde{\beta} - \beta = T^{-1}\beta_\perp B_T,$$

and hence the deviation between $\tilde{\beta}$ and β are contained in the space spanned by $\beta_\perp$ Note that $\tilde{\beta} = \beta_c$ for $c = \bar{\beta}$. We shall find the asymptotic distribution of

$$B_T = T\bar{\beta}_\perp'(\tilde{\beta} - \beta).$$

We then derive the well known result about the asymptotic distribution of $\tilde{\beta}$ in such a way that the proof can be applied in the $I(2)$ model. The likelihood equation for β for fixed α and Ω is

$$\alpha'\Omega^{-1}\sum_{t=1}^{T}(\Delta X_t - \alpha\tilde{\beta}' X_{t-1})X'_{t-1} = 0.$$

Now replace ΔX_t by $\alpha\beta' X_{t-1} + \epsilon_t$ and we find the equation

$$\alpha'\Omega^{-1}\alpha(\tilde{\beta} - \beta)'\sum_{t=1}^{T} X_{t-1}X'_{t-1} = \alpha'\Omega^{-1}\sum_{t=1}^{T}\epsilon_t X'_{t-1}.$$

Expressed in terms of B_T this becomes

$$\alpha'\Omega^{-1}\alpha B'_T T^{-2}\beta'_\perp \sum_{t=1}^{T} X_{t-1}X'_{t-1}\beta_\perp = \alpha'\Omega^{-1}T^{-1}\sum_{t=1}^{T}\epsilon_t X'_{t-1}\beta_\perp.$$

In the limit we find from Table 4 for B equal to the weak limit of B_T.

$$\alpha'\Omega^{-1}\alpha B'\beta'_\perp C\int_0^1 WW'dtC'\beta_\perp = \alpha'\Omega^{-1}\int_0^1 (dW)W'C'\beta_\perp.$$

Hence we find the result for the estimation of β that the limit distribution is mixed Gaussian:

$$T(\tilde{\beta} - \beta) \xrightarrow{w} \beta_\perp B = \bar{\beta}'_\perp\alpha_\perp\left[\alpha'_\perp\int_0^1 WW'dt\alpha_\perp\right]^{-1}\alpha'_\perp\int_0^1 W(dW)'\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1},$$

with asymptotic quadratic variation given by

$$\bar{\beta}'_\perp\alpha_\perp\left[\alpha'_\perp\int_0^1 WW'dt\alpha_\perp\right]^{-1}\alpha'_\perp\bar{\beta}_\perp \otimes (\alpha'\Omega^{-1}\alpha)^{-1}.$$

We have given here a very brief summary of the results known for $I(1)$ models, in order to present them in a way that can be generalized to the $I(2)$ case. Thus we shall first find derivatives of the likelihood function and discuss their order of magnitude. We then pick out for further study those parameters that are consistent of a higher order then usual, that is, of the order of T and higher.

7 Derivatives of the log likelihood function for the I(2) model

The likelihood function can be derived from (15) and (16) in terms of the new parameters and is proportional to

$$L(\tau,\rho,\xi) = (|\alpha'\alpha||\alpha_\perp'\alpha_\perp|\,|\Omega_1|\,|\Omega_2|)^{-\frac{1}{2}T}\exp\left\{-\frac{1}{2}tr(\Omega_1^{-1}\sum_{t=1}^{T}\epsilon_{1t}\epsilon_{1t}' + \Omega_2^{-1}\sum_{t=1}^{T}\epsilon_{2t}\epsilon_{2t}')\right\}, \tag{19}$$

where ϵ_{1t} and ϵ_{2t} depend on the new parameters. By expanding ϵ_{1t} and ϵ_{2t} we find the results in Table 9.

Table 9
Derivatives of ϵ_{1t} and ϵ_{2t} with respect to the parameters θ and their order of magnitude

$D_\alpha \epsilon_{1t}(a)$	$=$	$(D_\alpha \bar{\alpha}_\perp(a))'\Delta^2 X_t$	$I(0)$
$D_\kappa \epsilon_{1t}(k)$	$=$	$-k'\tau'\Delta X_{t-1}$	$I(0)$
$D_\tau \epsilon_{1t}(s)$	$=$	$-\kappa' s'\Delta X_{t-1}$	$I(1)$
$D_\alpha \epsilon_{2t}(a)$	$=$	$(D_\alpha(\bar{\alpha}(a) - \bar{\alpha}_\perp(a)\omega'))'\Delta^2 X_t$	$I(0)$
$D_\omega \epsilon_{2t}(b)$	$=$	$-b'\bar{\alpha}_\perp'\Delta^2 X_t$	$I(0)$
$D_\xi \epsilon_{2t}(h)$	$=$	$-h'\Delta X_{t-1}$	$I(1)$
$D_\rho \epsilon_{2t}(r)$	$=$	$-r'\tau' X_{t-2}$	$I(1)$
$D_\tau \epsilon_{2t}(s)$	$=$	$-\rho' s' X_{t-2}$	$I(2)$
$D^2_{\omega\alpha} \epsilon_{2t}(b,a)$	$=$	$-b'(D_\alpha \bar{\alpha}_\perp(a))'\Delta^2 X_t$	$I(0)$
$D^2_{\rho\tau} \epsilon_{2t}(r,s)$	$=$	$-r's'X_{t-2}$	$I(2)$
$D^2_{\kappa\tau} \epsilon_{1t}(k,s)$	$=$	$-k's'\Delta X_{t-1}$	$I(1)$.

We find, as in the $I(1)$ analysis, that asymptotic inference can be conducted independently in three blocks. The first block is defined by the variances Ω_1 and Ω_2. The second block is defined by the coefficients to the stationary variables, that is, (α, ω, κ) or in other words those parameters for which the derivative of the errors is $I(0)$. Finally the third block is defined by the parameters (τ, ρ, ξ) which are coefficients of non stationary variables, or those for which the derivative of the errors is non stationary, see Table 9.

We shall apply this result in the following and focus on the parameters (τ, ρ, ξ) , since these parameters are the ones for which "non standard" inference is needed.

We therefore give the results in Table 10:

Table 10
Derivatives of the log-likelihood function with respect to the parameters τ, ρ and ξ.

$T^{-1}D_\rho \ln L(r)$	$=$	$tr\{\Omega_2^{-1}r'\tau' M_{2\epsilon_2}\}$	$\in$	$O_P(1)$
$T^{-1}D_\tau \ln L(t)$	$=$	$tr\{\Omega_2^{-1}\rho' t' M_{2\epsilon_2}\} + tr\{\Omega_1^{-1}\kappa' t' M_{1\epsilon_1}\}$	$\in$	$O_P(T)$
$T^{-1}D_\xi \ln L(h)$	$=$	$tr\{\Omega_2^{-1}h' M_{1\epsilon_2}\}$	$\in$	$O_P(T)$
$-T^{-1}D^2_{\rho\rho} \ln L(r,r)$	$=$	$tr\{\Omega_2^{-1}r'\tau' M_{22}\tau r\}$	$\in$	$O_P(T)$
$-T^{-1}D^2_{\tau\rho} \ln L(r,t)$	$=$	$tr\{\Omega_2^{-1}r'\tau' M_{22}t\rho\} - tr\{\Omega_2^{-1}r't' M_{2\epsilon_2}\}$	$\in$	$O_P(T^{\frac{3}{2}})$
$-T^{-1}D^2_{\tau\tau} \ln L(t,t)$	$=$	$tr\{\Omega_2^{-1}\rho' t' M_{22}t\} - tr\{\Omega_1^{-1}\kappa' t' M_{11}t\kappa\}$	$\in$	$O_P(T^2)$
$-T^{-1}D^2_{\xi\xi} \ln L(h,h)$	$=$	$tr\{\Omega_2^{-1}h' M_{11}h\}$	$\in$	$O_P(T)$
$-T^{-1}D^2_{\xi\tau} \ln L(h,t)$	$=$	$tr\{\Omega_2^{-1}h' M_{12}t\rho\}$	$\in$	$O_P(T^{\frac{3}{2}})$
$-T^{-1}D^2_{\xi\rho} \ln L(h,r)$	$=$	$tr\{\Omega_2^{-1}h' M_{12}\tau r\}$	$\in$	$O_P(T)$

A formal argument for the above result that we need only consider some of the parameters, can be carried out as follows: If θ_i denotes a parameter for which $D_{\theta_i}\epsilon_t$ is $I(i)$, $i = 0, 1, 2$, then

$$D_{\theta_i} LogL(\theta) \in O_P(T^{i\vee\frac{1}{2}}),$$

$$D^2_{\theta_i\theta_j} LogL(\theta) \in O_P(T^{(i+j)\vee 1}),$$

see Table 9. This shows that we should normalize the first derivatives differently depending on their order of magnitude, and we should use a block diagonal matrix with diagonal elements $T^{-\frac{1}{2}}, T^{-1}$and T^{-2}, respectively for the derivatives. The $(i,j)'th$ block of the second derivative matrix is then normalized by $T^{-i\vee\frac{1}{2}-j\vee\frac{1}{2}}$ with the result that the order of magnitude of the normalized matrix of second derivatives becomes

$$\begin{pmatrix} & \theta_0 & \theta_1 & \theta_2 \\ \theta_0 & 1 & 0 & 0 \\ \theta_1 & 0 & 1 & 1 \\ \theta_2 & 0 & 1 & 1 \end{pmatrix},$$

which shows that the normalized matrix is block diagonal, and that inference concerning the parameters θ_1 and θ_2, which are coefficients of non-stationary variables, can be separated from inference for the parameter θ_0, which is a coefficient of a stationary variable. That is, when making inference on θ_0 we can assume that θ_1 and θ_2 are known, and vice versa.

In the asymptotic theory we can thus fix the parameters $(\omega, \alpha, \kappa, \Omega_1, \Omega_2)$ when deriving the asymptotic distribution of the maximum likelihood estimators for the parameters (τ, ρ, ξ).

Thus in the next section we analyse the likelihood function of the $I(2)$ model for fixed values of the parameters $(\alpha, \omega, \kappa, \Omega_1, \Omega_2)$, which means that the remaining parameters that have to be determined simultaneously are (τ, ρ, ξ). That is, the $C(2,1)$ vectors τ have to be determined and among those the β vectors that cointegrate with the differences, that is, the coefficients ρ. Finally the parameters ξ contain information about the linear combinations of the differences that are needed to make $\beta' X_t$ stationary, see (17).Thus we focus entirely on the parameters corresponding to the non-stationary variables.

8 The normalization of the estimators

The parameters in the new parametrization are varying freely, but turn out not to be identified. Hence it is important to discuss to what extent the model is overparametrized, and find suitable identifying normalizations.

Since the likelihood function of the original model only depends on the parameters through the products

$$\alpha\rho'\tau', \alpha\xi', \alpha_\perp\kappa'\tau', \alpha\omega\kappa'\tau',$$

see (17), the parameters are not identified. Specifically we find that for any non singular $\gamma(r+s)\times(r+s)$ the matrices $\tau\gamma^{-1}$, $\gamma\rho$ and $\gamma\kappa$ will give the same value of the likelihood function as τ, ρ and κ, and that for any $\psi(r\times r)$ the parameters $\alpha\psi$, $\rho\psi'^{-1}$, $\xi\psi'^{-1}$, $\psi^{-1}\omega$, will give the same value of the likelihood function.

It is therefore important to normalize the parameters in order to obtain estimates and asymptotic distributions. We introduce the two types of normalization as discussed in the $I(1)$ analysis. Thus for instance we define

$$\tau_c = \tau(c'\tau)^{-1},$$

which is normalized by the condition $c'\tau = I$, and

$$\tilde{\tau} = \hat{\tau}(\bar{\tau}'\hat{\tau})^{-1}.$$

Once τ has been identified then $sp(\rho)$ can be estimated, but ρ also needs a normalization. We choose to represent the deviations in the space spanned by τ and find the decomposition

$$\tau\hat{\rho} = \beta b_1 + \beta_1 b_2,$$

and define the estimator $\tilde{\rho} = \hat{\rho} b_1^{-1}$, so that

$$\tau\tilde{\rho} - \beta = \tau(\tilde{\rho} - \rho) = \beta_1 b_2 b_1^{-1},$$

which is contained in the space $sp(\beta_1)$. Finally we define

$$\tilde{\xi} = \hat{\xi}(\bar{\xi}'\hat{\xi})^{-1}.$$

Note that the derivative of ϵ_{2t} with respect to ξ in the direction h, that is, $-h'\Delta X_{t-1}$, is $I(0)$ if $sp(h) \subset sp(\tau)$, whereas it is $I(1)$ otherwise. This has the consequence that $\tau'(\tilde{\xi}-\xi)$ satisfy usual asymptotics, and we therefore focus on $\beta_2'(\tilde{\xi}-\xi)$ below.

The usual proof for the asymptotic distribution of the maximum likelihood estimator involves an expansion of the derivative of the likelihood function around the true value, keeping the terms involving derivatives and second derivatives. In the case of $I(2)$ variables it turns out the higher order terms are needed, and that a slightly different way of expanding is convenient. We shall simple write the equation for the definition of the maximum likelihood estimator and show by going to the limit that we can derive the limit distribution of the maximum likelihood estimator.

A detailed analysis in the next section of the likelihood function shows how the estimators should be normalized in order to give a limit distribution. We give below the results and show in section 10 how this choice leads to an asymptotic solution of the likelihood equations.

Table 11
Normalization of the estimators

B_{0T}	$=$	$T\beta_2'(\tilde{\xi}-\xi)$	$(p-r-s)\times r$
B_{1T}	$=$	$T\rho_\perp'(\tilde{\rho}-\rho)$	$s\times r$
B_{2T}	$=$	$T^2\beta_2'(\tilde{\tau}-\tau)(\rho+\bar{\rho}_\perp\rho_\perp'(\tilde{\rho}-\rho))$	$(p-r-s)\times r$
V_T	$=$	$T\beta_2'(\tilde{\tau}-\tau)\bar{\rho}_\perp\beta_1'\beta_1$	$(p-r-s)\times s$
U_T	$=$	$T^2\beta_2'(\tilde{\tau}-\tau)\rho$	$(p-r-s)\times r.$

From Table 11 we find that

$$\tilde{\tau}-\tau=T^{-2}\bar{\beta}_2U_T\bar{\rho}'+T^{-1}\bar{\beta}_2V_T(\beta_1'\beta_1)^{-1}\rho_\perp', \tag{20}$$

since $\tilde{\tau}-\tau\in sp(\beta_2)$. Further since $\tau(\tilde{\rho}-\rho)\in sp(\beta_1)$ it holds that

$$\tau(\tilde{\rho}-\rho)=\beta_1(\beta_1'\beta_1)^{-1}\beta_1'\tau(\tilde{\rho}-\rho)=T^{-1}\bar{\beta}_1B_{1T}, \tag{21}$$

since $\beta_1'\tau=\rho_\perp'\bar{\tau}'\tau=\rho_\perp'$. This can be interpreted as follows: The estimator $\tilde{\tau}$ deviates from the true value τ in the direction of $\beta_2=\tau_\perp$. In this direction there are still two different orders of magnitude. That is, $\beta_2'(\tilde{\tau}-\tau)\rho$ has to be normalized by T^2, whereas $\beta_2'(\tilde{\tau}-\tau)\rho_\perp$ should be normalized by T.

For the orthogonal complement of $\tilde{\rho}$ which will also be used below, we can use the expansion

$$\tilde{\rho}_\perp=\rho_\perp-\tau'\tau\rho(\beta'\beta)^{-1}(\tilde{\rho}-\rho)'\rho_\perp=\rho_\perp-T^{-1}\tau'\tau\rho(\beta'\beta)^{-1}B_{1T}'.$$

To see this note that from (21) it follows that

$$\rho'\tau'\tau(\tilde{\rho}-\rho)=T^{-1}\beta'\bar{\beta}_1B_{1T}=0,$$

so that

$$\tilde{\rho}'\tau'\tau\rho=\rho'\tau'\tau\rho=\beta'\beta.$$

Hence

$$\tilde{\rho}'\tilde{\rho}_\perp = \tilde{\rho}'\rho_\perp - \tilde{\rho}\tau'\tau\rho(\beta'\beta)^{-1}(\tilde{\rho}-\rho)'\rho_\perp = 0.$$

This choice of normalization has implications for the estimators of the other parameters of interest, in particular $\beta = \tau\rho$, $\beta_1 = \bar{\tau}\rho_\perp$, and $\beta_2 = \tau_\perp$. We derive expressions for the estimators of these parameters and expressions in terms of V_T, U_T and B_T.

8.1 Estimator for β

We then find

$$\tilde{\beta} - \beta = \tilde{\tau}\tilde{\rho} - \tau\rho = (\tilde{\tau}-\tau)(\tilde{\rho}-\rho) + \tau(\tilde{\rho}-\rho) + (\tilde{\tau}-\tau)\rho. \tag{22}$$

Since $\tilde{\tau} - \tau \in sp(\tau_\perp)$ it follows that $\beta'(\tilde{\tau}-\tau) = 0$ and $\beta_1'(\tilde{\tau}-\tau) = 0$ so that

$$\beta'(\tilde{\beta}-\beta) = \beta'\tau(\tilde{\rho}-\rho) = T\beta'\bar{\beta}_1 B_{1T} = 0,$$

and

$$T\beta_1'(\tilde{\beta}-\beta) = T\beta_1'\tau(\tilde{\rho}-\rho) = \beta_1'\bar{\beta}_1 B_{1T} = B_{1T},$$

see (21). Finally from (20).

$$T^2\beta_2'(\tilde{\beta}-\beta) = [T^{-1}U_T\bar{\rho}' + V_T(\beta_1'\beta_1)^{-1}\rho_\perp'][(\tau'\tau)^{-1}\rho_\perp(\beta_1'\beta_1)^{-1}B_{1T}] + U_T$$

$$= U_T + V_T(\beta_1'\beta_1)^{-1}B_{1T} + O_P(T^{-1}) = B_{2T} + O_P(T^{-1}),$$

where the last equality follows from the definition of B_{2T}. Note that (22) represents the estimator of β as a product of two terms which gives three terms in the expansion. Even though the first term is a product of two small terms and the others only contain one small term it still holds that there is a direction, β_2, in which the first and the third term have the same order of magnitude. Note also that the definition of $\tilde{\rho}$ implies that in the direction β the estimator $\tilde{\beta}-\beta$ has no component. This will turn out to be different for $\tilde{\beta}_1$ below.

8.2 Estimator for β_2

As an estimator of $\beta_2 = \tau_\perp$ we apply

$$\tilde{\beta}_2 = \tau_\perp - \tau(\tilde{\tau}'\tau)^{-1}\tilde{\tau}'\tau_\perp,$$

which satisfies $\tilde{\tau}'\tilde{\beta}_2 = \tilde{\tau}'\tau_\perp - \tilde{\tau}'\tau(\tilde{\tau}'\tau)^{-1}\tilde{\tau}'\tau_\perp = 0$, so that, since $\tilde{\tau}'\tau = \tau'\tau$,

$$\tilde{\beta}_2 - \beta_2 = \tilde{\tau}_\perp - \tau_\perp = -\tau(\tau'\tau)^{-1}(\tilde{\tau}-\tau)'\tau_\perp$$

$$= -\bar{\tau}(T^{-2}\bar{\rho}U_T' + T^{-1}\rho_\perp(\beta_1'\beta_1)^{-1}V_T').$$

From this expression we find the relations

$$\begin{aligned} T^2\beta'(\tilde{\beta}_2 - \beta_2) &= -U_T', \\ T\beta_1'(\tilde{\beta}_2 - \beta_2) &= -V_T' + O_P(T^{-1}), \\ \beta_2'(\tilde{\beta}_2 - \beta_2) &= 0. \end{aligned}$$

8.3 Estimator of β_1

The distribution of β_1 is found from $\tilde{\beta}_1 = \tilde{\tau}(\tilde{\tau}'\tilde{\tau})^{-1}\tilde{\rho}_\perp$. Since

$$\tilde{\tau}'\tilde{\tau} = \tau'\tau + (\tilde{\tau} - \tau)'(\tilde{\tau} - \tau)$$

it follows that

$$(\tilde{\tau}'\tilde{\tau})^{-1} = (\tau'\tau)^{-1} - (\tau'\tau)^{-1}(\tilde{\tau} - \tau)'(\tilde{\tau} - \tau)(\tau'\tau)^{-1} + O_P(T^{-4}).$$

Thus

$$\begin{aligned} &\tilde{\beta}_1 - \beta_1 \\ &= (\tilde{\tau} - \tau)(\tau'\tau)^{-1}\rho_\perp + \bar{\tau}(\tilde{\rho}_\perp - \rho_\perp) - \bar{\tau}(\tilde{\tau} - \tau)'(\tilde{\tau} - \tau)(\tau'\tau)^{-1}\rho_\perp \\ &\quad + (\tilde{\tau} - \tau)(\tau'\tau)^{-1}(\tilde{\rho}_\perp - \rho_\perp) + O_P(T^{-3}). \end{aligned}$$

This gives

$$T\beta'(\tilde{\beta}_1 - \beta_1) = T\beta'\bar{\tau}(\tilde{\rho}_\perp - \rho_\perp) + O_P(T^{-1}),$$

$$= T\rho'(\tilde{\rho}_\perp - \rho_\perp) + O_P(T^{-1}) = -B_{1T}' + O_P(T^{-1}),$$

and

$$T\beta_2'(\tilde{\beta}_1 - \beta_1) = T\beta_2'(\tilde{\tau} - \tau)(\tau'\tau)^{-1}\rho_\perp + O_P(T^{-1}) = V_T + O_P(T^{-1}).$$

In the direction β_1, however, the situation is different: For the first term we have β_1' $(\tilde{\tau} - \tau)(\tau'\tau)^{-1}\rho_\perp = 0$, since $\tilde{\tau} - \tau$ is contained in $sp(\beta_2)$. The second term becomes

$$T\beta_1'\bar{\tau}(\tilde{\rho}_\perp - \rho_\perp) = \rho_\perp'(\tau'\tau)^{-1}\tau'\tau(\tau'\tau)^{-1}(\tau'\tau)\rho(\beta'\beta)^{-1}B_{1T}' = 0.$$

Finally we find for the third term

$$\begin{aligned} &T^2\beta_1'\bar{\tau}(\tilde{\tau} - \tau)'(\tilde{\tau} - \tau)(\tau'\tau)^{-1}\rho_\perp \\ &= \rho_\perp'(\tau'\tau)^{-1}(T^{-1}\bar{\rho}U_T' + \rho_\perp(\beta_1'\beta)^{-1}V_T')\bar{\tau}_\perp'\bar{\tau}_\perp V_T(\beta_1'\beta_1)^{-1}\rho_\perp'(\tau'\tau)^{-1}\rho_\perp + O_P(T^{-1}) \\ &= V_T'(\beta_2'\beta_2)^{-1}V_T + O_P(T^{-1}). \end{aligned}$$

Thus we get the results for the three estimators in the three directions which we summarize in Table 12.

Table 12
The normalization of the estimators of the cointegrating relations in the various directions

	β	β_1	β_2
$\tilde{\beta}-\beta$	0	$T^{-1}B_{1T}$	$T^{-2}B_{2T}$
$\tilde{\beta}_1-\beta_1$	$-T^{-1}B'_{1T}$	$-T^{-2}V'_T(\beta'_2\beta_2)^{-1}V_T$	$T^{-1}V_T$
$\tilde{\beta}_2-\beta_2$	$T^{-2}B'_{1T}(\beta'_1\beta_1)^{-1}V'_T-T^{-2}B'_{2T}$	$-T^{-1}V'_T$	0

We have here discussed the normalization of the parameters so that they can be estimated. We have also given the results of how fast the various components of the parameters estimates converge to their limits, and in section 10 we will apply all that to the likelihood equations, with the purpose of deriving the asymptotic distribution of the estimated parameters ξ, β, β_1 and β_2, but first we consider the question of consistency.

9 The consistency of the maximum likelihood estimators

The likelihood function (19) derived from (15) and (16) depends on the parameters $\gamma = (\tau, \rho, \xi)$ and a maximum likelihood estimator $\hat{\gamma}$ satisfies $L(\hat{\gamma}) \geq L(\gamma_0)$ for all γ_0. Thus $\hat{\gamma}$ is a point in the set

$$S_T = \{\gamma | q(\gamma) = -2\log(L(\gamma)/L(\gamma_0)) \leq 0\}.$$

We can also consider the likelihood function a function of the parameters $\gamma^* = (\tau, \beta, \xi)$ where these parameters vary freely. In this case $-2\log L(\gamma^*)$ is quadratic in γ^* with a positive definite second derivative, so that

$$S_T^* = \{\gamma^* | q(\gamma^*) = -2\log(L(\gamma^*)/L(\gamma_0)) \leq 0\},$$

is convex and compact. Since $S_T \subset S_T^*$ also S_T is compact, and hence that the maximum likelihood estimator exists and is contained in S_T.

We prove below a result for a quadratic likelihood function, where it is shown that the order of magnitude of the second derivative determines how fast the estimator converges to its true value.

Lemma 1 *Let $A_T(p\times p)$, $B_T(p\times 1)$ and $C_T(1\times 1)$ be random matrices and D_T a non random $(p\times p)$ matrix with the property that*

$$D_T \to \infty, D_T'^{-1}A_T D_T^{-1} \xrightarrow{w} A,\ D_T'^{-1}B_T \xrightarrow{w} B,\ C_T \xrightarrow{w} C.$$

We further assume that A_T and A are positive definite. Assume that $\hat{\theta}_T$ is a sequence of estimators satisfying

$$\hat{\theta}'_T A_T \hat{\theta}_T + \hat{\theta}'_T B_T + C_T \leq 0,$$

then

$$T^{-\delta} D_T \hat{\theta}_T \xrightarrow{P} 0, \textit{ for all } \delta > 0.$$

Proof: From the inequality

$$T^{2\delta}(T^{-\delta}D_T\hat{\theta}_T)'(D_T'^{-1}A_TD_T^{-1})(T^{-\delta}D_T\hat{\theta}_T)+T^{\delta}(T^{-\delta}D_T\hat{\theta}_T)'(D_T'^{-1}B_T)+C_T\leq 0, \tag{23}$$

it follows that if $T^{-\delta}D_T\hat{\theta}_T$ does not converge to zero then the first term of (23) will tend to ∞, which violates the inequality. More precisely the first term is larger than

$$T^{2\delta}|T^{-\delta}D_T\hat{\theta}_T|^2\lambda_{\min}(D_T'^{-1}A_TD_T^{-1}).$$

If $|T^{-\delta}D_T\hat{\theta}_T|$ does not go to zero in probability then $T^{2\delta}|T^{-\delta}D_T\hat{\theta}_T|^2 \xrightarrow{P} \infty$, and by the weak convergence of $D_T'^{-1}A_TD_T^{-1}$ to a positive definite limit the third factor does not tend to zero. This shows that the first term of (23) tends to ∞ and since it dominates the other terms the inequality is violated for large enough values of T.

We apply this lemma to the likelihood function (19) derived from (15) and (16) by investigating the second derivative of $q(\gamma)$ with respect to the various parameters in the various directions.

Theorem 1 *Let $\hat{\tau}_T, \hat{\rho}_T$, and $\hat{\xi}_T$ be maximum likelihood estimators, which are in the set S_T, and normalize them as described above to $\tilde{\tau}_T, \tilde{\rho}_T$, and $\tilde{\xi}_T$. Let further θ_0 be the value under which probabilities are calculated. Then with respect to the probability measure P_{θ_0}, and for any $\delta > 0$*

$$T^{1-\delta}(\tilde{\rho}_T-\rho)\xrightarrow{P} 0, \tag{24}$$

$$T^{2-\delta}\beta_2'(\tilde{\tau}_T-\tau)\rho\xrightarrow{P} 0, \tag{25}$$

$$T^{1-\delta}\beta_2'(\tilde{\tau}_T-\tau)\rho_\perp\xrightarrow{P} 0, \tag{26}$$

$$T^{1-\delta}\beta_2'(\tilde{\xi}_T-\xi)\xrightarrow{P} 0. \tag{27}$$

These results show the superconsistency of the estimators of the parameters corresponding to the non stationary variables. Most of the results are of the same form as for $I(1)$ variables but (25) shows that for $I(2)$ variables we can get even faster convergence.

Proof: We consider the function $q(\gamma)$ at the point $\tilde{\gamma}_T = (\tilde{\tau}_T, \tilde{\rho}_T, \tilde{\xi}_T)$ and replace γ_0 by γ. Then

$$q(\tilde{\gamma}_T)=\begin{array}{c} tr\{\Omega_1^{-1}\sum_{t=1}^T(\epsilon_{1t}(\tilde{\gamma}_T)\epsilon_{1t}(\tilde{\gamma}_T)'-\epsilon_{1t}(\gamma)\epsilon_{1t}(\gamma)') \\ +\Omega_2^{-1}\sum_{t=1}^T(\epsilon_{2t}(\tilde{\gamma}_T)\epsilon_{2t}(\tilde{\gamma}_T)'-\epsilon_{2t}(\gamma)\epsilon_{2t}(\gamma)')\}, \end{array}$$

where

$$\epsilon_{1t}(\gamma)-\epsilon_{1t}(\tilde{\gamma}_T)=\kappa'(\tilde{\tau}_T-\tau)'\Delta X_{t-1},$$

$$\epsilon_{2t}(\gamma) - \epsilon_{2t}(\tilde{\gamma}_T) = (\tilde{\xi}_T - \xi)'\Delta X_{t-1} + \tilde{\rho}_T(\tilde{\tau}_T - \tau)'X_{t-2} + (\tilde{\rho}_T - \rho)'\tau' X_{t-2}$$

Thus $q(\tilde{\gamma}_T)$ is essentially quadratic in $\tilde{\gamma}_T - \gamma$. The quadratic term is easily found as expressed by M_{ij}, and we find that

$$\begin{aligned}
\beta_2' M_{11}\beta_2 &\in O_P(T) \text{ implies } T^{1-\delta}(\tilde{\tau}_T - \tau) \xrightarrow{P} 0,\\
\beta_2' M_{11}\beta_2 &\in O_P(T) \text{ implies } T^{1-\delta}\beta_2'(\tilde{\xi}_T - \xi) \xrightarrow{P} 0,\\
\tau' M_{22}\tau &\in O_P(T) \text{ implies } T^{1-\delta}(\tilde{\rho}_T - \rho) \xrightarrow{P} 0,\\
\beta_2' M_{22}\beta_2 &\in O_P(T^3) \text{ implies } T^{2-\delta}\beta_2'(\tilde{\tau} - \tau)\rho \xrightarrow{P} 0.
\end{aligned}$$

This completes the proof of the consistency of the estimators. The general rule that can be extracted from this is that the second derivative of the likelihood function determines the speed of convergence of the estimator.

10 The asymptotic distribution of the maximum likelihood estimators

Below we give the asymptotic distribution of the estimators for the parameters $\tilde{\xi}, \tilde{\beta}, \tilde{\beta}_1$, and $\tilde{\beta}_2$. These results will then be applied to find the distributions of the estimators of β, β_1 and β_2 normalized by matrices c, c_1 and c_2 respectively.

In order to describe the asymptotic distributions we define the mixed Gaussian distributions of $B = (B_0', B_1', B_2')'$

$$B = [\int_0^1 HH'dt]^{-1}\int_0^1 H(dW_2)' \tag{28}$$

and

$$V = [\int_0^1 H_0H_0'dt]^{-1}\int_0^1 H_0(dW_1)'\Omega_1^{-1}\kappa'\rho_\perp(\rho_\perp'\kappa\Omega_1^{-1}\kappa'\rho_\perp)^{-1}(\beta_1'\beta_1). \tag{29}$$

where $H = (H_0', H_1', H_2')'$ is given in Table 5.

Theorem 2 *The asymptotic distribution of the matrices $\tilde{\xi}$, $\tilde{\beta}$, $\tilde{\beta}_1$ and $\tilde{\beta}_2$ is given by*

$$\begin{aligned}
T\beta_2'(\tilde{\xi} - \xi) &\xrightarrow{w} B_0\\
T\beta_1'(\tilde{\beta} - \beta) &\xrightarrow{w} B_1\\
T^2\beta_2'(\tilde{\beta} - \beta) &\xrightarrow{w} B_2\\
T\beta'(\tilde{\beta}_1 - \beta_1) &\xrightarrow{w} -B_1'\\
T^2\beta_1'(\tilde{\beta}_1 - \beta_1) &\xrightarrow{w} -V'(\beta_2'\beta_2)^{-1}V\\
T\beta_2'(\tilde{\beta}_1 - \beta_1) &\xrightarrow{w} V
\end{aligned}$$

$$T\beta_1(\tilde{\beta}_2 - \beta_2) \xrightarrow{w} -V'$$

$$T^2\beta'(\tilde{\beta}_2 - \beta_2) \xrightarrow{w} -B_2' + B_1'(\beta_1'\beta_1)^{-1}V'.$$

Thus apart from $\tilde{\beta}_2$ in the direction β and $\tilde{\beta}_1$ in the direction β_1 the estimators are all asymptotic mixed Gaussian.

We only give the asymptotic distribution of $\tilde{\xi}$ in the direction β_2 since in the direction $\tau = (\beta, \beta_1)$ the normalization is $T^{\frac{1}{2}}$, which means that in this direction the asymptotic distribution should be determined simultaneously with the other parameters $(\alpha, \omega, \kappa, \Omega_1, \Omega_2)$.

Proof: The maximum likelihood estimator satisfies the first order condition $D_\tau \ln L(t) = 0$, for all t where $\tilde{\rho}$, $\tilde{\tau}$ and $\tilde{\xi}$ are inserted, that is, from Table 10 we get

$$T^{-1}D_\tau \ln L(t) = tr\{\Omega_2^{-1}M_{\epsilon_2 2}t\rho\} + tr\{\Omega_1^{-1}M_{\epsilon_1 1}t\kappa\}$$

$$= tr\{\rho\Omega_2^{-1}((\bar{\alpha}' - \omega\bar{\alpha}_\perp')M_{02} - \xi'M_{12} - \rho'\tau'M_{22})t\} + tr\{\kappa\Omega_1^{-1}(\bar{\alpha}_\perp' M_{01} - \kappa'\tau'M_{11})t\}.$$

At the maximum point we have

$$\begin{aligned} 0 &= \tilde{\rho}\Omega_2^{-1}((\bar{\alpha}' - \omega\bar{\alpha}_\perp')M_{02} - \tilde{\xi}'M_{12} - \tilde{\rho}'\tilde{\tau}'M_{22}) + \kappa\Omega_1^{-1}(\bar{\alpha}_\perp' M_{01} - \kappa'\tilde{\tau}'M_{11}) \\ &= \tilde{\rho}\Omega_2^{-1}(M_{\epsilon_2 2} - (\tilde{\xi} - \xi)'M_{12} - (\tilde{\beta} - \beta)'M_{22}) + \kappa\Omega_1^{-1}(M_{\epsilon_1 1} - \kappa'(\tilde{\tau} - \tau)'M_{11}), \end{aligned}$$

see (15) and (16). Now insert the expansion of $\tilde{\beta} - \beta$ and use Table 12:

$$\tilde{\beta} - \beta = \bar{\beta}_1\beta_1'(\tilde{\beta} - \beta) + \bar{\beta}_2\beta_2'(\tilde{\beta} - \beta) = T^{-1}\bar{\beta}_1 B_{1T} + T^{-2}\bar{\beta}_2 B_{2T},$$

$$\tilde{\xi} - \xi = \bar{\beta}_2\beta_2'(\tilde{\xi} - \xi) + \bar{\tau}\tau'(\tilde{\xi} - \xi) = T^{-1}\bar{\beta}_2 B_{0T} + \bar{\tau}\tau'(\tilde{\xi} - \xi).$$

We then find that

$$\begin{aligned} 0 = \tilde{\rho}\Omega_2^{-1} \left[M_{\epsilon_2 2} - T^{-1}B_{0T}'\bar{\beta}_2'M_{12} - (\tilde{\xi} - \xi)'\tau\bar{\tau}'M_{12} - T^{-1}B_{1T}'\bar{\beta}_1'M_{22} - T^{-2}B_{2T}'\bar{\beta}_2'M_{22}\right] \\ +\kappa\Omega_1^{-1}\left[M_{\epsilon_1 1} - \kappa'(\tilde{\tau} - \tau)'M_{11}\right]. \end{aligned} \tag{30}$$

We multiply by the matrix $T^{-1}\bar{\beta}_2$ from the right and find that some of the terms tend to zero in probability:

$$T^{-1}\tilde{\rho}\Omega_2^{-1}(\tilde{\xi} - \xi)'\tau\bar{\tau}'M_{12}\bar{\beta}_2 + T^{-1}\kappa\Omega_1^{-1}(M_{\epsilon_1 1}\bar{\beta}_2 - \kappa'(\tilde{\tau} - \tau)'M_{11}\bar{\beta}_2) \xrightarrow{P} 0,$$

since $\tilde{\xi}$, and $\tilde{\tau}$ are consistent, and $T^{-1}\bar{\tau}'M_{12}\bar{\beta}_2$, $M_{\epsilon_1 1}\bar{\beta}_2$ and $T^{-1}\bar{\beta}_2'M_{11}\bar{\beta}_2$ are bounded in probability, see Table 5. The remaining terms from (30) are

$$0 = [T^{-1}M_{\epsilon_2 2}\bar{\beta}_2] - B'_{0T}[T^{-2}\bar{\beta}_2' M_{12}\bar{\beta}_2] - B'_{1T}[T^{-2}\bar{\beta}_1' M_{22}\bar{\beta}_2] - B'_{2T}[T^{-3}\bar{\beta}_2' M_{22}\bar{\beta}_2] + o_P(1),$$

where we have cancelled $\tilde{\rho}'\Omega_2^{-1}$. Hence the limit $B = lim_{T\to\infty} B_T$ has to satisfy, see Table 5,

$$0 = \int_0^1 (dW_2)H_2' - B_0' \int_0^1 H_0 H_2' dt - B_1' \int_0^1 H_1 H_2' dt - B_2' \int_0^1 H_2 H_2' dt$$

$$= \int_0^1 (dW_2)H_2' - B' \int_0^1 H H_2' dt. \tag{31}$$

Multiplying by $\bar{\beta}_1$ from the right in (30) we obtain similarly the equation

$$0 = \int_0^1 (dW_2)H_1' - B' \int_0^1 H H_1' dt. \tag{32}$$

The derivative with respect to ξ is also zero at the maximum point and we get from Table 10

$$0 = \Omega_2^{-1} M_{\epsilon_2 1} = \Omega_2^{-1}(\bar{\alpha}' M_{01} - \omega \bar{\alpha}_\perp' M_{01} - \tilde{\xi}' M_{11} - \tilde{\beta}' M_{21})$$

$$= \Omega_2^{-1}(M_{\epsilon_2 1} - (\tilde{\xi} - \xi)' M_{11} - (\tilde{\beta} - \beta)' M_{21}).$$

We cancel Ω_2^{-1} and multiply by $\bar{\beta}_2$ from the right to get in the limit

$$0 = \int_0^1 (dW_2)H_0' - B' \int_0^1 H H_0' dt. \tag{33}$$

The equations (31), (32), and (33) prove that the limit of B_T is given by the mixed Gaussian distribution (28).

In order to find the limit distribution of V_T and U_T we multiply equation (30) by $\rho_\perp$ from the left such that the terms that were leading before now vanish. We then get

$$0 = \rho_\perp'(\tilde{\rho} - \rho)\Omega_2^{-1}\left[M_{\epsilon_2 2} - (\tilde{\xi} - \xi)' M_{12} - (\tilde{\beta} - \beta)' M_{22}\right]$$
$$+ \rho_\perp' \kappa \Omega_1^{-1}\left[M_{\epsilon_1 1} - \kappa'(\tilde{\tau} - \tau)' M_{11}\right].$$

Multiplying by $\bar{\beta}_2$ from the right we can apply the results from Table 5

$$M_{\epsilon_1 1}\bar{\beta}_2 \xrightarrow{w} \int_0^1 (dW_1)H_0',$$

and

$$T^{-1}\bar{\beta}_2' M_{11}\bar{\beta}_2 \xrightarrow{w} \int_0^1 H_0 H_0' dt,$$

and get in the limit

$$0 = B_1\Omega_2^{-1}(\int_0^1(dW_2)H_2' - B'\int_0^1 HH_2'dt)+$$
$$\rho_\perp'\kappa\Omega_1^{-1}(\int_0^1(dW_1)H_0' - \kappa'\rho_\perp(\beta_1'\beta_1)^{-1}V'\int_0^1 H_0H_0'dt).$$

The first term is zero by (33) and the relation can then be solved for V:

$$V = [\int_0^1 H_0H_0'dt]^{-1}\int_0^1 H_0(dW_1)'\Omega_1^{-1}\kappa'\rho_\perp[\rho_\perp'\kappa\Omega_1^{-1}\kappa'\rho_\perp]^{-1}(\beta_1'\beta_1).$$

The results of Theorem 2 now follow from Table 12.

Corollary 1 *The limit distribution of $\tilde{\beta}$ is found from*

$$\begin{pmatrix} T\bar{\beta}_1'(\tilde{\beta}-\beta) \\ T^2\bar{\beta}_2'(\tilde{\beta}-\beta) \end{pmatrix} \xrightarrow{w} \begin{pmatrix} (\beta_1'\beta_1)^{-1}B_1 \\ (\beta_2'\beta_2)^{-1}B_2 \end{pmatrix} = [\int_0^1 FF'du]^{-1}\int_0^1 F(dG)',$$

where $F = (F_1', F_2')'$,

$$F_i = (\beta_i'\beta_i)^{-1}(H_i - [\int_0^1 H_iH_0'du][\int_0^1 H_0H_0'du]^{-1}H_0),\ i = 1, 2,$$

and

$$G = (\bar{\alpha}' - \omega\bar{\alpha}_\perp')W = (\alpha\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}W.$$

This limit distribution is the same as the limit distribution for the reduced rank estimator of β suggested in Johansen ([7], Theorem 5 with $c = \tilde{\beta}$.)

Proof: We use $\bar{\beta}_1$ and $\bar{\beta}_2$ to normalize the estimator in order to make the result comparable with that of Theorem 5 in the above mentioned publication. The expression for F_i is now the same and the expression for G can be checked by the identity

$$\bar{\alpha}' - \omega\bar{\alpha}_\perp' = (\alpha\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1},$$

which is proved by multiplying by the full rank matrix $(\alpha, \Omega^{-1}\bar{\alpha}_\perp)$.

This result shows that the two step estimation procedure suggested in Johansen [7] is efficient for the estimation of β, that is, when estimating β one can disregard the second reduced rank condition and simply fit the $I(1)$ model as given by the restriction (5). Paruolo [10] contains a discussion of the efficiency of the two step procedure for the estimation of the remaining parameters.

Finally we give the result for the normalized parameters β_c, β_{1c}, and β_{2c} which are estimated from (4) with the constraints (5) and (6).

Theorem 3 *Let β, β_1, and β_2 be normalized by c, c_1 and c_2 respectively, and let $\hat{\beta}_c, \hat{\beta}_{1c}, \hat{\beta}_{2c}$ denote the maximum likelihood estimators in the I(2) model (4), then*

$$T(\hat{\beta}_c - \beta) \xrightarrow{w} (I - \beta c')\bar{\beta}_1 B_1,$$

$$T(\hat{\beta}_{1c} - \beta_1) \xrightarrow{w} (I - \beta_1 c_1')(\bar{\beta}_2 V - \bar{\beta} B_1'),$$

$$T(\hat{\beta}_{2c} - \beta_2) \xrightarrow{w} - (I - \beta_2 c_2')\bar{\beta}_1 V'.$$

Proof: The likelihood function derived from (4) can be concentrated with respect to $\Gamma_1, ..., \Gamma_{k-2}$ and give the regression equation

$$R_{0t} = \alpha(\rho'\tau' R_{2t} + \xi' R_{1t}) + (\alpha_\perp + \alpha\omega)\kappa'\tau' R_{1t} + \hat{\epsilon}_t,$$

where the residuals R_{it} are indexed by the order of integration. Since the residuals are corrected for stationary variables $\Delta^2 X_{t-1}, ...\Delta^2 X_{t-k+2}$ the asymptotic properties of product moments derived from the residuals R_{2t} and R_{1t} are the same as those derived from X_{t-2} and ΔX_{t-1}, and we can apply the results derived in previous sections. Consider first the estimation of β normalized by c so that $\beta' c = I$. We define the estimator $\hat{\beta}_c = \tilde{\beta}(c'\tilde{\beta})^{-1}$ and find the expansion

$$(\hat{\beta}_c - \beta) = (\beta + (\tilde{\beta} - \beta))(c'\beta + c'(\tilde{\beta} - \beta))^{-1} = (I - \beta c')(\tilde{\beta} - \beta) + O_P(T^{-2}),$$

since we have assumed that $c'\beta = I$. From the representation

$$(\tilde{\beta} - \beta) = \bar{\beta}_1 \beta_1'(\tilde{\beta} - \beta) + \bar{\beta}_2 \beta_2'(\tilde{\beta} - \beta) = \bar{\beta}_1 T^{-1} B_{1T} + \bar{\beta}_2 T^{-2} B_{2T},$$

we see that

$$T(\hat{\beta}_c - \beta) = (I - \beta c')\bar{\beta}_1 B_{1T} + o_P(1),$$

which shows the first result. The results for (β_1, β_2) follow in the same way. Note that for the results in Theorem 3 we have disregarded the component of the limit distribution which is of order T^{-2}. This gives a simpler formulation, but there may be cases where the relation between the normalization and the vectors $(\beta, \beta_1, \beta_2)$ is such that one needs the component of $(\tilde{\beta} - \beta)$ in the direction β_2, for instance. In such a case one can derive the distribution from the results in Theorem 2.

References

[1] Chan, N.H. and Wei, C.Z. (1988) Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16, 367-401.

[2] Granger, C.W.J. (1981) Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* 16, 121-130.

[3] Johansen, S. (1998) Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control* 12, 231-254.

[4] Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551-1580.

[5] Johansen, S. (1992) A representation of vector autoregressive processes integrated of order 2. *Econometric Theory* 8, 188-202.

[6] Johansen, S. (1990) An algorithm for estimating the cointegrating relations in vector autoregressive processes allowing for $I(2)$ variables, University of Copenhagen.

[7] Johansen, S. (1995) A statistical analysis of cointegration for I(2) variables. Forthcoming *Econometric Theory.*

[8] Johansen, S. (1992) An $I(2)$ cointegration analysis of the purchasing power parity between Australia and USA. In Colin Hargreaves (ed.): *Macroeconomic modelling in the long run.* Edward Elgar.

[9] Juselius, K. (1994) Do purchasing power parity and uncovered interest parity hold in the long run?-An example of likelihood inference in a multivariate time series model. *Journal of Econometrics.*

[10] Paruolo,P. (1994) Asymptotic efficiency of the 2 step estimator in I(2) VAR systems. University of Bologna.

[11] Paruolo,P. (1994) Testing for multicointegration in a two stage analysis of I(2) systems. University of Bologna.

[12] Phillips, P.C.B. and S.N. Durlauf, (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473-496.

[13] Phillips,P.C.B. (1991) Optimal inference in cointegrated systems. *Econometrica* 59, 256-271.

[14] Reinsel, G.C. and S.K. Ahn (1990) Vector autoregressive models with unit roots and reduced rank structure, estimation likelihood ratio test, and forecasting. *Journal of Time Series* 13, 283-295.

[15] Saikkonen, P. (1991) Asymptotically efficient estimation of cointegration regressions. *Econometric Theory* 7, 1-21.

[16] Stock, J. H. and M.W. Watson, (1993). A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica* 61, 783-820,

Preprints 1993

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, DK-2100 COPENHAGEN Ø, DENMARK. TELEPHONE +45 35 32 08 99.

No. 1 Hansen, Henrik and Johansen, Søren: Recursive Estimation in Cointegration VAR-Models.

No. 2 Stockmarr, A. and Jacobsen, M.: Gaussian Diffusions and Autoregressive Processes: Weak Convergence and Statistical Inference.

No. 3 Nishio, Atsushi: Testing for a Unit Root against Local Alternatives

No. 4 Tjur, Tue: StatUnit - An Alternative to Statistical Packages?

No. 5 Johansen, Søren: Likelihood Based Inference for Cointegration of Non-Stationary Time Series.

Preprints 1994

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, DK-2100 COPENHAGEN Ø, DENMARK. TELEPHONE 45 35 32 08 99, FAX 45 35 32 07 72.

No. 1 Jacobsen, Martin: Weak Convergence of Autoregressive Processes.

No. 2 Larsson, Rolf: Bartlett Corrections for Unit Root Test Statistics.

No. 3 Anthony W.F. Edwards, Anders Hald, George A. Barnard: Three Contributions to the History of Statistics.

No. 4 Tjur, Tue: Some Paradoxes Related to Sequential Situations.

No. 5 Johansen, Søren: The Role of Ancillarity in Inference for Non-Stationary Variables.

No. 6 Rahbek, Anders Christian: The Power of Some Multivariate Cointegration Tests.

No. 7 Johansen, Søren: A Likelihood Analysis of the I(2) Model.