Direct cointegration testing in error correction models

Frank Kleibergen, Herman K. van Dijk*

Econometric and Tinbergen Institute, 3063 DM Rotterdam, The Netherlands

Abstract

An error correction model is specified having only exact identified parameters, some of which reflect a possible departure from a cointegration model. Wald, likelihood ratio, and Lagrange multiplier statistics are derived to test for the significance of these parameters. The construction of the Wald statistic only involves linear regression, and under certain conditions the limiting distribution of the Wald statistic differs from the limiting distributions of the likelihood ratio and Lagrange multiplier statistics. A special ordering of the variables is recommended so that equal limiting distributions of the three different test statistics are obtained. The applicability of the derived testing procedures is illustrated using real demand for money, real GNP, and bond and deposit interest rates from Denmark.

Key words: Cointegration; Wald test; Limiting distributions; Error correction models; Two-step estimation
JEL classification: C12; C32; C50

1. Introduction

Recently, several models and estimation methods have been introduced for the analysis of multivariate economic time series where each of the individual

---

* Corresponding author.

series exhibits a possibly nonstationary character but a linear combination may be stationary (see, e.g., Box and Tiao, 1977; Engle and Granger, 1987; Stock and Watson, 1988; Boswijk, 1990; Johansen, 1991a, b; and Phillips, 1991). The distinction between the nonstationary individual series and the stationary linear combinations of these series is especially relevant for long-term prediction purposes. Examples of series which may be nonstationary individually and may behave stationary as linear combinations are long- and short-term interest rates (see Engle and Granger, 1987) and real demand for money, real GNP, and interest rates (see Johansen and Juselius, 1990).

The methods for analysing possibly cointegrated time series differ in several aspects. There are methods (Engle and Granger, 1987; Boswijk, 1990) which are based on single-equation analysis, while others (Box and Tiao, 1977; Stock and Watson, 1988; Johansen, 1991a; Phillips, 1991) make use of a complete system analysis. The distinction between testing for cointegration in single equations or multiple equations is crucial because in multiple-equations analysis one has to overcome a certain identification problem. The methods in the literature which test for cointegration in multiple-equations systems rely on 'data-parametric techniques' as principal components and canonical correlations. The method proposed in this paper solves the identification problem by a parametric restriction on the unknown parameters of the equation system (see Kleibergen and van Dijk, 1993a). The exact identification of the equation system parameters allows one to find Maximum Likelihood (ML) estimators and perform the likelihood-based testing procedures, i.e., Wald, Likelihood Ratio, and Lagrange Multiplier. The cointegration testing procedure is based on the addition of parameters to a cointegrating Error Correction Model and testing the significance of these additional parameters. The cointegration tests are therefore related to the 'Error Correction Tests' referred to by Davidson et al. (1978).

The contents of the paper are organized as follows. In Section 2 the parametric specification of a possibly cointegrating Error Correction Model is introduced. Wald, Likelihood Ratio (LR), and Lagrange Multiplier (LM) statistics to test the null hypothesis of cointegration against the alternative of stationarity are constructed in the third section. In Section 4 the limiting distributions of the different parameter estimators and test statistics involved in the construction of the Wald statistic are discussed. It is shown that the regression-based cointegrating vector estimator has a biased limiting distribution and that the limiting distributions of the regression-based cointegrating vector estimator and the Wald statistic change when there is a certain rank deficiency in the long-run multiplier matrix of the data-generating process. Apart from these limiting distributions, a test procedure which allows for sequential testing for the number of cointegrating relationships using the Wald statistic is constructed. The importance of Granger long-run causality which affects the interpretation of certain outcomes is discussed, and to prevent these interpretation issues, the series are suggested to be ordered according to their Granger long-run causal

relationships. The fifth section contains the limiting distribution of the LM statistic which is equal to the limiting distribution of the LR statistic. It is also shown that the equality of the limiting distributions of the Wald, LR, and LM statistics for cointegration testing does not always hold. In Section 6 the cointegration testing procedures are applied to the Danish series of real demand for money, real GNP, and bond and deposit interest rates for which another cointegration study by Johansen and Juselius (1990) exists. The sensitivity with respect to the chosen variable ordering and test statistic is investigated for these series. The seventh section concludes.

We end this introduction with a notational remark. To save on indices, the data series are depicted as row vectors in the formulas contained in this paper. Also the common expression for the long-run multiplier $\alpha\beta'$ is replaced by $\beta\alpha$, where β stands for the cointegrating vectors.

2. Specification of a parametric error correction model

Cointegration in Vector AutoRegressive (VAR) models amounts to a rank deficiency of the long-run multiplier matrix, which is defined as the sum of different parameter matrices (see, e.g., Engle and Granger, 1987). A class of models which is observationally equivalent with the standard VAR models and for which the long-run multiplier is directly estimable is the class of Error Correction Models (ECM). When there is cointegration, the ECMs are only defined in terms of stationary variables which is another attractive property. A standard formulation of a pth-order k-dimensional ECM (see, e.g., Johansen, 1991a) reads

$$\Delta x_t \Gamma(L) = c + td + x_{t-p}\Pi + \varepsilon_t, \qquad t = 1, \dots, T, \tag{1}$$

$$\Gamma(z) = I_k - \sum_{i=1}^{p-1} z^i \Gamma_i \Lambda, \tag{2}$$

where x_t, $t = 1, \dots, T$, denotes a $1 \times k$ *row vector* of time series generated by the ECM.

The cointegration literature focuses on the parameter matrix Π because, assuming that all roots of the implicit VAR polynomial lie outside the unit circle or are equal to 1.0 (nonexplosive finite cyclical behaviour), the nonstationary character of the analysed series x_t is determined by Π. If the rank of Π is equal to r which is less than k, the implicit VAR polynomial contains $k - r$ unit roots and Π can then be specified as the product of two full-rank $k \times r$ matrices,

$$\Pi = \beta\alpha, \qquad \beta, \alpha' : k \times r. \tag{3}$$

The time-series x_t are said to be cointegrated with cointegrating vector β, if $0 < r < k$. Cointegration implies that the variables X_t are nonstationary while

the variables $X_t\beta$ are stationary (see Engle and Granger, 1987). Because the number of parameters in $\beta\alpha$ $(2kr)$ does usually not equal the number of parameters in Π (k^2), one is confronted with an identification problem when testing for the number of cointegrating vectors or estimating the cointegrating vectors. The methods proposed in the literature so far overcome this identification problem in the manner outlined in the introduction. An alternative manner to overcome the identification problem is by choosing a suitable decomposition of the long-run multiplier Π (see Kleibergen and van Dijk, 1993a). The resulting specification of Π is strictly parametric and allows for the construction of the likelihood-based cointegration testing principles – Wald, LR, and LM. This model specification is closely related to the incomplete simultaneous-equations model (see Kleibergen, Urbain, and van Dijk, 1993). The likelihood analysed by Johansen (1991a) is also similar to the likelihood resulting from an incomplete simultaneous-equations model which shows that the result is quite natural. The decomposition of Π reads

$$\Pi = \beta\alpha, \qquad \beta = \begin{bmatrix} I_r & 0 \\ -\beta_2 & I_{k-r} \end{bmatrix}, \quad \alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ 0 & \alpha_{22} \end{bmatrix}, \tag{4}$$

$$= \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ -\beta_2\alpha_{11} & -\beta_2\alpha_{12} + \alpha_{22} \end{bmatrix}, \tag{5}$$

where β_2: $(k-r)\times r$, α_{11}: $r\times r$, α_{12}: $r\times(k-r)$, α_{22}: $(k-r)\times(k-r)$ are all unrestricted.

When $\alpha_{22} = 0$, Π has a lower rank value and the specification of $\Pi = \beta\alpha$ is similar to the specifications used before with $\beta = (I_r - \beta_2')'$ and $\alpha = (\alpha_{11} \ \alpha_{12})$. When $\alpha_{22} \neq 0$, the interpretation of β does not correspond with a cointegrating vector. Tests for the number of cointegrating vectors can be performed by testing whether $\alpha_{22} = 0$ for different values of r.

When Π_{11} has full rank, there is a one-to-one correspondence between the elements of Π and α, β, and ML estimators of the parameters of α and β can be obtained from the ML or Ordinary Least Squares (OLS) estimators of Π,

$$\hat{\Pi} = \begin{bmatrix} \hat{\Pi}_{11} & \hat{\Pi}_{12} \\ \hat{\Pi}_{21} & \hat{\Pi}_{22} \end{bmatrix} \Rightarrow \begin{array}{ll} \hat{\alpha}_{11} = \hat{\Pi}_{11}, & \hat{\alpha}_{12} = \hat{\Pi}_{12}, \\ \hat{\beta}_2 = -\hat{\Pi}_{21}\hat{\Pi}_{11}^{-1}, & \hat{\alpha}_{22} = \hat{\Pi}_{22} - \hat{\Pi}_{21}\hat{\Pi}_{11}^{-1}\hat{\Pi}_{12}. \end{array} \tag{6}$$

Estimators of the parameters of α and β can also be constructed using a two-step procedure. In the first step of this procedure, an estimator for the cointegrating vector is constructed. In the second step, the estimated cointegrating vector, $\hat{\beta}_2$, is used to determine the difference between the estimated model and a cointegration model. The two-step procedure also shows the change in interpretation which occurs with respect to some of the deterministic components once unit roots are assumed to be present.

(1) Estimate

$$\Delta X_1 = \iota c_1 + \tau d_1 + Z\Gamma_1 + (X_1 \ X_2)_{-p}\begin{bmatrix} \Pi_{11} \\ \Pi_{21} \end{bmatrix} + \varepsilon_1$$

$$= Z\Gamma_1 + [(X_1 + X_2\Pi_{21}\alpha_{11}^{-1})_{-p} + \iota c_1\alpha_{11}^{-1} + \tau d_1\alpha_{11}^{-1}]\alpha_{11} + \varepsilon_1$$

$$= Z\Gamma_1 + [(X_1 - X_2\beta_2)_{-p} - \iota\mu_1 - \tau\delta_1]\alpha_{11} + \varepsilon_1. \tag{7}$$

Construct

$$\hat{\beta}_2 = -\hat{\Pi}_{21}\hat{\Pi}_{11}^{-1}, \quad \hat{\mu}_1 = -\hat{c}_1\hat{\Pi}_{11}^{-1}, \quad \hat{\delta}_1 = -\hat{d}_1\hat{\Pi}_{11}^{-1}, \quad \hat{\alpha}_{11} = \hat{\Pi}_{11}. \tag{8}$$

(2) Estimate

$$\Delta X_2 = \iota c_2 + \tau d_2 + Z\Gamma_2 + (X_1 \ X_2)_{-p}\begin{bmatrix} \Pi_{12} \\ \Pi_{22} \end{bmatrix} + \varepsilon_2$$

$$= Z\Gamma_2 + [(X_1 - X_2\hat{\beta}_2)_{-p} - \iota\hat{\mu}_1 - \tau\hat{\delta}_1]\alpha_{12} + \iota(c_2 + \hat{\mu}_1\alpha_{12})$$

$$+ \tau(d_2 + \hat{\delta}_1\alpha_{12}) + (X_2)_{-p}(\Pi_{22} + \hat{\beta}_2\alpha_{12}) + \varepsilon_2$$

$$= Z\Gamma_2 + [(X_1 - X_2\hat{\beta}_2)_{-p} - \iota\hat{\mu}_1 - \tau\hat{\delta}_1]\alpha_{12} + \iota(c_2 - \hat{c}_1\hat{\alpha}_{11}^{-1}\alpha_{12})$$

$$+ \tau(d_2 - \hat{d}_1\hat{\alpha}_{11}^{-1}\alpha_{12}) + (X_2)_{-p}(\Pi_{22} - \hat{\Pi}_{21}\hat{\alpha}_{11}^{-1}\alpha_{12}) + \varepsilon_2$$

$$= Z\Gamma_2 + [(X_1 - X_2\hat{\beta}_2)_{-p} - \iota\hat{\mu}_1 - \tau\hat{\delta}_1]\alpha_{12} + \iota\mu_2 + \tau\delta_2$$

$$+ (X_2)_{-p}\alpha_{22} + \varepsilon_2, \tag{9}$$

where $Z = (\Delta X_{-1} \ldots \Delta X_{-(p-1)})$, $\iota = (1 \ldots 1)'$, $\tau = (1 \ldots T)'$, $X = (X_1 \ X_2)$, $\varepsilon = (\varepsilon_1 \ \varepsilon_2)$; ε_1, X_1: $T \times r$; ε_2, X_2: $T \times (k - r)$. When $\alpha_{22} = 0$, the constant term parameter in the second set of equations, μ_2, has the interpretation of a linear growth term parameter, while the trend parameter, δ_2, measures the 'quadratic growth'. More generally, the deterministic components estimated in the second step induce deterministic behaviour which is of a higher order when $\alpha_{22} = 0$, while the deterministic components estimated in the first step, (μ_1, δ_1), still have the same kind of interpretation. The subtle interpretation of the different deterministic component parameters can also be obtained by constructing the implied vector moving average representation of X_t when $\alpha_{22} = 0$.

Theorem 1. When $p = 1$ and $\alpha_{22} = 0$, the implied Vector Moving Average (VMA) representation of the system equation (1) becomes

$$X_t = C + tD + t^2 E + \sum_{i=1}^{t} \varepsilon_{2i}(\beta_2 \ I_{k-r}) + v_{1t}P(L), \tag{10}$$

where

$$C = (- \psi_1(I_r - \Phi_1^t)(I_r - \Phi_1)^{-2}\alpha_{12} + X_{20})(\beta_2\ I_{k-r})$$
$$+ (\mu_1 + \psi_1(I_r - \Phi_1)(I_r - \Phi_1)^{-1} + W_0\Phi_1^t 0),$$
$$D = (\mu_2 + \psi_1(I_r - \Phi_1)^{-1}\alpha_{12} + \tfrac{1}{2}\delta_2)(\beta_2\ I_{k-r}) + (\delta_1\ 0),$$
$$E = \tfrac{1}{2}\delta_2(\beta_2\ I_{k-r}),$$
$$\psi_2 = -\delta_2\beta_2,$$
$$v_{1t}P(L) = \sum_{n=1}^{t-1} v_{1n} \sum_{j=0}^{t-n-1} \Phi_1^{j-1}\alpha_{12}(\beta_2\ I_{k-r}) + \left(\sum_{i=0}^{t-1} v_{1t-i}\Phi_1^i\ 0 \right),$$
$$\Phi_1 = (I_r + \alpha_{11} - \alpha_{12}\beta_2),$$
$$\psi_1 = \delta_1(I_r + \alpha_{11} - \alpha_{12}\beta_2) - \mu_2\beta_2,$$
$$v_{1t} = \varepsilon_{1t} - \varepsilon_{2t}\beta_2.$$

Proof. Respecify the ECM, $\Delta X_t = (X_{1t} - X_{2t}\beta_2 - \mu_1 - t\delta_1)(\alpha_{11}\alpha_{12}) + \mu_2(0\ I_{k-r}) + t\delta_2(0\ I_{k-r}) + \varepsilon_t$, as the observationally equivalent invertible VAR model, $((X_{1t} - X_{2t}\beta_2 - \mu_1 - t\delta_1)\ \Delta X_{2t})(I_k - L((\Phi_1'\ 0)'\ (\alpha_{12}'\ 0)')) = (\psi_1\ \mu_2) + t(\psi_2\ \delta_2) + (v_{1t}\ \varepsilon_{2t})$, and invert this latter model to obtain the VMA specification.

3. Likelihood-based cointegration test statistics

In the proposed parametric cointegration model, tests for cointegration with r cointegrating vectors ($k - r$ unit roots) correspond with testing the hypothesis H_0: $\alpha_{22} = 0$ against the alternative hypothesis of stationarity H_1: $\alpha_{22} \neq 0$. Given that the hypothesis only concerns a directly estimable parameter, we can apply the three likelihood-based testing principles, i.e., Wald, LR, and LM, for testing the hypothesis of cointegration versus stationarity. Except for certain specific conditions, each of the three different testing procedures leads in large samples to the same kind of inference. We proceed with the construction of each of the three different test statistics.

First, the Wald statistic for testing for cointegration is constructed. A general formula of the Wald statistic to test the hypothesis H_0: $\theta = \theta_0$ against H_1: $\theta \neq \theta_0$ (see Engle, 1984) reads

$$t_W = (\hat{\theta} - \theta_0)'I(\hat{\theta})(\hat{\theta} - \theta_0), \tag{11}$$

where $I(\hat{\theta})$ stands for the information matrix evaluated in the ML parameter point $\hat{\theta}$. The model for testing the cointegrating relationships has an one-to-one correspondence with the linear ECM. When the disturbances $\varepsilon_t, t = 1, \ldots, T$, are independently normally distributed with mean zero and covariance matrix Ω (which is assumed throughout the whole paper), the information matrix of the

parameter Π in the ECM reads

$$I(\text{vec}(\hat{\Pi})) = (\hat{\Omega}^{-1} \otimes X'_{-p} M_{(Z\ W)} X_{-p}), \tag{12}$$

where $M_V = I_T - V(V'V)^{-1}V'$, $\hat{\Omega} = T^{-1} \Delta X' M_{(Z\ W\ X_{-p})} \Delta X$, $W = (\iota\ \tau)$, and $V = (Z\ W)$. The information matrix of the transformed parameters can be obtained by constructing the quadratic form of the Jacobian matrix of the transformation of the original parameters towards the transformed parameters with respect to the information matrix of the original parameters,

$$I(\theta) = -\mathscr{E}\left(\frac{\partial^2 \ln l(\theta)}{\partial\theta\ \partial\theta'}\right) = -\left(\frac{\partial\eta}{\partial\theta'}\right)' \mathscr{E}\left(\frac{\partial^2 \ln l(\eta(\theta))}{\partial\eta\ \partial\eta'}\right)\left(\frac{\partial\eta}{\partial\theta'}\right)$$

$$= \left(\frac{\partial\eta}{\partial\theta'}\right)' I(\eta(\theta))\left(\frac{\partial\eta}{\partial\theta'}\right). \tag{13}$$

The Jacobian matrix of the transformation from $(\text{vec}(\Pi_{11})'\ \text{vec}(\Pi_{12})'$ $\text{vec}(\Pi_{21})'\ \text{vec}(\Pi_{22})')'$ to $(\text{vec}(\alpha_{11})'\ \text{vec}(\alpha_{12})'\ \text{vec}(\beta_2)'\ \text{vec}(\alpha_{22})')'$ (see Magnus and Neudecker, 1988, for details on matrix differential calculus) reads

$$J(\Pi, \beta, \alpha) = \frac{\partial(\text{vec}(\Pi_{11})'\ \text{vec}(\Pi_{12})'\ \text{vec}(\Pi_{21})'\ \text{vec}(\Pi_{22})')'}{\partial(\text{vec}(\alpha_{11})'\ \text{vec}(\alpha_{12})'\ \text{vec}(\beta_2)'\ \text{vec}(\alpha_{22})')}$$

$$= \begin{bmatrix} I_{rr} & 0 & 0 & 0 \\ 0 & I_{r(k-r)} & 0 & 0 \\ -(I_r \otimes \beta_2) & 0 & -(\alpha'_{11} \otimes I_{k-r}) & 0 \\ 0 & -(I_{k-r} \otimes \beta_2) & -(\alpha'_{12} \otimes I_{k-r}) & I_{(k-r)(k-r)} \end{bmatrix}. \tag{14}$$

The information matrix of the parameters $(\alpha_{11}, \alpha_{12}, \beta_2, \alpha_{22})$ can be obtained by constructing the quadratic form of the Jacobian matrix in (14) with respect to a slightly respecified information matrix (12), which has to be respecified because the ordering in $\text{vec}(\Pi)$ does not correspond with the ordering used to construct the Jacobian matrix. The information matrices of the different parameters α_{11}, α_{12}, β_2, and α_{22} can then be constructed from the joint information matrix by calculating the inverses of the diagonal blocks of the inverse of the joint information matrix (partitioned inverses),

$$I(\hat{\alpha}_{11}) = (\hat{\Omega}_{11}^{-1} \otimes (X_1)'_{-p} M_{(Z\ W\ (X_2)_{-p})}(X_1)_{-p}) = I(\hat{\Pi}_{11}), \tag{15}$$

$$I(\hat{\alpha}_{12}) = (\hat{\Omega}_{22}^{-1} \otimes (X_1)'_{-p} M_{(Z\ W\ (X_2)_{-p})}(X_1)_{-p}) = I(\hat{\Pi}_{12}), \tag{16}$$

$$I(\hat{\beta}_2) = (\hat{\alpha}'_{11} \hat{\Omega}_{11}^{-1} \hat{\alpha}_{11} \otimes (X_2)'_{-p} M_{(Z\ W\ (X_1 - X_2\beta_2)_{-p})}(X_2)_{-p}), \tag{17}$$

$$I(\hat{\alpha}_{22}) = \left(\left(\begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix}' \hat{\Omega} \begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix}\right)^{-1}\right.$$

$$\left. \otimes (X_2)'_{-p} M_{(Z\ W\ (X_1 - X_2\beta_2)_{-p})}(X_2)_{-p}\right), \tag{18}$$

where $\hat{\Omega}_{11} = T^{-1} \Delta X_1' M_{(Z\ W\ X_{-p})} \Delta X_1$ and $\hat{\Omega}_{22} = T^{-1} \Delta X_2' M_{(Z\ W\ X_{-p})} \Delta X_2$. The Wald statistic to test the hypothesis of r cointegrating relationships, H_0: $\alpha_{22} = 0$, against H_1: $\alpha_{22} \neq 0$ then becomes

$$t_W = \text{vec}(\hat{\alpha}_{22})' I(\hat{\alpha}_{22})\ \text{vec}(\hat{\alpha}_{22})$$

$$= \text{tr}\left[\left(\begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix}' \hat{\Omega} \begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix}\right)^{-1}\right.$$

$$\left. \times \hat{\alpha}_{22}'(X_2)'_{-p} M_{(Z\ W\ (X_1 - X_2\beta_2)_{-p})}(X_2)_{-p}\hat{\alpha}_{22}\right]$$

$$= \text{tr}\left[\left(\begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix}' \hat{\Omega} \begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix}\right)^{-1}(\tilde{\varepsilon}_2'\tilde{\varepsilon}_2 - \hat{\varepsilon}_2'\hat{\varepsilon}_2)\right], \tag{19}$$

where $\hat{\varepsilon}_2$ are the residuals from the second set of $(k - r)$ equations when $\alpha_{22} \neq 0$ and $\tilde{\varepsilon}_2$ are the residuals when $\alpha_{22} = 0$. The formula in (19) shows close resemblance with the well-known F statistics in the general linear model and only differs from these F statistics in the sense that a different covariance matrix is used to account for the uncertainty of $\hat{\beta}_2$ used in the second set of equations.

The likelihood ratio statistic is straightforward to calculate and equals the Johansen trace statistic (Johansen and Juselius, 1990). Given normally distributed errors, one can derive that the LR statistic is equal to

$$t_{LR} = T \ln[|\tilde{\Omega}|/|\hat{\Omega}|], \tag{20}$$

where $\tilde{\Omega} = T^{-1} \Delta X' M_{(Z\ W\ (X_1 - X\beta_2)_{-p})} \Delta X$ and β_2 is estimated using the ML estimator of β_2 when $\alpha_{22} = 0$. We note that the ML estimator of β_2 when $\alpha_{22} = 0$ cannot be constructed by means of linear regression. It needs to be calculated using some other technique like for example canonical correlations (see Johansen, 1991a).

The Lagrange multiplier statistic for testing for cointegration is somewhat more difficult to construct analytically than the LR statistic. A standard formula of the LM statistic to test the hypothesis H_0: $\theta = \theta_0$ against H_1: $\theta \neq \theta_0$ (see Engle, 1984) reads

$$t_{LM} = \left[\frac{\partial\ \ln l(\theta_0)}{\partial\theta}\right]' I(\theta_0)^{-1} \left[\frac{\partial\ \ln l(\theta_0)}{\partial\theta}\right]. \tag{21}$$

The information matrix of α_{22} used in (21) has to be constructed in a different manner than the information matrix in (18) because $\alpha_{22} = 0$ in (21). The suitable information matrix of α_{22} for (21) is proportional to the information matrix of a parameter α_{22} stemming from an auxiliary regression of $\tilde{\varepsilon}_2$ on $(X_2)_{-p}$, where $\tilde{\varepsilon}_2$ are the residuals of the second set of $(k - r)$ equations resulting from the

estimated ECM with $\alpha_{22} = 0$. The information matrix of α_{22} reads

$$I(\hat{\alpha}_{22}) = (\tilde{\Omega}_{22}^{-1} \otimes (X_2)'_{-p} M_{(Z \ W \ (X_1 - X_2\beta_2)_{-p})} (X_2)_{-p}), \tag{22}$$

$$\frac{\partial \ln l(\alpha_{22} = 0)}{\partial \ vec(\alpha_{22})'} = vec[(X_2)'_{-p} M_{(Z \ W \ (X_1 - X_2\beta_2)_{-p})} \Delta X_2 \tilde{\Omega}_{22}^{-1}]'. \tag{23}$$

The LM statistic for testing for r cointegrating relationships then becomes

$$t_{LM} = tr \ [\tilde{\Omega}_{22}^{-1} \Delta X_2' M_{(Z \ W \ (X_1 - X_2\beta_2)_{-p})} (X_2)_{-p}$$
$$\times ((X_2)'_{-p} M_{(Z \ W \ (X_1 - X_2\beta_2)_{-p})} (X_2)_{-p})^{-1} (X_2)'_{-p}$$
$$\times M_{(Z \ W \ (X_1 - X_2\beta_2)_{-p})} \Delta X_2]$$
$$= T[tr \ \tilde{\Omega}_{22}^{-1} (\tilde{\Omega}_{22} - \hat{\Omega}_{22})]$$
$$= T[(k - r) - tr(\tilde{\Omega}_{22}^{-1} \hat{\Omega}_{22})], \tag{24}$$

where $\tilde{\Omega}_{22} = T^{-1} \Delta X_2' M_{(Z \ W \ (X_1 - X_2\beta_2)_{-p})} \Delta X_2$ and β_2 is estimated using ML for both equations simultaneously under the assumption that $\alpha_{22} = 0$. As shown in (24), the LM statistic can be computed from an auxiliary regression of $(X_2)_{-p}$ on $\tilde{\varepsilon}_2$.

For most stationary models it is known that the limiting distributions under H_0 are the same for the Wald, LR, and LM statistics (see Engle, 1984). In the following sections each of the test statistics is discussed in more detail, and it is shown that the LR and LM statistics have equal limiting distributions under H_0, while the limiting distribution of the Wald statistic is different when the decomposition in (4) is invalid, i.e., when Π_{11} has lower rank.

4. Wald statistic using two-step estimators

The construction of the Wald statistic (19) exactly follows the two-step estimation procedure (7)–(9). At first the cointegrating vector including the optional deterministic components is estimated using the first set of equations. Conditional on the estimated cointegrating vector (including its deterministic components), in the second step the departure from a cointegration model is estimated. Since the different estimators and test statistics are constructed in two distinct steps, we discuss the convergence behaviour of the estimators and test statistics for each step separately.

4.1. Limiting behaviour of the two-step cointegrating vector estimator

The limiting behaviour of the two-step cointegrating vector estimator is discussed for four different Data Generating Processes (DGP). These four DGP's cover the cases that a cointegrating vector is estimated while (*i*) the

process is stationary, (ii) the process is cointegrated and α_{11} has full rank, (iii) the process is cointegrated while $\alpha_{11} = 0$, and (iv) the process is difference-stationary.

(i) If the process is stationary, standard asymptotic theory can be applied.

$$\lim_{T \to \infty} T^{1/2}(\hat{\Pi} - \Pi) \Rightarrow n(0, \Omega \otimes Q^{-1}), \quad Q = \lim_{T \to \infty} T^{-1} X'_{-p} M_{(Z\ W)} X_{-p}, \quad (25)$$

where $\Rightarrow$ indicates weak convergence (see Billingsley, 1968); in fact, the convergence behaviour is even somewhat stronger as indicated here. The cointegrating vector estimator $\hat{\beta}_2(= -\hat{\Pi}_{21}\hat{\Pi}_{11}^{-1})$ equals the ratio of two estimators, both of which have a normal limiting distribution. We note that the asymptotic theory to be applied to the cointegrating vector estimator $\hat{\beta}_2$ is identical to the asymptotic theory developed by Phillips (1989) to describe the limiting behaviour of the estimators in an incomplete simultaneous-equations model. When Π_{11} has full rank, the limiting distribution of $\hat{\beta}_2$ becomes

$$\lim_{T \to \infty} T^{1/2}(\hat{\beta}_2 + \Pi_{21}\Pi_{11}^{-1}) \Rightarrow n(0, \Pi_{11}^{-1'}\Omega_{11}\Pi_{11}^{-1} \otimes P_{22}^{-1}), \quad (26)$$

where $P_{22} = \lim_{T \to \infty} T^{-1}(X_2)'_{-p} M_{(Z\ W\ (X_1 - X_2\beta_2)_{-p})}(X_2)_{-p}$. It follows from Phillips (1989) that certain linear combinations of $\hat{\beta}_2$ converge to noncentral Cauchy distributed random variables when Π_{11} has a lower rank value. The expression for the convergence behaviour of $\hat{\beta}_2$ when $\Pi_{11} = 0$ ($\Pi_{21} = -\beta_2\alpha_{11} = 0$) reads

$$\lim_{T \to \infty} \hat{\beta}_2 \Rightarrow n(0, \Omega_{11} \otimes Q_{22.1}^{-1})(n(0, \Omega_{11} \otimes Q_{11.2}^{-1}))^{-1}, \quad (27)$$

where $Q_{ii.j} = Q_{ii} - Q_{ij}Q_{jj}^{-1}Q_{ji}$, and the two normal distributed random variables are of course not independent. Because of the correlation between the two normal distributions, $\hat{\beta}_2$ converges to a noncentral Cauchy distribution. Note that $\Pi_{11} = 0$, $\Pi_{21} = 0$ implies a lower rank value of Π and consequently cointegration. The limiting distribution in (27) refers to a stationary model, however. Thus, it does not correspond with the case $\Pi_{11} = 0$, $\Pi_{21} = 0$. The importance of a full rank value of Π_{11} is, however, most easily understood under stationarity, and from the stationary case it can be extended to the case of nonstationarity (see below).

(ii) Under nonstationarity the asymptotic results become somewhat more complicated and amongst others depend on the deterministic components which are present in the DGP and the estimated model. The asymptotic distribution of the cointegrating vector estimator $(\hat{\beta}'_2 \ \hat{\mu}'_1 \ \hat{\delta}_1)'$ will therefore be discussed for three different structures of the deterministic components in the DGP.

Theorem 2. When the DGP (7)–(9) is such that $\alpha_{22} = 0$ and α_{11} full rank, the limiting distributions of the cointegrating vector estimator $(\hat{\beta}_2' \; \hat{\mu}_1' \; \hat{\delta}_1')'$ for the proper value of r become:

(1) *When $\mu_2 \neq 0$, $\delta_1 = 0$, $\delta_2 = 0$, and the estimated model does not contain a linear time trend, $d_1 = 0$, the limiting behaviour of $\hat{\beta}_2$ follows:*

$$\lim_{T \to \infty} T(\gamma \; T^{1/2}\mu_2')'\beta_\perp'\beta_\perp(\hat{\beta}_2 - \beta_2)$$

$$\Rightarrow -\begin{bmatrix} \gamma'\beta_\perp'\beta_\perp\gamma & 0 \\ 0 & \mu_2\beta_\perp'\beta_\perp\mu_2' \end{bmatrix}\left(\int_0^1 F'F\right)^{-1}$$

$$\times \left[\left(\int_0^1 F'dU_{1.2}\right) + \left(\int_0^1 F'dU_2\right)\Omega_{22}^{-1}\Omega_{21}\right]\alpha_{11}^{-1}.$$

(2) *When $\mu_2 = 0$, $\delta_1 = 0$, $\delta_2 = 0$, and the estimated model does not contain a linear time trend, $d_1 = 0$, the limiting behaviour of $(\hat{\beta}_2' \; \hat{\mu}_1')'$ follows:*

$$\lim_{T \to \infty} T\left(\begin{bmatrix} \hat{\beta}_2 \\ T^{-1/2}\hat{\mu}_1 \end{bmatrix} - \begin{bmatrix} \hat{\beta}_2 \\ T^{-1/2}\mu_1 \end{bmatrix}\right)$$

$$\Rightarrow -\begin{bmatrix} \Omega_{22}^{-1/2} & 0 \\ 0 & 1 \end{bmatrix}\left(\int_0^1 G'G\right)^{-1}$$

$$\times \left[\left(\int_0^1 G'dU_{1.2}\right) + \left(\int_0^1 G'dU_2\right)\Omega_{22}^{-1}\Omega_{21}\right]\alpha_{11}^{-1}.$$

(3) *When $\mu_2 \neq 0$, $\delta_1 \neq 0$, $\delta_2 = 0$, and the estimated model does contain a linear time trend, $d_1 \neq 0$, the limiting behaviour of $(\hat{\beta}_2' \; \hat{\mu}_1' \; \hat{\delta}_1')'$ follows:*

$$\lim_{T \to \infty} T\left(\begin{bmatrix} \hat{\beta}_2 \\ T^{1/2}\hat{\delta}_1 \\ T^{-1/2}\hat{\mu}_1 \end{bmatrix} - \begin{bmatrix} \hat{\beta}_2 \\ T^{1/2}\delta_1 \\ T^{-1/2}\mu_1 \end{bmatrix}\right)$$

$$\Rightarrow -\begin{bmatrix} \Omega_{22}^{-1/2} & 0 \\ 0 & I_2 \end{bmatrix}\left(\int_0^1 H'H\right)^{-1}\left[\left(\int_0^1 H'dU_{1.2}\right)\right.$$

$$\left. + \left(\int_0^1 H'dU_2\right)\Omega_{22}^{-1}\Omega_{21}\right]\alpha_{11}^{-1},$$

where $F = ((P - \int_0^1 P(t)\mathrm{d}t) \quad (Q - \int_0^1 Q(t)\mathrm{d}t))$, $\quad G = (S_{22}\, \iota)$, $\quad H = (S_{22}\, \tau\, \iota)$, $P = S_{21}\Sigma_{22}^{1/2}(\beta_2 \quad I_{k-r})(\beta_2 \quad I_{k-r})'\gamma$, $\quad Q = \tau\mu_2(\beta_2\, I_{k-r})(\beta_2 \quad I_{k-r})'\mu_2'$, $\quad S_{22} \sim$ $W(0, I_{k-r})$, $S_{21} \sim W(0, I_{k-r-1})$, $S_{22} = (S_{21}\, s_{22})$, $s_{22} \sim W(0, 1)$, $S_1 \sim W(0, I_r)$, $W(0, I_j)$ is a jth-dimensional Brownian motion with mean zero and covariance matrix I_j defined on $(0,1)$; $\tau(t) = t$, $\iota(t) = 1$, $0 < t$ < 1; $U = (U_1\, U_2) = S\Sigma^{1/2}$, $\Sigma = B'\Omega B$, $B = (\beta(0\, I_{k-r})')$, $S = (S_1\, S_{22})$, $U_{1.2} = U_1 - U_2\Sigma_{22}^{-1}\Sigma_{21}$, γ: $(k-r)\times(k-r-1)$, such that $\mu_2(\beta_2\, I_{k-r})$ $(\beta_2\, I_{k-r})'\gamma = 0$.

Note that just like in Phillips and Durlauf (1986) and Johansen (1991a), the time indices in most of the processes are skipped.

Proof. See Appendix.

When $\alpha_{22} = 0$, the OLS regression estimator of the cointegrating vector, $\hat{\beta}_2 = -\hat{\Pi}_{21}\hat{\Pi}_{11}^{-1}$, is, in general, not equal to the ML estimator of β_2. Phillips (1991) shows that only the ML estimators, under the imposition of the proper number of unit roots, have unbiased limiting distributions, while estimation of the unit root parameters leads to biased limiting distributions. The limiting distribution of the cointegrating vector estimator, $\hat{\beta}_2$, is biased as a consequence as can also be concluded from Theorem 2. Because F and U_2 in the first part of Theorem 2 are dependent, while F and $U_{1.2}$ are independent, there will be a bias in the limiting distribution of $\hat{\beta}_2$ unless $\Omega_{12} = 0$, i.e., X_1 and X_2 are weakly exogenous. The estimator $\hat{\beta}_2$ does therefore not belong to the so-called Locally Asymptotic Mixed Normal (LAMN) family, but is a member of the Limiting Gaussian Family (LGF). As a result of the biased limiting distribution of $\hat{\beta}_2$, test statistics using elements of $\hat{\beta}_2$ will not have a χ^2 limiting distribution as is the case for estimators belonging to the LAMN family. The test statistics using $\hat{\beta}_2$ will have noncentral χ^2 limiting distributions for which almost no tabulated critical values exist. As a consequence testing hypotheses formulated on elements of β_2 is recommended to be performed using ML estimates of β_2 obtained under the imposition of the hypothesized number of unit roots. These kind of test statistics namely have central χ^2 limiting distributions (see Johansen, 1991a; Phillips, 1991).

The first part of Theorem 2 also shows that the limiting distribution of $\hat{\beta}_2$ will be normal when a linear time trend is present and $r = k - 1$. The approximation of this limiting distribution by the empirical distribution of $\hat{\beta}_2$ in small samples is, however, quite inaccurate unless the linear time trend dominates the stochastic trend considerably.

Although the limiting distribution of $\hat{\beta}_2$ is biased, the order of convergence indicates that $\hat{\beta}_2$ is still a superconsistent estimator of β_2. This phenomenon is also visible in Fig. 1, where the distribution function of $\hat{\beta}_2$ for a nonstationary bivariate DGP with one cointegrating vector and $\alpha_{11} \neq 0$ is drawn for different
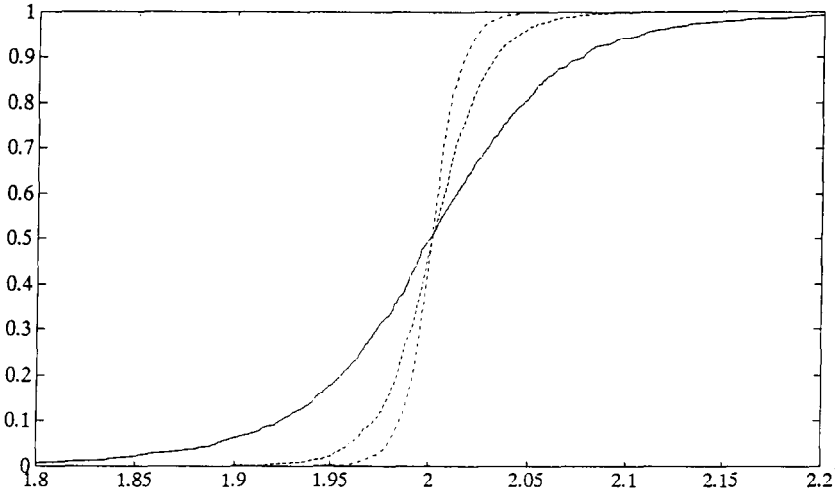
Fig. 1. Distribution functions $\hat{\beta}_2 = -\hat{\Pi}_{21}\hat{\Pi}_{11}^{-1}$ for bivariate DGP with one cointegrating vector and $T = 100$ (—), 250 (--), 500 (··); $\beta_2 = 2$

numbers of observations. Fig. 1 shows that the convergence of $\hat{\beta}_2$ towards β_2 is very fast, which indicates the superconsistent convergence of $\hat{\beta}_2$.

(iii) As shown in Theorem 2, the limiting distribution of $\hat{\beta}_2$ depends quite crucially on α_{11}. When α_{11} has a lower rank value, the convergence behaviour of $\hat{\beta}_2$ changes quite dramatically. For stationary processes we showed that, when $\alpha_{11} = 0$, $\hat{\beta}_2$ will be inconsistent and $\hat{\beta}_2$ converges to a Cauchy distributed random variable. For the stationary model with α_{11} full rank, the order of convergence of $\hat{\beta}_2$ is proportional to $T^{1/2}$. So when α_{11} converges from a full rank value to zero, the order of convergence of $\hat{\beta}_2$ decreases by a factor proportional to $T^{1/2}$. For nonstationary processes the decrease in the convergence behaviour turns out to be proportional to the same kind of factor, $T^{1/2}$. The order of convergence of $\hat{\beta}_2$ when α_{11} has full rank is proportional to T such that the resulting degree of convergence of $\hat{\beta}_2$ when $\alpha_{11} = 0$ will be proportional to $T^{1/2}$.

Theorem 3. When the DGP (7)–(9) is such that $\alpha_{22} = 0$ and $\alpha_{11} = 0$, the limiting distribution of $\hat{\beta}_2$ for the proper value of r becomes:

When $\mu_2 \neq 0$, $\delta_1 = 0$, $\delta_2 = 0$, and the estimated model does not contain a linear time trend, $d_1 = 0$, the limiting behaviour of $\hat{\beta}_2$ follows:

$$\lim_{T \to \infty} T^{1/2}(\gamma \ T^{1/2}\mu_2')'\beta_\perp'\beta_\perp(\hat{\beta}_2 - \beta_2)$$

$$\Rightarrow \ - \begin{bmatrix} \gamma'\beta_\perp'\beta_\perp\gamma & 0 \\ 0 & \mu_2\beta_\perp'\beta_\perp\mu_2' \end{bmatrix} \left(\int_0^1 F'F \right)^{-1}$$

$$\times \left[\left(\int_0^1 F'\mathrm{d}U_{1.2} \right) + \left(\int_0^1 F'\mathrm{d}U_2 \right) \Omega_{22}^{-1}\Omega_{21} \right]$$

$$\times \left[\left(\int_0^1 (X(t)\beta)'\mathrm{d}S \right) \Omega^{1/2}(I_r, \ 0)' \right]^{-1} (\mathrm{var}(X_t \ \beta)).$$

Proof. See Appendix.

Theorem 3 shows that $\hat{\beta}_2$ is no longer a superconsistent estimator of β_2, but a consistent one when α_{11} becomes equal to zero. Similar arguments can be derived when $\alpha_{11} \neq 0$, but has a lower rank value. The cointegrating vector estimator $\hat{\beta}_2$ is in that case superconsistent in certain directions, while in other directions it is only consistent (see also Phillips, 1989, where a similar phenomenon is discussed for the incomplete simultaneous-equations model). Note that also, when $\alpha_{11} = 0$, there is still one direction in which $\hat{\beta}_2$ converges superconsistently, i.e., the direction of the linear trend term μ_2.

Although $\hat{\beta}_2$ is still a consistent estimator of β_2 when $\alpha_{11} = 0$, the limiting distribution indicates a very slow convergence because it is very fat-tailed. The limiting distributions are namely equal to the product of a function of Brownian motions with nonzero mean (unless $\Omega_{12} = 0$) and the inverse of a normal distributed random matrix with mean zero. In case that both processes would be normal, the resulting product would have a noncentral Cauchy distribution, whose probability density function has fat tails resulting in an infinite mean and variance. The random process described by the limiting distribution will have even fatter tails than the mentioned noncentral Cauchy because the tail behaviour of the Brownian motion functional exceeds the tail behaviour of a normal probability density function. So the mean and variance of the limiting distribution will be infinite. Also the mode of the limiting distribution is unlikely to be equal to zero because of the nonzero mean of the Brownian motion functional. To show the convergence of $\hat{\beta}_2$ when $\alpha_{11} = 0$, we simulated series from a bivariate DGP with $\alpha_{11} = 0$ for different sample sizes. In Fig. 2 the distribution functions of $\hat{\beta}_2$ are shown for three different sample sizes. Through the fat tails of the limiting distribution there is only a very small degree of visible convergence.

One can conclude from the slow convergence of $\hat{\beta}_2$, when $\alpha_{11} = 0$, that one should try to have an ordering of the variables such that α_{11} is likely to have a full rank value for the hypothesized number of cointegrating vectors.
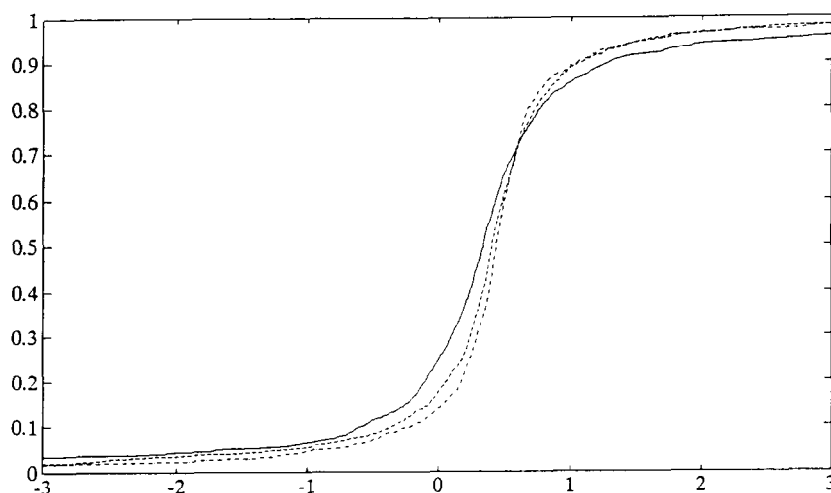
Fig. 2. Distribution functions $\hat{\beta}_2$ for bivariate DGP with one cointegrating vector and $\alpha_{11} = 0$, $T = 100$ (—), 250 (--), 500 (·-); $\beta_2 = \frac{1}{2}$

As shown in the Appendix, the limiting distribution of $\hat{\alpha}_{11}(= \hat{\Pi}_{11})$ is a normal probability density function in case of the proper number of unit roots even when $\alpha_{11} = 0$. To show the limiting behaviour of $\hat{\alpha}_{11}$, in Figs. 3 and 4, the distribution function of the 't-values' for the hypothesis, H_0: $\hat{\alpha}_{11} = \alpha_{11}$, for a bivariate DGP with one cointegrating relationship for $\alpha_{11} \neq 0$ (Fig. 3) and $\alpha_{11} = 0$ (Fig. 4) are shown. In these figures the 't-values' are shown because we want to focus on the normality issue of limiting distribution, which through the convergence of $\hat{\alpha}_{11}$ towards α_{11} becomes less visible if we choose to analyse the distribution of $\hat{\alpha}_{11}$.

Fig. 3 shows that the distribution of $\hat{\alpha}_{11}$ converges quite fast to a symmetric (normal) distribution function when $\alpha_{11} \neq 0$, while the distribution of $\hat{\alpha}_{11}$ converges much slower when $\alpha_{11} = 0$. Only for $T = 500$ does the distribution function of $\hat{\alpha}_{11}$ seem to be symmetric when $\alpha_{11} = 0$. This indicates that the limiting distribution of $\hat{\alpha}_{11}$ is indeed normal regardless of the true value of α_{11}, but that the small sample approximation of the limiting distribution does depend on the true value of α_{11}. The distribution functions do indicate however that an ordering of the variables using the t-values might be a good heuristic strategy for prevention of calculation of cointegrating vectors when $\alpha_{11} = 0$ or lower rank and that the limiting distributions of the test statistics for testing the rank of $\alpha_{11}(= \Pi_{11})$ are χ^2.

(iv) When the process is difference stationary, the long-run multiplier matrix, Π, equals zero. In case of no incorporation of any deterministic components in the
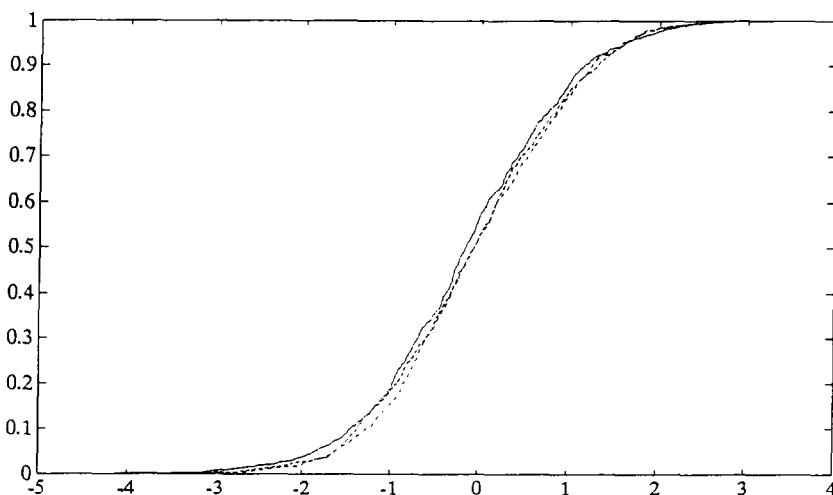
Fig. 3. Distribution functions $(\hat{\alpha}_{11} - \alpha_{11})(\text{var}(\hat{\alpha}_{11}))^{-1/2}$ for bivariate DGP with one cointegrating vector and $\alpha_{11} \neq 0$, $T = 100$ (—), 250 (--), 500 (··)

DGP and the estimated model, the limiting distribution of $\hat{\Pi}_1 = (\hat{\Pi}'_{11} \ \hat{\Pi}_{21})'$ then becomes (see Phillips and Durlauf, 1986)

$$\lim_{T \to \infty} T\hat{\Pi}_1 \Rightarrow \Omega^{-1/2\prime} \left(\int_0^1 S'S \right)^{-1} \left(\int_0^1 S'dS \right) \Omega^{1/2}(I_r, 0)', \tag{28}$$

where $S \sim W(0, I_k)$. These models, where $\Pi = 0$, are known as spurious regression models, and it is known that cointegrating vector estimators converge to random variables for these spurious regression models (see Phillips, 1989). Just like the other cointegrating vector estimators, the estimator $\hat{\beta}_2 = -\hat{\Pi}_{21}\hat{\Pi}_{11}^{-1}$ converges to a random variable when $\Pi = 0$,

$$\lim_{T \to \infty} \hat{\beta}_2 = \lim_{T \to \infty} -\hat{\Pi}_{21}\hat{\Pi}_{11}^{-1} = -\left[\lim_{T \to \infty} T\hat{\Pi}_{21} \right]\left[\lim_{T \to \infty} T\hat{\Pi}_{11} \right]^{-1}$$

$$\Rightarrow -\left[(0 \ I_{k-r})\Omega^{-1/2\prime} \left(\int_0^1 S'S \right)^{-1} \left(\int_0^1 S'dS \right) \Omega^{1/2}(I_r, 0)' \right]$$

$$\times \left[(I_r, 0)\Omega^{-1/2} \left(\int_0^1 S'S \right)^{-1} \left(\int_0^1 S'dS \right) \Omega^{1/2}(I_r, 0)' \right]^{-1}. \tag{29}$$
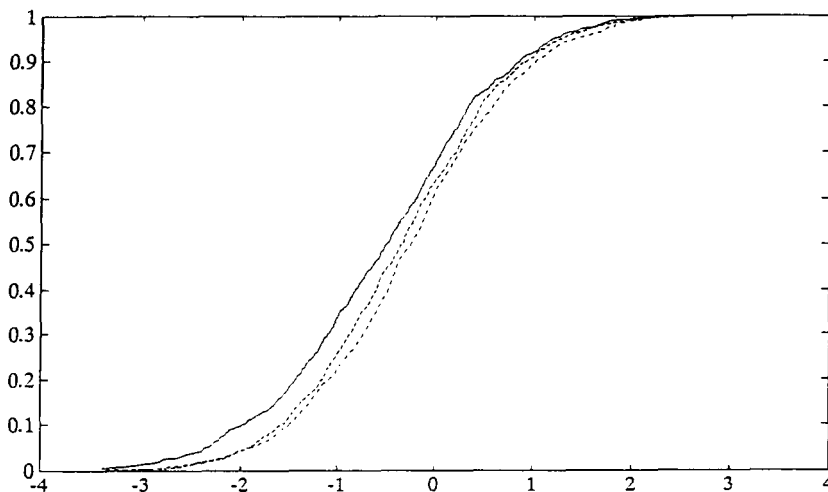
Fig. 4. Distribution functions $\hat{\alpha}_{11}$ $(\text{var}(\hat{\alpha}_{11}))^{-1/2}$ for bivariate DGP with one cointegrating vector and $\alpha_{11} = 0$, $T = 100$ (—), 250 (--), 500 (.-)

The basic convergence properties of the cointegrating vector estimator, $\hat{\beta}_2$, are covered with the four different cases discussed. The convergence properties of the cointegrating vector estimator $\hat{\beta}_2$ for other cases are usually equal to linear combinations of the discussed limiting distributions.

4.2. Limiting behaviour of the Wald statistic for cointegration

The discussion of the limiting distributions of the Wald statistic and the estimator $\hat{\alpha}_{22}$ is again performed along the lines of the four different DGP's discussed in the former section.

(i) When the process is stationary, $\alpha_{22} \neq 0$ and the limiting distribution of $\hat{\alpha}_{22} = \hat{\Pi}_{22} - \hat{\Pi}_{21}\hat{\Pi}_{11}^{-1}\hat{\Pi}_{12} = \hat{\Pi}_{22} + \hat{\beta}_2\hat{\Pi}_{12}$ is again normal in case of a full rank value of Π_{11} and noncentral Cauchy when $\Pi_{11} = 0$. The limiting distribution of the Wald statistic will be noncentral χ^2 in this case because $\text{vec}(\hat{\alpha}_{22})$ (19) is not taken in deviation from its mean $\text{vec}(\alpha_{22})$ but in deviation from zero.

(ii) The limiting distributions under a nonstationary DGP again depend on the deterministic components present in the DGP and the estimated model. Because of the change in the interpretation of certain deterministic components once unit roots become present $(\alpha_{22} = 0)$, it is common practice to jointly test for unit roots $(\alpha_{22} = 0)$ and equality to zero of certain deterministic component

parameters, μ_2, δ_2. The Wald statistic to test these hypotheses again stem from formula (19) where only $\tilde{\varepsilon}_2$ (= disturbances second set of equations when H_0 is true) and $\hat{\varepsilon}_2$ (= disturbances second set of equations when H_1 is true) need to be changed for (19) to be applicable to the case of joint testing for deterministic components and unit roots. The limiting distributions of the estimator measuring the departure from a cointegration model, $\hat{\alpha}_{22}$, and the Wald statistic, incorporating deterministic components for some instances, for the three different DGP's of Theorem 2 then become:

Theorem 4. When the DGP (7)–(9) is such that $\alpha_{22} = 0$ and α_{11} full rank, the limiting distributions of the parameters reflecting a departure from a cointegration model and the Wald statistic for testing such departures become:

(1) *In the DGP 1 from Theorem 2, the limiting behaviour of $\hat{\alpha}_{22}$ follows:*

$$\lim_{T \to \infty} T(\gamma \ T^{1/2}\mu_2')'\beta_\perp'\beta_\perp\hat{\alpha}_{22}$$

$$\Rightarrow \begin{bmatrix} \gamma'\beta_\perp'\beta_\perp\gamma & 0 \\ 0 & \mu_2\beta_\perp'\beta_\perp\mu_2' \end{bmatrix} \left(\int_0^1 F'F \right)^{-1} \left(\int_0^1 F'\mathrm{d}S \right) \Omega^{1/2}$$

$$\times \left[\begin{bmatrix} 0 \\ I_{k-r} \end{bmatrix} - \begin{bmatrix} I_r \\ 0 \end{bmatrix} \alpha_{11}^{-1}\alpha_{12} \right].$$

The limiting behaviour of the Wald statistic to test the hypothesis H_0: $\alpha_{22} = 0$ under the DGP 1 from Theorem 2 follows:

$$\lim_{T \to \infty} t_W \Rightarrow \mathrm{tr}\left(\int_0^1 V'\mathrm{d}S_{22} \right)'\left(\int_0^1 V'V \right)^{-1} \left(\int_0^1 V'\mathrm{d}S_{22} \right).$$

(2) *In the DGP 2 from Theorem 2, the limiting behaviour of $(\hat{\alpha}_{22}' \ \hat{\mu}_2')'$ follows:*

$$\lim_{T \to \infty} T \begin{bmatrix} \hat{\alpha}_{22} \\ T^{-1/2}\hat{\mu}_2 \end{bmatrix} \Rightarrow \begin{bmatrix} \Omega_{22}^{-1/2} & 0 \\ 0 & 1 \end{bmatrix} \left(\int_0^1 G'G \right)^{-1} \left(\int_0^1 G'\mathrm{d}S \right) \Omega^{1/2}$$

$$\times \left[\begin{bmatrix} 0 \\ I_{k-r} \end{bmatrix} - \begin{bmatrix} I_r \\ 0 \end{bmatrix} \alpha_{11}^{-1}\alpha_{12} \right].$$

The limiting behaviour of the Wald statistic to test the hypothesis H_0: $\alpha_{22} = 0$, $\mu_2 = 0$ *under the DGP 2 from Theorem 2 follows:*

$$\lim_{T \to \infty} t_W \Rightarrow \text{tr}\left(\int_0^1 G'dS_{22} \right)' \left(\int_0^1 G'G \right)^{-1} \left(\int_0^1 G'dS_{22} \right).$$

(3) *In the DGP 3 from Theorem 2, the limiting behaviour of* $(\hat{\alpha}'_{22} \ \hat{\delta}'_2)'$ *follows:*

$$\lim_{T \to \infty} T\begin{bmatrix} \hat{\alpha}_{22} \\ T^{1/2}\hat{\delta}_2 \end{bmatrix} \to \begin{bmatrix} \Omega_{22}^{-1/2} & 0 \\ 0 & 1 \end{bmatrix} \left(\int_0^1 E'E \right)^{-1} \left(\int_0^1 E'dS \right) \Omega^{1/2}$$

$$\times \left[\begin{bmatrix} 0 \\ I_{k-r} \end{bmatrix} - \begin{bmatrix} I_r \\ 0 \end{bmatrix} \alpha_{11}^{-1}\alpha_{12} \right].$$

The limiting behaviour of the Wald statistic to test the hypothesis H_0: $\alpha_{22} = 0$, $\delta_2 = 0$ *under the DGP 3 from Theorem 2 follows:*

$$\lim_{T \to \infty} t_W \Rightarrow \text{tr}\left(\int_0^1 E'dS_{22} \right)' \left(\int_0^1 E'E \right)^{-1} \left(\int_0^1 E'dS_{22} \right),$$

where $V = ((S_{21} - \int_0^1 S_{21}(t)dt) \quad (\tau - \int_0^1 \tau(t)dt)), \quad E = ((S_{22} - \int_0^1 S_{22}(t)dt) \ (\tau - \int_0^1 \tau(t)dt)).$

Proof. See Appendix.

Theorem 4 shows that, when α_{11} has full rank, the limiting distribution of t_W under H_0 equals the limiting distribution of the Johansen trace statistic for testing the same kind of hypothesis (see Johansen, 1991a,b). Thus the asymptotic critical values of the Johansen trace statistic tabulated by Johansen and Juselius (1990) can be used as asymptotic critical values for the Wald statistic.

The approximation of the small sample distribution of $\hat{\alpha}_{22}$ under DGP 1 by the limiting distribution of $\hat{\alpha}_{22}$ from Theorem 4 is often inaccurate. In the limit the deterministic trend dominates the stochastic trend, but this does in general not hold for small samples. The small sample distribution of the Wald statistic to test the hypothesis H_0: $\alpha_{22} = 0$ for series whose stochastic trend is not dominated considerably by their linear trend, is as a consequence often much better approximated by $(\int \bar{S}'_{22}d\bar{S}_{22})'(\int \bar{S}'_{22} \ \bar{S}_{22})^{-1} (\int S'_{22}d\bar{S}_{22})$, where $\bar{S}_{22} = S_{22} - \int S_{22}(t)dt$ (see also Johansen and Juselius, 1990).

The same kind of arguments made to explain the bias in the limiting distribution of $\hat{\beta}_2$ can again be made concerning the limiting distribution of $\hat{\alpha}_{22}$. The

bias in the limiting distribution of $\hat{\alpha}_{22}$ is of course very natural given that $\hat{\alpha}_{22}$ is an unit root estimator and that estimation of unit roots leads to biased limiting distributions (see Phillips, 1991). Just like $\hat{\beta}_2$, $\hat{\alpha}_{22}$ is a superconsistent estimator of its true value $\alpha_{22}(=0)$ when α_{22} is estimated for the proper number of cointegrating vectors.

(iii) As shown in Theorem 4, the limiting distribution of α_{22} depends on α_{11}, while the limiting distribution of the Wald statistic, t_W, seems to be insensitive to α_{11}. The question is therefore: what happens with the limiting distribution of $\hat{\alpha}_{22}$ and t_W when α_{11} has a lower rank value or is equal to zero?

Theorem 5. When the DGP (7)–(9) is such that $\alpha_{22} = 0$ and $\alpha_{11} = 0$, the limiting distribution of $\hat{\alpha}_{22}$ and t_W for the proper value of r can be characterized as:

When $\mu_2 \neq 0$, $\delta_1 = 0$, $\delta_2 = 0$, $r = k - r$, and the estimated model does not contain a linear time trend, $d_1 = 0$, the limiting behaviour of $\hat{\alpha}_{22}$ follows:

$$\lim_{T \to \infty} T^{1/2}(\gamma \ T^{1/2}\mu_2)'\beta'_\perp\beta_\perp\hat{\alpha}_{22}$$

$$\Rightarrow \begin{bmatrix} \gamma'\beta'_\perp\beta_\perp\gamma & 0 \\ 0 & \mu_2\beta'_\perp\beta_\perp\mu'_2 \end{bmatrix} \left(\int_0^1 F'F\right)^{-1} \left(\int_0^1 F'dS\right) \Omega^{1/2}(I_r, 0)'$$

$$\times \left[\int_0^1 (X(t)\beta)'dS)\,\Omega^{1/2}(I_r, 0)'\right]^{-1} (\text{var}(X_t\beta))\alpha_{12}.$$

The limiting behaviour of the Wald statistic to test the hypothesis H_0: $\alpha_{22} = 0$ under the DGP outlined above follows:

$$\lim_{T \to \infty} t_W \Rightarrow \text{tr}\left(\int_0^1 W'dS_1\right)' \left(\left(\int_0^1 W'W\right) + \left(\int_0^1 W'dS_1\right)\left[\int_0^1 Y(t)'dS_1\right]^{-1}\right)$$

$$\times \left[\int_0^1 Y(t)'dS_1\right]^{-1'}\left(\int_0^1 W'dS_1\right)\right)^{-1}\left(\int_0^1 W'dS_1\right).$$

Proof. See Appendix.

Theorem 5 shows that α_{22} is no longer estimated superconsistently because the order of convergence is now proportional to $T^{1/2}$, except for the direction of the linear trend term. As shown in the proof of Theorem 5, the limiting
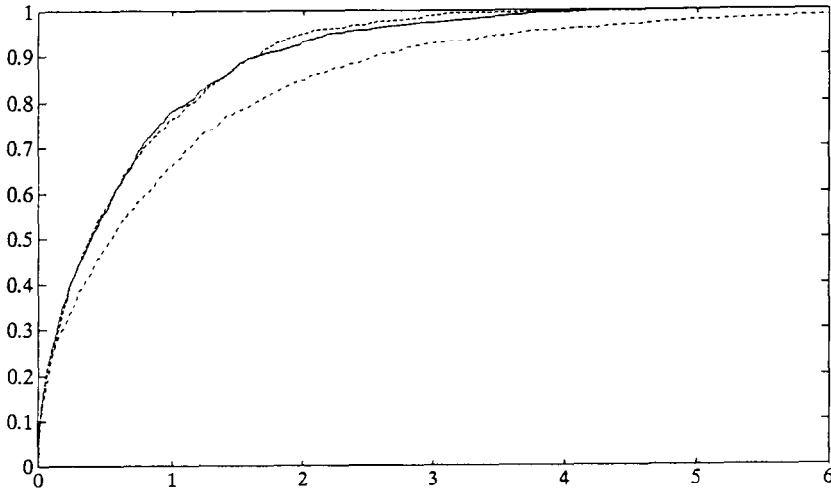
Fig. 5. Distribution functions Wald statistic, t_W, for bivariate DGP with $\mu_1 \neq 0$, $\mu_2 = 0$, $\alpha_{11} = 0$, $T = 100$ (--), 500 (--), limiting (—)

distribution of $\hat{\alpha}_{22} = \hat{\Pi}_{22} + \hat{\beta}_2 \hat{\Pi}_{12}$ is proportional to the limiting distribution of $\hat{\beta}_2$ when $\alpha_{11} = 0$ and $r = k - r$. This results from the consistent estimation of β_2 when $\alpha_{11} = 0$, while β_2 is estimated superconsistently when α_{11} has full rank. As a consequence the limiting distribution of t_W is also affected by the limiting distribution of β_2, which does not occur when β_2 is estimated superconsistently. In the former section we showed that the limiting behaviour of $\hat{\beta}_2$ is even more volatile than a noncentral Cauchy random variable when $\alpha_{11} = 0$. The same holds as a consequence for $\hat{\alpha}_{22}$ such that, although $\hat{\alpha}_{22}$ is a consistent estimator, it will converge only very slowly. The limiting distribution of the Wald statistic in Theorem 4 does not depend on α_{11}, but Theorem 5 shows that the limiting distribution of the Wald statistic changes when $\alpha_{11} = 0$ and $r = k - r$. The inverse normal random variables, already appearing in the limiting distributions of $\hat{\beta}_2$ and $\hat{\alpha}_{22}$ when $\alpha_{11} = 0$, also enter the limiting distribution of the Wald statistic when $\alpha_{11} = 0$. The same kind of results can be derived for α_{11} lower rank and/or $r \neq k - r$, in which cases the limiting distributions will be equal to a mixture of the limiting distributions stated in Theorems 4 and 5. As a consequence the Wald statistic for cointegration shows the well-known property that the Wald statistic is not invariant under parameter transformations (see Cox and Hinkley, 1974). Different orderings of the analysed series may lead to different limiting distributions of the Wald statistic for cointegration.

To show the consequences of these results for the small sample distribution of the Wald statistic, in Fig. 5 the empirical and limiting distributions of the Wald

statistic to test the hypothesis H_0: $\alpha_{22} = 0$, $\mu_2 = 0$ in DGP 2 from Theorem 2 are shown. The distributions shown in Fig. 5 indicate that the limiting and empirical distributions of the Wald statistic indeed change considerably when $\alpha_{11} = 0$ or has a lower rank value compared with the distributions associated with full rank values of α_{11} shown in Fig. 7.

The problems arising with the empirical and limiting distributions of the Wald statistic, t_W, again emphasize the importance of a proper ordering of the observed variables such that the resulting α_{11} parameter will have a full rank value for the proper number of unit roots.

(iv) Unlike the cointegrating vector estimator, $\hat{\beta}_2$, for a difference-stationary process, the unit root estimator, $\hat{\alpha}_{22}$, is still (super)consistent. The cointegrating vector estimator, $\hat{\beta}_2$, tries to estimate the nonexisting stationary relationships, while $\hat{\alpha}_{22}$ estimates the existing unit roots in the process. The (in)consistency of the two different estimators is explained by these arguments. The limiting distribution of $\hat{\alpha}_{22}$ can directly be obtained from Theorem 5 when neither the process nor the estimated model contain any deterministic components,

$$
\lim_{T \to \infty} T\hat{\alpha}_{22} \Rightarrow (0 \ I_{k-r}) \left(\Omega^{-1/2\prime} \left(\int_0^1 S'S \right)^{-1} \left(\int_0^1 S'\mathrm{d}S \right) \Omega^{1/2} (0 \ I_{k-r})' \right)
$$

$$
- (0 \ I_{k-r}) \Omega^{-1/2\prime} \left(\int_0^1 S'S \right)^{-1} \left(\int_0^1 S'\mathrm{d}S \right) \Omega^{1/2} (I_r \ 0)'
$$

$$
\times \left[(I_r \ 0) \Omega^{-1/2\prime} \left(\int_0^1 S'S \right)^{-1} \left(\int_0^1 S'\mathrm{d}S \right) \Omega^{1/2} (I_r \ 0)' \right]^{-1}
$$

$$
\times (I_r \ 0) \left(\Omega^{-1/2\prime} \left(\int_0^1 S'S \right)^{-1} \left(\int_0^1 S'\mathrm{d}S \right) \Omega^{1/2} (0 \ I_{k-r})' \right). \tag{30}
$$

The discussion of the two-step estimation/testing procedure so far has been focused on theoretical issues. In the following section practical aspects of the two-step procedure will briefly be discussed.

4.3. Two-step procedure in practice

In Fig. 6 a flow diagram of the two-step estimation/testing procedure is shown. The flow diagram shows the different steps of a sequential testing procedure for the number of unit roots, cointegrating vectors, and the resulting structure for the deterministic components. The optional structure for the
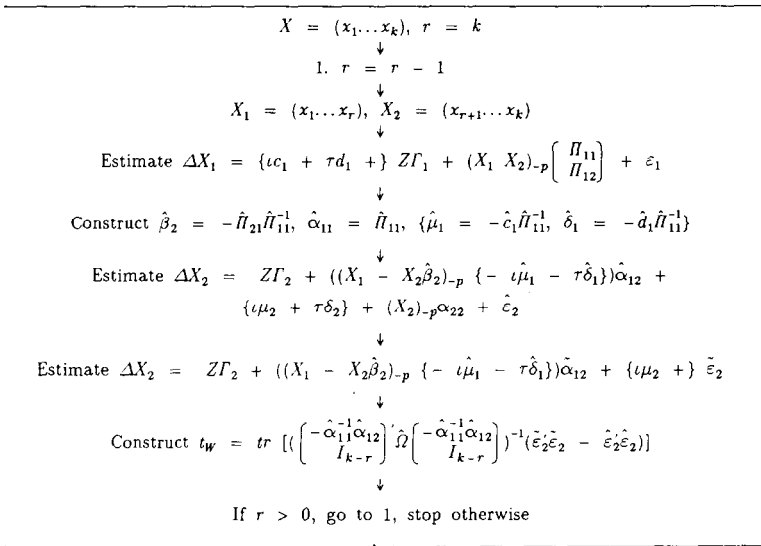
$$X = (x_1 \ldots x_k), \quad r = k$$

$$\downarrow$$

$$1. \quad r = r - 1$$

$$\downarrow$$

$$X_1 = (x_1 \ldots x_r), \quad X_2 = (x_{r+1} \ldots x_k)$$

$$\downarrow$$

$$\text{Estimate } \Delta X_1 = \{\iota c_1 + \tau d_1 +\} Z\Gamma_1 + (X_1 \ X_2)_{-p} \begin{bmatrix} \Pi_{11} \\ \Pi_{12} \end{bmatrix} + \varepsilon_1$$

$$\downarrow$$

$$\text{Construct } \hat{\beta}_2 = -\hat{\Pi}_{21}\hat{\Pi}_{11}^{-1}, \quad \hat{\alpha}_{11} = \hat{\Pi}_{11}, \quad \{\hat{\mu}_1 = -\hat{c}_1\hat{\Pi}_{11}^{-1}, \quad \hat{\delta}_1 = -\hat{d}_1\hat{\Pi}_{11}^{-1}\}$$

$$\downarrow$$

$$\text{Estimate } \Delta X_2 = Z\Gamma_2 + ((X_1 - X_2\hat{\beta}_2)_{-p} \{- \iota\hat{\mu}_1 - \tau\hat{\delta}_1\})\hat{\alpha}_{12} +$$

$$\{\iota\mu_2 + \tau\delta_2\} + (X_2)_{-p}\alpha_{22} + \hat{c}_2$$

$$\downarrow$$

$$\text{Estimate } \Delta X_2 = Z\Gamma_2 + ((X_1 - X_2\hat{\beta}_2)_{-p} \{- \iota\hat{\mu}_1 - \tau\hat{\delta}_1\})\tilde{\alpha}_{12} + \{\iota\mu_2 +\} \tilde{\varepsilon}_2$$

$$\downarrow$$

$$\text{Construct } t_W = tr \left[\left\{ \begin{pmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{pmatrix}' \hat{\Omega} \begin{pmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{pmatrix} \right\}^{-1} (\tilde{\varepsilon}_2'\tilde{\varepsilon}_2 - \hat{\varepsilon}_2'\hat{\varepsilon}_2) \right]$$

$$\downarrow$$

$$\text{If } r > 0, \text{ go to 1, stop otherwise}$$

Fig. 6. Flow diagram for sequential cointegration testing

deterministic components is put between braces ($\{ \ \}$) to indicate that one can choose the proper specification for the analysed series.

The two-step estimation procedure developed and discussed in the former sections and shown in Fig. 6 is by no means unique. In case of testing for r cointegrating relationships in k different series,

$$\begin{bmatrix} k \\ r \end{bmatrix} \left(= \frac{k!}{r!(k-r)!} \right),$$

different variable orderings exist, each of which leads to different two-step estimators. The question now is in which instances these two-step estimators could lead to different conclusions. In the former section we saw already that α_{11} lower rank could give different kind of results, and that one should order the variables such that α_{11} has full rank for the hypothesized number of cointegrating vectors. As shown, test statistics testing hypotheses on α_{11} have limiting χ^2 distributions such that tests on α_{11} are easily performed. The remark that α_{11} should have full rank for the presumed number of cointegrating vectors is quite important because, if one starts a sequential testing procedure for testing for unit roots, one starts with testing whether the hypothesis $H_0: r = k - 1$ (number of cointegrating vectors equal to total number of series minus one) can be rejected and, if not, one proceeds with testing the hypothesis $H_0: r = k - 2$. If the true number of cointegrating vectors is less than $k - 1$, the true value of α_{11} used in

this first step of the sequential testing procedure has a lower rank value. These situations should not give any problems for the suggested testing procedure.

Another difficulty which may arise concerns the Granger long-run causality relationships between the different analysed variables. For example, if $k = 2r$, two possible decompositions of the long-run multiplier Π are

$$\Pi = \beta\alpha, \quad \Pi = \gamma\kappa, \quad \gamma = \begin{bmatrix} I_{k-r} & -\gamma_1 \\ 0 & I_r \end{bmatrix}, \quad \kappa = \begin{bmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{bmatrix},$$

$$\begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ -\beta_2\alpha_{11} & \alpha_{22} - \beta_2\alpha_{12} \end{bmatrix} = \begin{bmatrix} \kappa_{11} - \gamma_1\kappa_{21} & -\gamma_1\kappa_{22} \\ \kappa_{21} & \kappa_{22} \end{bmatrix}. \quad (31)$$

Testing the hypothesis of r cointegrating relationships or $k - r$ unit roots can now be performed using α_{22} and κ_{11} [there exist even $\begin{bmatrix} k \\ r \end{bmatrix} - 2$ different other parameters on which the hypothesis of r cointegrating relationships can be formulated, but because of expository purposes we just focus on α_{22} and κ_{11}]. In the former sections we saw that, when κ_{22} has lower rank and α_{11} full rank, it is preferred to test the hypothesis using α_{22} and vice versa. This is, however, not the only circumstance in which a certain specification is preferred. Another reason for preferring a particular specification depends on the Granger long-run causality relationships. Assume for (31) that X_2 Granger long-run causes $X_1, \Pi_{21} \neq 0$, but X_1 does not long-run Granger cause $X_2, \Pi_{12} = 0$ (see Urbain, 1992). The parameters used in (31) then become

$$\Pi = \begin{bmatrix} \Pi_{11} & 0 \\ \Pi_{21} & \Pi_{22} \end{bmatrix} \Rightarrow$$

$$\alpha_{11} = \Pi_{11}, \quad \alpha_{12} = 0, \quad \beta_2 = -\Pi_{21}\Pi_{11}^{-1}, \quad \alpha_{22} = \Pi_{22}, \quad (32)$$
$$\kappa_{22} = \Pi_{22}, \quad \gamma_1 = 0, \quad \kappa_{21} = \Pi_{21}, \quad \kappa_{11} = \Pi_{11}.$$

Testing hypotheses using $\alpha_{22} = \Pi_{22}$ is still a proper cointegration test. When testing hypotheses using κ_{11}, nonrejection of r cointegrating vectors again raises some problems. When $\kappa_{11} = \Pi_{11}$ becomes equal to zero, there is no cointegration between X_1 and X_2 because X_1 is difference-stationary and X_2 is stationary when Π_{22} has full rank. To check for these kind of problems one should therefore always test whether the cointegrating vector estimators $\hat{\beta}_2$ and $\hat{\gamma}_1$ are different from zero. If $\hat{\beta}_2$ and $\hat{\gamma}_1$ are not different from zero, also α_{11} and κ_{22} need to be tested and have nonnormal limiting distributions in that case. The tests on $\hat{\beta}_2$, $\hat{\gamma}_1$ can be performed using the standard formula of the Wald statistic, (11), and the definition of the information matrix of $\hat{\beta}_2$, (17). As pointed out in the former section, the limiting distributions of the Wald statistics for $\hat{\beta}_2$ are noncentral χ^2, but large values of the Wald statistics will indicate significant values of β_2, γ_1. To partly circumvent the problem of both the lower rank values of α_{11} and the Granger long-run causality, a heuristic ordering of the variables using the t-values and parameter estimates of the long-run multiplier Π can be

very practical. We suggest therefore to order the variables in $X = (x_1 \ldots, x_k)$ such that each variable Granger long-run causes at least one of the variables in the preceding columns and each variable is Granger long-run caused by at least one variable from a latter column. So X_k should at least Granger long-run cause one of the variables in $X_1, \ldots, X_{k-1}$ and X_2 should Granger long-run cause X_1. Also X_1 should at least be Granger long-run caused by one variable from $X_2, \ldots, X_k$ and X_{k-1} should be Granger long-run caused by X_k. The Granger long-run causality condition can also be reflected as: $\pi_{ji} \neq 0$ for at least one i and one j with, $j > i, j = 2, \ldots, k$. To finish with a cointegrated system for which α_{11} has full rank, the variables in X should be ordered such that not only the Granger long-run causality ordering is kept but that also the upper diagonal elements of Π are likely to be different from zero.

The 'data-parametric' cointegration testing techniques discussed in a former section, like Johansen's procedure, do not suffer from these causality issues because they do not normalize the system (and consequently put a causality ordering on the variables) a priori. These methods work with non-identified cointegrating vectors however, and if one wants to put an identifying structure on the cointegrating vectors, the same kind of problems arise.

5. Limiting distributions of the likelihood ratio and Lagrange multiplier statistics

The asymptotic theory to be used for the estimators and test statistics used in the construction of the LR and LM statistics for cointegration has to a large extent been documented in the literature so far. The LR statistic, for example, equals the Johansen trace statistic whose limiting distributions are discussed in Johansen (1991a, b) and equal the limiting distributions of the Wald statistic, t_W, when α_{11} has full rank. The limiting distribution of the cointegrating vector estimator when $\alpha_{22} = 0$ is discussed in Johansen (1991a) and Phillips (1991). The cointegrating vector estimator used in both the LR and LM statistic is estimated under the imposition of the tested number of unit roots such that it belongs to the LAMN family. As a consequence test statistics which test hypotheses on the cointegrating vector estimator have central χ^2 limiting distributions contrary to the limiting distributions of the test statistics using the two-step cointegrating vector estimator which are noncentral χ^2. The only limiting distributions which are not documented in the literature concern the limiting distribution of the LM statistic and the estimator measuring the departure from the cointegration model used for the construction of the LM statistic.

Theorem 6. When the DGP (7)–(9) is such that $\alpha_{22} = 0$, the limiting distributions of the parameters resulting from an auxiliary regression reflecting departures from a cointegration model and the Lagrange multiplier statistic for testing such departures become:

(1) *In the DGP 1 from Theorem 2, the limiting behaviour of $\tilde{\alpha}_{22}$ follows:*

$$\lim_{T \to \infty} T(\gamma \ T^{1/2}\mu_2)'\tilde{\alpha}_{22} \Rightarrow (\gamma \ \mu_2)' \left(\int_0^1 F'F \right)^{-1} \left(\int_0^1 F'\mathrm{d}S \right)\Omega^{1/2} \begin{bmatrix} 0 \\ I_{k-r} \end{bmatrix}.$$

The limiting behaviour of the LM statistic to test the hypothesis H_0: $\alpha_{22} = 0$, under the DGP 1 from Theorem 2 follows:

$$\lim_{T \to \infty} t_{\mathrm{LM}} \Rightarrow \mathrm{tr}\left(\int_0^1 V'\mathrm{d}S_{22} \right)' \left(\int_0^1 V'V \right)^{-1} \left(\int_0^1 V'\mathrm{d}S_{22} \right).$$

(2) *In the DGP 2 from Theorem 2, the limiting behaviour of $(\tilde{\alpha}'_{22} \ \tilde{\mu}'_2)'$ follows:*

$$\lim_{T \to \infty} T \begin{bmatrix} \tilde{\alpha}_{22} \\ T^{-1/2}\tilde{\mu}_2 \end{bmatrix} \Rightarrow \begin{bmatrix} \Omega_{22}^{-1/2} & 0 \\ 0 & 1 \end{bmatrix} \left(\int_0^1 G'G \right)^{-1} \left(\int_0^1 G'\mathrm{d}S \right)\Omega^{1/2} \begin{bmatrix} 0 \\ I_{k-r} \end{bmatrix}.$$

The limiting behaviour of the LM statistic to test the hypothesis H_0: $\alpha_{22} = 0$, $\mu_2 = 0$ under the DGP 2 from Theorem 2 follows:

$$\lim_{T \to \infty} t_{\mathrm{LM}} \Rightarrow \mathrm{tr}\left(\int_0^1 G'\mathrm{d}S_{22} \right)' \left(\int_0^1 G'G \right)^{-1} \left(\int_0^1 G'\mathrm{d}S_{22} \right).$$

(3) *In the DGP 3 from Theorem 2, the limiting behaviour of $(\tilde{\alpha}'_{22} \ \tilde{\delta}'_2)'$ follows:*

$$\lim_{T \to \infty} T \begin{bmatrix} \tilde{\alpha}_{22} \\ T^{1/2}\tilde{\delta}_2 \end{bmatrix} \Rightarrow \begin{bmatrix} \Omega_{22}^{-1/2} & 0 \\ 0 & 1 \end{bmatrix} \left(\int_0^1 E'E \right)^{-1} \left(\int_0^1 E'\mathrm{d}S \right)\Omega^{1/2} \begin{bmatrix} 0 \\ I_{k-r} \end{bmatrix}.$$

The limiting behaviour of the LM statistic to test the hypothesis H_0: $\alpha_{22} = 0$, $\delta_2 = 0$ under the DGP 3 from Theorem 2 follows:

$$\lim_{T \to \infty} t_{\mathrm{LM}} \Rightarrow \mathrm{tr}\left(\int_0^1 E'\mathrm{d}S_{22} \right)' \left(\int_0^1 E'E \right)^{-1} \left(\int_0^1 E'\mathrm{d}S_{22} \right).$$

Proof. Straightforward using Appendix, and auxiliary regression using the model $\tilde{\varepsilon}_2 = \{\iota\mu_2 + \tau\delta_2\} + (X_2)_{-p}\alpha_{22} + v_2$, where $\tilde{\varepsilon}_2$: $T \times (k - r)$ are the $(k - r)$ residuals of the equations of ΔX_2 resulting from the estimated cointegration model containing r cointegrating vectors $[(k - r)$ unit roots] and $\tilde{\alpha}_{22}$, $\tilde{\mu}_2$, and $\tilde{\delta}_2$ are the OLS estimators of the parameters in the auxiliary regression model.
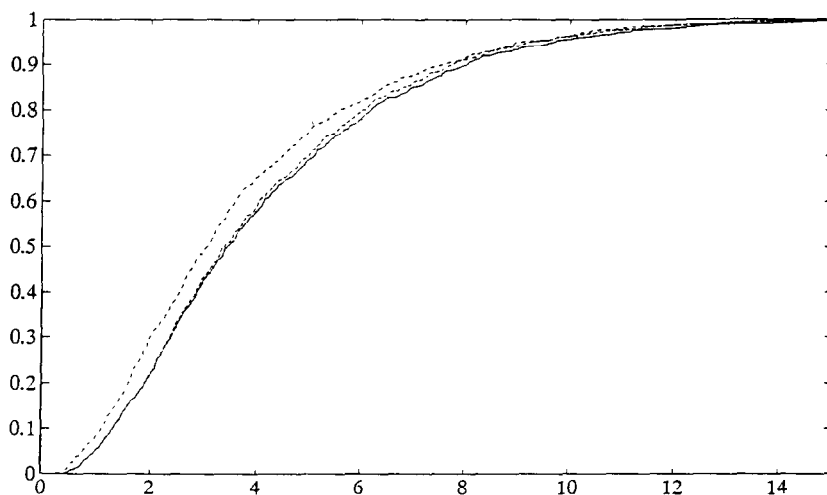
Fig. 7. Empirical distribution function t_W (—), t_{LR} (--), t_{LM} (-·) for bivariate model corresponding with DGP 2, with $\alpha_{11} \neq 0$

(X_2 not necessarily needs to correspond with the variables multiplied by β_2 in the cointegrating relationship.)

Theorem 6 shows that the well-known phenomenon that the Wald, LR, and LM statistics have the same limiting distribution under hypotheses which imply stationary series extend towards hypotheses which reflect a certain amount of nonstationarity when α_{11} has full rank. The limiting distribution of $\tilde{\alpha}_{22}$ shown in Theorem 6 is unlike the limiting distribution of $\hat{\alpha}_{22}$ discussed in Theorems 4 and 5 independent from the values of α_{11} and α_{12}. This is the result of the independent estimation of $\tilde{\alpha}_{22}$ from α_{11} and α_{12}, while $\hat{\alpha}_{22} = \hat{\Pi}_{22} - \hat{\Pi}_{21}\hat{\Pi}_{11}^{-1}\hat{\Pi}_{12}$ partly is a function of the variables $\alpha_{11}(= \Pi_{11})$ and $\alpha_{12}(= \Pi_{12})$.

Although the limiting distributions of the Wald, LR, and LM statistics for cointegration are the same when α_{11} has full rank, in small samples the results from the three statistics will differ. To get an idea to what extent the different cointegration testing statistics could differ we simulated the empirical distribution of the three different test statistics for a bivariate DGP with one cointegrating vector (DGP 2 from Theorem 2) and 100 observations, with $\alpha_{11} \neq 0$ (Fig. 7) and $\alpha_{11} = 0$ (Fig. 8). Fig. 7 shows that for $\alpha_{11} \neq 0$ the empirical distribution functions of the three different statistics are almost the same (and almost equal to their limiting distribution), while the empirical distribution of the Wald statistic differs from the other two distributions when $\alpha_{11} = 0$, as was to be expected. Fig. 7 also shows that, when α_{11} has full rank, the well-known
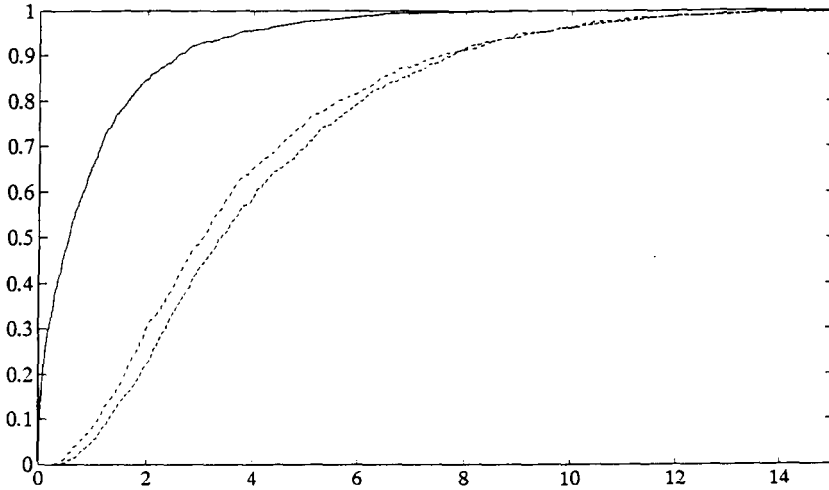
Fig. 8. Empirical distribution function t_W (—), t_{LR} (--), t_{LM} (-·) for bivariate model corresponding with DGP 2, with $\alpha_{11} = 0$

inequality relationship for standard linear models, $t_{LM} \lessgtr t_{LR} \lessgtr t_W$, also tends to hold for the cointegration tests. Although the results obtained from the three different statistics will not differ much when the DGP corresponds with the model under H_0, the power functions of the three different statistics could differ considerably. Because of the vast number of reasonable alternatives for which the power of the different statistics could be calculated, we did not perform such an exercise.

6. Application

To show the applicability of the derived testing procedures, we applied the method described in the former sections on a set of data series. The data set consists of real demand for money, real GNP, and bond and deposit interest rates in Denmark, starting in the first quarter of 1974 and ending in the last quarter of 1987 (55 observations). These series are used to be able to compare the results from the testing procedure proposed in this paper with the results obtained by Johansen and Juselius (1990) in their cointegration study of these four series.

The results in Table 1 are obtained under the same conditions, lag length ($p = 2$), and seasonal dummies as in Johansen and Juselius (1990). Also the hypotheses, which are tested by the Wald, LR, and LM statistics, are the same as

Table 1
Cointegration characteristics for real demand for money ($m2$), real GNP (y), bankdeposit rate (i_d), bond rate (i_b) in Denmark

$$\hat{\Pi} = \begin{bmatrix} \begin{array}{cccc} \Delta m2 & \Delta y & \Delta i_b & \Delta i_d \\ -0.18 & 0.19 & 0.014 & -0.0037 \\ (-2.03) & (-2.04) & (0.41) & (-0.16) \\ 0.11 & -0.31 & -0.018 & 0.020 \\ (0.91) & (2.51) & (-0.37) & (0.67) \\ -1.04 & 0.66 & 0.082 & 0.14 \\ (-2.95) & (1.82) & (0.58) & (1.60) \\ 0.64 & -0.65 & -0.17 & -0.31 \\ (1.47) & (-1.46) & (-0.96) & (-2.87) \end{array} \end{bmatrix} \begin{matrix} \\ m2 \\ \\ \\ y \\ \\ i_b \\ \\ i_d \end{matrix}$$

Number of cointegrating relationships:

	0	1	2	3
t_W	61.1	17.8	13.1	0.06
t_{LR}	49.1	19.1	8.69	2.35
t_{LM}	40.6	15.9	2.73	0.005
Crit. (95%)	53.4	35.1	20.1	9.09

Estimated cointegrating vector:

	$m2$	y	i_b	i_d	con
Two-step	1.0	0.61	-5.76	3.53	8.76
		(1.53)	(-3.72)	(1.61)	(3.70)
ML	1.0	1.03	-5.19	4.19	6.05
		(5.29)	(-6.12)	(2.65)	(5.01)

The critical values are obtained from Johansen and Juselius (1990, Table A3).

tested in Johansen and Juselius, i.e., cointegrating relationships containing a constant term while the series do not contain a linear time trend (DGP 2, Theorem 2).

The testing procedure is started with the construction of an ordering of the analysed series using the estimated value of the Π matrix. The constructed ordering corresponds with the ordering of the series in the estimated Π matrix. The constructed ordering assures the condition that each series is likely to Granger long-run cause at least one series from a preceding column, while the α_{11} matrix for the presumable number of cointegrating vectors (one or two?) probably has a full rank value. Using the mentioned ordering, the suggested testing procedure in Fig. 6 is started with decreasing values of the number of cointegrating relationships. The estimated values of the Wald, LR (= Johansen

Table 2
Sensitivity cointegration statistics for chosen variable ordering for Danish series

Ordering		0	1	2	3
$m2, y, i_b, i_d$	t_W	61.1	17.8	13.1	0.06
	t_{LM}	40.6	15.9	2.73	0.005
$y, m2, i_b, i_d$	t_W	61.1	28.5	13.1	0.06
	t_{LM}	40.6	15.7	2.67	0.003
$i_b, y, m2, i_d$	t_W	61.1	17.1	7.48	0.06
	t_{LM}	40.6	15.0	3.24	0.004
$i_d, i_b, y, m2,$	t_W	61.1	26.8	5.00	0.10
	t_{LM}	40.6	15.3	5.73	0.57

trace), and LM statistics lead to a certain amount of ambiguity with respect to the (non)rejection of the hypothesis of zero cointegrating vectors. Using the asymptotic critical value, the Wald statistic clearly rejects the hypothesis of zero cointegrating vectors with 95% significance, while the LR and LM statistics are both below their 95% asymptotic significance levels. So, although the asymptotic distributions of the Wald, LR, and LM statistics under the hypotheses of cointegration are the same, in small samples the results from the three different statistics can differ substantially. This shows again how careful one has to be with the application of asymptotic theory in small samples. The weak foundation of the cointegration hypothesis indicated by the LR and LM statistics is of course not surprising given the small 't-values' of the elements of the long-run multiplier Π.

In Table 1 an ordering of the variables was used which was derived from 'pre-testing'. In the analysed example the Granger long-run causality relationships are rather clear because the real variables ($m2, y$) are not likely to Granger long-run cause the monetary variables (i_b, i_d), while the opposite relation is much more likely. The question is how sensitive the results are for the chosen ordering of the analysed series because for most other series these relationships tend to be much less clear. To show this sensitivity for the analysed series we calculated the statistics for four different variable orderings which are shown in Table 2.

The LR statistic is not contained in Table 2 because of its invariance with respect to the chosen ordering of the variables. Although the values of the Wald statistic tend to be sensitive to the ordering chosen, the conclusions with respect to the (non)rejection of certain hypotheses are not affected by the chosen ordering. The conclusions obtained from the use of the Wald statistic for the analysed series seem to be quite robust with respect to the chosen variable ordering. Yet, one has to bear in mind that Fig. 8 shows that lower rank values of α_{11} tend to decrease the value of the Wald statistic. This may also explain the small values of certain Wald statistics. The LM statistic seems to be almost insensitive with respect to the chosen ordering, as was to be expected because

the limiting distribution of the LM statistic is independent from the value of other parameters (except α_{22} of course).

7. Conclusions

The paper shows how one can construct Wald, LR, and LM statistics to test for cointegration using a parametric ECM. The calculation of the Wald statistic only involves linear regression, while the LM statistic stems from an auxiliary regression once one has estimated a cointegration model. The LM statistic is therefore an interesting and easy to calculate diagnostic. For the calculation of the critical values, the limiting distributions of these statistics were constructed, which turn out to be equal to one another when α_{11} has full rank, while the limiting distribution of the Wald statistic changes when α_{11} has a lower rank value. The application of the three cointegration statistics to a Danish data set shows, however, that although the limiting distributions are the same, the conclusions stemming from the use of the different statistics can be quite different from one another. This shows how careful one has to be with the use of asymptotic theory in small samples.

In future work, the parametric cointegration model will be examined for series which have a more complicated nature like, for instance, series with heteroscedastic or nonnormal distributed disturbances. The limiting distributions of the Wald, LR, and LM statistics are invariant under a certain degree of heteroscedasticity/nonnormality, but the testing procedures will lose power when one neglects to model the heteroscedasticity/nonnormality properly. In the parametric cointegration model it seems less difficult to jointly test for cointegration and model heteroscedasticity/nonnormality than in the 'data-parametric' cointegration techniques used in the literature. We emphasize this point because in recent work on unit roots in univariate models the presence of unit roots turned out to be quite sensitive to the incorporation of heteroscedastic or nonnormal disturbances (see Kleibergen and Van Dijk, 1993b). It is shown there that the probability of an unit root in the US treasury bill rate increases considerably when heteroscedasticity and nonnormality are properly modelled. Another interesting issue which also seems to be more straightforward to conduct in the parametric cointegration model concerns the issue of the recursive calculation of the test statistics. This is a topic of further research.

Appendix

Some of the proofs in this appendix are informal (heuristic). For more formal definitions of certain arguments we refer to Billingsley (1968), Phillips and Durlauf (1986), Phillips (1992), and Johansen (1991a, b).

Proof of Theorem 2 (only the first part is proved)

To construct the limiting distribution of the cointegrating vector estimator, $\hat{\beta}_2 = -\hat{\Pi}_{21}\hat{\Pi}_{11}^{-1}$, at first we construct the limiting distribution of the least-squares estimator, $\hat{\Pi}_1 = (X'_{-1}M_t X_{-1})^{-1} X'_{-1}M_t \Delta X_1$ of $\Pi_1 = (\Pi'_{11}\ \Pi'_{21})'$ in the model (for simplicity $p = 1$ which does not affect the limiting distributions).

$$\Delta X_1 = c_1 + X_{-1}\Pi_1 + \varepsilon_1, \quad \mathscr{E}((\varepsilon_{1t}\varepsilon_{2t})'(\varepsilon_{1t}\varepsilon_{2t})) = \Omega, \quad v_{1t} = \varepsilon_{1t} - \varepsilon_{2t}\beta_2,$$

$$\Delta X_2 = c_2 + X_{-1}\Pi_2 + \varepsilon_2, \quad \mathscr{E}((v_{1t}\varepsilon_{2t})'(v_{1t}\varepsilon_{2t})) = \Sigma = B'\Omega B,$$

$$\Pi = (\Pi_1\ \Pi_2) = \begin{bmatrix} I_r \\ -\beta_2 \end{bmatrix}(\alpha_{11}\ \alpha_{12}), \quad \beta = \begin{bmatrix} I_r \\ -\beta_2 \end{bmatrix}, \quad \beta_\perp = \begin{bmatrix} \beta'_2 \\ I_{k-r} \end{bmatrix},$$

$$\hat{\Pi}_1 = (X'_{-1}M_t X_{-1})^{-1} X'_{-1}M_t \Delta X_1$$

$$= (\beta\ \beta_\perp\gamma\ D)((\beta\ \beta_\perp\gamma\ D)'X'_{-1}M_t X_{-1}(\beta\ \beta_\perp\gamma\ D))^{-1}$$

$$(\beta\ \beta_\perp\gamma\ D)'X'_{-1}M_t\Delta X_1,$$

$$B = (\beta(0\ I_{k-r})'), \qquad D = \beta_\perp\mu'_2, \quad D'\beta_\perp\gamma = 0,$$

$$(\beta\ \beta_\perp\gamma\ D)'X'_{-1}M_t X_{-1}(\beta\ \beta_\perp\gamma\ D) = \begin{bmatrix} M_{11}(1) & M_{12}(1) & M_{13}(1\tfrac{1}{2}) \\ M_{21}(1) & M_{22}(2) & M_{23}(2\tfrac{1}{2}) \\ M_{31}(1\tfrac{1}{2}) & M_{32}(2\tfrac{1}{2}) & M_{33}(3) \end{bmatrix},$$

where $M_{ij}(k)$, $Q_{ij}(k)$, $O_i(k)$ represent the T^kth-order convergence behaviour of the M_{ij}, Q_{ij}, O_i block matrices. Using the convergence behaviour of the different block matrices, the limiting distribution of the inverse of moment matrix $(\beta\ \beta_\perp\gamma\ D)'X'_{-1}M_t X_{-1}(\beta\ \beta_\perp\gamma\ D)$ can be constructed,

$$((\beta\ \beta_\perp\gamma\ D)'X'_{-1}M_t X_{-1}(\beta\ \beta_\perp\gamma\ D))^{-1}$$

$$= \begin{bmatrix} Q_{11}(-1) & Q_{12}(-2) & Q_{13}(-2\tfrac{1}{2}) \\ Q_{21}(-2) & Q_{22}(-2) & Q_{23}(-2\tfrac{1}{2}) \\ Q_{31}(-2\tfrac{1}{2}) & Q_{32}(-2\tfrac{1}{2}) & Q_{33}(-3) \end{bmatrix}.$$

The limiting behaviour of $(\beta\ \beta_\perp\gamma\ D)'X'_{-1}M_t\Delta X_1$ can be represented as

$$(\beta\ \beta_\perp\gamma\ D)'X'_{-1}M_t\Delta X_1 = (\beta\ \beta_\perp\gamma\ D)'X'_{-1}M_t(X_{-1}\beta\alpha_{11} + \varepsilon_1)$$

$$= \begin{bmatrix} M_{11}(1)\alpha_{11} + O_1(\tfrac{1}{2}) \\ M_{21}(1)\alpha_{11} + O_2(1) \\ M_{31}(1\tfrac{1}{2})\alpha_{11} + O_3(1\tfrac{1}{2}) \end{bmatrix}.$$

Because the first part of the limiting matrix of $(\beta\ \beta_\perp\gamma\ D)'X'_{-1}M_i\Delta X_1$ is proportional to the first column of the moment matrix $(\beta\ \beta_\perp\gamma\ D)' \times X'_{-1}M_iX_{-1}(\beta\ \beta_\perp\gamma\ D)$, the limiting form of the product of $((\beta\ \beta_\perp\gamma\ D)'X'_{-1} \times M_iX_{-1}(\beta\ \beta_\perp\gamma\ D))^{-1}$ and $(\beta\ \beta_\perp\gamma\ D)'X'_{-1}M_i\Delta X_1$ becomes

$$\begin{bmatrix}\alpha_{11}\\0\\0\end{bmatrix}+\begin{bmatrix}Q_{11}(-1)O_1(\tfrac12)\\Q_{22}(-2)O_2(1)\ +\ Q_{23}(-2\tfrac12)O_3(1\tfrac12)\\Q_{32}(-2\tfrac12)O_2(1)\ +\ Q_{33}(-3)O_3(1\tfrac12)\end{bmatrix}.$$

The limiting form of the least squares estimator $\hat\Pi_1$ then becomes

$$(\beta\ \beta_\perp\gamma\ D)\left[\begin{bmatrix}\alpha_{11}\\0\\0\end{bmatrix}+\begin{bmatrix}Q_{11}(-1)O_1(\tfrac12)\\Q_{22}(-2)O_2(1)\ +\ Q_{23}(-2\tfrac12)O_3(1\tfrac12)\\Q_{32}(-2\tfrac12)O_2(1)\ +\ Q_{33}(-3)O_3(1\tfrac12)\end{bmatrix}\right]$$

$$=\left[\left[\begin{array}{c}\alpha_{11}+Q_{11}(-1)O_1(\tfrac12)\\-\beta_2(\alpha_{11}+Q_{11}(-1)O_1(\tfrac12))\end{array}\right.\right.$$

$$\left.\left.+(\gamma\ \mu_2')\begin{bmatrix}Q_{22}(-2)&Q_{23}(-2\tfrac12)\\Q_{32}(-2\tfrac12)&Q_{33}(-3)\end{bmatrix}\begin{bmatrix}O_2(1)\\O_3(1\tfrac12)\end{bmatrix}\right]\right],$$

while the limiting form of the cointegrating vector estimator $\hat\beta_2=-\hat\Pi_{21}\hat\Pi_{11}^{-1}$ becomes equal to

$$\hat\beta_2=-\hat\Pi_{21}\hat\Pi_{11}^{-1}$$

$$=\left[\beta_2(\alpha_{11}+Q_{11}(-1)O_1(\tfrac12))\right.$$

$$\left.-(\gamma\ \mu_2')\begin{bmatrix}Q_{22}(-2)&Q_{23}(-2\tfrac12)\\Q_{32}(-2\tfrac12)&Q_{33}(-3)\end{bmatrix}\begin{bmatrix}O_2(1)\\O_3(1\tfrac12)\end{bmatrix}\right]$$

$$\times(\alpha_{11}+Q_{11}(-1)O_1(\tfrac12))^{-1}$$

$$=\beta_2-(\gamma\ \mu_2')\begin{bmatrix}Q_{22}(-2)&Q_{23}(-2\tfrac12)\\Q_{32}(-2\tfrac12)&Q_{33}(-3)\end{bmatrix}\begin{bmatrix}O_2(1)\\O_3(1\tfrac12)\end{bmatrix}(\alpha_{11}+Q_{11}(-1)O_1(\tfrac12))^{-1}$$

$$=\beta_2-(\gamma\ \mu_2')\begin{bmatrix}Q_{22}(-2)&Q_{23}(-2\tfrac12)\\Q_{32}(-2\tfrac12)&Q_{33}(-3)\end{bmatrix}\begin{bmatrix}O_2(1)\\O_3(1\tfrac12)\end{bmatrix}\alpha_{11}^{-1}.$$

To construct the exact functional expression for the limiting distribution of $\hat\beta_2$ we need to construct the functional forms of Q_{22}, Q_{23}, Q_{33}, and O_2, O_3 (see also Phillips and Durlauf, 1986; Johansen, 1991a),

$$\lim_{T \to \infty} T^{-2} (\beta_\perp \gamma\ T^{-1/2} D)' X'_{-1} M_\iota X_{-1} (\beta_\perp \gamma\ T^{-1/2} D) \Rightarrow \begin{bmatrix} \int_0^1 \bar{P}' \bar{P} & \int_0^1 \bar{P}' \bar{Q} \\ \int_0^1 \bar{Q}' \bar{P} & \int_0^1 \bar{Q}' \bar{Q} \end{bmatrix},$$

$$\lim_{T \to \infty} T^{-1} (\beta_\perp \gamma\ T^{-1/2} D)' X'_{-1} M_\iota \varepsilon_1 \Rightarrow \begin{bmatrix} \left(\int_0^1 \bar{P}' dS \right) \\ \left(\int_0^1 \bar{Q}' dS \right) \end{bmatrix} \Sigma^{1/2} (I_r\ \beta_2')',$$

$$\lim_{T \to \infty} T (\gamma\ T^{1/2} \mu_2')' \beta_\perp' \beta_\perp (\hat{\beta}_2 - \beta_2) = -\lim_{T \to \infty} T (\gamma\ T^{1/2} \mu_2')' \beta_\perp' \beta_\perp (\gamma\ \mu_2')$$

$$\times \begin{bmatrix} Q_{22}(-2) & Q_{23}(-2\tfrac{1}{2}) \\ Q_{32}(-2\tfrac{1}{2}) & Q_{33}(-3) \end{bmatrix} \begin{bmatrix} O_2(1) \\ O_3(1\tfrac{1}{2}) \end{bmatrix} \alpha_{11}^{-1}$$

$$\Rightarrow - \begin{bmatrix} \gamma' \beta_\perp' \beta_\perp \gamma & 0 \\ 0 & \mu_2 \beta_\perp' \beta_\perp \mu_2' \end{bmatrix} \begin{bmatrix} \int_0^1 \bar{P}' \bar{P} & \int_0^1 \bar{P}' \bar{Q} \\ \int_0^1 \bar{Q}' \bar{P} & \int_0^1 \bar{Q}' \bar{Q} \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 \bar{P}' dS \\ \int_0^1 \bar{Q}' dS \end{bmatrix} \Sigma^{1/2} (I_r\ \beta_2')' \alpha_{11}^{-1}$$

$$\Rightarrow - \begin{bmatrix} \gamma' \beta_\perp' \beta_\perp \gamma & 0 \\ 0 & \mu_2 \beta_\perp' \beta_\perp \mu_2' \end{bmatrix} \left(\int_0^1 F'F \right)^{-1} \left(\int_0^1 F' dS \right) \Sigma^{1/2} (I_r\ \beta_2')' \alpha_{11}^{-1}$$

$$\Rightarrow - \begin{bmatrix} \gamma' \beta_\perp' \beta_\perp \gamma & 0 \\ 0 & \mu_2 \beta_\perp' \beta_\perp \mu_2' \end{bmatrix} \left(\int_0^1 F'F \right)^{-1} \left(\int_0^1 F' dS \right) \Sigma^{1/2} K_\Sigma K_\Sigma^{-1} (I_r\ \beta_2')' \alpha_{11}^{-1}$$

$$\Rightarrow - \begin{bmatrix} \gamma' \beta_\perp' \beta_\perp \gamma & 0 \\ 0 & \mu_2 \beta_\perp' \beta_\perp \mu_2' \end{bmatrix} \left(\int_0^1 F'F \right)^{-1} \left[\left(\int_0^1 F' dU_{1.2} \right) \right.$$

$$\left. + \left(\int_0^1 F' dU_2 \right) \left(\Sigma_{22}^{-1} \Sigma_{21} + \beta_2 \right) \right] \alpha_{11}^{-1}$$

$$\Rightarrow - \begin{bmatrix} \gamma' \beta_\perp' \beta_\perp \gamma & 0 \\ 0 & \mu_2 \beta_\perp' \beta_\perp \mu_2' \end{bmatrix} \left(\int_0^1 F'F \right)^{-1} \left[\left(\int_0^1 F' dU_{1.2} \right) \right.$$

$$\left. + \left(\int_0^1 F' dU_2 \right) \Omega_{22}^{-1} \Omega_{21} \right] \alpha_{11}^{-1},$$

where

$$U = (S_1 \ S_2)\Sigma^{1/2}, \qquad U_{1.2} = U_1 - U_2\Sigma_{22}^{-1}\Sigma_{21},$$

$$\bar{P} = P - \int_0^1 P(t)\mathrm{d}t, \qquad \bar{Q} = Q - \int_0^1 Q(t)\mathrm{d}t,$$

$$K_\Sigma = \begin{bmatrix} I_r & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I_{k-r} \end{bmatrix}, \qquad K_\Sigma^{-1} = \begin{bmatrix} I_r & 0 \\ \Sigma_{22}^{-1}\Sigma_{21} & I_{k-r} \end{bmatrix},$$

$$\Sigma^{1/2}(I_r \ \beta_2')' = \Omega^{1/2}(I_r \ 0).$$

Proof of Theorem 3

When $\alpha_{11} = 0$,

$$\hat{\beta}_2 = \left[\beta_2 Q_{11}(-1)O_1(\tfrac{1}{2}) - (\gamma \ \mu_2') \begin{bmatrix} Q_{22}(-2) & Q_{23}(-2\tfrac{1}{2}) \\ Q_{32}(-2\tfrac{1}{2}) & Q_{33}(-3) \end{bmatrix} \right.$$

$$\left. \times \begin{bmatrix} O_2(1) \\ O_3(1\tfrac{1}{2}) \end{bmatrix} (Q_{11}(-1)O_1(\tfrac{1}{2}))^{-1} \right]$$

$$= \left[\beta_2 - (\gamma \ \mu_2') \begin{bmatrix} Q_{22}(-2) & Q_{23}(-2\tfrac{1}{2}) \\ Q_{32}(-2\tfrac{1}{2}) & Q_{33}(-3) \end{bmatrix} \begin{bmatrix} O_2(1) \\ O_3(1\tfrac{1}{2}) \end{bmatrix} (Q_{11}(-1)O_1(\tfrac{1}{2}))^{-1} \right],$$

$$\lim_{T\to\infty} T^{-1}\beta'X_{-1}'M_t X_{-1}\beta \Rightarrow \mathrm{var}(X_t\beta) \Leftarrow \left(\lim_{T\to\infty} TQ_{11}(-1) \right)^{-1},$$

$$\lim_{T\to\infty} T^{-1/2}O_1(\tfrac{1}{2}) = \lim_{T\to\infty} T^{-1/2}\beta'X_{-1}'M_t\varepsilon_1 \Rightarrow \mathrm{n}(0, \Omega_{11} \otimes \mathrm{var}(X_t\beta))$$

$$\Rightarrow \int_0^1 ((X(t)\beta)'\mathrm{d}S)\Omega^{1/2}(I_r \ 0)',$$

$$\lim_{T\to\infty} T^{1/2}Q_{11}(-1)O_1(\tfrac{1}{2}) \Rightarrow \mathrm{n}(0, \Omega_{11} \otimes (\mathrm{var}(X_t\beta))^{-1})$$

$$\Rightarrow (\mathrm{var}(X_t\beta))^{-1}\left(\int_0^1 (X(t)\beta)'\mathrm{d}S \right)\Omega^{1/2}(I_r \ 0)'$$

$$\Leftarrow \lim_{T\to\infty} T^{1/2}(\hat{\alpha}_{11} - \alpha_{11}).$$

The limiting distribution of $\hat{\beta}_2$ then becomes

$$\lim_{T \to \infty} T^{1/2}(\gamma \; T^{1/2}\mu_2')'\beta_\perp'\beta_\perp(\hat{\beta}_2 - \beta_2)$$

$$\Rightarrow - \begin{bmatrix} \gamma'\beta_\perp'\beta_\perp\gamma & 0 \\ 0 & \mu_2\beta_\perp'\beta_\perp\mu_2' \end{bmatrix} \left(\int_0^1 F'F \right)^{-1}$$

$$\times \left[\left(\int_0^1 F'\mathrm{d}U_{1.2} \right) + \left(\int_0^1 F'\mathrm{d}U_2 \right) \Omega_{22}^{-1}\Omega_{21} \right]$$

$$\times [n(0, \Omega_{11} \otimes (\mathrm{var}(X_{1t} - X_{2t}\beta_2 - \mu_1))^{-1})]^{-1}$$

$$\Rightarrow - \begin{bmatrix} \gamma'\beta_\perp'\beta_\perp\gamma & 0 \\ 0 & \mu_2\beta_\perp'\beta_\perp\mu_2' \end{bmatrix} \left(\int_0^1 F'F \right)^{-1}$$

$$\times \left[\left(\int_0^1 F'\mathrm{d}U_{1.2} \right) + \left(\int_0^1 F'\mathrm{d}U_2 \right) \Omega_{22}^{-1}\Omega_{21} \right]$$

$$\times \left[\left(\int_0^1 (X(t)\beta)'\mathrm{d}S \right) \Omega^{1/2}(I_r, 0)' \right]^{-1} (\mathrm{var}(X_t\beta)).$$

Proof of Theorem 4 (only the first part is proved)

$$\lim_{T \to \infty} T(\gamma \; T^{1/2}\mu_2')'\beta_\perp'\beta_\perp(\hat{\Pi}_{22} + \beta_2\alpha_{12})$$

$$\Rightarrow \begin{bmatrix} \gamma'\beta_\perp'\beta_\perp\gamma & 0 \\ 0 & \mu_2\beta_\perp'\beta_\perp\mu_2' \end{bmatrix} \left(\int_0^1 F'F \right)^{-1} \left(\int_0^1 F'\mathrm{d}S \right) \Omega^{1/2}(0 \; I_{k-r})',$$

$$\lim_{T \to \infty} T(\gamma \; T^{1/2}\mu_2')'\beta_\perp'\beta_\perp(\hat{\beta}_2\hat{\alpha}_{12} - \beta_2\alpha_{12})$$

$$\Rightarrow - \begin{bmatrix} \gamma'\beta_\perp'\beta_\perp\gamma & 0 \\ 0 & \mu_2\beta_\perp'\beta_\perp\mu_2' \end{bmatrix} \left(\int_0^1 F'F \right)^{-1} \left(\int_0^1 F'\mathrm{d}S \right) \Omega^{1/2}(I_r, 0)'\alpha_{11}^{-1}\alpha_{12},$$

$$\lim_{T \to \infty} T(\gamma \; T^{1/2}\mu_2')'\beta_\perp'\beta_\perp\hat{\alpha}_{22} = \lim_{T \to \infty} T(\gamma \; T^{1/2}\mu_2')'\beta_\perp'\beta_\perp[(\hat{\Pi}_{22} + \beta_2\alpha_{12})$$

$$+ (\hat{\beta}_2\hat{\alpha}_{12} - \beta_2\alpha_{12})]$$

$$\Rightarrow -\begin{bmatrix} \gamma'\beta'_\perp\beta_\perp\gamma & 0 \\ 0 & \mu_2\beta'_\perp\beta_\perp\mu'_2 \end{bmatrix} \left(\int_0^1 F'F \right)^{-1} \left(\int_0^1 F'dS \right) \Omega^{1/2}$$

$$\times \left[\begin{bmatrix} 0 \\ I_{k-r} \end{bmatrix} - \begin{bmatrix} I_r \\ 0 \end{bmatrix} \alpha_{11}^{-1}\alpha_{12} \right],$$

$$\lim_{T\to\infty} T^{-2}(\gamma\ T^{1/2}\mu'_2)'(X_2)'_{-1}M_{(\iota\ (X_1-X\beta_2)\ -1)}(X_2)(\gamma\ T^{1/2}\mu'_2) \Rightarrow \left(\int_0^1 F'F \right),$$

$$\lim_{T\to\infty} t_W = \lim_{T\to\infty} \mathrm{tr}\left(\begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix}' \hat{\Omega} \begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix} \right)^{-1}$$

$$\times \hat{\alpha}'_{22}(X_2)'_{-1}M_{(\iota\ (X_1-X\beta_2)_{-1})}(X_2)_{-1}\hat{\alpha}_{22}$$

$$= \mathrm{tr}\left[\lim_{T\to\infty} \left(\begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix}' \hat{\Omega} \begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix} \right)^{-1} \right]$$

$$\times \left[\lim_{T\to\infty} T(\gamma\ T^{1/2}\mu'_2)'\beta'_\perp\beta_\perp\hat{\alpha}_{22} \right]'\left[\lim_{T\to\infty} T^{-2}(\gamma\ T^{1/2}\mu'_2)'(X_2)'_{-1} \right.$$

$$\times M_{(\iota\ (X_1-X\beta_2)_{-1})}(X_2)(\gamma\ T^{1/2}\mu'_2) \left] \right[\lim_{T\to\infty} T(\gamma\ T^{1/2}\mu'_2)'\beta'_\perp\beta_\perp\hat{\alpha}_{22} \right]$$

$$\Rightarrow \mathrm{tr}\left(\int_0^1 F'dS_{22} \right)'\left(\int_0^1 F'F \right)^{-1}\left(\int_0^1 F'dS_{22} \right)$$

$$\Rightarrow \mathrm{tr}\left(\int_0^1 V'dS_{22} \right)'\left(\int_0^1 V'V \right)^{-1}\left(\int_0^1 V'dS_{22} \right).$$

Proof of Theorem 5

When $\alpha_{11} = 0$ and $r = k - r$,

$$\lim_{T\to\infty} T^{1/2}(\gamma\ T^{1/2}\mu'_2)'\beta'_\perp\beta_\perp(\hat{\beta}_2\hat{\alpha}_{12} - \beta_2\alpha_{12})$$

$$\Rightarrow -\begin{bmatrix} \gamma'\beta'_\perp\beta_\perp\gamma & 0 \\ 0 & \mu_2\beta'_\perp\beta_\perp\mu'_2 \end{bmatrix} \left(\int_0^1 F'F \right)^{-1}$$

$$\times \left[\left(\int_0^1 F'dS \right) \Omega^{1/2}(I_r, 0)'[\mathrm{n}(0, \Omega_{11} \otimes (\mathrm{var}(X_t\beta))^{-1})]^{-1} \right] \alpha_{12}$$

$$\Rightarrow -\begin{bmatrix} \gamma'\beta'_\perp\beta_\perp\gamma & 0 \\ 0 & \mu_2\beta'_\perp\beta_\perp\mu'_2 \end{bmatrix}\left(\int_0^1 F'F\right)^{-1}\left[\left(\int_0^1 F'dS\right)\Omega^{1/2}(I_r, 0)'\right]$$

$$\times\left[\int_0^1 ((X(t)\beta)'dS)\Omega^{1/2}(I_r, 0)'\right]^{-1}(\mathrm{var}(X_t\beta))\alpha_{12},$$

$$\lim_{T\to\infty} T(\gamma\, T^{1/2}\mu'_2)'\beta'_\perp\beta_\perp(\hat{\Pi}_{22} + \beta_2\alpha_{12})$$

$$\Rightarrow \begin{bmatrix} \gamma'\beta'_\perp\beta_\perp\gamma & 0 \\ 0 & \mu_2\beta'_\perp\beta_\perp\mu'_2 \end{bmatrix}\left(\int_0^1 F'F\right)^{-1}\left(\int_0^1 F'dS\right)\Omega^{1/2}(0\, I_{k-r})',$$

$$\lim_{T\to\infty} T^{1/2}(\gamma\, T^{1/2}\mu_2)'\beta'_\perp\beta_\perp\hat{\alpha}_{22}$$

$$= \lim_{T\to\infty} T^{1/2}(\gamma\, T^{1/2}\mu_2)'[(\hat{\Pi}_{22} + \beta_2\alpha_{12}) + (\hat{\beta}_2\hat{\alpha}_{12} - \beta_2\alpha_{12})]$$

$$\Rightarrow -\begin{bmatrix} \gamma'\beta'_\perp\beta_\perp\gamma & 0 \\ 0 & \mu_2\beta'_\perp\beta_\perp\mu'_2 \end{bmatrix}\left(\int_0^1 F'F\right)^{-1}\left(\int_0^1 F'dS\right)\Omega^{1/2}(I, 0)'$$

$$\times [\mathrm{n}(0, \Omega_{11} \otimes (\mathrm{var}(X_t\beta))^{-1})]^{-1}\alpha_{12}$$

$$\Rightarrow -\begin{bmatrix} \gamma'\beta'_\perp\beta_\perp\gamma & 0 \\ 0 & \mu_2\beta'_\perp\beta_\perp\mu'_2 \end{bmatrix}\left(\int_0^1 F'F\right)^{-1}\left(\int_0^1 F'dS\right)\Omega^{1/2}(I_r, 0)'$$

$$\times\left[\int_0^1 (X(t)\beta)'dS)\Omega^{1/2}(I_r, 0)'\right]^{-1}(\mathrm{var}(X_t\beta))\alpha_{12},$$

$$\lim_{T\to\infty} T^{-1}(X_1 - X_2\hat{\beta}_2)'_{-1}M_t(X_1 - X_2\hat{\beta}_2)_{-1}$$

$$= \mathrm{var}(X_t\beta) + \lim_{T\to\infty} T^{-1}(\beta_2 - \hat{\beta}_2)'(X_2)'_{-1}M_t(X_2)_{-1}(\beta_2 - \hat{\beta}_2)$$

$$\Rightarrow \mathrm{var}(X_t\beta) + [\mathrm{n}(0, \Omega_{11} \otimes (\mathrm{var}(X_t\beta))^{-1})]^{-1'}(I_r, 0)\Omega^{1/2'}$$

$$\times\left(\int_0^1 F'dS\right)'\left(\int_0^1 F'F\right)^{-1}\left(\int_0^1 F'dS\right)$$

$$\times \Omega^{1/2}(I_r, 0)'[\mathrm{n}(0, \Omega_{11} \otimes (\mathrm{var}(X_t\beta))^{-1})]^{-1}$$

$$\Rightarrow \text{var}(X_t\beta) + (\text{var}(X_t\beta)) \left[\int_0^1 (X(t)\beta)' dS) \Omega^{1/2} (I_r, 0)' \right]^{-1'} (I_r, 0) \Omega^{1/2'}$$

$$\times \left(\int_0^1 F' dS \right)' \left(\int_0^1 F'F \right)^{-1} \left(\int_0^1 F' dS \right) \Omega^{1/2} (I_r, 0)$$

$$\times \left[\int_0^1 (X(t)\beta)' dS) \Omega^{1/2} (I_r, 0)' \right]^{-1} (\text{var}(X_t\beta)),$$

$$\lim_{T \to \infty} T^{-2} (\gamma T^{-1/2} \mu_2')' (X_2)_{-1}' M_t (X_2)_{-1} (\gamma T^{-1/2} \mu_2') \Rightarrow \left(\int_0^1 F'F \right),$$

$$\lim_{T \to \infty} T^{-1(1/2)} (\gamma T^{-1/2} \mu_2')' (X_2)_{-1}' M_t (X_1 - X_2 \hat{\beta}_2)_{-1}$$

$$= \lim_{T \to \infty} T^{-1(1/2)} (\gamma T^{1/2} \mu_2')' (X_2)_{-1}' M_t (X_1 - X_2 \beta_2)_{-1}$$

$$+ \lim_{T \to \infty} T^{-1(1/2)} (\gamma T^{1/2} \mu_2)' (X_2)_{-1}' M_t (X_2)_{-1} (\beta_2 - \hat{\beta}_2)$$

$$= \lim_{T \to \infty} T^{-1(1/2)} (\gamma T^{-1/2} \mu_2')' (X_2)_{-1}' M_t (X_2)_{-1} (\beta_2 - \hat{\beta}_2)$$

$$\Rightarrow \left(\int_0^1 F'F \right) \left(\int_0^1 F'F \right)^{-1} \left(\int_0^1 F' dS \right) \Omega^{1/2} (I_r, 0)'$$

$$\times [n(0, \Omega_{11} \otimes (\text{var}(X_t\beta))^{-1})]^{-1}$$

$$\Rightarrow \left(\int_0^1 F' dS \right) \Omega^{1/2} (I_r, 0)' \left[\int_0^1 (X(t)\beta)' dS) \Omega^{1/2} (I_r, 0)' \right]^{-1} (\text{var}(X_t\beta)),$$

$$\lim_{T \to \infty} T^{-2} (\gamma T^{-1/2} \mu_2')' [(X_2)_{-1}' M_{(t}{}_{(X_1 - X\beta_2)_{-1})} (X_2)_{-1} (\gamma T^{-1/2} \mu_2')]$$

$$= \lim_{T \to \infty} T^{-2} (\gamma T^{-1/2} \mu_2')' [(X_2)_{-1}' M_t (X_2)_{-1} - (X_2)_{-1}' M_t (X_2)_{-1} (\beta_2 - \hat{\beta}_2)$$

$$\times [(X_1 - X_2 \beta_2)_{-1}' M_t (X_1 - X_2 \beta_2)_{-1}$$

$$+ (\beta_2 - \hat{\beta}_2)' (X_2)_{-1}' M_t (X_2)_{-1} (\beta_2 - \hat{\beta}_2)]^{-1}$$

$$\times (\beta_2 - \hat{\beta}_2)' (X_2)_{-1}' M_t (X_2)_{-1}] (\gamma T^{-1/2} \mu_2')$$

$$\Rightarrow \left(\int_0^1 F'F\right) - \left(\int_0^1 F'\mathrm{d}S\right)\Omega^{1/2}(I_r,0)'\left[\int_0^1 (X(t)\beta)'\mathrm{d}S)\,\Omega^{1/2}(I_r,0)'\right]^{-1}$$

$$\times (\mathrm{var}(X_t\beta))\left\{\mathrm{var}(X_t\beta) + (\mathrm{var}(X_t\beta))\left[\int_0^1 (X(t)\beta)'\mathrm{d}S)\,\Omega^{1/2}(I_r,0)'\right]^{-1'}\right.$$

$$\times (I_r,0)\,\Omega^{1/2'}\left(\int_0^1 F'\mathrm{d}S\right)'\left(\int_0^1 F'F\right)^{-1}\left(\int_0^1 F'\mathrm{d}S\right)\Omega^{1/2}(I_r,0)$$

$$\times \left[\int_0^1 (X(t)\beta)'\mathrm{d}S)\,\Omega^{1/2}(I_r,0)'\right]^{-1}(\mathrm{var}(X_t\beta))\right\}^{-1}$$

$$\times (\mathrm{var}(X_t\beta))\left[\int_0^1 (X(t)\beta)'\mathrm{d}S)\,\Omega^{1/2}(I_r,0)'\right]^{-1'}(I_r,0)\,\Omega^{1/2'}\left(\int_0^1 F'\mathrm{d}S\right)',$$

$$\left(\int_0^1 F'F\right)^{-1}\left[\left(\int_0^1 F'F\right) - \left(\int_0^1 F'\mathrm{d}S\right)\Omega^{1/2}(I_r,0)'\right.$$

$$\times \left[\int_0^1 (X(t)\beta)'\mathrm{d}S)\,\Omega^{1/2}(I_r,0)'\right]^{-1}(\mathrm{var}(X_t\beta))$$

$$\times \left\{\mathrm{var}(X_t\beta) + (\mathrm{var}(X_t\beta))\left[\int_0^1 (X(t)\beta)'\mathrm{d}S)\,\Omega^{1/2}(I_r,0)'\right]^{-1'}(I_r,0)\,\Omega^{1/2'}\right.$$

$$\times \left(\int_0^1 F'\mathrm{d}S\right)'\left(\int_0^1 F'F\right)^{-1}\left(\int_0^1 F'\mathrm{d}S\right)\Omega^{1/2}(I_r,0)$$

$$\times \left[\int_0^1 (X(t)\beta)'\mathrm{d}S)\,\Omega^{1/2}(I_r,0)'\right]^{-1}(\mathrm{var}(X_t\beta))\right\}^{-1}(\mathrm{var}(X_t\beta))$$

$$\times \left[\int_0^1 (X(t)\beta)'\mathrm{d}S)\,\Omega^{1/2}(I_r,0)'\right]^{-1'}(I_r,0)\,\Omega^{1/2'}\left(\int_0^1 F'\mathrm{d}S\right)'\right]\left(\int_0^1 F'F\right)^{-1}$$

$$= \left(\left(\int_0^1 F'F \right) + \left(\int_0^1 F' \mathrm{d}S \right) \Omega^{1/2}(I_r,0)' \left[\int_0^1 (X(t)\beta)' \mathrm{d}S) \Omega^{1/2}(I_r,0)' \right]^{-1} \right.$$

$$\times (\mathrm{var}(X_t\beta)) \left[\int_0^1 (X(t)\beta)' \mathrm{d}S) \Omega^{1/2}(I_r,0)' \right]^{-1'} (I_r,0) \Omega^{1/2'} \left(\int_0^1 F' \mathrm{d}S \right)' \right)^{-1}$$

$$= \left(\left(\int_0^1 F'F \right) + \left(\int_0^1 F' \mathrm{d}V \right) \left[\int_0^1 Y(t)' \mathrm{d}V \right]^{-1} \right.$$

$$\times \left[\int_0^1 Y(t)' \mathrm{d}V \right]^{-1'} \left(\int_0^1 F' \mathrm{d}V \right)' \right)^{-1},$$

where $Y(t) = (\mathrm{var}(X_t\beta))^{-1/2}) X(t)\beta,\quad \mathrm{var}(Y(t)) = I_r,\quad \int_0^1 Y(t)'Y(t)\mathrm{d}t = I_r,$ and $V = S\Omega^{1/2}(I_r,0)'.$

$$\lim_{T \to \infty} T^{-1} \left(\begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix}' \hat{\Omega} \begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix} \right)$$

$$\Rightarrow \alpha_{12}' [n(0,\Omega_{11} \otimes ((\mathrm{var}\,X_t\beta))^{-1})]^{-1'} \Omega_{11} [n(0,\Omega_{11} \otimes (\mathrm{var}\,X_t\beta))^{-1})]^{-1} \alpha_{12}$$

$$\Rightarrow \alpha_{12}' (\mathrm{var}(X_t\beta))' \left[\int_0^1 (X(t)\beta)' \mathrm{d}S) \Omega^{1/2}(I_r,0)' \right]^{-1'} (I_r,0) \Omega (I_r,0)'$$

$$\times \left[\int_0^1 (X(t)\beta)' \mathrm{d}S) \Omega^{1/2}(I_r,0)' \right]^{-1} (\mathrm{var}(X_t\beta)) \alpha_{12},$$

$$\lim_{T \to \infty} t_W = \lim_{T \to \infty} \mathrm{tr} \left(\begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix}' \hat{\Omega} \begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix} \right)^{-1}$$

$$\times \hat{\alpha}_{22}' (X_2)'_{-1} M_{((X_1 - X\beta_2)_{-1})} (X_2)_{-1} \hat{\alpha}_{22}$$

$$= \mathrm{tr} \left[\lim_{T \to \infty} T^{-1} \left(\begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix}' \hat{\Omega} \begin{bmatrix} -\hat{\alpha}_{11}^{-1}\hat{\alpha}_{12} \\ I_{k-r} \end{bmatrix} \right) \right]^{-1}$$

$$\times \left[\lim_{T \to \infty} T^{1/2} (\gamma \, T^{1/2}\mu_2')' \beta_\perp' \beta_\perp' \hat{\alpha}_{22} \right]'$$

$$\times \left[\lim_{T \to \infty} T^{-2} (\gamma\, T^{-1/2} \mu_2')' (X_2)_{-1}' M_{(t\, x_1 - x\beta_2)_{-1}} (X_2)_{-1} (\gamma\, T^{-1/2} \mu_2') \right]$$

$$\times \left[\lim_{T \to \infty} T^{1/2} (\gamma\, T^{1/2} \mu_2')' \beta_\perp' \beta_\perp' \hat{\alpha}_{22} \right]$$

$$\Rightarrow \mathrm{tr} \left(\int_0^1 F'\mathrm{d}V \right)' \left(\left(\int_0^1 F'F \right) + \left(\int_0^1 F'\mathrm{d}V \right) \left[\int_0^1 Y(t)'\mathrm{d}V \right]^{-1} \right.$$

$$\left. \times \left[\int_0^1 Y(t)'\mathrm{d}V \right]^{-1'} \left(\int_0^1 F'\mathrm{d}V \right)' \right)^{-1} \left(\int_0^1 F'\mathrm{d}V \right)$$

$$\Rightarrow \mathrm{tr} \left(\int_0^1 W'\mathrm{d}S_1 \right)' \left(\left(\int_0^1 W'W \right) + \left(\int_0^1 W'\mathrm{d}S_1 \right) \left[\int_0^1 Y(t)'\mathrm{d}S_1 \right]^{-1} \right.$$

$$\left. \times \left[\int_0^1 Y(t)'\mathrm{d}S_1 \right]^{-1'} \left(\int_0^1 W'\mathrm{d}S_1 \right)' \right)^{-1} \left(\int_0^1 W'\mathrm{d}S_1 \right).$$

References

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

Boswijk, H.P., 1990, Testing for cointegration in conditional models (Tinbergen Institute, Amsterdam).

Box, G.E.P. and G.C. Tiao, 1977, A canonical analysis of multiple time series, Biometrika 64, 355–365.

Cox, D.R. and D.V. Hinkley, 1974, Theoretical statistics (Chapman and Hall, London).

Davidson, J.E.H., D.F. Hendry, F. Srba, and S. Yeo, 1978, Econometric modelling of the aggregate time-series relationship between consumer's expenditure and income in the United Kingdom, Economic Journal 88, 661–692.

Engle, R.F., 1984, Wald, likelihood ratio and Lagrange multiplier tests in econometrics, Ch.9 in: Z. Griliches and M. Intriligator, eds., Handbook of econometrics, Vol. 2 (North-Holland, Amsterdam).

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

Johansen, S., 1991a, Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models, Econometrica 59, 1551–1581.

Johansen, S., 1991b, The role of the constant term in cointegration analysis of nonstationary variables, Preprint (University of Copenhagen, Copenhagen).

Johansen, S. and K. Juselius, 1990, Maximum likelihood estimation and inference on cointegration - With applications to the demand for money, Oxford Bulletin of Economics and Statistics 52, 169–210.

Kleibergen, F. and H.K. van Dijk, 1993a, On the shape of the likelihood/posterior in cointegration models, Econometric Theory, forthcoming.

Kleibergen, F. and H.K. van Dijk, 1993b, Nonstationarity in GARCH models: A Bayesian analysis, Journal of Applied Econometrics 8, S41–S61.

Kleibergen, F., J.P. Urbain, and H.K. van Dijk, 1993, A cointegration study of aggregate imports using likelihood testing principles, Unpublished manuscript (Tinbergen Institute, Rotterdam).

Magnus, J. and H. Neudecker, 1988, Matrix differential calculus with applications in statistics and economics (Wiley, Chichester).

Phillips, P.C.B., 1989, Partially identified econometric models, Econometric Theory, 181–240.

Phillips, P.C.B., 1991, Optimal inference in cointegrated systems, Econometrica 59, 283–307.

Phillips, P.C.B. and S.N. Durlauf, 1986, Multiple time series regression with integrated processes, Review of Economic Studies 53, 473–495.

Stock, J.H. and M.W. Watson, 1988, Testing for common trends, Journal of the American Statistical Association 83, 1097–1107.

Urbain, J.P., 1992, Exogeneity in error correction models, Doctoral dissertation (Université de Liège, Liège).