Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models

Author(s): Søren Johansen

Source: *Econometrica*, Nov., 1991, Vol. 59, No. 6 (Nov., 1991), pp. 1551-1580

Published by: The Econometric Society

Stable URL: https://www.jstor.org/stable/2938278

Econometrica, Vol. 59, No. 6 (November, 1991), 1551–1580

ESTIMATION AND HYPOTHESIS TESTING OF COINTEGRATION VECTORS IN GAUSSIAN VECTOR AUTOREGRESSIVE MODELS

BY SØREN JOHANSEN

The purpose of this paper is to present the likelihood methods for the analysis of cointegration in VAR models with Gaussian errors, seasonal dummies, and constant terms. We discuss likelihood ratio tests of cointegration rank and find the asymptotic distribution of the test statistics. We characterize the maximum likelihood estimator of the cointegrating relations and formulate tests of structural hypotheses about these relations. We show that the asymptotic distribution of the maximum likelihood estimator is mixed Gaussian. Once a certain eigenvalue problem is solved and the eigenvectors and eigenvalues calculated, one can conduct inference on the cointegrating rank using some nonstandard distributions, and test hypotheses about cointegrating relations using the χ^2 distribution.

KEYWORDS: Cointegration, error correction models, maximum likelihood estimation, likelihood ratio test, Gaussian VAR models.

1. INTRODUCTION AND SUMMARY

A LARGE NUMBER OF PAPERS are devoted to the analysis of the concept of *cointegration* defined first by Granger (1981, 1983), Granger and Weiss (1983), and studied further by Engle and Granger (1987). Under this heading the topic has been studied by Stock (1987), Phillips and Ouliaris (1988), Phillips (1988, 1990), Johansen (1988b), Johansen and Juselius (1990, 1991). The main statistical technique that has been applied is *regression with integrated regressors*, which has been studied by Phillips (1988), Phillips and Park (1988), Park and Phillips (1988, 1989), Phillips and Hansen (1990), Park (1988), and Sims, Stock, and Watson (1990). Similar problems have been studied under the name *common trends* (see Stock and Watson (1988)).

The purpose of this paper is to present some new results on maximum likelihood estimators and likelihood ratio tests for cointegration in Gaussian vector autoregressive models which allow for constant term and seasonal dummies. This brings in the technique of *reduced rank regression* (see Anderson (1951), Velu, Reinsel, and Wichern (1986), Ahn and Reinsel (1990), and Reinsel and Ahn (1990)), as well as the notion of *canonical analysis* (Box and Tiao (1981), Velu, Wichern, and Reinsel (1987), Pena and Box (1987), and the very elegant paper by Tso (1981)). In Johansen (1988b) the likelihood based theory was presented for such a model without constant term and seasonal dummies, but it turns out that the constant plays a crucial role for the interpretation of the model, as well as for the statistical and the probabilistic analysis.

A detailed statistical analysis illustrating the techniques by data on money demand from Denmark and Finland is given in Johansen and Juselius (1990), and the present paper deals mainly with the underlying probability theory that allows one to make asymptotic inference.

The structure of the paper is the following: Section 2 describes the cointegration model and the tests for cointegration rank. The asymptotic distribution of the likelihood ratio test statistic for the hypothesis of r cointegration vectors is given. In Section 3 it is shown that the cointegration model with linear restrictions on the cointegrating relations and the adjustment coefficients allows explicit estimation. The likelihood ratio test statistic of this hypothesis is given. Section 4 gives a simple proof of Granger's representation theorem which clarifies the role of the constant term and gives a condition for the process to be integrated of order 1. In Section 5 the asymptotic distribution of the maximum likelihood estimator for the cointegrating relations is given together with an estimate of its "variance" to be used in constructing Wald tests. The presence of the trend gives rise to some new limit distributions. Section 6 contains a brief discussion of the relation of the present work to the results of Phillips, Stock, and Watson and others, and the appendices contain technical details as well as results for inference concerning smooth hypotheses on the cointegrating relations.

2. THE STATISTICAL ANALYSIS OF THE VAR MODEL FOR COINTEGRATION AND THE TEST FOR COINTEGRATION RANK

Consider a general VAR model with Gaussian errors written in the error correction form

$$(2.1) \qquad \Delta X_t = \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Pi X_{t-k} + \Phi D_t + \mu + \varepsilon_t \qquad (t = 1, \dots, T),$$

where D_t are seasonal dummies orthogonal to the constant term. Further, ε_t $(t = 1, \dots, T)$ are independent p-dimensional Gaussian variables with mean zero and variance matrix Λ. The first k data points $X_{1-k}, \dots, X_0$ are considered fixed and the likelihood function is calculated for given values of these. The parameters $\Gamma_1, \dots, \Gamma_{k-1}$, Φ, μ, and Λ are assumed to vary without restrictions, and we formulate the hypotheses of interest as restrictions on Π.

In this section we analyze the likelihood function conditional on the initial values. There are two reasons for this. Firstly we shall discuss nonstationary processes, for which only the conditional likelihood can be defined, and secondly the conditional likelihood function gives the usual least squares regression estimators in the unrestricted model, and hence gives tractable estimators. When it comes to discussing the properties of the process X_t, as will be necessary for the asymptotic analysis, it is convenient (see Theorem 4.1) to consider some linear combinations of X_t as well as ΔX_t as stationary processes under the conditions stated there. Thus the likelihood function described in this section is the conditional likelihood function for the observations of $X_1, \dots, X_T$ from the process, described in detail in Theorem 4.1, conditional on the initial values, that is, conditional on the first k observations.

Model (2.1) is denoted by H_1 and we formulate the hypothesis of (at most) r cointegration vectors as

$$H_2\colon \Pi = \alpha\beta', \tag{2.2}$$

where β, the cointegrating vectors, and α, the adjustment coefficients, are $p \times r$ matrices. Sometimes we compare models with different numbers of cointegration vectors, and we then use the notation $H_2(r)$.

The purpose of the analysis of this paper is to conduct inference on the number of cointegrating relations as well as the structure of these without imposing a priori structural relations. This is accomplished by fitting the general VAR model (2.1), which is used to describe the variation of the data, and then formulating questions concerning structural economic relations as hypotheses on parameters of the VAR model. These hypotheses are tested using likelihood ratio statistics, and allow the researcher to check interesting economic hypotheses against the data.

It is seen that the parameters α and β are not identified in model H_2, since for any choice of an $r \times r$ matrix ξ, the matrices $\alpha\xi'^{-1}$ and $\beta\xi$ imply the same distribution. What can be determined by the model is the space spanned by β, the cointegration space $\mathrm{sp}(\beta)$, and the space spanned by α, the adjustment space $\mathrm{sp}(\alpha)$. Note that the space spanned by β is the row space of Π, and the adjustment space is the column space of Π.

It turns out that the role of the constant term is crucial for the statistical analysis as well as for the probabilistic analysis. It is proved in Theorem 4.1 that under certain conditions on the parameters the process given by (2.1) is integrated of order 1. In this model the constant term μ can be decomposed into two parts, $\alpha(\alpha'\alpha)^{-1}\alpha'\mu$ which contributes to the intercept in the cointegrating relation (see (4.9)), and $\alpha_\perp(\alpha_\perp'\alpha_\perp)^{-1}\alpha_\perp'\mu$ which determines a linear trend. Here $\alpha_\perp$ is a $p \times (p-r)$ matrix of full rank consisting of vectors orthogonal to the vectors in α. The presence of the linear trend changes the analysis and it is therefore convenient to define a model H_2^* where the * indicates that apart from the restriction imposed under H_2 we also impose the restriction $\mu = \alpha\beta_0$, where the $(r \times 1)$ vector $-\beta_0$ has the interpretation as an intercept in the cointegration relations. In this case clearly $\alpha_\perp'\mu = 0$, and the linear trend is absent.

In order to facilitate the presentation of the main result of this section we first introduce some notation.

Let $Z_{0t} = \Delta X_t$, $Z_{1t} = (\Delta X_{t-1}', \dots, \Delta X_{t-k+1}', D_t', 1)'$, and $Z_{kt} = X_{t-k}$, and let Γ consist of the parameters $(\Gamma_1, \dots, \Gamma_{k-1}, \Phi, \mu)$. Then the model becomes

$$Z_0 = \Gamma Z_{1t} + \alpha\beta' Z_{kt} + \varepsilon_t. \tag{2.3}$$

With this notation define the product moment matrices

$$M_{ij} = T^{-1}\sum_{t=1}^{T} Z_{it}Z_{jt}' \qquad (i, j = 0, 1, k), \tag{2.4}$$

the residuals

$$R_{it} = Z_{it} - M_{i1}M_{11}^{-1}Z_{1t} \qquad (i = 0, k),$$

and the residual sums of squares

$$S_{ij} = M_{ij} - M_{i1}M_{11}^{-1}M_{1j} \qquad (i, j = 0, k). \tag{2.5}$$

The estimate of Γ for fixed values of α, β, and Λ is found to be

$$\hat{\Gamma}(\alpha, \beta) = (M_{01} - \alpha\beta' M_{k1})M_{11}^{-1}. \tag{2.6}$$

Thus the residuals are found by regressing ΔX_t and X_{t-k} on the lagged differences, the dummies and the constant. This gives the likelihood function concentrated with respect to the parameters $\Gamma_1, \dots, \Gamma_{k-1}$, Φ, and μ:

$$\begin{aligned} &L_{\max}^{-2/T}(\alpha, \beta, \Lambda) \\ &\quad = |\Lambda| \exp\left\{ T^{-1} \sum_{t=1}^{T} (R_{0t} - \alpha\beta' R_{kt})' \Lambda^{-1} (R_{0t} - \alpha\beta' R_{kt}) \right\}. \end{aligned} \tag{2.7}$$

This function is easily minimized for fixed β to give

$$\hat{\alpha}(\beta) = S_{0k}\beta(\beta' S_{kk}\beta)^{-1}, \tag{2.8}$$

$$\hat{\Lambda}(\beta) = S_{00} - S_{0k}\beta(\beta' S_{kk}\beta)^{-1}\beta' S_{k0}, \tag{2.9}$$

together with

$$L_{\max}^{-2/T}(\beta) = |S_{00}|\,\left|\beta'(S_{kk} - S_{k0}S_{00}^{-1}S_{0k})\beta\right| / |\beta' S_{kk}\beta|. \tag{2.10}$$

This again is minimized by the choice $\hat{\beta} = (\hat{v}_1, \dots, \hat{v}_r)$, where $\hat{V} = (\hat{v}_1, \dots, \hat{v}_p)$ are the eigenvectors of the equation

$$\left|\lambda S_{kk} - S_{k0}S_{00}^{-1}S_{0k}\right| = 0, \tag{2.11}$$

normed by $\hat{V}' S_{kk} \hat{V} = I$, and ordered by $\hat{\lambda}_1 > \cdots > \hat{\lambda}_p > 0$. The maximized likelihood function is found from

$$L_{\max}^{-2/T}(r) = |S_{00}| \prod_{i=1}^{r} \left(1 - \hat{\lambda}_i\right). \tag{2.12}$$

This procedure is given in Johansen (1988b) for the model without constant term and seasonal dummies, and consists of well known multivariate techniques from the theory of partial canonical correlations and reduced rank regression (see Anderson (1951) and Tso (1981)).

To give an intuition for the above analysis, consider the estimate of Π in the unrestricted VAR model given by $\hat{\Pi} = S_{0k}S_{kk}^{-1}$. Since the hypothesis of cointegration is the hypothesis of reduced rank of Π, it is intuitively reasonable to calculate the eigenvalues of the $\hat{\Pi}$ and check whether they are close to zero. This is the approach of Fountis and Dickey (1989). Another possibility is to calculate singular values, i.e. eigenvalues of $\hat{\Pi}'\hat{\Pi}$, since they are real and positive. It is interesting to see that the maximum likelihood estimation involves solving (2.11) which amounts to calculating the singular values of $S_{00}^{-1/2}S_{0k}S_{kk}^{-1/2} = S_{00}^{-1/2}\hat{\Pi}S_{kk}^{1/2}$, that is, a normalized version of the intuitively reasonable solution. The normalization of the problem given by the likelihood method guarantees a simple asymptotic distribution of the likelihood ratio tests, which only depends on the dimension of the problem considered and not on nuisance parameters. The distributions of the test statistics are nonstandard and have to be tabulated by simulation. They are natural generalizations of the Dickey-Fuller or unit root distributions.

The relation (2.12) gives the maximized likelihood function for all values of r. For $r = p$, $H_2(r) = H_1$, so that the ratio $L_{\max}(r)/L_{\max}(p)$ is the likelihood ratio test statistic of the hypothesis $H_2(r)$ in H_1, and $L_{\max}(r)/L_{\max}(r+1)$ is the likelihood ratio statistic for $H_2(r)$ in $H_2(r+1)$.

The main result that is given here and proved in Appendix B is that the asymptotic distribution of the test statistic has a limit distribution that only depends on the dimension of the problem $(p - r)$, and on whether $\alpha'_\perp \mu = 0$ or not. The relation of this result to that of Stock and Watson (1988) is discussed in Section 6.

THEOREM 2.1: *The likelihood ratio test statistic for hypothesis H_2: $\Pi = \alpha\beta'$ versus H_1 is given by*

$$-2\ln(Q; H_2|H_1) = -T\sum_{i=r+1}^{p}\ln\left(1-\hat{\lambda}_i\right), \tag{2.13}$$

whereas the likelihood ratio test statistic of $H_2(r)$ versus $H_2(r+1)$ is given by

$$-2\ln(Q; r|r+1) = -T\ln\left(1-\hat{\lambda}_{r+1}\right). \tag{2.14}$$

The statistic $-2\ln(Q; H_2|H_1)$ has a limit distribution which, if $\alpha'_\perp \mu \neq 0$, can be expressed in terms of a $(p-r)$-dimensional Brownian motion B with i.i.d. components as

$$\operatorname{tr}\left\{\int (dB)F'\left[\int FF'\,du\right]^{-1}\int F(dB)'\right\}, \tag{2.15}$$

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

$$F_{1i}(t) = B_i(t) - \int B_i(u)\,du \qquad (i = 1, \dots, p-r-1) \tag{2.16}$$

and

$$F_2(t) = t - \tfrac{1}{2}. \tag{2.17}$$

The test statistic $-2\ln(Q; r|r+1)$ is asymptotically distributed as the maximum eigenvalue of the matrix in (2.15).

If in fact $\alpha'_\perp \mu = 0$, then the asymptotic distributions of $-2\ln(Q; H_2|H_1)$ and $-2\ln(Q; r|r+1)$ are given as the trace and the maximal eigenvalue respectively of the matrix in (2.15) *with $F(t) = B(t) - \int B(u)\,du$.*

Here, and in the following, the integrals are all on the unit interval, where the Brownian motions are defined. Note that integrals of the form $\int FF' du$ are ordinary Riemann integrals of continuous functions and the result is a matrix of stochastic variables. The integral $\int F(dB)'$, however, is a matrix of stochastic integrals, defined as L_2 limits of the corresponding Riemann sums.

This section is concluded by pointing out how one can analyze the model H_2^*. First note that if $\mu = \alpha\beta_0$, then

$$\alpha\beta' X_{t-k} + \mu = \alpha\beta' X_{t-k} + \alpha\beta_0 = \alpha\beta^{*\prime} X_{t-k}^*,$$

for $\beta^* = (\beta', \beta_0)'$ and $X_{t-k}^* = (X_{t-k}', 1)'$. In these new variables the model is written

$$\Delta X_t = \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \alpha\beta^{*\prime} X_{t-k}^* + \Phi D_t + \varepsilon_t \qquad (t = 1, \ldots, T).$$

The analysis is now performed as above by defining $Z_{0t}^* = \Delta X_t$, $Z_{1t}^* = (\Delta X_{t-1}', \ldots, \Delta X_{t-k+1}', D_t')'$, and $Z_{kt}^* = X_{t-k}^*$, as well as the corresponding moment matrices M_{ij}^* and S_{ij}^* (see (2.4) and (2.5)).

THEOREM 2.2: *Under hypothesis H_2^*: $\Pi = \alpha\beta'$ and $\mu = \alpha\beta_0$, the likelihood ratio statistics $-2\ln(Q; H_2^* | H_1)$ and $-2\ln(Q; H_2^*(r) | H_2^*(r+1))$ are distributed as the trace and maximal eigenvalue respectively of the matrix in* (2.15), *with* $F = (B', 1)'$.

Finally we test the hypothesis H_2^* in H_2 by a likelihood ratio test, i.e., test that the trend is absent under the assumption that there are r cointegrating relations.

THEOREM 2.3: *The asymptotic distribution of the likelihood ratio test $-2\ln(Q; H_2^* | H_2)$ for the hypothesis H_2^* given the hypothesis H_2, i.e. $\alpha_\perp' \mu = 0$, when there are r cointegration vectors, is asymptotically distributed as χ^2 with $p - r$ degrees of freedom.*

The distributions derived in this section have been tabulated by simulation in Johansen and Juselius (1990) ($p - r = 1, \ldots, 5$) in connection with an application of the methods to money demand in Denmark and Finland. The tables have been extended ($p - r = 1, \ldots, 10$) by Osterwald-Lenum (1992); see also Reinsel and Ahn (1990).

3. THE TEST OF HYPOTHESES ON THE COINTEGRATING RELATIONS AND THE ADJUSTMENT COEFFICIENTS

The purpose of fitting the VAR model and determining the cointegrating rank is that one gets the opportunity to formulate and test interesting hypotheses about the cointegrating relations and their adjustment coefficients.

Since the parameters α and β are not identified we can only test restrictions that cannot be satisfied by normalization.

We consider here in detail a simple but important model for linear restrictions of the cointegrating space and the adjustment space that allows explicit maximum likelihood estimation:

$$H_3\colon \beta = H\varphi \quad \text{and} \quad \alpha = A\psi. \tag{3.1}$$

Note that H_3 is a submodel of H_2. The likelihood ratio test of the restrictions H_3 in the model H_2 will be discussed below. There are of course many other possible hypotheses on the cointegrating relations but the ones chosen here are simple to analyze, and have a wide variety of applications; see Johansen and Juselius (1990), Hoffman and Rasche (1989), and Kunst and Neusser (1990). Another class of hypotheses of the form $\beta = (H\varphi, \psi)$ can be solved with similar methods (see Johansen and Juselius (1991), and Mosconi and Giannini (1992)). For more general hypotheses on β of the form $h(\beta) = 0$ one can of course not prove the existence and uniqueness of the maximum likelihood estimator, but such hypotheses can be tested by likelihood ratio or Wald tests using the asymptotic distribution of $\hat{\beta}$ derived in Appendix C.

Under hypothesis H_3 we transform the matrices S_{ij} some more. Together with A $(p \times m)$ we consider a $(p \times (p-m))$ matrix $B = A_\perp$ of full rank, such that $B'A = 0$, and introduce the notation

$$S_{hh.b} = H'S_{kk}H - H'S_{k0}B(B'S_{00}B)^{-1}B'S_{0k}H,$$

$$S_{aa.b} = A'S_{00}A - A'S_{00}B(B'S_{00}B)^{-1}B'S_{00}A,$$

and similarly for $S_{ha.b}, S_{ah.b}, \Lambda_{ab}, S_{bb}$, etc.

THEOREM 3.1: *Under hypothesis H_3: $\beta = H\varphi$ and $\alpha = A\varphi$ where H is $p \times s$ and A is $p \times m$, the maximum likelihood estimators are found as follows: First solve*

$$\left|\lambda S_{hh.b} - S_{ha.b}S_{aa.b}^{-1}S_{ah.b}\right| = 0, \tag{3.2}$$

for eigenvalues $\hat{\lambda}_1 > \cdots > \hat{\lambda}_s > 0$, and eigenvectors $\hat{v}_1, \ldots, \hat{v}_s$. Then

$$\hat{\beta} = H(\hat{v}_1, \ldots, \hat{v}_r), \tag{3.3}$$

$$\hat{\alpha} = A(A'A)^{-1}S_{ak.b}\hat{\beta}. \tag{3.4}$$

The estimate of Λ is found from

$$\hat{\Lambda}_{bb} = S_{bb}, \tag{3.5}$$

$$\hat{\Lambda}_{ab} = S_{ab} - A'\hat{\alpha}\hat{\beta}'S_{kb}, \tag{3.6}$$

$$\hat{\Lambda}_{aa.b} = S_{aa} - A'\hat{\alpha}\hat{\alpha}'A. \tag{3.7}$$

The estimate for Γ is found from (2.6) *and the maximized likelihood function is*

$$L_{\max}^{-2/T} = |S_{00}| \prod_{i=1}^{r} \left(1 - \hat{\lambda}_i\right). \tag{3.8}$$

THEOREM 3.2: *The likelihood ratio test statistic of the restriction $\beta = H\varphi$ and $\alpha = A\psi$ versus H_2 is given by*

$$-2\ln\left(Q; H_3 | H_2\right) = T \sum_{i=1}^{r} \ln\left\{\left(1 - \hat{\lambda}_i(H_3)\right) / \left(1 - \hat{\lambda}_i(H_2)\right)\right\}, \tag{3.9}$$

which is asymptotically χ^2 distributed with $(p-m)r + (p-s)r$ degrees of freedom.

The proofs of these results are given in Appendix B and C.

4. GRANGER'S REPRESENTATION THEOREM

When we want to investigate the distributional properties of the estimators and the test statistics we have to make more assumptions about the process in order to rule out various types of nonstationarity. The basic assumption is that for the characteristic polynomial derived from model (2.1),

$$\Pi(z) = (1-z)I - \sum_{i=1}^{k-1} \Gamma_i(1-z)z^i - \Pi z^k, \tag{4.1}$$

it holds that $|\Pi(z)| = 0$ implies that either $|z| > 1$ or $z = 1$. This guarantees that the nonstationarity of X_t can be removed by differencing.

If Π has full rank it is well known that under the above condition the equations (2.1) determine X_t as a stationary process provided the initial values are given their invariant distribution. If we start with a doubly infinite sequence $\{\varepsilon_t\}$, we can represent the initial values and hence the whole process as $X_t = A^{-1}(L)(\varepsilon_t + \mu + \Phi D_t)$, that is, as an infinite linear combination of the ε_t's (see Anderson (1971, p. 170)).

If Π has reduced rank we want to prove that some linear combinations of X_t have stationary distributions for a suitable choice of initial distribution, whereas others are nonstationary.

The model defined by (2.1) is rewritten as

$$\Pi(L)X_t = -\Pi X_t + \Psi(L)\,\Delta X_t = \varepsilon_t + \mu + \Phi D_t \qquad (t = 1, \dots, T), \tag{4.2}$$

where we have introduced $\Psi(L) = (\Pi(L) - \Pi(1))/(1-L)$. Note that $-\Pi = \Pi(1)$ is the value of $\Pi(z)$ for $z = 1$, and that $-\Psi = -\Psi(1)$ is the derivative of $\Pi(z)$ for $z = 1$.

The result that we want to prove is the fundamental result about error correction models of order 1 and their structure. The basic result is due to Granger (1983) (see also Engle and Granger (1987) or Johansen (1988a)). We give a very simple proof here. In addition we provide an explicit condition for the process to be integrated of order 1 (see (4.4) below) and we clarify the role of the constant term.

THEOREM 4.1: (Granger's Representation Theorem). *Let the process X_t satisfy the equation* (4.2) *for* $t = 1, 2, \ldots,$ *and let*

$$\Pi = \alpha\beta' \tag{4.3}$$

for α and β of dimension $p \times r$ and rank r, and let

$$\alpha'_\perp \Psi \beta_\perp \tag{4.4}$$

have full rank $p - r$. We define

$$C = \beta_\perp (\alpha'_\perp \Psi \beta_\perp)^{-1} \alpha'_\perp . \tag{4.5}$$

Then ΔX_t and $\beta' X_t$ can be given initial distributions, such that

(4.6) ΔX_t *is stationary*,

(4.7) $\beta' X_t$ *is stationary*,

(4.8) X_t *is nonstationary, with linear trend* $\tau t = C\mu t$.

Further

$$E\beta' X_t = -(\alpha'\alpha)^{-1}\alpha'\mu + (\alpha'\alpha)^{-1}\alpha'\Psi\beta_\perp(\alpha'_\perp \Psi \beta_\perp)^{-1}\alpha'_\perp \mu, \tag{4.9}$$

$$E\Delta X_t = \tau, \tag{4.10}$$

apart from terms involving the seasonal dummies. If $\alpha'_\perp \mu = 0$, then $\tau = 0$ and the linear trend disappears. If the initial distributions are expressed in terms of the doubly infinite sequence $\{\varepsilon_t\}$, then ΔX_t has a representation

$$\Delta X_t = C(L)(\varepsilon_t + \mu + \Phi D_t)$$

with $C(1) = C$. For $C_1(L) = (C(L) - C(1))/(1 - L)$, so that $C(L) = C(1) + (1 - L)C_1(L)$, the process X_t has the representation

$$X_t = X_0 + C \sum_{i=1}^{t} \varepsilon_i + \tau t + C(L)\Phi \sum_{i=1}^{t} D_i + S_t - S_0, \tag{4.11}$$

where $S_t = C_1(L)\varepsilon_t$, and $\beta' X_0 = \beta' S_0$.

PROOF: If we multiply the equation (4.2) by α' and $\alpha'_\perp$, respectively, we get the equations

$$-\alpha'\alpha\beta' X_t + \alpha'\Psi(L)\,\Delta X_t = \alpha'(\varepsilon_t + \mu + \Phi D_t),$$

$$\alpha'_\perp \Psi(L)\,\Delta X_t = \alpha'_\perp(\varepsilon_t + \mu + \Phi D_t).$$

To discuss the properties of the process X_t we solve the equations for X_t and express it in terms of the ε_t's. The problem is of course that since Π is singular the system is not invertible, and we therefore introduce the new variables $Z_t = (\beta'\beta)^{-1}\beta' X_t$ and $Y_t = (\beta'_\perp \beta_\perp)^{-1}\beta'_\perp \Delta X_t$, where $\beta_\perp$ is a $p \times (p - r)$ matrix of full rank such that $\beta'\beta_\perp = 0$. It is also convenient with the notation $\bar{a} =$

$a(a'a)^{-1}$ for any matrix a of full rank. With this notation note that $a'\bar{a} = I$ and $\bar{a}a' = a\bar{a}' = a(a'a)^{-1}a'$ which is just the projection onto the space spanned by the columns of a. The process ΔX_t can be recovered from Z_t and Y_t:

$$\Delta X_t = \left(\beta_\perp \bar{\beta}_\perp' + \beta\bar{\beta}'\right)\Delta X_t = \beta_\perp Y_t + \beta\,\Delta Z_t.$$

This gives the equations for Z_t and Y_t

$$-\alpha'\alpha\beta'\beta Z_t + \alpha'\Psi(L)\beta\,\Delta Z_t + \alpha'\Psi(L)\beta_\perp Y_t = \alpha'(\varepsilon_t + \mu + \Phi D_t), \tag{4.12}$$

$$\alpha_\perp'\Psi(L)\beta\,\Delta Z_t + \alpha_\perp'\Psi(L)\beta_\perp Y_t = \alpha_\perp'(\varepsilon_t + \mu + \Phi D_t). \tag{4.13}$$

The idea of the proof is now to show that the equations for the processes Z_t and Y_t constitute an invertible autoregressive model.

We write the equations for Z_t and Y_t as

$$\tilde{A}(L)(Z_t', Y_t')' = (\alpha, \alpha_\perp)'(\varepsilon_t + \mu + \Phi D_t)$$

with

$$\tilde{A}(z) = \begin{pmatrix} -\alpha'\alpha\beta'\beta + \alpha'\Psi(z)\beta(1-z) & \alpha'\Psi(z)\beta_\perp \\ \alpha_\perp'\Psi(z)\beta(1-z) & \alpha_\perp'\Psi(z)\beta_\perp \end{pmatrix}.$$

For $z = 1$ this has determinant

$$\left|\tilde{A}(1)\right| = |\alpha'\alpha|\;|\beta'\beta|\;|\alpha_\perp'\Psi\beta_\perp|,$$

which is nonzero by assumptions (4.3) and (4.4). Hence $z = 1$ is not a root. For $z \neq 1$ we use the representation

$$\tilde{A}(z) = (\alpha, \alpha_\perp)'\Pi(z)\left(\beta, \beta_\perp(1-z)^{-1}\right),$$

which gives the determinant as

$$\left|\tilde{A}(z)\right| = |(\alpha, \alpha_\perp)|\;|\Pi(z)|\;|(\beta, \beta_\perp)|(1-z)^{-(p-r)}.$$

This shows that all roots of $|\tilde{A}(z)| = 0$ are outside the unit disk, by the assumption about $\Pi(z)$; see (4.1).

It follows that the system defined by (4.12) and (4.13) is invertible, and hence that Z_t and Y_t can be given such initial distributions that they become stationary. Hence also $\Delta X_t = \beta_\perp Y_t + \beta\,\Delta Z_t$ is stationary apart from the contribution from the centered dummies. This proves (4.6) and (4.7).

If these initial distributions are expressed in terms of a doubly infinite sequence $\{\varepsilon_t\}$, then the process (Z_t', Y_t') has the representation

$$(Z_t', Y_t')' = \tilde{A}(L)^{-1}(\alpha, \alpha_\perp)'(\varepsilon_t + \mu + \Phi D_t).$$

The expectation of Z_t and Y_t can be found from $\tilde{A}(1)^{-1}(\alpha, \alpha_\perp)'\mu$. From the representation of the processes Z_t and Y_t we get a representation of ΔX_t by multiplying by the matrix $(\beta\Delta, \beta_\perp)$. Hence

$$C(L) = (\beta\Delta, \beta_\perp)\tilde{A}(L)^{-1}(\alpha, \alpha_\perp)'.$$

For $L = 1$ we get (4.5). By summation of ΔX_t we find that X_t has the representation (4.11) and hence contains the nonstationary component $C\Sigma_{s=1}^{t}\varepsilon_s$ together with a linear trend $\tau t = C\mu t$, which proves (4.8) and completes the proof of Theorem 4.1.

Note that μ enters the linear trend only through $\alpha_{\perp}'\mu$, and that the linear trend τ is contained in the span of $\beta_{\perp}$, and hence cancels if we consider the components $\beta' X_t$. The seasonal dummies are so constructed that they remain bounded even after summation over t and hence do not contribute to the linear trend.

Strictly speaking the processes ΔX_t and $\beta' X_{t-k}$ equal a stationary process plus the term involving the seasonal dummies, but we shall call such a process stationary. One can also make the seasonal dummies stationary by initial random assignment of a season.

Relations (4.3) and (4.5) display an interesting symmetry between the singularity of the "impact" matrix Π for the autoregressive representation and the singularity of the "impact" matrix for the moving average representation. The null space for C' is the range space for Π and the range space for Π' is the null space for C. It is this symmetry that allows the results for $I(1)$ process to be relatively simple.

Note also that the condition (4.4) is what is needed for the process to be integrated of order 1. If this matrix has reduced rank, the process X_t will be integrated of higher order then 1. Thus a theory of $I(2)$ processes in the context of a VAR model can be based on the reduced rank of the first two matrices in the expansion of $\Pi(z)$ at the point $z = 1$. The mathematical and statistical theory for such processes has been worked out in Johansen (1988a, 1990, 1991). It is of course easy at this point to give a representation of the process for the model with a linear term added to (2.1). Such a term gives rise to a quadratic trend in general. The asymptotic analysis of such a model, however, becomes somewhat more complicated because there are more directions that will require special normalization.

5. ASYMPTOTIC PROPERTIES OF THE ESTIMATORS UNDER THE ASSUMPTION OF COINTEGRATION AND LINEAR RESTRICTIONS ON α AND β

The asymptotic properties of the estimators are here given under the hypothesis H_3 where restrictions are imposed on both α and β. The maximum likelihood estimators under H_3 are denoted by $\hat{}$. The results are corollaries of Theorem C.1 which gives the asymptotic distribution of the estimator under a smooth hypothesis on the parameters. An important result is that inference concerning β can be conducted as if the other parameters are fixed, and vice versa. Thus inference concerning α and the short term dynamics represented by $\Gamma_1, \dots, \Gamma_{k-1}$ follow the usual results for stationary processes. We therefore concentrate on the results for β.

The first result concerns $\hat{\beta}$ normalized by c, i.e. $\hat{\beta}_c = \hat{\beta}(c'\hat{\beta})^{-1}$, such that $\beta' c$ has full rank. The results will be expressed in the natural coordinate system in $\operatorname{sp}(H)$. We choose, apart from $\beta \in \operatorname{sp}(H)$, the projection of τ onto $\operatorname{sp}(H)$,

$\tau_H = P_H\tau$, which is also orthogonal to β, since $\beta'\tau_H = \beta' P_H \tau = \beta'\tau = 0$. Next supplement with $s-r-1$ vectors $\gamma_H = (\gamma_{H1}, \dots, \gamma_{Hs-r-1}) \in \mathrm{sp}(H)$, such that $(\beta, \gamma_H, \tau_H)$ consists of s mutually orthogonal vectors spanning $\mathrm{sp}(H)$. Similarly let $\gamma(p \times (p-r-1))$ be chosen orthogonal to (β, τ) such that (β, τ, γ) span R^p.

THEOREM 5.1: *Suppose hypothesis H_3: $\alpha = A\psi$ and $\beta = H\varphi$ is satisfied. If $\tau_H \neq 0$, the limit distribution of $T(\hat\beta_c - \beta_c)$ is given by*

$$(5.1) \qquad (I - \beta_c c')\gamma_H \left(\gamma_H'\gamma \int G_{1.2}G_{1.2}'\,du\,\gamma'\gamma_H\right)^{-1} \gamma_H'\gamma \int G_{1.2}(dV_\alpha)'(c'\beta)^{-1}$$

where

$$(5.2) \qquad V_\alpha = (\alpha'\Lambda^{-1}\alpha)^{-1}\alpha'\Lambda^{-1}W$$

is independent of $G' = (G_1', G_2)$, defined as

$$(5.3) \qquad G_1(t) = \bar\gamma' C\left(W(t) - \int W(u)\,du\right) \qquad (t \in [0,1]),$$

$$(5.4) \qquad G_2(t) = t - \tfrac{1}{2} \qquad (t \in [0,1]),$$

$$G_{1.2} = G_1 - \left(\int G_1 G_2\,du\right)\left(\int G_2 G_2\,du\right)^{-1} G_2,$$

that is, the Brownian motion corrected for level and trend. The asymptotic conditional variance is

$$(5.5) \qquad (I - \beta_c c')\gamma_H \left(\gamma_H'\gamma \int G_{1.2}G_{1.2}'\,du\,\gamma'\gamma_H\right)^{-1} \gamma_H'(I - c\beta_c') \otimes (c'\Pi'\Lambda^{-1}\Pi c)^{-1},$$

which is consistently estimated by

$$(5.6) \qquad T\left(I - \hat\beta_c c'\right)H\hat v\hat v' H'\left(I - c\hat\beta_c'\right) \otimes \left(c'\hat\Pi'\hat\Lambda^{-1}\hat\Pi c\right)^{-1},$$

where $\hat v = (\hat v_{r+1}, \dots, \hat v_p)$; see (3.2). *One can replace the matrix $\hat v\hat v'$ by S_{hh}^{-1} in* (5.6), *and apply the identity*

$$c'\hat\Pi'\hat\Lambda^{-1}\hat\Pi c = c'\hat\beta\left(\mathrm{diag}\left\{\hat\lambda_1^{-1} - 1, \dots, \hat\lambda_r^{-1} - 1\right\}\right)^{-1}\hat\beta' c.$$

If $\tau_H = 0$ then (5.1) *and* (5.5) *holds with $\gamma_H(p \times (p-s))$ orthogonal to β such that (β, γ_H) span* $\mathrm{sp}(H)$, *and $G_{1.2}$ replaced by G_1.*

The proof is given in Appendix C. The stochastic integrals are all taken on the unit interval. Since the constant term is included in the regressors the process X_t is corrected for its mean in the preliminary regressions. This is seen to be reflected in the asymptotics by subtracting $\int W(u)\,du$ and $\frac{1}{2} = \int u\,du$. Since the process contains a linear trend only $p-r-1$ components $(\gamma' C)$ of the process W enter the result. The trend is described by defining the last component of G by t.

Note that the limiting distribution for fixed G is Gaussian with mean zero and variance

$$(5.7)\qquad (I-\beta_c c')\gamma_H\left(\gamma_H'\gamma\int G_{1.2}G_{1.2}'\,du\,\gamma'\gamma_H\right)^{-1}\gamma_H'(I-c\beta_c')$$
$$\otimes(c'\Pi'\Lambda^{-1}\Pi c)^{-1},$$

which we call the *limiting conditional variance*. Thus the limiting distribution of $T(\hat{\beta}_c-\beta_c)$ is a mixture of Gaussian distributions. See Jeganathan (1988) for a general theory of locally asymptotically mixed normal models.

The result shows that if the β's are normalized by c one can find the limiting distribution of any of the coefficients and hence of any smooth function of the coefficients. Note, however, that if we were interested in the linear combination $T\tau_H'(\hat{\beta}_c-\beta_c)$ then the limiting distribution degenerates to zero, since $\tau_H'(I-\beta(c'\beta)^{-1}c')\gamma_H=0$. A different normalization by $T^{3/2}$ is needed in this case. The result can be determined from the proof of Theorem 5.1 in Appendix C, but will not be explicitly formulated here.

Without proof we give the corresponding result for model H_3^*, i.e. when $\mu=\alpha\beta_0$, $\beta=H\varphi$, and $\alpha=A\psi$. Introduce γ_H such that β and γ_H span H and define $\gamma_H^*=(\gamma_H',0)'$ and $\xi=(0,1)'$. The normalization by the $p\times r$ matrix c is now done as follows: $\beta_c^*=((\beta'c)^{-1}\beta',(\beta'c)^{-1}\beta_0)'=(\beta_c',\beta_{0c})'$.

THEOREM 5.2: *Suppose hypothesis H_3^*: $\alpha=A\varphi$, $\beta=H\varphi$, and $\mu=\alpha\beta_0$ is satisfied; then the limit distribution of $T(\hat{\beta}_c-\beta_c)$ is given by* (5.1) *with $G_{1.2}$ replaced by G_1.*

A consistent estimate of the asymptotic conditional variance is given by

$$(5.8)\qquad T\left(I-\hat{\beta}_c c'\right)(H,0)\hat{v}^*\hat{v}^{*\prime}(H,0)'\left(I-c\hat{\beta}_c'\right)\otimes\left(c'\hat{\Pi}'\hat{\Lambda}^{-1}\hat{\Pi}c\right)^{-1}.$$

The estimator of the constant term behaves differently: Let $G_1^*=\bar{\gamma}'CW$, $G_2^*(t)=1$, and

$$G_{2.1}^*=G_2^*-\int G_2^*G_1^{*\prime}\,du\left(\int G_1^*G_1^{*\prime}\,du\right)^{-1}G_1^*;$$

then

$$(5.9)\qquad T^{\frac{1}{2}}\left(\hat{\beta}_{0c}'-\beta_{0c}'\right)\xrightarrow{w}\left(\int G_{2.1}^*G_{2.1}^{*\prime}\,du\right)^{-1}\int G_{2.1}^*(dV_\alpha)'(c'\beta)^{-1},$$

and a consistent estimator for the asymptotic conditional variance is

$$(5.10)\qquad T(0,1)\hat{v}^*\hat{v}^{*\prime}(0,1)'\otimes\left(c'\hat{\Pi}'\hat{\Lambda}^{-1}\hat{\Pi}c\right)^{-1}.$$

Here $(\hat{\beta}^*,\hat{v}^*)$ are the eigenvectors from (3.2) *with $S_{ij.b}$ replaced by $S_{ij.b}^*$.*

Note that $T(\hat{\beta}_c - \beta_c)$ has the same limit distribution as that given by (5.1) for $\tau_H = 0$.

As an example of an application of the results given above consider the following simple situation where $r = 1$, and where we want to test a linear constraint $K'\beta = 0$ on the cointegration relation $\beta' = (\beta_1, \dots, \beta_p)$.

We formulate the result as a Corollary.

COROLLARY 5.3: *If only* 1 *cointegration vector* β *is present* $(r = 1)$, *and if we want to test the hypothesis* $K'\beta = 0$, *then the test statistic*

$$T\left(K'\hat{\beta}\right)^2\left(\left(\hat{\lambda}_1^{-1} - 1\right)\left(K'\hat{v}\hat{v}'K\right)\right)^{-1} \tag{5.11}$$

is asymptotically χ^2 *with* 1 *degree of freedom. Here* $\hat{\lambda}_1$ *is the maximal eigenvalue and* $\hat{\beta}$ *the corresponding eigenvector of the equation*

$$\left|\lambda S_{kk} - S_{k0}S_{00}^{-1}S_{0k}\right| = 0.$$

The remaining eigenvectors form $\hat{v}$. *A similar result holds for the model with no trend.*

Note that the test statistic (5.11) is very easy to calculate once the basic eigenvalue problem (2.11) has been solved. After having picked out the eigenvector that estimates the cointegrating relation one can apply the remaining eigenvectors to estimate the "variance" of the coefficients of the cointegrating relations.

Thus if there is only one cointegration vector $\hat{\beta}$ one can think of the matrix $(\hat{\lambda}_1^{-1} - 1)\hat{v}\hat{v}'/T$ as giving an estimate of the asymptotic "variance" of $\hat{\beta}$.

If we want to derive a confidence interval for the parameter $\rho = \beta_2/\beta_1$ we define $K' = (\rho_0, -1, 0, \dots, 0)$, such that $K'\beta = \rho_0\beta_1 - \beta_2$, which is zero if $\rho = \rho_0$. Theorem 5.1 yields the result that

$$\omega(\rho_0) = T\left(\rho_0\hat{\beta}_1 - \hat{\beta}_2\right)^2\left\{\left(\hat{\lambda}_1^{-1} - 1\right)\sum_{j=2}^{p}\left(\rho_0\hat{v}_{1j} - \hat{v}_{2j}\right)^2\right\}^{-1}$$

is asymptotically $\chi^2(1)$ if $\rho = \rho_0$. Then the set

$$\left\{\rho \mid |\omega(\rho)| < \chi_{1-\varepsilon}^2\right\}$$

will be an asymptotic $1 - \varepsilon$ confidence set for the parameter ρ, where $\chi_{1-\varepsilon}^2$ is the $1 - \varepsilon$ quantile in the distribution of $\chi^2(1)$. A simpler interval can be obtained by inserting the estimate $\hat{\rho} = \hat{\beta}_2/\hat{\beta}_1$ for ρ_0 in the denominator which gives the interval

$$\hat{\rho}\left(1 \pm \chi_{1-\varepsilon}\left\{\left(\hat{\lambda}_1^{-1} - 1\right)\sum_{j=2}^{p}\left(\hat{v}_{1j}/\hat{\beta}_1 - \hat{v}_{2j}/\hat{\beta}_2\right)^2\right\}^{\frac{1}{2}} T^{-\frac{1}{2}}\right). \tag{5.12}$$

6. DISCUSSION

This paper addresses three issues: first the problem of finding the number of cointegrating relations in nonstationary data, next the problem of estimating the cointegrating relations, and finally that of testing interesting economic hypotheses about their structure. The approach is model based in the sense that we assume that a VAR model describes the data adequately, but no economic structure is imposed on the model in the initial analysis. The VAR model is analyzed using likelihood methods in order to answer the above problems.

The method has the advantage that, once the eigenvalue problem (2.11) is solved, the inference can be based entirely on the eigenvalues and eigenvectors found. The successive tests for the rank are all based on the eigenvalues from (2.11). For any value r of the cointegrating rank the estimate of the cointegrating relations is the subset of the eigenvectors corresponding to the r largest eigenvalues. Finally the remaining eigenvectors corresponding to $p-r$ smallest eigenvalues contain information about the "variance" of the estimators. In view of this summary let us now consider some of the methods that have been proposed before.

We first consider the estimation of β, assuming that the cointegrating rank is known. The original method of Engle and Granger for estimating the long-run parameters consisted of regressing some of the variables on the others. This gives consistent estimators as shown by Stock (1987), but the asymptotic distribution theory is complicated, which makes inference on structural hypotheses difficult (see Phillips (1990) for a discussion of these problems). Very briefly one can say that the simple regression estimator has a limiting distribution that is composed of a mixed Gaussian distribution, a unit root distribution, and a constant. One can get rid of the constant by including the lags, and one can get rid of the unit root distribution component by analyzing the full system rather than single equations.

A number of other methods have been proposed for estimating cointegration relations: Stock and Watson (1988) have suggested the smallest principal components of $\Sigma_t(X_t-\bar{X})(X_t-\bar{X})'$ as estimator of the cointegrating relations, and the orthogonal complement as estimator of the common trends. Bossaerts (1988) has suggested canonical variates between X_t and X_{t-1}, and there are a number of single equation methods. Common to all these methods is that they give asymptotic inference for the cointegrating relations that have the same problem with the asymptotic inference as indicated above for the regression estimator. A simulation of a number of these methods has been performed by Gonzalo (1989). He finds, not surprisingly, that if the data are generated by an error correction model, like the one we have analyzed here, the likelihood methods, which derive methods by analyzing the model, have a better performance.

Other methods give mixed Gaussian limit distributions. Phillips (1988) has suggested a nonparametric spectral regression method which permits the estimation of long-run equilibrium relationships in the frequency domain. Phillips (1990) contains a discussion of maximum likelihood estimation in a number of models, including the VAR. Park (1988) and Phillips and Hansen (1990) suggest a regression estimator where the regressors are corrected using a spectral estimate of the long-run variance matrix. Finally Engle and Yoo (1989) have suggested a three stage estimator for the error correction model, which starts with the original regression estimator, calculates the remaining parameters by OLS, and then performs one step in a Newton-Raphson algorithm in order to approach the maximum of the likelihood function. This estimator is asymptotically equivalent to the maximum likelihood estimator in the VAR context.

Next consider the problem of determining the cointegration rank r. A systematic approach to finding the cointegration rank is proposed by Stock and Watson (1988). They determine the rank of the cointegrating space by first estimating β by a given number of principal components, then filtering the common trends $\beta'_\perp \Delta X_t$ by fitting an autoregressive process, and finally regressing the residuals on the summed residuals. The estimated coefficient matrix is then investigated for unit roots. The procedure is repeated until the correct number of cointegrating relations is found.

In order to facilitate the comparison between their method and the method derived from the likelihood function in the VAR model, we present in an informal manner their calculations in our notation. The principle components are almost the same as the smallest eigenvectors in S_{kk}, and the subsequent fitting of an autoregressive model to $\Delta\beta'_\perp X_t$ is similar to the autoregressions performed in this paper. Thus the final matrix that they analyze is analogous to $\beta'_\perp S_{0k}\beta_\perp(\beta'_\perp S_{kk}\beta_\perp)^{-1}$. An investigation of eigenvalues in this matrix is analogous to an investigation of eigenvalues of $\hat{\Pi} = S_{0k}S_{kk}^{-1}$. The limiting distribution of the matrix is of the form encountered here, that is $(\int BB'\,du)^{-1}\int B(dB)'$ for a $(p-r)$-dimensional Brownian motion B. Thus many of the calculations are similar to those based on the likelihood methods for the error correction model, but the setup is different, in that the present approach is model dependent through the use of the VAR model and the likelihood analysis.

Throughout this paper we have assumed the Gaussian distribution in order to be able to analyze the likelihood function, with the purpose of developing new methods that are presumably optimal for this distribution. These methods clearly depend on the VAR model assumptions, and major departures from these assumptions would require new models. It would, however, be interesting to see how robust the methods derived are to minor departures from the assumptions.

The assumption of a Gaussian distribution is not so serious, as long as the process $\Sigma_0^t \varepsilon_i$ can be approximated by a Brownian motion, since it is not difficult to see that the asymptotic analysis gives the same results.

The choice of lag length is more important, but simulations indicate that for moderate departures (which would not be detected in the initial statistical analysis) the inference does not seem to change too much; see Gonzalo (1989).

It is the advantage of the model based inference presented here that one can check whether the model fits the data, and one can give a precise formulation of the economic hypotheses to be tested, but the methods are clearly model dependent. If major departures from the model assumptions underlying the present analysis are relevant, a new model should be formulated and analyzed.

It is important to note that for VAR models that allow integration of higher order, the likelihood analysis is more complicated. It turns out, however, that the present methods can be applied to processes that are $I(2)$ with only minor modifications; see Johansen (1991).

Institute of Mathematical Statistics, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark

Manuscript received January, 1989; final revision received March, 1991.

APPENDIX A. Some Technical Results

Since we have proved in Theorem 4.1 that under certain conditions ΔX_t and $\beta' X_t$ can be considered stationary, it follows that the stochastic components of $Z_{1t}' = (\Delta X_{t-1}', \ldots, \Delta X_{t-k-1}', D_t', 1)$ are stationary. Let $\Sigma_{kk}^t = \mathrm{Var}(X_{t-k} | Z_{1t})$. Since the process X_{t-k} is nonstationary, this variance clearly depends on t, but since $\beta' X_{t-k}$ is stationary, then $\beta' \Sigma_{kk}^t \beta$ does not depend on t. We shall indicate this by leaving out the dependence on t and defining

$$\mathrm{Var}\begin{pmatrix} \Delta X_t \\ \beta' X_{t-k} \end{pmatrix} \cdot | Z_{1t}\Bigg) = \begin{pmatrix} \Sigma_{00} & \Sigma_{0k}\beta \\ \beta'\Sigma_{k0} & \beta'\Sigma_{kk}\beta \end{pmatrix}.$$

The first result concerns the relations between these variance-covariance matrices and the parameters α and β in the model H_2.

Lemma A.1: *The following relations hold*:

$$\Sigma_{00} = \alpha\beta'\Sigma_{k0} + \Lambda, \tag{A.1}$$

$$\Sigma_{0k}\beta = \alpha\beta'\Sigma_{kk}\beta, \tag{A.2}$$

and hence

$$\Sigma_{00} = \alpha(\beta'\Sigma_{kk}\beta)\alpha' + \Lambda. \tag{A.3}$$

These relations imply that

$$\left(\alpha'\Sigma_{00}^{-1}\alpha\right)^{-1}\alpha'\Sigma_{00}^{-1} = (\alpha'\Lambda^{-1}\alpha)^{-1}\alpha'\Lambda^{-1}, \tag{A.4}$$

$$\begin{aligned} \Sigma_{00}^{-1} - \Sigma_{00}^{-1}\alpha\left(\alpha'\Sigma_{00}^{-1}\alpha\right)^{-1}\alpha'\Sigma_{00}^{-1} &= \alpha_\perp(\alpha_\perp'\Sigma_{00}\alpha_\perp)^{-1}\alpha_\perp' \\ &= \alpha_\perp(\alpha_\perp'\Lambda\alpha_\perp)^{-1}\alpha_\perp' \\ &= \Lambda^{-1} - \Lambda^{-1}\alpha(\alpha'\Lambda^{-1}\alpha)^{-1}\alpha'\Lambda^{-1}, \end{aligned} \tag{A.5}$$

$$\beta'\Sigma_{kk}\beta\left(\beta'\Sigma_{k0}\Sigma_{00}^{-1}\Sigma_{0k}\beta\right)^{-1}\beta'\Sigma_{kk}\beta - \beta'\Sigma_{kk}\beta = (\alpha'\Lambda^{-1}\alpha)^{-1}. \tag{A.6}$$

Proof: From equation (2.3),

$$\Delta X_t = \Gamma Z_{1t} + \Pi Z_{kt} + \varepsilon_t,$$

one finds immediately the results (A.1), (A.2), and (A.3). To prove (A.4) multiply first by α from the right, and both sides become the identity; then multiply by $\Sigma_{00}\alpha_\perp = \Lambda\alpha_\perp$ and both sides reduce to zero. Since the $p \times p$ matrix $(\alpha, \Lambda\alpha_\perp)$ has full rank the relation (A.4) has been proved.

The first relation in (A.5) is proved the same way, by multiplying by $(\alpha, \Sigma_{00}\alpha_\perp)$. The second equality in (A.5) follows from (A.3) since $\alpha_\perp' \Sigma_{00} = \alpha_\perp' \Lambda$, and the third is proved as the first.

Finally (A.6) is proved by inserting $\alpha = \Sigma_{0k}\beta(\beta'\Sigma_{kk}\beta)^{-1}$ such that (A.6) becomes

$$\left(\alpha'\Sigma_{00}^{-1}\alpha\right)^{-1} - \beta'\Sigma_{kk}\beta = (\alpha'\Lambda^{-1}\alpha)^{-1}.$$

This relation can be proved by multiplying (A.4) by

$$\Sigma_{00}\alpha(\alpha'\alpha)^{-1} = (\alpha\beta'\Sigma_{kk}\beta\alpha' + \Lambda)\alpha(\alpha'\alpha)^{-1}.$$

This completes the proof of Lemma A.1.

The asymptotic properties of the nonstationary process X_t are described by a Brownian motion W in p dimensions on the unit interval. This Brownian motion is the limit of the random walk $\Sigma_{i=0}^t \varepsilon_i$, which appears in the representation (4.11) and can be found by rescaling the time axis and the variables as follows:

$$T^{-\frac{1}{2}} \sum_{i=0}^{[Tu]} \varepsilon_i \xrightarrow{w} W(u) \qquad (u \in [0,1]).$$

From the representation (4.11) it follows that X_t is composed of a random walk, a linear trend, and a stationary process. The asymptotic properties of the process therefore depend on the linear combination of the process we consider. If we consider $\tau' X_t$ it is clear that the process is dominated by the linear trend, whereas if we take vectors γ which are orthogonal to τ and linearly independent of β, then the dominating term is the random walk. Finally if we take the linear combinations $\beta' X_t$, then both the trend and the random walk are multiplied by $\beta' C = 0$, and the process becomes stationary. Thus let $\gamma(p \times (p-r-1))$ be chosen orthogonal to τ and β, such that (β, γ, τ) span all of R^p. The properties of the process are then summarized in the following lemma.

LEMMA A.2: *Let $T \to \infty$ and $u \in [0,1]$;*

(A.7) $$T^{-\frac{1}{2}}\gamma' X_{[Tu]} = T^{-\frac{1}{2}}\gamma' C \sum_{i=0}^{[Tu]} \varepsilon_t + o_P(1) \xrightarrow{w} \gamma' C W(u),$$

(A.8) $$T^{-1}\tau' X_{[Tu]} = T^{-1}\tau' C \sum_{i=0}^{[Tu]} \varepsilon_i + \tau'\tau[Tu]T^{-1} + o_P(1) \xrightarrow{P} \tau'\tau u.$$

Note that the limiting behavior of the nonstationary part of the process is completely described by the matrix C (see (4.5)), the direction $\tau = C\mu$, and the variance matrix of the errors Λ.

Using these results one can describe the asymptotic properties of the product moment matrices and S_{ij} defined in Section 2, which are basic for the properties of the estimators and tests. We also need the product moment matrices when only the intercept is corrected for

$$\tilde{M}_{kk} = T^{-1} \sum_{t=1}^{T} \left(X_{t-k} - \bar{X}_{-k}\right)\left(X_{t-k} - \bar{X}_{-k}\right)',$$

$$\tilde{M}_{k0} = T^{-1} \sum_{t=1}^{T} \left(X_{t-k} - \bar{X}_{-k}\right)(\varepsilon_t - \bar{\varepsilon})'.$$

We do not give the asymptotic results in detail here since the proofs are similar to those in Johansen (1988b), which are based on the results in Phillips and Durlauf (1986), and are simple consequences of the representation (4.11), but we summarize the results in two lemmas.

LEMMA A.3: *For $\tau = C\mu \neq 0$ define $B_T = (\bar{\gamma}, T^{-\frac{1}{2}}\bar{\tau})$ and define $G' = (G_1', G_2)$, as in Theorem* 5.1; *then*

$$T^{-1}B_T' S_{kk} B_T \xrightarrow{w} \int GG'\,du, \tag{A.9}$$

$$B_T'(S_{k0} - S_{kk}\beta\alpha') \xrightarrow{w} \int G(dW)'. \tag{A.10}$$

Finally $\beta' S_{kk}\beta \xrightarrow{\text{a.s.}} \beta'\Sigma_{kk}\beta$, $\beta' S_{k0} \xrightarrow{\text{a.s.}} \beta'\Sigma_{k0}$ and $S_{00} \xrightarrow{\text{a.s.}} \Sigma_{00}$.

If $\tau = 0$, we choose $\gamma(p \times (p-r)) = \beta_\perp$, delete the terms involving τ, and then (A.9) *and* (A.10) *hold with G replaced by G_1. The same results hold if S_{kk} is replaced by $\tilde{M}_{kk}$ and S_{k0} is replaced by $\tilde{M}_{k0}$.*

Under hypothesis H_2^*, which also restricts μ to have no trend component, i.e. $\tau = 0$ or $\mu = \alpha\beta_0$, we get different asymptotic results. Estimates are calculated using the matrices S_{ij}^* (see Section 2), and we define $\beta^{*\prime} = (\beta', \beta_0)$, choose $\xi' = (0,1)$, and $\gamma^{*\prime} = (\beta_\perp', 0)$. As β has full rank the vectors (β^*, γ^*, ξ) are $r + (p-r) + 1 = p+1$ linearly independent vectors spanning R^{p+1}. We then have the following lemma.

LEMMA A.4: *Let $B_T^* = (\bar{\gamma}^*, T^{\frac{1}{2}}\xi)$, and define $G^{*\prime} = (G_1^{*\prime}, G_2^*)$, as in Theorem* 5.2; *then $T^{-1}\Sigma_{t=1}^T \beta' X_{t-k} \xrightarrow{\text{a.s.}} -\beta_0'$, and*

$$T^{-1}B_T^{*\prime} S_{kk}^* B_T^* \xrightarrow{w} \int G^* G^{*\prime}\,du, \tag{A.11}$$

$$B_T^{*\prime}(S_{k0}^* - S_{kk}^*\beta^*\alpha') \xrightarrow{w} \int G^*(dW)'. \tag{A.12}$$

Finally $\beta^{*\prime} S_{kk}^*\beta^* \xrightarrow{\text{a.s.}} \beta'\Sigma_{kk}\beta$, $\beta^{*\prime} S_{k0}^* \xrightarrow{\text{a.s.}} \beta'\Sigma_{k0}$ and $S_{00}^* \xrightarrow{\text{a.s.}} \Sigma_{00}$.

By moving the constant term to the vector X_{t-k}, we no longer correct for the mean in the process W and the added 1 gives an extra dimension to the matrix S_{kk}^*. It is seen that the constant term plays an important role for the formulation of the limiting results, either because it implies a linear trend for the nonstationary part of the process, or because it enters the cointegration vector. The two cases require a different normalization. The seasonal dummies do not play an equally important role once they have been orthogonalized to the constant term. The reason for this is that quantities like $T^{-1}\Sigma_{t=1}^T D_t\,\Delta X_t'$ and $T^{-1}\Sigma_{t=1}^T D_t X_{t-k}'$ remain bounded in probability as $T \to \infty$. The crucial property, which is applied to see this, is that the partial sums of D_t remain bounded.

APPENDIX B. PROOF OF THE RESULTS IN SECTION 2 AND SECTION 3

Proof of Theorem 2.1

The likelihood ratio test statistic of H_2 in H_1 is given in the form

$$-2\ln(Q; H_2|H_1) = -T\sum_{i=r+1}^{p} \ln\left(1 - \hat{\lambda}_i\right), \tag{B.1}$$

where the eigenvalues $\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_p$ are the smallest solutions to the equation

$$\left|\lambda S_{kk} - S_{k0}S_{00}^{-1}S_{0k}\right| = 0; \tag{B.2}$$

see (2.11). Let $S(\lambda) = \lambda S_{kk} - S_{k0}S_{00}^{-1}S_{0k}$. We apply Lemma A.3 to investigate the asymptotic properties of $S(\lambda)$ and apply the fact that the ordered solutions of (B.2) are continuous functions of the coefficient matrices.

As in Lemma A.3 we let γ be orthogonal to β and τ, such that (β, γ, τ) span R^p. We then find from Lemma A.3, that for $B_T = (\bar{\gamma}, T^{-\frac{1}{2}}\bar{\tau})$ and $A_T = (\beta, T^{-\frac{1}{2}}B_T)$ we get

$$(B.3) \qquad |A_T'(S(\lambda))A_T| \xrightarrow{w} \left| \lambda \begin{bmatrix} \beta'\Sigma_{kk}\beta & 0 \\ 0 & \int GG'\,du \end{bmatrix} - \begin{bmatrix} \beta'\Sigma_{k0}\Sigma_{00}^{-1}\Sigma_{0k}\beta & 0 \\ 0 & 0 \end{bmatrix} \right|$$

$$= \left|\lambda\beta'\Sigma_{kk}\beta - \beta'\Sigma_{k0}\Sigma_{00}^{-1}\Sigma_{0k}\beta\right| \left|\lambda \int GG'\,du\right|$$

which has r positive roots and $(p-r)$ zero roots. This shows that the r largest solutions of (B.2) converge to the roots of (B.3) and that the rest converge to zero.

Next consider the decomposition

$$|(\beta, B_T)'S(\lambda)(\beta, B_T)| = |\beta'S(\lambda)\beta| \left|B_T'\{S(\lambda) - S(\lambda)\beta[\beta'S(\lambda)\beta]^{-1}\beta'S(\lambda)\}B_T\right|,$$

and let $T \to \infty$ and $\lambda \to 0$ such that $\rho = T\lambda$ is fixed. From Lemma A.3 it follows that

$$\beta'S(\lambda)\beta = \rho T^{-1}\beta'S_{kk}\beta - \beta'S_{k0}S_{00}^{-1}S_{0k}\beta \xrightarrow{\text{a.s.}} -\beta'\Sigma_{k0}\Sigma_{00}^{-1}\Sigma_{0k}\beta,$$

which shows that in the limit the first factor has no roots. In order to investigate the next factor we note the following consequences of Lemma A.3:

$$B_T'S(\lambda)\beta = \rho T^{-1}B_T'S_{kk}\beta - B_T'S_{k0}S_{00}^{-1}S_{0k}\beta = -B_T'S_{k0}\Sigma_{00}^{-1}\Sigma_{0k}\beta + o_P(1),$$

and

$$B_T'\{S(\lambda) - S(\lambda)\beta[\beta'S(\lambda)\beta]^{-1}\beta'S(\lambda)\}B_T = \rho T^{-1}B_T'S_{kk}B_T - B_T'S_{k0}NS_{0k}B_T + o_P(1),$$

where N is a notation for the matrix

$$\Sigma_{00}^{-1} - \Sigma_{00}^{-1}\Sigma_{0k}\beta\left[\beta'\Sigma_{k0}\Sigma_{00}^{-1}\Sigma_{0k}\beta\right]^{-1}\beta'\Sigma_{k0}\Sigma_{00}^{-1}.$$

By Lemma A.1 this matrix equals $\alpha_\perp(\alpha_\perp'\Lambda\alpha_\perp)^{-1}\alpha_\perp'$, which shows that the limit distribution of $B_T'S_{k0}\alpha_\perp = B_T'(S_{k0} - S_{kk}\beta\alpha')\alpha_\perp$ can be found from Lemma A.3.

The above results imply that the $p-r$ smallest solutions of (B.2) normalized by T converge to those of the equation

$$(B.4) \qquad \left|\rho\int GG'\,du - \int G(dW)'\alpha_\perp(\alpha_\perp'\Lambda\alpha_\perp)^{-1}\alpha_\perp'\int(dW)G'\right| = 0,$$

where $G_1(t) = \bar{\gamma}'C(W(t) - \int W(u)\,du)$ and $G_2(t) = t - \frac{1}{2}$, and $G' = (G_1', G_2)$.

In order to simplify this expression introduce the $(p-r)$-dimensional Brownian motion $U = (\alpha_\perp'\Lambda\alpha_\perp)^{-\frac{1}{2}}\alpha_\perp'W$, which has variance matrix I, and the $(p-r+1)$-dimensional process $\tilde{F}(t) = (U(t)' - \int U'(u)\,du, (t-\frac{1}{2}))'$. We can then write (B.4) as

$$(B.5) \qquad \left|L\left(\rho\int\tilde{F}\tilde{F}'\,du - \int\tilde{F}(dU)'\int(dU)\tilde{F}'\right)L'\right| = 0,$$

where the $(p-r)\times(p-r+1)$ matrix L has the form

$$L = \begin{pmatrix} L_{11} & 0 \\ 0 & 1 \end{pmatrix},$$

and $L_{11} = \bar{\gamma}'\beta_\perp(\alpha_\perp'\Psi\beta_\perp)^{-1}(\alpha_\perp'\Lambda\alpha_\perp)^{\frac{1}{2}}$, applying the representation (4.5) for the matrix C. The process $\tilde{F}$ enters into the integrals with the factor L_{11} which are $p-r-1$ linearly independent combinations of the components of $U - \int U(u)\,du$. By multiplying by $(L_{11}L_{11}')^{-\frac{1}{2}}$ we can turn these into orthonormal components and by supplementing these vectors with an extra orthonormal vector, which is proportional to $(\alpha_\perp'\Lambda\alpha_\perp)^{-\frac{1}{2}}\alpha_\perp'\mu$, we can transform the process U by an orthornormal matrix O to the process $B = OU$. Then the equation can be written as

$$(B.6) \qquad \left|\rho\int FF'\,du - \int F(dB)'\int(dB)F'\right| = 0,$$

where F is given by (2.16) and (2.17). This equation has $p-r$ roots. Thus we have seen that the $p-r$ smallest roots of (B.2) decrease to zero at the rate T^{-1} and that $T\lambda$ converge to the roots of (B.6). From the expression for the likelihood ratio test statistics we find that

$$-2\ln(Q;H_2|H_1) = T\sum_{i=r+1}^{p}\hat{\lambda}_i + o_P(1)$$

$$\xrightarrow{w} \sum_{i=r+1}^{p}\hat{\rho}_i = \operatorname{tr}\left\{\int(dB)F'\left[\int FF'\,du\right]^{-1}\int F(dB)'\right\}.$$

Note that if $\tau=0$, i.e. the linear trend is missing, then again applying Lemma A.3 we can choose $\gamma=\beta_\perp$, and the results have to be modified by leaving out the terms containing τ. The matrix L_{11} is $(p-r)\times(p-r)$ and cancels in (B.5) so that the test of H_2 in H_1 is distributed as

$$T_2 = \operatorname{tr}\left\{\int(dU)F'\left[\int FF'\,du\right]^{-1}\int F(dU)'\right\} \tag{B.7}$$

with $F(t)=U(t)-\int U(u)\,du$. This completes the proof of Theorem 2.1.

Proof of Theorem 2.2

The estimation under H_2^* involved the solution of the equation

$$\left|\lambda S_{kk}^* - S_{k0}^* S_{00}^{*-1} S_{0k}^*\right| = 0; \tag{B.8}$$

see (2.11) with the S_{ij} replaced by S_{ij}^*. Let $A_T^* = (\beta^*, T^{-\frac{1}{2}}B_T^*)$ and multiply the matrix in (B.8) by A_T and its transpose (see Lemma A.4), and let $T\to\infty$. The roots of (B.8) converge to the roots of the equation

$$\left|\begin{bmatrix}\lambda\beta'\Sigma_{kk}\beta - \beta'\Sigma_{k0}\Sigma_{00}^{-1}\Sigma_{0k}\beta & 0\\ 0 & \lambda\int G^*G^{*\prime}\,du\end{bmatrix}\right| = 0.$$

This shows that the r largest solutions of (B.8) converge to the roots of the same limiting equation as before (see (B.3)). Now multiply instead by B_T^* and its transpose and let $\rho=T\lambda$ and $\lambda\to 0$; then we obtain, by an argument similar to that given in the proof of Theorem 2.1, that in the limit the $p-r+1$ smallest roots normalized by T will converge in distribution to the roots of the equation

$$\left|\rho\int G^*G^{*\prime}\,du - \int G^*(dW)'\alpha_\perp(\alpha_\perp\Lambda\alpha_\perp)^{-1}\alpha_\perp\int(dW)G^{*\prime}\right| = 0.$$

Again we can introduce the $p-r$ dimensional process $U=(\alpha_\perp'\Lambda\alpha_\perp)^{-\frac{1}{2}}\alpha_\perp' W$ and cancel the matrix $(\alpha_\perp'\Psi\beta_\perp)^{-1}(\alpha_\perp'\Lambda\alpha_\perp)^{\frac{1}{2}}$ to see that the test statistic has a limit distribution which is given by

$$T_2^* = \operatorname{tr}\left\{\int(dU)\begin{pmatrix}U\\1\end{pmatrix}'\left[\int\begin{pmatrix}U\\1\end{pmatrix}\begin{pmatrix}U\\1\end{pmatrix}'du\right]^{-1}\int\begin{pmatrix}U\\1\end{pmatrix}(dU)'\right\}. \tag{B.9}$$

The result for the maximal eigenvalue follows similarly. This completes the proof of Theorem 2.2.

Proof of Theorem 2.3

From the relation

$$\begin{pmatrix}U\\1\end{pmatrix}'\left[\int\begin{pmatrix}U\\1\end{pmatrix}\begin{pmatrix}U\\1\end{pmatrix}'du\right]^{-1}\begin{pmatrix}U\\1\end{pmatrix} = 1 + (U-\bar{U})'\left[\int(U-\bar{U})(U-\bar{U})'\,du\right]^{-1}(U-\tilde{U})$$

where $\bar{U}=\int U(u)\,du$, it follows that $T_2^* = U(1)'U(1) + T_2$ (see (B.7) and (B.9)). The likelihood ratio test statistic of H_2^* in H_2 is the difference of the two test statistics considered in Theorem 2.1 and Theorem 2.2. Furthermore the test statistics have the same variables entering the asymptotic expansions and hence the distribution can be found by subtracting the above random variables T_2 and T_2^*, but $U(1)'U(1)$ is $\chi^2(p-r)$.

Proof of Theorem 3.1

We multiply (2.3) by A' and B' respectively and insert $\alpha = A\psi$ and $\beta = H\varphi$ and obtain

$$A'Z_{0t} = A'\Gamma Z_{1t} + A'A\psi\varphi'H'Z_{kt} + A'\varepsilon_t, \tag{B.10}$$

$$B'Z_{0t} = B'\Gamma Z_{t1} + \qquad\qquad B'\varepsilon_t, \tag{B.11}$$

since $B'A\psi = 0$. These equations are analyzed by considering the contribution to the likelihood function from (B.11) and then the contribution from (B.10) given (B.11).

The contribution from (B.11) after having maximized with respect to the parameters $(\Gamma_1, \dots, \Gamma_{k-1}, \Phi, \mu)$ is

$$L_{\max}^{-2/T}(\Lambda_{bb}) = |\Lambda_{bb}| \exp\left\{T^{-1} \sum_{t=1}^{T} R'_{bt}\Lambda_{bb}^{-1}R_{bt}\right\} / |B'B|,$$

where $R_{bt} = B'R_{0t}$ and $\Lambda_{bb} = B'\Lambda B$. Maximizing we find

$$\hat{\Lambda}_{bb} = S_{bb} = B'S_{00}B,$$

which proves (3.5). The relevant part of the maximized likelihood function is

$$L_{\max}^{-2/T} = |S_{bb}| / |B'B|.$$

The contribution from (B.10) given (B.11) is, after the initial maximization,

$$L_{\max}^{-2/T}(\psi, \varphi, \Lambda_{aa.b}, \Lambda_{ab}\Lambda_{bb}^{-1}) = |\Lambda_{aa.b}| \exp\left\{T^{-1} \sum_{t=1}^{T} R'_t \Lambda_{aa.b}^{-1} R_t\right\} / |A'A|,$$

where

$$R_t = R_{at} - \Lambda_{ab}\Lambda_{bb}^{-1}R_{bt} - A'A\psi\varphi'R_{ht},$$

and $R_{at} = A'R_{0t}$ and $R_{ht} = H'R_{kt}$. Minimizing with respect to the parameter $\Lambda_{ab}\Lambda_{bb}^{-1}$ gives rise to yet another regression of R_{at} and R_{ht} on R_{bt}, and the estimate

$$\hat{\Lambda}_{ab}\hat{\Lambda}_{bb}^{-1}(\varphi, \psi) = (S_{ab} - A'A\psi\varphi'S_{hb})S_{bb}^{-1}.$$

This proves (3.6) and gives the new residuals

$$R_{a.bt} = R_{at} - S_{ab}S_{bb}^{-1}R_{bt}$$

and

$$R_{h.bt} = R_{ht} - S_{hb}S_{bb}^{-1}R_{bt}.$$

For $\tilde{\psi} = A'A\psi$, the likelihood function is reduced to the form (2.7) in terms of $R_{a.bt}$, $R_{h.bt}$, φ, $\tilde{\psi}$, and $\Lambda_{aa.b}$. Hence the solution can be found by solving (3.2) and using $\alpha = A(A'A)^{-1}\psi$ together with the relations (2.8), (2.9), (2.12), and (2.6), which completes the proof of Theorem 3.1.

Proof of Theorem 3.2

The limit result follows from Theorem C.1, proved in Appendix C. What remains is to calculate the degrees of freedom. The matrix $\Pi = \alpha\beta' = A\psi\varphi'H'$ is identified, as is the matrix $\psi\varphi'$. Now normalize φ to be of the form $\varphi' = (I, \varphi'_0)$ with φ_0 of dimension $r \times (s - r)$. Then there are $rm + r(s - r)$ free parameters under the assumptions of H_3. For $m = s = p$ we get the result for H_2 and the difference is the degrees of freedom for the test.

APPENDIX C. Asymptotic Inference

Proof of Theorem C.1

There is a qualitative difference between inference for β and that for the other parameters. It was proved by Stock (1987) that the regression estimate for β was superconsistent. This has consequences for the usual proof for asymptotic normality, as later exploited by Phillips (1990).

The idea can be illustrated as follows: Let $\vartheta = (\vartheta_1, \vartheta_2)$ denote all the parameters, and ϑ_2 the parameters in β. Let q denote the log likelihood function normalized by T, and q_1, q_{12}, etc. denote the derivatives with respect to ϑ_1, ϑ_1 and ϑ_2, etc. The usual asymptotic representation for the maximum likelihood estimator is

$$\begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}\begin{bmatrix} \hat{\vartheta}_1 - \vartheta_1 \\ \hat{\vartheta}_2 - \vartheta_2 \end{bmatrix} \approx \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}.$$

If one can now prove that q_{11}, q_{12}, q_{21} are $O_P(1)$ whereas q_{22} is $O_P(T^{\nu})$ for some $\nu > 0$, then it is convenient to normalize as follows:

$$\text{(C.1)} \qquad \begin{bmatrix} q_{11} & T^{-\frac{1}{2}\nu}q_{12} \\ T^{-\frac{1}{2}\nu}q_{21} & T^{-\nu}q_{22} \end{bmatrix}\begin{bmatrix} T^{\frac{1}{2}}\left(\hat{\vartheta}_1 - \vartheta_1\right) \\ T^{\frac{1}{2}(\nu+1)}\left(\hat{\vartheta}_2 - \vartheta_2\right) \end{bmatrix} \approx \begin{bmatrix} T^{\frac{1}{2}}q_1 \\ T^{\frac{1}{2}(1-\nu)}q_2 \end{bmatrix}.$$

Since by assumption $T^{-\frac{1}{2}\nu}q_{12} \xrightarrow{P} 0$, the equation (C.1) splits into

$$q_{11}T^{\frac{1}{2}}\left(\hat{\vartheta}_1 - \vartheta_1\right) \approx T^{\frac{1}{2}}q_1$$

and

$$\left[T^{-\nu}q_{22}\right]T^{\frac{1}{2}(\nu+1)}\left(\hat{\vartheta}_2 - \vartheta_2\right) \approx T^{\frac{1}{2}(1-\nu)}q_2.$$

These equations are the ones we would get when conducting inference about ϑ_1 for fixed ϑ_2 and ϑ_2 for fixed ϑ_1. The same expansion will show that the likelihood ratio test statistic for a simple hypothesis about ϑ is

$$\begin{aligned} -2\ln Q &\approx T\begin{bmatrix} \hat{\vartheta}_1 - \vartheta_1 \\ \hat{\vartheta}_2 - \vartheta_2 \end{bmatrix}'\begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}\begin{bmatrix} \hat{\vartheta}_1 - \vartheta_1 \\ \hat{\vartheta}_2 - \vartheta_2 \end{bmatrix} \\ &\approx \begin{bmatrix} T^{\frac{1}{2}}\left(\hat{\vartheta}_1 - \vartheta_1\right) \\ T^{\frac{1}{2}(\nu+1)}\left(\hat{\vartheta}_2 - \vartheta_2\right) \end{bmatrix}'\begin{bmatrix} q_{11} & T^{-\frac{1}{2}\nu}q_{12} \\ T^{-\frac{1}{2}\nu}q_{21} & T^{-\nu}q_{22} \end{bmatrix}\begin{bmatrix} T^{\frac{1}{2}}\left(\hat{\vartheta}_1 - \vartheta_1\right) \\ T^{\frac{1}{2}(\nu+1)}\left(\hat{\vartheta}_2 - \vartheta_2\right) \end{bmatrix} \\ &\approx T^{\frac{1}{2}}\left(\hat{\vartheta}_1 - \vartheta_1\right)'q_{11}T^{\frac{1}{2}}\left(\hat{\vartheta}_1 - \vartheta_1\right) \\ &\quad + T^{\frac{1}{2}(\nu+1)}\left(\hat{\vartheta}_2 - \vartheta_2\right)'\left[T^{-\nu}q_{22}\right]T^{\frac{1}{2}(\nu+1)}\left(\hat{\vartheta}_2 - \vartheta_2\right). \end{aligned}$$

This shows that the test statistic decomposes into a test for ϑ_1 and an independent test for ϑ_2.

The above argument indicates that inference about ϑ_2 can be conducted as if ϑ_1 were known, and vice versa.

We can prove the above property about the second derivatives of the likelihood function concentrated with respect to μ and we therefore deduce that inference about $(\Gamma_1, \dots, \Gamma_{k-1}, \alpha, \Phi, \Lambda)$ can be conducted for fixed β; hence one can apply the well known results for asymptotic inference for the stationary processes ΔX_t and $\beta' X_t$. See Dunsmuir and Hannan (1976) for a general treatment of smooth hypotheses for stationary processes. The asymptotic distribution of μ is somewhat more complicated, and will not be treated in detail here.

The asymptotic properties of estimators and test statistics are discussed here for a general smooth hypothesis on the cointegrating relations: $\beta = \beta(\vartheta)$, $\vartheta \in \theta \subset R^k$, leaving the remaining parameters unrestricted. Let $D\beta(u)$ denote the derivative of $\beta(\vartheta)$ with respect to ϑ in the direction $u \in R^k$, i.e. the $p \times r$ matrix, with elements $\sum_{s=1}^{k} u_s \, \partial\beta_{ij}(\vartheta)/\partial\vartheta_s$. We assume throughout that $D\beta(u)$ has full rank for $u \neq 0$, and that $\beta' D\beta(u) = 0$ for all u, where β is the value of the parameter for which the results are derived. This last condition can always be achieved by normalizing $\beta(\vartheta)$ by β, i.e. by considering $\beta(\vartheta)(\beta'\beta(\vartheta))^{-1}$. We also let $D\beta$ denote the $pr \times k$ matrix with element $((i,j),s)$ equal to

$$\partial\beta_{ij}(\vartheta)/\partial\vartheta_s \qquad (i = 1, \dots, p;\ j = 1, \dots, r;\ s = 1, \dots, k),$$

so that $(D\beta)u = \operatorname{vec}(D\beta(u))$. The natural coordinate system in R^p is given by (β, γ, τ) and the corresponding coordinate system in R^k is defined by choosing $u_1, \ldots, u_k$ orthogonal in R^k such that

$$\tau' D\beta(u_i) = 0 \qquad (i = 1, \ldots, k_1) \tag{C.2}$$

and

$$\tau' D\beta(u_i) \neq 0 \qquad (i = k_1 + 1, \ldots, k). \tag{C.3}$$

The behavior of the estimate of ϑ depends on which linear combination $u_i'(\hat{\vartheta} - \vartheta)$ is considered and we define N_T as the $k \times k$ matrix with ith column given by $Tu_i(u_i'u_i)^{-1}$, if $i = 1, \ldots, k_1$ and $T^{3/2}u_i(u_i'u_i)^{-1}$, if $i = k_1 + 1, \ldots, k$.

Furthermore we need a $(p-r)r \times k$ matrix $D\tilde{\beta}$ with ith column given by

$$D\tilde{\beta}_i = \operatorname{vec}\{(\gamma, 0)' D\beta(u_i)\} \qquad (i = 1, \ldots, k_1), \tag{C.4}$$

$$D\tilde{\beta}_i = \operatorname{vec}\{(0, \tau)' D\beta(u_i)\} \qquad (i = k_1 + 1, \ldots, k). \tag{C.5}$$

We also define the variable $Y_t = (\Delta X_{t-1}', \ldots, \Delta X_{t-k+1}', X_{t-k}'\beta, D_t')'$, as well as the parameters

$$\mu_Y = E(Y_t)$$

and

$$\Omega = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} (Y_t - \bar{Y})(Y_t - \bar{Y})'.$$

The results below are derived under the assumption that the maximum likelihood estimator for ϑ exists and is consistent.

THEOREM C.1: *Under the assumption $\Pi = \alpha\beta(\vartheta)'$ the asymptotic distribution of*

$$T^{\frac{1}{2}}\left(\left(\hat{\Gamma}_1, \ldots, \hat{\Gamma}_{k-1}, \hat{\alpha}, \hat{\Phi}\right) - (\Gamma_1, \ldots, \Gamma_{k-1}, \alpha, \Phi)\right)$$

is Gaussian with mean zero and variance matrix

$$\Lambda \otimes \begin{pmatrix} \Omega^{-1} & 0 \\ 0 & 2\Lambda \end{pmatrix}.$$

The asymptotic distribution of $\hat{\vartheta}$ is given by

$$N_T'(\hat{\vartheta} - \vartheta) \xrightarrow{w} \left\{D\tilde{\beta}'\left(\int GG'\,du \otimes (\alpha'\Lambda^{-1}\alpha)\right)D\tilde{\beta}\right\}^{-1} D\tilde{\beta}' \operatorname{vec}\left\{\int G(dV)'\right\}, \tag{C.6}$$

that is, Gaussian for fixed G with variance

$$\left\{D\tilde{\beta}'\left(\int GG'\,du \otimes (\alpha'\Lambda^{-1}\alpha)\right)D\tilde{\beta}\right\}^{-1}$$

which we call the asymptotic conditional variance. Here $V = \alpha'\Lambda^{-1}W$. It follows that the limit distribution of $(T\gamma, T^{3/2}\tau)'(\beta(\hat{\vartheta}) - \beta)$ is given by

$$D\tilde{\beta}\left\{D\tilde{\beta}'\left(\int GG'\,du \otimes (\alpha'\Lambda^{-1}\alpha)\right)D\tilde{\beta}\right\}^{-1} D\tilde{\beta}' \operatorname{vec}\left\{\int G(dV)'\right\}. \tag{C.7}$$

If $k_1 = k$, that is if $\tau' D\beta(u) = 0$ for all u, then the results should be modified by replacing G by G_1; see Theorem 5.1.

The likelihood ratio test statistic of a smooth hypothesis $\vartheta = \vartheta(\eta)$, $\eta \in R^s$, $s < k$ is asymptotically distributed as χ^2 with $k - s$ degrees of freedom. A similar result holds if $\alpha'\mu = 0$; only G should be replaced by G^* (see Theorem 5.2).

PROOF: The likelihood function concentrated with respect to μ and divided by T is given by

$$q(\vartheta, \alpha, \Lambda) = -\tfrac{1}{2}\ln|\Lambda| - \tfrac{1}{2}\operatorname{tr}\left\{\Lambda^{-1}T^{-1}\Sigma_t(\varepsilon_t - \bar{\varepsilon})(\varepsilon_t - \bar{\varepsilon})'\right\}$$

for

$$\varepsilon_t - \bar{\varepsilon} = \Delta X_t - \Delta \bar{X} - \sum_{i=1}^{k-1} \Gamma_i \left(\Delta X_{t-i} - \Delta \bar{X}_{-i} \right) - \alpha \beta(\theta)' \left(X_{t-k} - \bar{X}_{-k} \right) - \Phi\left(D_t - \bar{D} \right).$$

Here the bar denotes average. The derivatives are most easily found by a Taylor expansion. Thus if $q_\vartheta(u)$ denotes the derivative of q with respect to ϑ in the direction u, we can find the derivative from the expansion $q(\vartheta + u, \alpha, \Lambda) = q(\vartheta, \alpha, \Lambda) + q_\vartheta(u) + O(|u|^2)$. We then get derivatives

$$q_{\Gamma_i}(g) = \operatorname{tr}\left\{ \Lambda^{-1} T^{-1} \Sigma_t (\varepsilon_t - \bar{\varepsilon}) \left(\Delta X_{t-i} - \Delta \bar{X}_{-i} \right)' g' \right\},$$

$$q_\alpha(a) = \operatorname{tr}\left\{ \Lambda^{-1} T^{-1} \Sigma_t (\varepsilon_t - \bar{\varepsilon}) \left(X_{t-k} - \bar{X}_{-k} \right)' \beta a' \right\},$$

$$q_\Phi(f) = \operatorname{tr}\left\{ \Lambda^{-1} T^{-1} \Sigma_t (\varepsilon_t - \bar{\varepsilon}) \left(D_t - \bar{D} \right)' f' \right\},$$

$$q_\Lambda(l) = \tfrac{1}{2} \operatorname{tr}\left\{ \Lambda^{-1} l \Lambda^{-1} T^{-1} \Sigma_t (\varepsilon_t - \bar{\varepsilon})(\varepsilon_t - \bar{\varepsilon})' \right\} - \tfrac{1}{2} \operatorname{tr}\{\Lambda^{-1} l\},$$

$$q_\vartheta(u) = \operatorname{tr}\left\{ \Lambda^{-1} \tilde{M}_{0k} D\beta(u) \alpha' \right\},$$

and second derivatives

$$q_{\Gamma_i \Gamma_i}(g, g) = -\operatorname{tr}\left\{ \Lambda^{-1} g T^{-1} \Sigma_t \left(\Delta X_{t-i} - \Delta \bar{X}_{-i} \right) \left(\Delta X_{t-i} - \Delta \bar{X}_{-i} \right)' g' \right\},$$

$$q_{\alpha\alpha}(a, a) = -\operatorname{tr}\left\{ \Lambda^{-1} a \beta' T^{-1} \Sigma_t \left(X_{t-k} - \bar{X}_{-k} \right) \left(X_{t-k} - \bar{X}_{-k} \right) \beta a' \right\},$$

$$q_{\Phi\Phi}(f, f) = -\operatorname{tr}\left\{ \Lambda^{-1} f T^{-1} \Sigma_t \left(D_t - \bar{D} \right) \left(D_t - \bar{D} \right)' f' \right\},$$

$$q_{\Lambda\Lambda}(l, l) = -\operatorname{tr}\left\{ \Lambda^{-1} l \Lambda^{-1} l \Lambda^{-1} T^{-1} \Sigma_t (\varepsilon_t - \bar{\varepsilon})(\varepsilon_t - \bar{\varepsilon})' \right\} + \tfrac{1}{2} \operatorname{tr}\{\Lambda^{-1} l \Lambda^{-1} l\},$$

$$q_{\vartheta\vartheta}(u, u) = -\operatorname{tr}\left\{ \Lambda^{-1} T^{-1} \Sigma_t (\varepsilon_t - \bar{\varepsilon}) \left(X_{t-k} - \bar{X}_{-k} \right)' D^2 \beta(u, u) \alpha' \right\}$$
$$- \operatorname{tr}\left\{ \Lambda^{-1} \alpha D\beta(u)' \tilde{M}_{kk} D\beta(u) \alpha' \right\}$$

together with similar expressions for mixed second derivatives. It follows from Lemma A.3, that all terms in the above derivatives are $O_P(1)$, except the expression $-\operatorname{tr}\{\Lambda^{-1} \alpha D\beta(u)' \tilde{M}_{kk} D\beta(u) \alpha'\}$ in $q_{\vartheta\vartheta}$, and that this tends to infinity, since the columns of $D\beta(u)$ are orthogonal to β. Thus we apply the above general argument, even though we have two different normalizations, and have shown that inference for β can be conducted for fixed value of the other parameters, and vice versa.

It is not difficult to see by the central limit theorem for stationary ergodic processes (see White (1984)) that the derivatives $T^{\frac{1}{2}}(q_{\Gamma_1}, \ldots, q_{\Gamma_{k-1}}, q_\alpha, q_\Phi, q_\Lambda)$ are asymptotically Gaussian with mean zero and variance matrix

$$\Lambda^{-1} \otimes \begin{pmatrix} \Omega & 0 \\ 0 & \frac{1}{2} \Lambda^{-1} \end{pmatrix},$$

which is also the limit of the matrix of the second derivatives with opposite sign with respect to these parameters, such that the first conclusion of the Theorem holds and the variance is given by (C.7); see also Lütkepohl and Reimers (1989).

To find the asymptotic distribution of the estimate of ϑ we expand the likelihood function around the point $\hat{\vartheta}$, such that the other parameters are kept fixed. The relation (C.4) now takes the form

$$\operatorname{tr}\left\{ D\beta(u)' \tilde{M}_{kk} D\beta(\hat{\vartheta} - \vartheta) \alpha' \Lambda^{-1} \alpha \right\} \approx \operatorname{tr}\left\{ D\beta(u)' \tilde{M}_{k0} \Lambda^{-1} \alpha \right\}$$

for all $u \in R^k$. In vectorized form this becomes

$$\text{(C.8)} \qquad \left[D\beta' \left(\tilde{M}_{kk} \otimes \alpha' \Lambda^{-1} \alpha \right) D\beta \right] (\hat{\vartheta} - \vartheta) \approx D\beta' \operatorname{vec}\left\{ \tilde{M}_{0k} \Lambda^{-1} \alpha \right\}.$$

The limiting behavior of the various matrices is given by Lemma A.3: From the identity $B_T(\gamma, T^{\frac{1}{2}}\tau)' = \bar{\gamma}\gamma' + \bar{\tau}\tau' = I - P_\beta$, we get since $\beta' D\beta(u) = 0$ that for $u, v \in R^k$,

$$D\beta(u)' \tilde{M}_{kk} D\beta(v) = D\beta(u)' \left(\gamma, T^{\frac{1}{2}}\tau \right) B_T' \tilde{M}_{kk} B_T \left(\gamma, T^{\frac{1}{2}}\tau \right)' D\beta(v) = M(u, v),$$

say. The matrix $\tilde{M}_{kk}$ should be normalized by T^{-1} and by B_T to get convergence (see Lemma A.3), but of course the factor $T^{\frac{1}{2}}\tau$ has to be taken care of. Now introduce the coordinate system $w_i = T^{-\frac{1}{2}}u_i$, $i = 1,\dots,k_1$, $w_i = T^{-1}u_i$, $i = k_1+1,\dots,k$. Then $M(w_i, w_j)$ is weakly convergent towards

$$u_i' D\tilde{\beta}'\left[\int GG'\,du\right]D\tilde{\beta}u_j$$

(see (C.4) and (C.5)). Similarly

$$T^{\frac{1}{2}}D\beta(u)'\tilde{M}_{k0}\Lambda^{-1}\alpha = T^{\frac{1}{2}}D\beta(u)'\left(\gamma, T^{\frac{1}{2}}\tau\right)B_T'\tilde{M}_{k0}\Lambda^{-1}\alpha$$

converges weakly for $u = w_t$ towards

$$u_t D\tilde{\beta}'\int G(dV)'.$$

With this notation we can replace (C.8) by the result (C.6). By a Taylor's expansion we find (C.7). By expanding the likelihood function around a fixed value of ϑ one finds that the test statistic for a simple hypothesis for ϑ is

$$\begin{aligned}-2\ln Q &\simeq T\,\mathrm{tr}\left\{D\beta(\hat{\vartheta}-\vartheta)'\tilde{M}_{kk}D\beta(\hat{\vartheta}-\vartheta)\alpha'\Lambda^{-1}\alpha\right\}\\ &\xrightarrow{w}\left\{\mathrm{vec}\left(\int G(dV)'\right)\right\}'D\tilde{\beta}\left\{D\tilde{\beta}'\left(\int GG'\,du\otimes\alpha'\Lambda^{-1}\alpha\right)D\tilde{\beta}\right\}^{-1}D\tilde{\beta}'\\ &\quad\times\left\{\mathrm{vec}\left(\int G(dV)'\right)\right\}\end{aligned}$$

which for given G is χ^2 distributed with k degrees of freedom. Now if $\vartheta = \vartheta(\eta)$ one finds the same result except that $D\tilde{\beta}$ is replaced by $(D\tilde{\beta})(D\vartheta)$, i.e. the $pr \times k$ matrix $D\tilde{\beta}$ is multiplied by the $k \times s$ matrix of derivatives $D\vartheta$. The likelihood ratio test statistic for the hypothesis $\vartheta = \vartheta(\eta)$ is then asymptotically distributed as

$$Q'\left(D\tilde{\beta}\left\{D\tilde{\beta}'V(Q|G)D\tilde{\beta}\right\}^{-1}D\tilde{\beta}' - D\tilde{\beta}D\vartheta\left(D\vartheta'D\tilde{\beta}'V(Q|G)D\tilde{\beta}D\vartheta\right)^{-1}D\vartheta'D\tilde{\beta}'\right)Q,$$

where $Q = \mathrm{vec}(\int G(dV)')$ and $V(Q|G) = \int GG'\,du\otimes\alpha'\Lambda^{-1}\alpha$. For fixed G the variable Q is Gaussian with covariance matrix $V(Q|G)$; hence this statistic is χ^2 distributed with $k-s$ degrees of freedom and as this distribution does not depend on G the result holds unconditionally.

This completes the proof of Theorem C.1, but let us just see how these results can be applied to find the asymptotic distribution of $\hat{\mu}$. The estimating equation is

$$\Delta\overline{X}_0 = \sum_{i=1}^{k-1}\hat{\Gamma}_i\Delta\overline{X}_{-i} + \hat{\alpha}\hat{\beta}'\overline{X}_{-k} + \hat{\Phi}\overline{D} + \hat{\mu},$$

which together with the identity

$$\Delta\overline{X}_0 = \sum_{i=1}^{k-1}\Gamma_i\Delta\overline{X}_{-i} + \alpha\beta'\overline{X}_{-k} + \Phi\overline{D} + \mu + \bar{\varepsilon}$$

shows that

$$\begin{aligned}T^{\frac{1}{2}}(\hat{\mu}-\mu) &= -\left\{T^{\frac{1}{2}}\left(\hat{\Gamma}_1-\Gamma_1,\dots,\hat{\Gamma}_{k-1}-\Gamma_{k-1},\hat{\alpha}-\alpha,\hat{\Phi}-\Phi\right)\overline{Y} - T^{\frac{1}{2}}\bar{\varepsilon}\right\}\\ &\quad-\hat{\alpha}T^{\frac{1}{2}}\left(\hat{\beta}-\beta\right)'\overline{X}_{-k}.\end{aligned}$$

The first term converges towards a Gaussian distribution with mean zero and variance matrix $\Lambda(\mu_Y'\Omega^{-1}\mu_Y + 1)$ and the second is normalized just right, so that the distribution can be found from the second part of Theorem C.1. Note that the asymptotic distribution of $\alpha_\perp'\hat{\mu}$ is Gaussian, but the component which goes into the cointegrating relation has a more complicated limit distribution.

Proof of Theorem 5.1

The proof consists of simplifying the expressions given in Theorem C.1. A point in $\mathrm{sp}(H)$ can be represented as $\beta + (\gamma_H, \tau_H)\vartheta$, where $(\beta, \gamma_H, \tau_H)$ are orthogonal and span $\mathrm{sp}(H)$, and where $\vartheta = \{\vartheta_{i,j},\ i = 1, \dots, s-r,\ j = 1, \dots, r\}$. It follows that $D\beta(u) = (\gamma_H, \tau_H)u$ and that the equation

$$\tau' D\beta(u) = \tau'(\gamma_H, \tau_H)u = (0, \tau_H'\tau_H)u = 0$$

(see (C.2) and (C.3)), can be solved by choosing $u_{ij} = o_i e_j'$, $i = 1, \dots, s-r-1$, $j = 1, \dots, r$, where o_i are unit vectors in R^{s-r} and e_j unit vectors in R^r. Thus in this case $k = (s-r)r$ and $k_1 = (k-s-1)r$. The (i,j)th column of $D\tilde{\beta}$ (see (C.4) and (C.5)), is given by

$$\mathrm{vec}\left\{(\gamma, 0)'(\gamma_H, 0)o_i e_j'\right\} \qquad (i = 1, \dots, s-r-1,\ j = 1, \dots, r),$$

$$\mathrm{vec}\left\{(0, \tau)'(0, \tau_H)o_i e_j'\right\} \qquad (i = s-r,\ j = 1, \dots, r).$$

This shows that

$$D\tilde{\beta}'\left(\int GG'\,du \otimes (\alpha'\Lambda^{-1}\alpha)\right)D\tilde{\beta} = P \otimes \alpha'\Lambda^{-1}\alpha,$$

where P is a notation for

$$P = \begin{pmatrix} \gamma_H'\gamma\int G_1G_1'\,du\,\gamma'\gamma_H & \gamma_H'\gamma\int G_1G_2\,du\,\tau'\tau_H \\ \tau_H'\tau\int G_2G_1'\,du\,\gamma'\gamma_H & \tau_H'\tau\int G_2G_2\,du\,\tau'\tau_H \end{pmatrix}.$$

The left hand side of (C.7) becomes $(\mathrm{tr}(u_{ij}'u_{ij}))^{-1}\,\mathrm{tr}(u_{ij}'\hat{\vartheta}) = \hat{\vartheta}_{ij}$ multiplied by T if $i = 1, \dots, s-r-1$ and $T^{3/2}$ if $i = s-r$. The matrix $D\tilde{\beta}'\,\mathrm{vec}\{\int G(dV)'\}$ has (i,j)th element equal to

$$\mathrm{tr}\left\{e_j o_i'(\gamma_H, 0)'(\gamma, 0)\int G(dV)'\right\} = \gamma_{Hi}'\gamma\int G_1(dV_j) \qquad (i = 1, \dots, s-r-1,\ j = 1, \dots, r)$$

and

$$\mathrm{tr}\left\{e_j o_i'(0, \tau_H)'(0, \tau)\int G(dV)'\right\} = \tau_H'\tau\int G_2(dV_j)' \qquad (i = s-r,\ j = 1, \dots, r).$$

Collecting these results we obtain

$$(T\gamma_H, T^{3/2}\tau_H)\hat{\vartheta} \xrightarrow{w} (\gamma_H, \tau_H)P^{-1}(\gamma_H, \tau_H)'(\gamma, \tau)\int G(dV)'(\alpha'\Lambda^{-1}\alpha)^{-1}.$$

The normalized estimate $\hat{\beta}_c = \beta_c(\hat{\vartheta})$ has the expansion around $\vartheta = 0$:

$$\beta_c(\hat{\vartheta}) - \beta_c = \left(I - \beta(c'\beta)^{-1}c'\right)\hat{\vartheta}(c'\beta)^{-1} + O_P(|\hat{\vartheta}|^2)$$

which shows that

$$\text{(C.9)} \qquad T\left(\hat{\beta}_c - \beta_c\right) \xrightarrow{w} (I - \beta_c c')(\gamma_H, 0)P^{-1}(\gamma_H, \tau_H)'(\gamma, \tau)\int G(dV)'(\alpha'\Lambda^{-1}\alpha)^{-1}(c'\beta)^{-1}.$$

We can simplify the expression

$$\begin{aligned}(\gamma_H, 0)P^{-1}(\gamma_H, \tau_H)'(\gamma, \tau)G &= \gamma_H P^{11}\gamma_H'\gamma G_1 + P^{12}\tau_H'\tau G_2 \\ &= \gamma_H P^{11}\left(\gamma_H'\gamma G_1 - P_{12}P_{22}^{-1}\tau_H'\tau G_2\right) \\ &= \gamma_H\left(\gamma_H'\gamma\int G_{1:2}G_{1:2}\,du\,\gamma'\gamma_H\right)^{-1}\gamma_H'\gamma G_{1:2},\end{aligned}$$

which inserted into (C.9) gives the result (5.1).

The consistent estimator (5.6) for the asymptotic conditional variance is found from Lemma C.2 below, for the choice $K = I - \beta_c c'$, which satisfies $K'\hat{\beta} = 0$.

Finally we note that the normalization $\hat{V}'S_{hh.b}\hat{V} = I$ (see (3.2)) implies that $HS_{hh.b}^{-1}H' = H\hat{V}\hat{V}'H' = \hat{\beta}\hat{\beta}' + H\hat{v}\hat{v}'H'$. Since $(I - \beta_c c')\beta = 0$, $(I - \beta_c c')\hat{\beta}$ is $O_P(T^{-1})$ and $(I - \beta_c c')'H\hat{v}$ is $O_P(T^{-\frac{1}{2}})$, such that

$$T(I - \beta_c c')' HS_{hh}^{-1}H(I - c\beta_c') = T(I - \beta_c c')' HS_{hh.b}^{-1}H(I - c\beta_c') + o_P(1)$$

$$= T(I - \beta_c c')' H\hat{v}\hat{v}'H'(I - c\beta_c') + o_P(1).$$

Hence one can apply either of these in the consistent estimation of the asymptotic conditional variance. The relation (A.6) from Lemma A.1 gives the identity $(\hat{\alpha}'\hat{\Lambda}^{-1}\hat{\alpha})^{-1} = \text{diag}(\hat{\lambda}_1 - 1, \ldots, \hat{\lambda}_r - 1\}$. This completes the proof of Theorem 5.1.

Consistent Estimates of the Asymptotic Conditional Variance

The next results are needed for the consistent estimation of the limiting conditional variance in the limiting distribution of $\hat{\beta}_c$. We take the coordinates $(\beta, \gamma_H, \tau_H)$ where $\gamma_H(p \times (p - s - 1))$ is chosen such that $(\beta, \gamma_H, \tau_H)$ span $\text{sp}(H)$. We let $\hat{v} = (\hat{v}_{r+1}, \ldots, \hat{v}_p)$ (see (3.2) with the normalization $\hat{v}'S_{hh.b}\hat{v} = I$).

LEMMA C.2: *If $K'\beta = 0$, and $\tau_H \neq 0$, then*

$$TK'H\hat{v}\hat{v}'H'K \xrightarrow{w} K'\gamma_H\left(\gamma_H'\gamma\int G_{1.2}G_{1.2}'\,du\,\gamma'\gamma_H\right)^{-1}\gamma_H'K.$$

If $\tau_H = 0$ then this results holds with $\gamma_H(p \times (p - s))$ chosen such that (β, γ_H) span $\text{sp}(H)$ and $G_{1.2}$ is replaced by G_1.

PROOF: We first expand

$$H\hat{v} = \beta e + \gamma_H g + \tau_H f \tag{C.10}$$

and then note that from

$$\hat{v}'S_{ha.b}S_{aa.b}^{-1}S_{ah.b}\hat{v} = \text{diag}\left(\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_s\right) \in O_P(T^{-1}),$$

it follows that $\hat{v}$ and hence also the coordinates (e, g, f) are $O_P(T^{-1/2})$. From the normalization $\hat{v}'S_{hh.b}\hat{v} = I$, it even follows that $f \in O_P(T^{-1})$ and that, since $e \xrightarrow{P} 0$, we have

$$(\gamma_H g + \tau_H f)' S_{kk.b}(\gamma_H g + \tau_H f) \xrightarrow{P} I$$

and hence that

$$(\gamma_H g + \tau_H f)' S_{kk}(\gamma_H g + \tau_H f) \xrightarrow{P} I. \tag{C.11}$$

Note finally that from (C.10) we have, since $K'\beta = 0$, that

$$K'H\hat{v} = K'(\gamma_H g + \tau_H f) = K'(\gamma_H, \tau_H)(g', f')'. \tag{C.12}$$

Now insert (C.11) and (C.12) into $TK'H\hat{v}\hat{v}'H'K$ and we get

$$\begin{aligned} &TK'(\gamma_H g + \tau_H f)\{(\gamma_H g + \tau_H f)' S_{kk}(\gamma_H g + \tau_H f)\}^{-1}(\gamma_H g + \tau_H f)'K \\ &= TK'(\gamma_H, \tau_H)\{(\gamma_H, \tau_H)' S_{kk}(\gamma_H, \tau_H)\}^{-1}(\gamma_H, \tau_H)'K \\ &= K'B_T\left(T^{-1}B_T'S_{kk}B_T\right)^{-1}B_T'K, \end{aligned} \tag{C.13}$$

for $B_T = (\bar{\gamma}_H, T^{-\frac{1}{2}}\bar{\tau}_H)$. The terms involving $T^{-\frac{1}{2}}\tau_H$ are of smaller order of magnitude than the terms involving γ_H and hence (C.13) converges to

$$K'\gamma_H\left(\gamma_H'\gamma\left\{\int G_1G_1'\,du-\left(\int G_1G_2\,du\right)\left(\int G_2G_2\,du\right)^{-1}(G_2G_1'\,du)\right\}\gamma'\gamma_H\right)^{-1}\gamma_H'K.$$

If $\tau_H=0$ we can drop the terms involving τ_H and choose γ_H orthogonal to β such that they span $\mathrm{sp}(H)$ and apply Lemma A.3 again.

REFERENCES

AHN, S. K., AND G. C. REINSEL (1990): "Estimation for Partially Non-stationary Multivariate Autoregressive Models," *Journal of the American Statistical Association*, 85, 813–823.

ANDERSON, T. W. (1951): "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *Annals of Mathematical Statistics*, 22, 327–351.

——— (1971): *The Statistical Analysis of Time Series*. New York: Wiley.

BOSSAERT, P. (1988): "Common Nonstationary Components of Asset Prices," *Journal of Economic Dynamics and Control*, 12, 347–364.

BOX, G. E. P., AND G. C. TIAO (1981): "A Canonical Analysis of Multiple Time Series with Applications," *Biometrika*, 64, 355–365.

DUNSMUIR, W., AND E. J. HANNAN (1976): "Vector Linear Time Series Models," *Advances in Applied Probability*, 8, 339–364.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–276.

ENGLE, R. F., AND B. S. YOO (1989): "Cointegrated Economic Time Series: A Survey with New Results," Discussion Paper 89-38, University of California, San Diego.

FOUNTIS, N. G., AND D. A. DICKEY (1989): "Testing for Unit Root Nonstationarity in Multivariate Autoregressive Time Series," *Annals of Statistics*, 17, 419–428.

GONZALO, J. (1989): "Comparison of Five Alternative Methods of Estimating Long-Run Equilibrium Relationships," Discussion Paper 89-55, University of California, San Diego.

GRANGER, C. W. J. (1983): "Cointegrated Variables and Error Correction Models," Discussion Paper, 83-13a, University of California, San Diego.

——— (1981): "Some Properties of Time Series Data and their Use in Econometric Model Specification," *Journal of Econometrics*, 16, 121–130.

GRANGER, C. W. J., AND A. A. WEISS (1983): "Time Series Analysis of Error Correcting Models," in *Studies in Econometrics, Time Series and Multivariate Statistics*, ed. by S. Karlin, T. Amemiya, and L. A. Goodman. New York: Academic Press, 255–278.

HOFFMAN, D., AND R. H. RASCHE (1989): "Long-run Income and Interest Elasticities of Money Demand in the United States," National Bureau of Economic Research Discussion Paper No. 2949.

JEGANATHAN, P. (1988): "Some Aspects of Asymptotic Theory with Applications to Time Series Models," The University of Michigan.

JOHANSEN, S. (1988a): "The Mathematical Structure of Error Correction Models," *Contemporary Mathematics*, 80, 259–386.

——— (1988b): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

——— (1990): "A Representation of Vector Autoregressive Processes Integrated of Order 2," to appear in *Econometric Theory*.

——— (1991): "The Statistical Analysis of $I(2)$ Variables," University of 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.

——— (1991): "Some Structural Hypotheses in a Multivariate Cointegration Analysis of the Purchasing Power Parity and the Uncovered Interest Parity for UK," to appear in *Journal of Econometrics*.

LÜTKEPOHL, H., AND H.-E. REIMERS (1989): "Impulse Response Analysis of Cointegrated Systems with an Investigation of German Money Demand," Christian-Albrechts Universitat Kiel.

KUNST, R., AND K. NEUSSER (1990): "Cointegration in a Macro-economic System," *Journal of Applied Econometrics*, 5, 351–365.

Osterwald-Lenum, M. (1992): "A Note with Fractiles of the Asymptotic Distribution of the Likelihood Cointegration Rank Test Statistics: Four Cases," to appear in *Oxford Bulletin of Economics and Statistics*.
Mosconi, R., and C. Giannini (1992): "Non-Causality in Cointegrated Systems: Representation, Estimation and Testing," to appear in *Oxford Bulletin of Economics and Statistics*.
Park, J. Y. (1988): "Canonical Cointegrating Regressions," Cornell University.
Park, J. Y., and P. C. B. Phillips (1988): "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 4, 468–497.
——— (1989): "Statistical Inference in Regressions with Integrated Processes: Part 2," *Econometric Theory*, 5, 95–131.
Pena, D., and G. E. P. Box (1987): "Identifying a Simplifying Structure in Time Series," *Journal of the American Statistical Association*, 82, 836–843.
Phillips, P. C. B. (1988): "Spectral Regression for Cointegrated Time Series," Cowles Foundation No. 872.
——— (1990): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.
Phillips, P. C. B., and S. N. Durlauf (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–495.
Phillips, P. C. B., and S. Ouliaris (1988): "Testing for Cointegration using Principal Components Methods," *Journal of Economic Dynamics and Control*, 12, 1–26.
Phillips, P. C. B., and Y. J. Park (1988): "Asymptotic Equivalence of OLS and GLS in Regression with Integrated Regressors," *Journal of the American Statistical Association*, 83, 111–115.
Phillips, P. C. B., and B. E. Hansen (1990): "Statistical Inference with $I(1)$ Processes," *Review of Economic Studies*, 57, 99–124.
Reinsel, G. C., and S. K. Ahn (1990): "Vectors AR Models with Unit Roots and Reduced Rank Structure: Estimation, Likelihood Ratio Test, and Forecasting," University of Wisconsin.
Sims, C. A., J. H. Stock, and M. W. Watson (1990): "Inference in Linear Time Series Models with some Unit Roots," *Econometrica*, 58, 113–144.
Stock, J. H. (1987): "Asymptotic Properties of Least Squares Estimates of Cointegration Vectors," *Econometrica*, 55, 1035–1056.
Stock, J. H., and M. W. Watson (1988): "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097–1107.
Tso, M.K.-S. (1981): "Reduced-Rank Regression and Canonical Analysis," *Journal of the Royal Statistical Society, Series B*, 43, 183–189.
Velu, R. P., G. C. Reinsel, and D. W. Wichern (1986): "Reduced Rank Models for Multiple Time Series," *Biometrika*, 73, 105–118.
Velu, R. P., D. W. Wichern, and G. C. Reinsel (1987): "A Note on Non-stationary and Canonical Analysis of Multiple Time Series Models," *Journal of Time Series Analysis*, 8, 479–487.
White, H. (1984): *Asymptotic Theory for Econometricians*. New York: Academic Press.