Søren Johansen

Cointegration in Partial Systems and the Efficiency of Single Equation Analysis

Preprint
September
1990

4

Institute of Mathematical Statistics
University of Copenhagen

ISSN 0902-8846

Søren Johansen*

COINTEGRATION IN PARTIAL SYSTEMS AND THE EFFICIENCY OF SINGLE EQUATION ANALYSIS.

Preprint 1990 No. 4

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

September 1990

* This paper is part of a project supported by the Danish Social Science Research Council, acc. 14-5387.

Abstract

It is shown how one can estimate cointegration relations in a partially modelled system by the method of maximum likelihood. The estimator is compared with the estimator based on the full system, and it is shown that the two estimators are identical if the conditioning variables are weakly exogenous for the cointegrating relations and their adjustment coefficients. Suggestions are made for analysing the partial system, when there is no weak exogeneity.

0. *Introduction*

The general VAR model is often used to describe the statistical variation of economic time series. Economic systems, however, often have so many potentially useful variables that the system gets very large and it is tempting to consider partially specified systems, where only some of the variables are treated as endogenous, and model these conditionally on the remaining variables. If we are interested in the cointegrating relations and their adjustment coefficients, this will in general imply a loss of efficiency, and the purpose of this paper is to discuss this problem in a quantitative manner.

The results are close to those of Phillips (1990) who considered a slightly different model. The present paper is an attempt to apply similar ideas within the framework of the VAR model. Since the full details of the analysis of the likelihood function are rather tedious we here only state the neccesary results and refer the reader to the papers by Johansen (1990) and Johansen and Juselius (1990) where the full system has been treated in detail.

The paper starts with a discussion of weak exogeneity and we show that if the equations which have not been modelled, have no cointegration then the partial estimator is equal to the full maximum likelihood estimators and in this sense efficient.

In section 2 we show that if the partial system has more equations than cointegrating relations, then a suitable eigenvalue problem solves the estimation problem, and if there are fewer equations in the partial system than cointegrating relations, then the cointegrating relations can be determined by regression. Finally we givein section 3, without proof, the asymptotic distributions following Johansen (1990) and apply these to discuss the efficiency in section 4. Section 3 also contains some suggestions for analysing partial systems.

1. *Partial models and weak exogeneity*

The VAR model for cointegration can be written in the form

$$\Delta X_t = \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \alpha\beta' X_{t-k} + \phi D_t + \mu + \epsilon_t,\ t = 1,\ldots,T \tag{1.1}$$

where $X_{-k+1},\ldots,X_0$ are fixed and $\epsilon_1,\ldots,\epsilon_T$ are independent p–dimensional Gaussian variables with mean zero and variance matrix Λ. The vector D_t denotes seasonal dummies, centered at zero. The parameters in the model are composed by the short–run effects $\Gamma_1,\ldots,\Gamma_{k-1}$, the seasonal coefficients ϕ, the constant term μ, the covariance matrix Λ and the p×r matrices α (the adjustment coefficients) and β (the cointegrating relations) all of which vary unrestrictedly.

The maximum likelihood estimation has been treated in Johansen (1988), for the model without a constant term, and by Johansen (1990) and Johansen and Juselius (1990) for the above model.

A different method for estimation of a long–run steady state relation is to consider the first equation, say, given the other equations and in this *partial model* to estimate the parameters by regression analysis. We shall call such an analysis a *single equation analysis.* The estimate of β is called "the static long–run solution" of the autoregressive model, see Hendry (1989). It is clear that if there are more than one cointegration relation we shall by this analysis only determine a suitable combination of the cointegrating relations. It is also clear that even if there is only one cointegration relation the conditional analysis will in general be inefficient. In order to be able to discuss this problem in more detail we derive expressions for the asymptotic conditional covariance matrices of the estimable parameters such that the question of efficiency can be discussed quantitatively.

Let therefore a be a known p×m matrix of rank m and let $b = a_\perp$ be a (p×(p–m)) full rank matrix of vectors orthogonal to a. It is not difficult to show that

$$E(a'\Delta X_t | b'\Delta X_t, \Delta X_{t-1},\ldots,\Delta X_{t-k+1}, X_{t-k}) = \tag{1.2}$$

$$(a - b\Lambda_{bb}^{-1}\Lambda_{ba})'(\sum_{i=1}^{k-1}\Gamma_i\Delta X_{t-i} + \alpha\beta' X_{t-k} + \phi D_t + \mu) + \Lambda_{ab}\Lambda_{bb}^{-1}b'\Delta X_t,$$

and that

$$\text{Var}(a'\Delta X_t \mid b'\Delta X_t, \Delta X_{t-1}, \ldots, \Delta X_{t-k+1}, X_{t-k}) = \Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba} = \Lambda_{aa.b}. \tag{1.3}$$

where we have applied the notation $\Lambda_{aa} = a'\Lambda a$, $\Lambda_{ab} = a'\Lambda b$ etc.

The model

$$a'\Delta X_t = (a - b\Lambda_{bb}^{-1}\Lambda_{ba})'(\sum_{i=1}^{k-1}\Gamma_i\Delta X_{t-i} + \alpha\beta' X_{t-k} + \phi D_t + \mu) + \Lambda_{ab}\Lambda_{bb}^{-1}b'\Delta X_t + u_t,\ t = 1,\ldots,T. \tag{1.4}$$

Here u_t are independent Gaussian variables with mean zero and variance matrix $\Lambda_{aa.b}$ will be called a *partial* VAR *model* or a conditional model for $a'\Delta X_t$ given $b'\Delta X_t$ and past information.

If in particular $a' = (1,0,\ldots,0)$ we get a model for ΔX_{1t} given $\Delta X_{2t},\ldots,\Delta X_{pt}$ and for $a' = (I_{m\times m},0)$ we get a model for $\Delta X_{1t},\ldots,\Delta X_{mt}$ given $\Delta X_{m+1t},\ldots,\Delta X_{pt}$. We show in section 2 that for $m \geq r$ all cointegration vectors can be estimated, but for $m < r$ only certain linear combinations of the cointegration vectors. By formulating partial models in this way we are considering them as derived from a general fully specified VAR model. Thus even though only the partial model is analysed in detail for its cointegrating relations, the properties of the conditioning variables are given by the full model (1.1).

In general the marginal distribution of $b'\Delta X_t$ given the past contains information on the cointegrating relations, since they enter in the form

$$b'\alpha\beta' X_{t-k}.$$

Thus in order to get efficient estimation one will have to analyse both the marginal and the conditional model, i.e. the full model. There is one situation, however, where the conditional model has full information on the cointegrating relations, namely if $b'\Delta X_t$ is weakly exogenous for (α,β).

The basic result on weak exogeneity is given in

THEOREM 1.1: *If $b'\alpha = 0$, then $b'\Delta X_t$ is weakly exogenous for the parameter (α,β). Hence the maximum likelihood estimator for (α,β) in the full model (1.1) is the same as the maximum partial likelihood estimator in the partial model (1.4).*

Proof: The proof is a consequence of Corollary 6.2 in Johansen and Juselius (1990), but will be given in detail here. If $b'\alpha = 0$, then (α,β) does not appear in the model for $b'\Delta X_t$ given the past and hence the parameters of this distribution are

(1.5) $$\{b'\Gamma_i,\ i = 1,\dots,k-1,\ \Lambda_{bb}, b'\phi, b'\mu\}$$

whereas the conditional model of $a'\Delta X_t$ given $b'\Delta X_t$ (and the past) contains the parameters

(1.6) $$\{(a-\Lambda_{ab}\Lambda_{bb}^{-1}b)'\Gamma_i,\ i = 1,\dots,k-1,\ (a-\Lambda_{ab}\Lambda_{bb}^{-1}b)'(\alpha,\phi,\mu),\ \beta,\ \Lambda_{ab}\Lambda_{bb}^{-1},\ \Lambda_{aa.b}\}.$$

It is a well–known property of the multivariate Gaussian distribution that the parameters

(1.7) $$\{\Lambda_{bb},\ \Lambda_{ab}\Lambda_{bb}^{-1},\ \Lambda_{aa.b}\}$$

are variation free. In order to prove that (1.5) and (1.6) are variation free we must check that any choice of the quantities (1.5) and (1.6), compatible with the obvious restriction that variance matrices are symmetric and positive definite, makes it possible to reconstruct the parameters of the original model $\{\Gamma_i,\ i = 1,\dots,k-1,\ \alpha,\ \beta,\ \Lambda,\ \phi,\mu\}$.

That Λ can be reconstructed is clear from (1.7), and β can be found directly since it enters into (1.6). Since $b'\alpha = 0$ we have that $m \geq r$, which implies that α can be found from $a'\alpha$. Finally $b'\Gamma_i$ is given, and $a'\Gamma_i$ can be found from the first component in (1.6), which shows that also Γ_i, $i = 1,\dots,k-1$, can be found. The argument for ϕ and μ is similar. Thus the parameters in (1.5) and (1.6) are variation free, and since the joint density of $a'\Delta X_t$ and $b'\Delta X_t$ given the past factorizes into the product of the conditional and the marginal density it follows that $b'X_t$ is weakly exogenous for (α,β) if $b'\alpha = 0$, i.e. if the equations we condition on do not contain cointegration. The weak exogeneity of $b'\Delta X_t$ with respect to (α,β) implies that the maximum likelihood estimator in the full system (1.1) is identical to the (partial) maximum likelihood estimator derived from the partial system, since the likelihood function only depends on the parameter (α,β) through the partial likelihood function.

2. *Estimation of partial models*

The estimation of model (1.4) is most easily discussed if we first concentrate out the parameters $(a - b\Lambda_{bb}^{-1}\Lambda_{ba})'\Gamma_i$, $i = 1,...,k-1$, $(a - b\Lambda_{bb}^{-1}\Lambda_{ba})'\phi$, $(a - b\Lambda_{bb}^{-1}\Lambda_{ba})'\mu$ and $\Lambda_{ab}\Lambda_{bb}^{-1}$ by regressing $a'\Delta X_t$ and X_{t-k} on $b'\Delta X_t, \Delta X_{t-1},...,\Delta X_{t-k+1}, D_t, 1$. This gives residuals $R_{a.bt}$ and $R_{k.bt}$ respectively, and the model can be writtten as

(2.1) $$R_{a.bt} = \alpha_a\beta' R_{k.bt} + \text{error},$$

where

(2.2) $$\alpha_a = (a - b\Lambda_{bb}^{-1}\Lambda_{ba})'\alpha.$$

We define the product moment matrices:

(2.3) $$S_{ij.b} = T^{-1}\sum_{t=1}^{T} R_{i.bt}R_{j.bt}', \qquad i,j = a,k.$$

We can then prove

THEOREM 2.1. *If $m \leq r$ the cointegrating relations are the rows of the matrix $\Pi_a = (a-b\Lambda_{bb}^{-1}\Lambda_{ba})'\alpha\beta'$. This matrix is estimated by regression: $\hat{\Pi}_a = S_{ak.b}S_{kk.b}^{-1}$.*

Proof: If $m \leq r$, the matrix

(2.4) $$\Pi_a = \alpha_a\beta' = (a - b\Lambda_{bb}^{-1}\Lambda_{ba})'\alpha\beta'$$

varies in the space of all $m\times r$ matrices, and hence $\hat{\Pi}_a$ can be found by regression:

(2.5) $$\hat{\Pi}_a = S_{ak.b}S_{kk.b}^{-1},$$

which completes the proof.

Thus if $m \leq r$ the maximum likelihood estimation in the partial model is just ordinary least squares. If in particular $a' = (1,0,...,0)$ and Λ and α are partitioned accordingly into

$$\Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}, \ \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix},$$

then the 1×r matrix α_a is given by

$$(a - b\Lambda_{bb}^{-1}\Lambda_{ba})'\alpha = \begin{bmatrix} 1 \\ -\Lambda_{22}^{-1}\Lambda_{21} \end{bmatrix}' \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \alpha_1 - \Lambda_{12}\Lambda_{22}^{-1}\alpha_2,$$

which shows how the adjustment coefficients in equations 2,...,p (α_2) are combined with the adjustment coefficients in the first equation (α_1) to define the linear combination of β that is estimated. Note that even though one chooses equations without cointegration i.e. $\alpha_1 = 0$, then conditioning on ΔX_{2t} brings back cointegration into the first equation, if the error terms are correlated.

Another case of interest is to let m = r, in which case all cointegration vectors or rather the cointegration space sp(β) can be found by simple regression. We emphasize in particular

COROLLARY 2.2: *Single equation analysis is equivalent to maximum likelihood if the remaining equations contain no cointegration.*

Next consider the case m > r,

THEOREM 2.3: *If m > r, the cointegration vectors are found as* $\hat{\beta} = (\hat{v}_1,...,\hat{v}_r)$, *where* $(\hat{v}_1,...,\hat{v}_p)$ *are the eigenvectors in the eigenvalue problem*

(2.6) $$|\lambda S_{kk.b} - S_{ka.b}S_{aa.b}^{-1}S_{ak.b}| = 0,$$

which has solutions

$$\hat{\lambda}_1 > ... > \hat{\lambda}_m > \hat{\lambda}_{m+1} = ... = \hat{\lambda}_p = 0,$$

The estimate of $\alpha_a = (a - b\Lambda_{bb}^{-1}\Lambda_{ba})'\alpha$ *is given by*

$$(2.7) \qquad \hat{\alpha}_a = S_{ak.b}\hat{\beta}$$

Proof: The proof is identical to the proof in Johansen (1988) and exploits reduced rank regression as solved by Anderson (1951). The p–m last eigenvalues are zero since the rank of $S_{ka.b}S_{aa.b}^{-1}S_{ak.b}$ is m. An economical way of using this fact is to solve the dual problem instead, which only involves an eigenvalue routine for m×m matrices:

$$|\lambda S_{aa.b} - S_{ak.b}S_{kk.b}^{-1}S_{ka.b}| = 0,$$

which has the same positive eigenvalues and eigenvectors $w_1,...,w_m$. Then $\hat{u}_i = \lambda_i^{-\frac{1}{2}}S_{kk.b}^{-1}S_{ka.b}w_i$, will be the eigenvectors of (2.6) normalized by $\hat{u}'S_{kk.b}\hat{u} = I$.

One can then estimate the remaining parameters by ordinary least squares, keeping $\beta = \hat{\beta}_a$ fixed, since, once the optimal β is found, there only remains to solve a usual linear regression problem.

Thus for $m \leq r$ we can do with regression, and for $m \geq r$ an eigenvalue routine is needed. For m = r the two methods coincide. Note that by conditioning on simultaneous values of $b'\Delta X_t$ we reduce the dimensionality of the eigenvalue problem (2.6) by forcing p–m eigenvalues to be zero, thereby reducing the possible choice of the number of cointegrating relations from p to m. Note also that although the algebra of the solution is similar to that of maximum likelihood estimation in the full model (2.1) one only obtains these estimators if $b'\Delta X_t$ is weakly exogenous, i.e. if $b'\alpha = 0$.

3. *The asymptotic distribution of the estimators*

In order to derive the asymptotic properties of the estimators we first specify the properties of the process in terms of its characteristic polynomial:

$$A(z) = I(1-z) - \sum_{i=1}^{k-1} \Gamma_i(1-z)z^i - \alpha\beta'z^k.$$

We assume that it has all roots outside the unit circle or possibly at the point z = 1. This condition excludes explosive processes. Further we assume the condition that $\alpha_\perp' \Psi \beta_\perp$ has full rank, see Johansen (1990), which guarantees that X_t is integrated of order 1. Here $-\Psi$ denotes the derivative of A(z) at the point z = 1. We define $C = \beta_\perp(\alpha_\perp' \Psi \beta_\perp)^{-1}\alpha_\perp'$, which has an important role in the Granger representation of X_t, since the non–stationary part of X_t is given by $C \sum_{s=1}^{t} \epsilon_s + C\mu t = C \sum_{s=0}^{t} \epsilon_s + \tau t$.

The limit

$$T^{-\frac{1}{2}} \sum_{s=1}^{[Tt]} \epsilon_s \xrightarrow{w} W(t),$$

defines a Brownian motion W in p dimensions with covariance matrix Λ. We use this to describe the asymptotic properties of the estimators. The asymptotic distribution of X_t is different in the direction τ, where the linear trend is dominating, and in the direction γ orthogonal to τ and β such that (β,γ,τ) spans R^p. It follows from Grangers representation theorem that

$$T^{-\frac{1}{2}}\gamma' X_{[Tt]} \xrightarrow{w} \gamma' CW(t) \text{ and } T^{-1}\ \tau' X_{[Tt]} \xrightarrow{P} \tau'\tau t$$

The parameter β is not identified, since $\alpha\xi$ and $\beta\xi'^{-1}$ give the same value of Π and hence are observationaly equivalent. We can therefore normalize β as follows: Let c be any p×r matrix such that $\beta' c$ has full rank, and define $\beta_c = \beta(c'\beta)^{-1}$ and correspondingly $\alpha_c = \alpha(\beta' c)$ such that $\alpha\beta' = \alpha_c\beta_c'$. For r = 1, one often chooses $c' = (-1,0,\ldots,0)$ such that β_c is the normalization by the coefficient to the first component of β. Thus the normalization by c corresponding to " solving" the relation $\beta' X = 0$ for $c'X$.

If r > 1 we can sometimes choose $c = (-I_{r\times r}, 0)$ so that we isolate certain variables $X_{1t},\ldots,X_{rt}$ and express the stable relations in the form that $X_{1t},\ldots,X_{rt}$ are linear functions of $X_{r+1,t},\ldots,X_{pt}$. This clearly requires that we can assume that the variables $X_{r+1,t},\ldots,X_{pt}$ are not cointegrated. If $\beta' = (\beta_1',\beta_2')$ with β_1 r×r then $\beta_c' = (-I,B')$ with B

$= -\beta_2\beta_1^{-1}$. Thus if X = (Y_1, Y_2) this corresponds to solving the relation $\beta'X = 0$ as $Y_1 = BY_2$. This identifies the coefficients to Y_2 and below we give the asymptotic distribution of these coefficients.

THEOREM 3.1: *For $m \geq r$, and α_a and β of rank r the asymptotic distribution of $\hat{\beta}_c = \hat{\beta}(c'\hat{\beta})^{-1}$ is given by*

$$T(\hat{\beta}_c - \beta_c) \overset{w}{\to} (I - \beta_c c')(\gamma'\gamma)^{-1}\gamma(\int G(u)G(u)'du)^{-1}\int G(dV)'(c'\beta)^{-1}, \tag{3.1}$$

where

$$G(t) = (\gamma'\gamma)^{-1}\gamma'C[W(t) - \int W(u)du - (t-1/2)\int W(u)(u-1/2)du/\int(u-1/2)^2 du], \tag{3.2}$$

i.e. the Brownian motion corrected for a linear trend.

Further,

$$V = (\alpha'(\Lambda^{-1} - b\Lambda_{bb}^{-1}b')\alpha)^{-1}\alpha'(\Lambda^{-1} - b\Lambda_{bb}^{-1}b')W. \tag{3.3}$$

The asymptotic distribution of $\hat{\alpha}_a$ normalized by c is given by

$$T^{\frac{1}{2}}(\hat{\alpha}_{ac} - \alpha_{ac}) \overset{w}{\to} N_{p\times r}(0, \Lambda_{aa.b} \otimes (c'\beta)(\beta\Sigma_{kk.b}\beta)^{-1}(\beta'c)). \tag{3.4}$$

The integral $\int G(u)G(u)'du$ is a $(p-r)\times(p-r)$ matrix of stochastic variables found as ordinary Riemann integrals of the continuous functions $G_i(u)G_j(u)$, whereas the integral $\int G(dV)'$ is a matrix where the (i,j)'th element is given by a stochastic integral of the process G_i with respect to the process V_j. Finally $\Lambda_{aa.b} = a'\Lambda a - a'\Lambda b(b'\Lambda b)^{-1}b'\Lambda a$ and $\Sigma_{kk.b} = \Sigma_{kk} - \Sigma_{k0}b(b'\Sigma_{00}b)^{-1}b'\Sigma_{0k}$.

Proof. The proof of this is identical to the proof given in Johansen (1990) from which we get the expression for V

$$V = (\alpha_a'\Lambda_{aa.b}^{-1}\alpha_a)^{-1}\alpha_a'\Lambda_{aa.b}^{-1}(a - b\Lambda_{bb}^{-1}\Lambda_{ba})'W.$$

From the identity

$$\Lambda^{-1} = (\mathrm{a},\mathrm{b})\left[\begin{bmatrix}\mathrm{a}'\\ \mathrm{b}'\end{bmatrix}'\Lambda\,(\mathrm{a},\mathrm{b})\right]^{-1}\begin{bmatrix}\mathrm{a}'\\ \mathrm{b}'\end{bmatrix} =$$

$$\mathrm{b}\Lambda_{\mathrm{bb}}^{-1}\mathrm{b}' + (\mathrm{a}-\mathrm{b}\Lambda_{\mathrm{bb}}^{-1}\Lambda_{\mathrm{ba}})\Lambda_{\mathrm{aa.b}}^{-1}(\mathrm{a}-\mathrm{b}\Lambda_{\mathrm{bb}}^{-1}\Lambda_{\mathrm{ba}})',$$

the result follows. Note that in general V and G are dependent, which makes it difficult to conduct inference on the structure of β, see below.

Of special interest is the case where $\mathrm{b}'X_t$ is weakly exogenous for β, or $\mathrm{b}'\alpha = 0$.

COROLLARY 3.2: *If $b'\alpha = 0$ then estimation of β from (1.4) is the maximum likelihood estimation in the full model (1.1) and its asymptotic distribution is for $m \geq r$ given by by (3.1), (3.2) and (3.3) with*

(3.5) $$V = (\alpha'\Lambda^{-1}\alpha)^{-1}\alpha'\Lambda^{-1}W,$$

which is stochasticaly independent of G.

A consistent estimator of the asymptotic conditional variance of $T(\hat{\beta}_c-\beta_c)$ is given by

(3.6) $$(I-\hat{\beta}_c c')\hat{v}\hat{v}'(I-c\hat{\beta}_c') \otimes (\hat{\beta}'c)^{-1}(\hat{D}^{-1}-I)(c'\hat{\beta})^{-1},$$

where $\hat{v} = (\hat{v}_{r+1},\ldots,\hat{v}_p)$, see (2.6)

The proof follows Johansen (1990). Thus under weak exogeneity of b'X for β the asymptotic distribution of β is mixed Gaussian and asymptotic inference concerning hypotheses on β can be performed by the χ^2 distribution see Johansen (1990), Phillips (1990) and Jeganathan (1989).

If $\alpha'\mathrm{b} = 0$, i.e. we have weak exogeneity, then estimation conditionally on b'X gives efficient estimation of β. If $\alpha'\mathrm{b}$ is not zero, then the asymptotic distribution of β is a complicated distribution, since G and V are not independent. It seems therefore of interest to derive a test that the hypothesis $\alpha'\mathrm{b} = 0$ is actually satisfied. This is nothing but a linear restriction on α, and the test can be conducted as a likelihood ratio test in the full model (1.1) using the likelihood ratio test as given in Johansen and Juselius (1990). If m $\geq$

r and if we want to avoid the full model analysis, especially the eigenvalue routine for estimating β in the full model, we can estimate the cointegrating vectors from the partial system. The estimate $\hat{\beta}$ is superconsistent, and inserting it into the equations for $b'\Delta X_t$, one can, for fixed $\beta = \hat{\beta}$, test the hypothesis that the coefficient to $\beta' X_{t-k}$ is zero by an approximate F–test. Thus the full analysis of the model (1.1) is replaced by:

(3.7) *A cointegration analysis of the (small) partial model for* $a'\Delta X_t$ *given* $b'\Delta X_t$ *and the past.*

(3.8) *A regression analysis, for fixed* $\hat{\beta}$, *of the (large) marginal model for* $b'\Delta X_t$ *given the past, and a misspecification test to check the weak exogeneity.*

In order to illustrate the difficulties met in the case of no weak exogeneity we consider the situation where $m \geq r$, and a simple hypothesis on β :

$$\text{H: } \beta = \beta_0.$$

A Taylors expansion of the likelihood function shows that when $\beta = \beta_0$ it holds that as $T \to \infty$,

(3.9) $$-2\ln Q \xrightarrow{w} \mathrm{tr}\{\mathrm{Var}(V)^{-1}\int(dV)G'[\int GG'du]^{-1}\int G(dV)'\},$$

with G and V given by (3.2) and (3.3).

It is clear that the limiting distribution (3.9) is invariant under multiplication of either V or G by non–singular matrices. This means that the limiting distribution only depends on the canonical correlations between the linear combinations of the process W given by (3.3) and

$$\tilde{G} = (\gamma'\gamma)^{-1}\gamma' CW,$$

and hence on the solutions $(\rho_1,\ldots,\rho_r)$ of the equation

$$|\rho \mathrm{Var}(V) - \mathrm{Cov}(V,\tilde{G})\mathrm{Var}(\tilde{G})^{-1}\mathrm{Cov}(G,\tilde{V})| = 0,$$

or equivalently, after some reductions,

$$|\rho\alpha_a'\Lambda_{aa.b}^{-1}\alpha_a - \alpha'b\Lambda_{bb}^{-1}b'\Lambda C'\gamma(\gamma'C\Lambda C'\gamma)^{-1}\gamma'C\Lambda b\Lambda_{bb}^{-1}b'\alpha| = 0. \tag{3.10}$$

In principle the analysis of the partial system can thus be supplemented by the last step:

(3.11) *If weak exogeneity fails then estimate the canonical correlation from (3.10) and simulate the distributions involved in the asymptotic tests of hypotheses concerning β.*

We finally give some results on the estimation of Π

THEOREM 3.3: *The asymptotic distribution of* $\hat{\Pi}_a = \hat{\alpha}_a\hat{\beta}' = \hat{\alpha}_{ac}\hat{\beta}_c'$ *is given by*

$$T^{\frac{1}{2}}(\hat{\Pi}_a - \Pi_a) \overset{W}{\to} N_{p\times p}(0,\Lambda_{aa.b} \otimes \beta(\beta\Sigma_{kk.b}\beta)^{-1}\beta') \tag{3.12}$$

$$T(\hat{\Pi}_a - \Pi_a)\gamma \overset{W}{\to} \alpha \int (dV)G'[\int GG'du]^{-1}\gamma'\gamma \tag{3.13}$$

Note that the relation (3.12) is not very useful for linear combinations like $T^{\frac{1}{2}}(\hat{\Pi}_a - \Pi_a)\gamma$ since it only implies that the limit is zero. A different normalization is needed in the direction γ. A similar remark holds for $T^{3/2}(\hat{\Pi}_a - \Pi_a)\tau$.

Proof. The proof follows for m $\geq$ r from Theorem 3.1, and for m $<$ r one can apply the same proof.

The asymptotic distribution of $\hat{\Pi}_a$ thus depends on which direction we are interested in, and this has consequences for hypothesis testing concerning β. If we consider for instance linear restrictions $K'\beta = 0$, then the test will involve $K'\hat{\beta}$, or possibly $K'\hat{\Pi}'$. The first result (3.12) is not useful for this since the asymptotic distribution of such linear combinations is degenerate. This is where the second result comes in. The third result about $\tau'\hat{\beta}$ would only be of interest if K happend to be proportional to τ, which is unlikely, and the result is therefore not given.

4. *Efficiency results*

When we analyse the conditional system for information on β and $m \geq r$, then at least we get superconsistency for β, but the limit distribution is rather complicated and there seems to be no easy way to make inference on hypotheses on β. If we nevertheless compare $\hat{\beta}_{part}$ and $\hat{\beta}_{full}$ we see that in both cases we get a limit distribution of the type (3.1) with the same function G, but with different V. From (3.3) we find

$$V_{part} = (\alpha'(\Lambda^{-1}-b\Lambda_{bb}^{-1}b')\alpha)^{-1}\alpha'(\Lambda^{-1}-b\Lambda_{bb}^{-1}b')W$$
$$V_{full} = (\alpha'\Lambda^{-1}\alpha)^{-1}\alpha'\Lambda^{-1}W.$$

It is easy to see that

$$Var(V_{part}) \geq Var(V_{full})$$

in the sense of positive definite matrices, and that equality only holds for $\alpha'b = 0$.

Hence we loose some efficiency by increasing the variation in the limiting distribution, and only when we have weak exogeneity do we get no loss of efficiency in this sense. In case of weak exogeneity we also get the possibility to make easy inference since the limit distribution is a mixture of Gaussian distributions, and asymptotic inference concerning β can be conducted via the χ^2 distribution, see Johansen (1990).

If $m < r$ the first sense in which we loose efficiency is that we can not estimate all the cointegrating vectors, only certain linear combinations of them. Since the estimation of β is really the estimation of the Π matrix the results are therefore given by Theorem 3.2.

Now consider the estimation of α. By analysing a partial system we can only estimate

$$\alpha_a = (a - \Lambda_{ab}\Lambda_{bb}^{-1})'\alpha$$

i.e. some linear combinations of the α's. This is not so important, since evidently the equations we are analysing are the interesting ones.

The asymptotic distribution is Gaussian and has a limiting variance given by

$$\Omega_{part} = \Lambda_{aa.b} \otimes (\beta'\Sigma_{kk.b}\beta)^{-1}$$

compared to the variance of the estimate of α_a we would obtain from the full system:

$$\Omega_{\text{full}} = \Lambda_{\text{aa.b}} \otimes (\beta'\Sigma_{\text{kk}}\beta)^{-1}$$

It is seen that $\Omega_{\text{part}} \geq \Omega_{\text{full}}$, since

$$\beta'\Sigma_{\text{kk.b}}\beta = \beta'\Sigma_{\text{kk}}\beta - \beta'\Sigma_{\text{k0}}\text{b}(\text{b}'\Sigma_{00}\text{b})^{-1}\text{b}'\Sigma_{0\text{k}}\beta \leq \beta'\Sigma_{\text{kk}}\beta.$$

Now $\beta'\Sigma_{\text{k0}} = \beta'\Sigma_{\text{kk}}\beta\alpha'$, such that equality only holds if $\text{b}'\alpha = 0$. Hence in the case of weak exogeneity we loose no information in the analysis of the partial system as long as we are only interested in the coefficients in the partial system.

The efficiency question can also be asked in relation to how much we loose in the estimation of Π and Theorem 3.2 shows that it can be answered by the above considerations concerning α and β.

5. *References*

Anderson, T.W. (1951): "Estimating linear restrictions on regression coefficients for multivariate normal distributions," *Annals of Mathematical Statistics,* 22, 327–351.

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

Granger, C. J. (1981): " Some Properties of Time Series Data and their Use in Econometric Model Specification,"*Journal of Econometrics*, 16, 251–276.

Granger, C. W. J., and A. A. Weiss (1983): "Time Series Analysis of Error Correction Models," in *Studies in economic time series and multivariate statistics*, ed. by S. Karlin, T. Amemiya and L. A. Goodman. New York: Academic Press.

Hendry D. (1989): "PC–GIVE an interactive econometric modelling system," University of Oxford.

Jeganathan, P. (1988): "Some aspects of asymptotic theory with applications to time series models, " Technical report The University of Michigan.

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic*

Dynamics and Control, 12, 231–254.

Johansen, S. and K. Juselius (1990): "Maximum Likelihood estimation and Inference on Cointegration – with Applications to the demand for money," *Oxford Bulletin of Economics and Statistics* 52, 169–210.

Johansen, S. (1990): " Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models." Forthcoming in *Econometrica.*

Phillips, P. C. B. (1990): "Optimal inference in cointegrated systems," Cowles Foundation discussion paper No.866.

PREPRINTS 1989

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK, TELEPHONE + 45 1 35 31 33 .

No. 1 Bertelsen, Aksel: Asymptotic Expansion of a Complex Hypergeometric Function.

No. 2 Davidsen, Michael and Jacobsen, Martin: Weak Convergence of Twosided Stochastic Integrals, with an Application to Models for Left Truncated Survival Data.

No. 3 Johansen, Søren: Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models.

No. 4 Johansen, Søren and Juselius, Katarina: The Full Information Maximum Likelihood Procedure for Inference on Cointegration - with Applications.

No. 5 Thompson, Steven K.: Adaptive Cluster Sampling.

No. 6 Thompson, Steven K.: Adaptive Cluster Sampling: Designs with Primary and Secondary Units.

No. 7 Thompson, Steven K.: Stratified Adaptive Cluster Sampling.

No. 8 Johansen, Søren: The Power Function of the Likelihood Ratio Test for Cointegration.

No. 9 Andersson, S.A. and Perlman, M.D.: Lattice-Ordered Conditional Independence Models for Missing Data.

PREPRINTS 1990

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, DK-2100 COPENHAGEN Ø, DENMARK, TELEPHONE + 45 31 35 31 33.

No. 1 Johansen, Søren and Juselius, Katarina: Some Structural Hypotheses in a Multivariate Cointegration Analysis of the Purchasing Power Parity and the Uncovered Interest Parity for UK.

No. 2 Tjur, Tue: Analysis of Variance and Design of Experiments.

No. 3 Johansen, Søren: A Representation of Vector Autoregressive Processes Integrated of Order 2.

No. 4 Johansen, Søren: Cointegration in Partial Systems and the Efficiency of Single Equation Analysis.