Paolo Paruolo

Asymptotic Inference on the Moving Average Impact Matrix in Cointegrated I(1) VAR Systems

Preprint
May
1992
2

Institute of Mathematical Statistics
University of Copenhagen

ISSN 0902-8846

Paolo Paruolo

ASYMPTOTIC INFERENCE ON THE MOVING AVERAGE IMPACT MATRIX IN COINTEGRATED I(1) VAR SYSTEMS

Preprint 1992 No. 2

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

May 1992

ASYMPTOTIC INFERENCE ON THE MOVING AVERAGE IMPACT MATRIX IN COINTEGRATED I(1) VAR SYSTEMS

Paolo Paruolo°

Abstract

The asymptotic distributions of the maximum likelihood estimators of the moving average impact matrix and of the common trends linear combinations in vector auto-regressive processes integrated of order one are presented. In order to derive the results, a relation between properly normalized orthogonal complements is obtained, which may be of separate interest. Finally Wald type tests on the moving average impact matrix and of the common trends linear combinations are discussed.

° Dipartimento di Scienze Statistiche "Paolo Fortunati", Università di Bologna, Via Belle Arti 41, 40126 Bologna, Italy.
Current address until May 15 1992:
Institute of Mathematical Statistics, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark.

1 Introduction

Integrated processes have been the focus of much recent research. For these processes the ideas of cointegration, Engle and Granger (1987), and of common trends, Stock and Watson (1988), are dual concepts, which are related to the rank of the autoregressive and of the moving average impact matrices respectively. The moving average impact matrix $C = C(1)$ plays an important role in the definition of common trends in integrated systems of order one, I(1), see e.g. Stock Watson (1988); in particular the row space of C, here indicated as $\alpha_\perp'$, determines the linear combinations of the innovations that form the random walk component of the system.

In this paper we define maximum likelihood estimators of C and $\alpha_\perp$ in the vector autoregressive case and determine their asymptotic distributions. As it is to be expected, $\hat{C}$ and $\hat{\alpha}_\perp$ are functions of the autoregressive parameters, so that inference can be based on the likelihood analysis of Johansen (1991). These estimators involve quantities that are defined to be orthogonal to some matrices related to the autoregressive parameters, i. e. orthogonal complements to functions of the autoregressive coefficients. For this kind of problem it is useful to relate inference about a basis of a linear space of interest with respect to inference on the basis of the linear space orthogonal to it. To this end we state a simple linear relation between normalized bases of orthogonal spaces; this relation shows explicitly how the problem of inference about two orthogonal bases is really a *single* problem.

These ideas are exploited in the definition of a properly normalized estimator for $\alpha_\perp$ and in the derivation of the asymptotic distributions; the distributions are gaussian and simple consistent estimators of the asymptotic covariance matrices are readily available by modifying standard output least squares covariance matrices, and lead to conventional χ^2 asymptotic Wald tests for general smooth hypotheses.

Throughout the paper $\xrightarrow{w}$ will denote weak convergence, the *vec* operator will denote the column stacking operator and $\otimes$ the Kronecker product, i.e. $A \otimes B = [a_{ij}B]$. The rest of the paper is organized as follows: section 2 present the model; section 3 discusses an alternative equivalent parametrization;

section 4 states the relation between properly normalized bases of orthogonal spaces; in section 5 the asymptotic distributions of C and $\alpha_\perp$ are derived, while section 6 discusses Wald tests on $\alpha_\perp$ and section 7 reports some remarks on the Wald tests on C; section 8 concludes.

2 The model

Consider the gaussian $p \times 1$ vector autoregressive process X_t

(2.1) $$A(L)X_t = \mu + \theta D_t + \varepsilon_t$$

where $A(L) = I - A_1 L - \ldots - A_k L^k$ is a finite matrix polynomial, L is the backward shift operator, $LX_t = X_{t-1}$, μ is a vector of constants, D_t is a vector of seasonal dummies orthogonal to the constant and ε_t is i.i.d. $N(0, \Omega)$. Model (2.1) has been studied by several authors, e.g. Ahn and Reinsel (1990) and Johansen (1991), under the assumption that all the roots of $|A(z)| = 0$ lie either outside the unit circle in the complex plane or at the point $z = 1$; we will also adopt the same assumption here. As it is well known, the roots at the point $z = 1$ are responsible for the non-stationarity of the system. The following reparametrization of the process (2.1) is of interest for the statistical analysis considered in this paper

(2.2) $$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \mu + \theta D_t + \varepsilon_t$$

where $\Pi = -A(1)$ is the autoregressive impact matrix with a sign change. Parametrization (2.2) is equivalent to the one adopted in Johansen (1991); as already noted by several authors, the levels term X_{t-n} can be specified for any $n = 1, ..., k$ with the only effect of changing the definition of the Γ_i matrices[1]. In the next section we will consider an equivalent parametrization which is also relevant. In the following it will be helpful to let $\Psi = -\dot{A}(1) - \Pi = I - \Sigma_{i=1}^{k-1} \Gamma_i$, where $\dot{A}(1) = (dA(z)/dz)_{z=1}$ is the first derivative of the $A(z)$ polynomial at the point $z = 1$, and by $b_\perp$ the orthogonal complement of a $p \times r$, $p > r$, matrix b of full column rank, i.e. a $p \times (p-r)$ matrix such that $b' b_\perp = 0$ and $(b, b_\perp)$ span $\Re^p$.

[1] For an application of this property to lag length determination see Marzocchi, Mulargia and Paruolo (1991).

As it is well known, see Johansen's (1991) Granger representation theorem[2], if Π has reduced rank r, $\Pi = \alpha\beta'$ for α, β full column rank $p \times r$ matrices, and if $\alpha_\perp'\Psi\beta_\perp$ has full rank $p - r$, the process (2.1) is non-stationary with representation

(2.3) $$X_t = C\sum_{i=1}^{t}\varepsilon_i + C\mu t + C(L)\theta\sum_{i=1}^{t}D_t + X_0 + Y_t - Y_0$$

where

(2.4) $$C = \beta_\perp(\alpha_\perp'\Psi\beta_\perp)^{-1}\alpha_\perp'$$

and Y_t is a stationary process, $\beta'X_0 = \beta'Y_0$. Representation (2.4) makes clear that X_t is non-stationary while the first difference process ΔX_t, $\Delta \equiv 1 - L$, is stationary, that is X_t is integrated of order one, as the application of the Δ operator once induces stationarity. The linear combinations β' of the process are stationary, as can be verified by pre-multiplying (2.3) by β', so that X_t is co-integrated in the sense of Engle and Granger (1987). Moreover the differenced process has representation

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

where $C = C(1)$. C is therefore the impact matrix in the moving average representation of the differenced process. Note also that the random walk term $\sum_i \varepsilon_i$ in (2.3) that determines the stochastic non-stationarity of X_t enters into the process through C. Stock and Watson (1988) call $\alpha_\perp'\sum_{i=1}^{t}\varepsilon_t + \alpha_\perp'\mu t$ the *common trends* in the system.

Definition (2.4) makes clear how the moving average impact matrix is related to the autoregressive impact matrix $\Pi \equiv \alpha\beta'$ through the orthogonal complements of α and β. It will be the focus of section 5 and 7 to discuss inference on C; there are many related works on this topic, including Park (1990), Lutkepohl Reimers (1988), Warne (1990).

The likelihood based statistical analysis of model (2.2) under the above assumptions is performed through reduced rank regression, see Anderson (1951), Johansen (1991) and reference therein. Let

---

[2] In Johansen (1991), theorem 4.1 is stated in terms of the matrix $\alpha_\perp'(-dA(z)/dz)_{z=1}\beta_\perp$ instead of $\alpha_\perp'\Psi\beta_\perp$, see Johansen's eq. (4.4). Nevertheless the two conditions are equivalent, as

$$\alpha_\perp'\Psi\beta_\perp = \alpha_\perp'(-\dot{A}(1))\beta_\perp - \alpha_\perp'\Pi\beta_\perp = -\alpha_\perp'(\dot{A}(1))\beta_\perp$$

$y_t \equiv \Delta X_t$, $x_t \equiv X_{t-1}$, $z_{1t} = (\Delta X_{t-1}', \ldots, \Delta X_{t-k+1}')'$, $z_{2t} \equiv (1, D_t')'$, $z_t \equiv (z_{1t}', z_{2t}')'$; equation (2.2) can now be rewritten as

(2.6)
$$y_t = \alpha\beta' x_t + (\Gamma, \delta) z_t + \varepsilon_t$$

where $\Gamma \equiv (\Gamma_1, \ldots, \Gamma_{k-1})$ and $\delta \equiv (\mu, \theta)$. The maximum of the likelihood function is obtained solving the eigenvalue problem

(2.7)
$$| \lambda M_{xx.z} - M_{xy.z} M_{yy.z}^{-1} M_{yx.z} |$$

where $M_{ab.c} \equiv M_{ab} - M_{ac} M_{cc}^{-1} M_{cb}$, $M_{ab} = T^{-1} \sum_{t=1}^{T} a_t b_t'$, $a, b = y, x, z$ are respectively conditional and unconditional sample product moments. Let $(u_1, \ldots, u_p)$ be the ordered eigenvectors associated with the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$ of (2.7). For any fixed value of r, the maximum likelihood estimators are then given by

(2.8)
$$\hat{\beta} = (u_1, \ldots, u_r) \qquad (\hat{\Gamma}, \hat{\delta}) = (M_{yz} - \hat{\alpha}\hat{\beta}' M_{xz}) M_{zz}^{-1}$$
$$\hat{\alpha} = M_{yx.z} \hat{\beta} (\hat{\beta}' M_{xx.z} \hat{\beta})^{-1} \qquad \hat{\Omega} = M_{yy.z} - M_{yx.z} \hat{\beta} (\hat{\beta}' M_{xx.z} \hat{\beta})^{-1} \hat{\beta}' M_{xy.z}$$

see Johansen (1991) eq. (3.6) through (3.12). It is useful to slightly rearrange the computation with respect to Γ. Since the estimator of Γ is just a regression estimator for fixed $\beta = \hat{\beta}$, it is useful to rewrite (2.6) substituting $\beta = \hat{\beta}$ as follows

$$\begin{aligned} y_t &= (\alpha, \Gamma) \begin{pmatrix} \hat{\beta}' x_t \\ z_{1t} \end{pmatrix} + \delta z_{2t} + \varepsilon_t \\ &= (\alpha, \Gamma) v_t + \delta z_{2t} + \varepsilon_t \end{aligned}$$

The estimator of (α, Γ) can then be obtained in two steps[3]: first regress y_t and $v_t \equiv (X_{t-1}'\hat{\beta}, z_{1t}')'$ on z_{2t} obtaining residuals $y_{t.z2}$ and $v_{t.z2}$ respectively; next estimate (α, Γ) by regressing $y_{t.z2}$ on $v_{t.z2}$. In this final regression it is convenient to save the standard output covariance matrix produced by the least squares program $T^{-1} M_{vv.z2}^{-1} \otimes \hat{\Omega}$ for further calculations, see section 5 [4].

[3] This follows from the known properties of the least squares.

[4] We assume that a multivariate regression least squares program provides as an output the variance of vec(τ) for a model of the form $y_t = \tau x_t + \eta_t$.

In order to obtain the maximum likelihood estimator of the matrix C in eq. (2.4), the maximum likelihood estimators of $\alpha_\perp$ and $\beta_\perp$ are needed. A particular normalization of $\hat{\alpha}_\perp$ and $\hat{\beta}_\perp$ will be proposed in section 4. Once these estimators are defined, the resulting estimator for C is, by the invariance property of maximum likelihood estimators, given by

$$\hat{C} = \hat{\beta}_\perp(\hat{\alpha}_\perp{}'\hat{\Psi}\hat{\beta}_\perp)^{-1}\hat{\alpha}_\perp{}' \tag{2.10}$$

where $\hat{\Psi} = I_p - \Sigma_{i=1}^{k-1}\hat{\Gamma}_i = I - \hat{\Gamma}(i_{k-1} \otimes I_p)$ and $i_n = (1, \ldots, 1)'$ is an $n \times 1$ vector of ones, so that the matrix $(i_{k-1} \otimes I_p)$ simply sums blocks of size p. Note also that $\hat{\Psi} - \Psi = -(\hat{\Gamma} - \Gamma)(i_{k-1} \otimes I_p)$.

3 An equivalent parametrization

Although the parametrization of the model and the statistical calculations reviewed in the previous section are adequate for making inference on C and $\alpha_\perp$, in this section the following equivalent parametrization of the model is presented, in which the matrix $\Phi \equiv I - \Psi$ can be directly estimated as a regression coefficient matrix:

$$\Delta X_t = \Pi X_{t-1} + \Phi \Delta X_{t-1} + \sum_{i=1}^{k-2} \phi_i \Delta^2 X_{t-i} + \mu + \theta D_t + \varepsilon_t \tag{3.1}$$

Such a parametrization entails a different, although equivalent, set of calculations, which may be found easier to implement. One of the possible ways of deriving (3.1) is to consider the polynomial $\Gamma(L) = \Gamma_1 + \Gamma_2 L + \ldots + \Gamma_{k-1}L^{k-2}$ in (2.2) and decompose it as $\Gamma(L) = \Phi + (1-L)\phi(L)$, where $\Phi \equiv \Gamma(1) = \Sigma_{i=1}^{k-1}\Gamma_i$. Now, since $\Psi \equiv I - \Sigma_{i=1}^{k-1}\Gamma_i$, it is clear how Ψ and Φ are related, i.e. $\Psi = I - \Phi$. Note also that $\hat{\Psi} - \Psi = -(\hat{\Phi} - \Phi)$ [5].

Other parametrizations are possible; for instance the first difference term ΔX_{t-n} in the right-hand-side of (3.1) could be specified at $n = 1, \ldots, k-1$ with the only effect of changing the definition of the ϕ_i matrices. Moreover the other parametrization

$$\Delta^2 X_t = \Pi X_{t-1} + (\Phi - I)\Delta X_{t-1} + \sum_{i=1}^{k-2} \phi_i \Delta^2 X_{t-i} + \mu + \theta D_t + \varepsilon_t$$

[5] When considering the distributions of $\hat{\Psi}$ and $\hat{\Phi}$, the above equality just states that the two distributions are mirror images.

could be used with no effect on the statistical analysis of the model under the mantained assumptions $\Pi = \alpha\beta'$, rank $(\alpha_{\perp}'\Psi\beta_{\perp}) = p - r$. The latter formulation is the one adopted by Johansen (1992) in the I(2) analysis under different assumptions.

Parametrization (3.1) is therefore a way of focusing on the parameters of interest, as C depends on $\Pi \equiv \alpha\beta'$ and Φ and does not depend on the other parameters, which can be concentrated out of the likelihood function in the statistical analysis. Note that (3.1) is just a rearrangement of the regressors in z_t in (2.6). For the purpose of the statistical calculations we define $w_{1t} \equiv \Delta X_{t-1}$, $w_{2t} \equiv (\Delta^2 X_{t-1}', \ldots, \Delta^2 X_{t-k+2}', 1, D_t')'$, $w_t \equiv (w_{1t}', w_{2t}')'$, $\phi \equiv (\phi_1, \ldots, \phi_{k-2})$; can one thus rewrite (3.1) as

$$y_t = \alpha\beta' x_t + \Phi w_{1t} + (\phi, \delta) w_{2t} + \varepsilon_t \qquad (3.2)$$

$$= \alpha\beta' x_t + (\Phi, \phi, \delta) w_t + \varepsilon_t$$

The statistical calculations in the reparametrized model can then be arranged in the following way:

1. Maximize the likelihood function with respect to β; the calculations are exactly the same as in (2.7), (2.8), as the projection of y_t and x_t on z_t or w_t are identical, for the known properties of the least squares; one thus obtains $\hat{\beta}$ and $\hat{\Omega}$ in (2.8).
2. Next consider the maximization of the likelihood function with respect to Φ, for fixed $\beta = \hat{\beta}$. From eq. (3.2) one sees that one can define the vector of variables $s_t \equiv (X_{t-1}'\hat{\beta}, w_{1t}')'$ and regress y_t and s_t on w_{2t}, obtaining respectively residuals $y_{t.w2}$ and $s_{t.w2} = (X_{t.w2}'\hat{\beta}, w_{1t.w2}')'$; the maximum likelihood estimate of Φ is then obtained by regressing $y_{t.w2}$ on $s_{t.w2}$. In this final regression the usual output covariance matrix provided by the least squares program is $T^{-1}M_{ss.w2}^{-1} \otimes \hat{\Omega}$, which must then be stored for the further calculations, see section 5.
3. Finally the estimates of (ϕ, δ) can be obtained by regression substituting the estimates of $\Pi = \alpha\beta'$ and Φ back in (3.2).

It is important to stress that the estimates obtained by the above procedure are exactly the same as the ones obtained in section 2. The purpose of the above reparametrization consists in obtaining the estimate of Φ directly as a regression coefficient; one thus can also save the estimate of the covariance matrix of $(\hat{\alpha}, \hat{\Phi})$ directly as a standard computer output.

Therefore in the following no distinction will be made between the estimators or the models of section 2 and the present section. Note, also, that a single set of calculations is needed, and it will be a matter of convenience which one to use in applications. Finally note that for $k = 2$ (2.2) and (3.1) coincide.

4 Orthogonal spaces and the relation between bases

In this section we analyze the relation between a matrix and its orthogonal complement. Consider in general a full column rank $p \times r$ matrix β, $p > r$ and the linear space of all the vectors obtained as $\beta\tau$ for any τ; denote the resulting linear space as $\mathcal{B} \subset \Re^{p \times r}$. Consider now all the possible vectors that are orthogonal to the vectors in $\mathcal{B}$; they can be represented as $\beta_\perp\gamma$, where $\beta_\perp$ is a $p \times (p - r)$ matrix such that $\beta'\beta_\perp = 0$ and that $(\beta, \beta_\perp)$ span $\Re^p$. The space of all vectors orthogonal to the ones in $\mathcal{B}$ is called the orthogonal space and will be denoted by $\mathcal{B}_\perp$. The two matrices β and $\beta_\perp$ are bases of the two spaces $\mathcal{B}$ and $\mathcal{B}_\perp$.

The choice of the basis in a space is arbitrary; any other matrix $\beta^* = \beta\xi$, for ξ square and non-singular, can be chosen as a basis of $\mathcal{B}$, since any vector $\beta\tau$ has equivalent representation $\beta^*\xi^{-1}\tau = \beta^*\tau^*$. Note that for the case of the matrix C in eq. (2.4) any choice of basis of the spaces spanned by $\alpha_\perp$ and by $\beta_\perp$ would lead to the same C, that is the function C is invariant to the choice of basis in these spaces. In this section we will discuss some useful normalization of the basis of a space for the purpose of statistical inference.

An intuitive choice of a basis for $\mathcal{B}$, for instance, is the one that restricts the column vectors of the basis to have either unit or zero loadings on some selected set of r rows, that is transform a square block of the basis into the identity matrix of order r:

$$\beta_a = \begin{pmatrix} I_r \\ A \end{pmatrix} \tag{4.1}$$

where A is a $(p - r) \times r$ matrix.

Note that not *every* set of r rows of β can be transformed into the identity matrix, as the corresponding minor of β could be zero. Nevertheless at least one such feasible set exists from the assumption that β has full column rank; in fact if there did not exist at least one non-null minor of dimension r, the original basis would be of rank $r - 1$ or less.

Given normalization (4.1) for the basis of the space $\mathcal{B}$, a natural normalization of the basis of the orthogonal space $\mathcal{B}_{\perp}$ is seen to be

(4.2)
$$\beta_{a\perp}=\begin{pmatrix}-A'\\ I_{p-r}\end{pmatrix}$$

In fact one has $\beta_a'\beta_{a\perp}=A'-A'=0$. Note that the *same* matrix A enters into the bases of the two spaces; therefore eq. (4.1) and (4.2) make clear that β and $\beta_{\perp}$ are linearly related when properly normalized. This basic argument leads to a single estimation problem for β and $\beta_{\perp}$.

The choices (4.1) and (4.2) are special cases of a more general set of normalizations. Consider in fact

(4.3)
$$\beta_c \equiv \beta(c'\beta)^{-1}$$

where c is a $p\times r$ matrix such that $c'\beta$ is non-singular. It is straightforward to verify that (4.1) can be obtained from (4.3) by setting $c=a\equiv(I_r,0_{r\times(p-r)})'$. In the following we will denote by $\bar{b}$ the matrix $b_b\equiv b(b'b)^{-1}$, which is a special case of (4.3). We will also indicate as P_b the projection matrix $b\bar{b}'=\bar{b}b'=b(b'b)^{-1}b'$.

Given the matrix c, it is possible to find its orthogonal complement. There are many ways of obtaining a matrix $c_{\perp}$ that satisfies the requirements $c'c_{\perp}=0$, $span(c,c_{\perp})=\Re^p$. For instance, define a square $p\times p$ matrix with the block of the first r rows equal to c' and the second block of rows q' such that the full matrix is non singular; usually blocks of I_p can be used to specify q. Since the full matrix $(c,q)'$ is non-singular, one can calculate its inverse, which we partition in two blocks of r and $p-r$ columns, $(c,q)'^{-1}=(a,b)$. From the definition of the inverse

$$\begin{pmatrix}c'\\ q'\end{pmatrix}(a,b)=I_p$$

so that one has $c'b=0_{r\times(p-r)}$, and therefore b is a possible choice of $c_{\perp}$.

An alternative possible procedure to derive the orthogonal complement of c is to consider $I_p-\bar{c}c'$. From the theory of orthogonal projections $I_p-\bar{c}c'=\bar{c}_{\perp}c_{\perp}'$, so that a spectral decomposition of the matrix $I_p-\bar{c}c'$ supplies the matrix of eigenvectors corresponding to the non-null eigenvalues of $I_p-\bar{c}c'$ as a possible choice of $c_{\perp}$. One other way to calculate the orthogonal complement is through a QR decomposition; in fact c has decomposition $c=(\underline{Q}_1,\underline{Q}_2)(\underline{R}',0)'=\underline{Q}_1\underline{R}$ where $(\underline{Q}_1,\underline{Q}_2)$ is an orthogonal matrix and $\underline{R}$ is a triangular matrix of dimension r. It is easy to see from the orthogonality of $(\underline{Q}_1,\underline{Q}_2)$ that $\underline{Q}_2'c = \underline{Q}_2'\underline{Q}_1\underline{R} = 0$, i.e. that $\underline{Q}_2$ is a possible choice for $c_\perp$. Note that all the above methods in general specify different sets of vectors for $c_\perp$, which non-the-less span the same space $span(c_\perp)$, and therefore are possible alternative bases for that space.

Often, as for the choice $c = a \equiv (I_r, 0_{r\times(p-r)})'$, the orthogonal complement $c_\perp$ is very simple, and does not require any computation; it is easy to see in fact that $a_\perp = (0_{(p-r)\times r}, I_{p-r})'$.

Given β, c and $c_\perp$, a possible direct normalization of $\beta_\perp$ with respect to the normalization β_c (4.3) of β is given by

(4.4) $$\beta_{c\perp} \equiv (I - c\beta_c')\bar{c}_\perp$$

It is easy to verify that $\beta_c'\beta_{c\perp} = 0$ and that for the particular choice of $c = a \equiv (I_r, 0_{r\times(p-r)})'$ and $c_\perp = (0_{(p-r)\times r}, I_{p-r})'$ (4.4) reduces to (4.2).

Normalizations (4.3) and (4.4) have a number of properties, which are summarized in the following lemma, a proof of which is deferred to the appendix.

Lemma 1

Given two orthogonal spaces $\mathcal{B}$ and $\mathcal{B}_\perp$ with respective bases β and $\beta_\perp$, and two matrices, c and $c_\perp$, such that $c'\beta$ and $c_\perp'\beta_\perp$ are square and non-singular, the following properties hold for the matrices defined in (4.3) *and* (4.4)*:*

1.*a*) *β_c and $\beta_{c\perp}$ are bases of the spaces $\mathcal{B}$ and $\mathcal{B}_\perp$;*

1.*b*) *β_c and $\beta_{c\perp}$ are normalized to be the identity matrix in directions c and $c_\perp$ respectively,*

$$c'\beta_c = I_r \qquad c_\perp'\beta_{c\perp} = I_{p-r}$$

1.*c*) *the following recursive relations hold*

$$\beta_c = (\beta_d)_c \qquad \beta_{c\perp} = (\beta_{d\perp})_{c\perp}$$

1.*d*) *β_c and $\beta_{c\perp}$ have the symmetric representation*

(4.5) $$\beta_{c\perp} = \beta_\perp(c_\perp'\beta_\perp)^{-1}$$

(4.6) $$\beta_c = (I - c_\perp\beta_{c\perp}')\bar{c}$$

1.*e*) *consider some other pair of orthogonal spaces $\mathcal{D}$, $\mathcal{D}_\perp$ with the same dimensions of $\mathcal{B}$, $\mathcal{B}_\perp$, and* *respective bases* δ *and* $\delta_\perp$, *for which* $c'\delta$ *and* $c_\perp{}'\delta_\perp$ *are square and non-singular; then*

$$\beta_{c\perp}{}'(\delta_c - \beta_c) = \bar{c}_\perp{}'(\delta_c - \beta_c) \tag{4.7}$$

$$\beta_c{}'\left(\delta_{c\perp} - \beta_{c\perp}\right) = \bar{c}'\left(\delta_{c\perp} - \beta_{c\perp}\right) \tag{4.8}$$

Lemma 1.*a*) makes clear that (4.3) and (4.4) are possible choices of basis of the spaces $\mathcal{B}$ and $\mathcal{B}_\perp$. Property 1.*b*) illustrates the spirit of the above normalizations; note that it also states that (4.4) specifies the basis of the orthogonal complement so that it is already normalized along $c_\perp$; therefore in the following no distinction is made between the two ways of obtaining $\beta_{c\perp}$, that is either by (4.4) or by normalizing $\beta_\perp$ along $c_\perp$. Property 1.*d*) shows, moreover, that definitions (4.3) and (4.4) are symmetric; in other words, given c and $c_\perp$ and either β or $\beta_\perp$, (4.3) and (4.4) lead to the same normalized bases. Finally property 1.*c*) guarantees that the chosen normalizations can be obtained applying (4.3) and (4.4) either to β (or $\beta_\perp$) or to any other already normalized basis. The usefulness of 1.*e*) will become clear in the following.

As the above discussion suggests, in the context of statistical inference there is really only one estimation problem for the bases of both a space and its orthogonal complement. It is easy to observe, in fact, that once an estimator is obtained for a space, say $\hat{\beta}$, an estimator of the basis of the orthogonal space is also already specified, e.g. as through (4.4). Therefore consider, for any estimator $\hat{\beta}$ of β, the normalization $\hat{\beta}_c$ and the corresponding normalization of the orthogonal complement $\hat{\beta}_{c\perp}$. Some special choice of c is of some interest; consider for instance $c = \bar{\beta}$, and let $\tilde{\beta} \equiv \hat{\beta}_{\bar{\beta}} = \hat{\beta}(\beta'\hat{\beta})^{-1}(\beta'\beta)$. It is easy to see that the projection of $\tilde{\beta}$ onto the space β is equal to β itself. Therefore when one decomposes $\tilde{\beta}$ onto its projection on β and $\beta_\perp$, one finds $\tilde{\beta} = (P_\beta + P_{\beta_\perp})\tilde{\beta} = \beta + \beta_\perp d$ or $\tilde{\beta} - \beta = \beta_\perp d$, which shows that the sampling variation of $\tilde{\beta} - \beta$, contained in d, all lies in the space of the orthogonal complement $\beta_\perp$. The present case $c = \bar{\beta}$, although highly interesting in view of the above property, does not seem to be a viable final choice of normalization of $\hat{\beta}$, as it assumes knowledge of the true β. Nevertheless such a choice can be valuable in an intermediate stage, as from lemma 1.*c*) any other normalization can be obtained from it.

Note that for the choice $c = a$ (the special case of (4.1) and (4.2)) the difference $\hat{\beta}_a - \beta_a$ turns out to be

$$\hat{\beta}_a - \beta_a = \begin{pmatrix} 0 \\ \hat{A} - A \end{pmatrix}$$

thus reflecting the fact that $\hat{A}$ summarizes the sampling variation of the estimator $\hat{\beta}_a$. The term $\hat{A} - A$ can be isolated by pre-multiplying the difference $\hat{\beta}_a - \beta_a$ by $\beta_{a\perp}' = (-A, I_{(p-r)})$ that is $\beta_{a\perp}'(\hat{\beta}_a - \beta_a) = \hat{A} - A$. On the other hand the difference $(\hat{\beta}_{a\perp} - \beta_{a\perp})'$ is just $(\hat{\beta}_{a\perp} - \beta_{a\perp})' = (-(\hat{A} - A), 0)$. Note that in this case $\hat{A} - A$ can be isolated post-multiplying by $\beta_a = (I_r, A')'$, that is $(\hat{\beta}_{a\perp} - \beta_{a\perp})'\beta_a = -(\hat{A} - A)$. For the special case (4.1), (4.2) one therefore has

(4.8) $$\beta_{a\perp}'(\hat{\beta}_a - \beta_a) = -(\hat{\beta}_{a\perp} - \beta_{a\perp})\beta_a$$

The following lemma shows that (4.8) in fact holds for any choice of c and not just for $c = a$.

Lemma 2

Given any estimator $\hat{\beta}$ of β and the hypotheses of lemma 1 *with respect to both $\mathcal{B}$ and $\hat{\mathcal{B}}$, then the following equality is true*

(4.9) $$\beta_{c\perp}'(\hat{\beta}_c - \beta_c) = -(\hat{\beta}_{c\perp} - \beta_{c\perp})'\beta_c$$

Let $\tilde{\beta} \equiv \hat{\beta}_{\overline{\beta}}$ and $\tilde{\beta}_\perp \equiv (\hat{\beta}_\perp)_{\overline{\beta}_\perp}$, which correspond to $c = \overline{\beta}$ and $c_\perp \equiv \overline{\beta}_\perp$ in (4.3) (4.4); *for this choice*

(4.10) $$\beta_\perp'(\tilde{\beta} - \beta) = -(\tilde{\beta}_\perp - \beta_\perp)'\beta$$

PROOF From the definition (4.4)

$$\hat{\beta}_{c\perp} - \beta_{c\perp} = (I_p - c\hat{\beta}_c')\overline{c}_\perp - (I_p - c\beta_c')\overline{c}_\perp = -c(\hat{\beta}_c - \beta_c)'\overline{c}_\perp$$

Pre-multiplying by β_c' and remembering from lemma 1.*b*) that $\beta_c'c = I_r$, one obtains

$$\beta_c'(\hat{\beta}_{c\perp} - \beta_{c\perp}) = -(\hat{\beta}_c - \beta_c)'\overline{c}_\perp$$

We need only to show that $(\hat{\beta}_c - \beta_c)'\overline{c}_\perp = (\hat{\beta}_c - \beta_c)'\beta_{c\perp}$; this follows directly from lemma 1.*e*), thus proving (4.9); in order to obtain (4.10) just consider the choice $c = \overline{\beta}$, $c_\perp \equiv \overline{\beta}_\perp$ in (4.9).

Q.E.D.

It may be helpful to illustrate the above result by a graphical example. Let us consider $p = 2, r = 1, p - r = 1$. For simplicity normalize the β space along the second unit coordinate vector, $c = (0, c_1)'$ so that $c_\perp = (c_2, 0)'$ lies on the first coordinate axis. With respect to the $(c, c_\perp)$ coordinate system, the pairs $(\beta, \beta_\perp)$ and $(\hat\beta, \hat\beta_\perp)$ represent perpendicular vectors centered in the origin (see fig. 1).

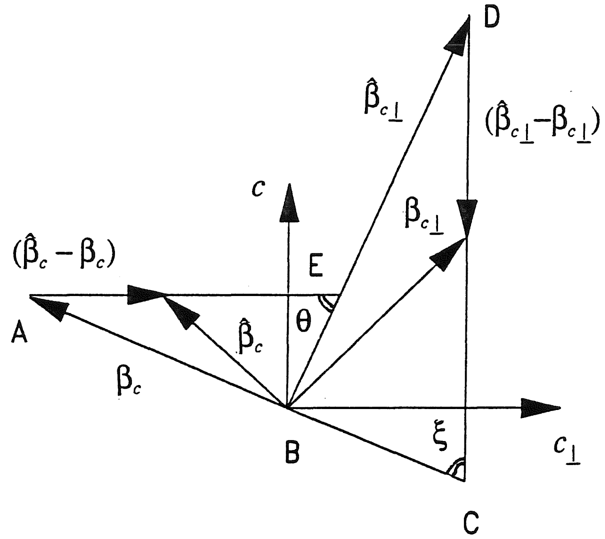

Fig. 1

The normalizations β_c and $\hat\beta_c$ for β and $\hat\beta$ set the lengths of the two vectors so that their projections on the c axis coincide; analogously normalization (4.4) for $\beta_\perp$ and $\hat\beta_\perp$ makes the projections of $\beta_{c\perp}$ and $\hat\beta_{c\perp}$ on the $c_\perp$ axis equal. The difference vector $\hat\beta_c - \beta_c$ therefore lies only in the $c_\perp$ subspace and the difference $\hat\beta_{c\perp} - \beta_{c\perp}$ lies in the c subspace. If c and $c_\perp$ are chosen to have the same norm ($c_1 = c_2$), the lengths of $\hat\beta_c - \beta_c$ and $\hat\beta_{c\perp} - \beta_{c\perp}$ become equal, as fig. 1 suggests.

It also turns out that in the $p = 2$ case the product of the lengths of β_c and $\hat{\beta}_{c\perp} - \beta_{c\perp}$ is equal to the product of lengths of $\hat{\beta}_c - \beta_c$ and $\beta_{c\perp}$, that is $\| \beta_c \| \cdot \| \hat{\beta}_{c\perp} - \beta_{c\perp} \| = \| \beta_{c\perp} \| \cdot \| \hat{\beta}_c - \beta_c \| = 1/k$ see appendix; we can then interpret the left-hand-side of (4.9) as $k \cos\theta$ and the right-hand-side as $k \cos\xi$. In other words (4.9) expresses the fact that the angles θ and ξ are equal, which is clear from the similarity of the rectangular triangles ABE and BDC. Finally one can illustrate equation (4.10) again using fig. 1, only rotating the axes so to make $c = \beta$ and $c_\perp = \beta_\perp$.

5 The asymptotic distributions of the estimators of C and $\alpha_\perp$

In this section we return to the problem of inference in model (2.2) or (3.1). It is useful to adopt the following notation:

$\Sigma_{vv.z2} = Var(v_t \mid z_{2t})$ (variance of v_t calculated from its seasonal mean),

where $v_t \equiv (X_{t-1}'\beta, \Delta X_{t-1}', \ldots, \Delta X_{t-k+1}')'$, $z_{2t} \equiv (1, D_t')'$, see section 2;

$\Sigma_{ss.w2} = Var(s_t \mid w_{2t})$ where $s_t \equiv (X_{t-1}'\beta, \Delta X_{t-1}')'$ and $w_{2t} \equiv (\Delta^2 X_{t-1}', \ldots, \Delta^2 X_{t-k+2}', 1, D_t')'$

see section 3;

$\beta'\Sigma_{xx.z}\beta \equiv Var(\beta' X_{t-1} \mid z_t)$ where $z_t = (\Delta X_{t-1}', \ldots, \Delta X_{t-k+1}', 1, D_t')'$.

The following lemma reports some asymptotic distributions which are connected to the estimator (2.10) of C.

Lemma 3

In the model (2.2) (3.1)

$$(5.1) \qquad T(\hat{\beta}_c - \beta_c) \xrightarrow{w} (I - \beta_c c')\bar{\gamma}\left(\int_0^1 G_{1.2} G_{1.2}' du\right)^{-1} \int_0^1 G_{1.2} (dV_\alpha)'(c'\beta)^{-1}$$

$$(5.2a) \qquad \sqrt{T}\,\mathrm{vec}([\hat{\alpha}, \hat{\Gamma}] - [\alpha, \Gamma]) \xrightarrow{w} N(0, \Sigma_{vv.z2}^{-1} \otimes \Omega)$$

$$(5.2b) \qquad \sqrt{T}\,\mathrm{vec}([\hat{\alpha}, \hat{\Phi}] - [\alpha, \Phi]) \xrightarrow{w} N(0, \Sigma_{ss.w2}^{-1} \otimes \Omega)$$

$$(5.2c) \qquad \sqrt{T}\,\mathrm{vec}[(\hat{\alpha} - \alpha)'] \xrightarrow{w} N(0, \Omega \otimes (\beta'\Sigma_{xx.z}\beta)^{-1})$$

where γ *is a* $p \times (p-r-1)$ *matrix such that* $(\gamma, C\mu) = \beta_{\perp}$, $G_{1.2} = G_1 - (\int G_1 G_2 du)(\int G_2 G_2 du)^{-1} G_2$, $G_1 = \bar{\gamma}' C(W - \int W du)$, $G_2 = u - 1/2$, $V_\alpha \equiv (\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}W$, W *is a Brownian motion on* $[0,1]$ *with covariance* Ω *and* V_α *and* $G \equiv (G_1', G_2)'$ *are independent*[6].

If $\mu = \alpha\beta_0$ *the previous results are still valid substituting* G_1 *for* $G_{1.2}$ *in* (5.1).

Results (5.1) (5.2*a*, *c*) are given in Johansen (1991); specifically (5.1) corresponds to Johansen's eq. (5.1) in theorems 5.1 and 5.2 and (5.2*a*) to Johansen's theorem *C*.1; (5.2*c*) is just a rewriting of the same result and (5.2*b*) is derived analogously.

The above results are sufficient to derive the asymptotic distribution of the estimator (2.10) of C when (4.4) has been applied to $\hat{\alpha}$ and $\hat{\beta}$ to obtain estimators of their orthogonal complements, respectively normalized along $b_{\perp}$ and $c_{\perp}$; note that $\alpha_{\perp}$ is therefore normalized so that $b_{\perp}'\alpha_{b\perp} = I_{p-r}$.

The following result holds:

Theorem 4

The asymptotic distribution of the estimator (2.10) *of* C *is gaussian*

$$\sqrt{T}\,\mathrm{vec}(\hat{C} - C) \xrightarrow{w} N(0, \aleph) \tag{5.3a}$$

where the asymptotic covariance matrix has the equivalent representations

$$\aleph = Q\Sigma_{vv.z2}^{-1}Q' \otimes C\Omega C' \tag{5.3b}$$

$$\aleph = Q^{*}\Sigma_{ss.w2}^{-1}Q^{*}{}' \otimes C\Omega C' \tag{5.3c}$$

where $Q \equiv (Q_1, Q_2)$, $Q_1 \equiv (C'\Psi' - I_p)\bar{\alpha} = (C' - C'\Phi' - I_p)\bar{\alpha}$, $Q_2 = C'(i_{k-1}' \otimes I_p)$, *and* $Q^{*} = (Q_1, C')$.

The covariance matrix $\aleph$ *can be consistently estimated either by*

$$\hat{Q}M_{vv.z2}^{-1}\hat{Q}' \otimes \hat{C}\hat{\Omega}\hat{C}'$$

that is by pre-multiplying the covariance matrix obtained in the regression of $y_{t.z2}$ *on* $v_{t.z2}$ *(see section* 2) *by* $(\hat{Q} \otimes \hat{C})$ *and post-multiplying it by its transpose, or by*

$$\hat{Q}^{*}M_{ss.w2}^{-1}\hat{Q}^{*}{}' \otimes \hat{C}\hat{\Omega}\hat{C}'$$

[6] See Johansen (1991) theorem 5.1.

that is pre-multiplying the covariance matrix obtained in the regression of $y_{t.w2}$ *on* $s_{t.w2}$ *(see section* 3) *by* $(\hat{Q}^* \otimes \hat{C})$ *and post-multiplying it by its transpose.*

Moreover the limit distribution of $\hat{\alpha}_{b\perp}$ *is also normal* [7]

(5.4) $$\sqrt{T}\,\mathrm{vec}[\alpha_b'(\hat{\alpha}_{b\perp}-\alpha_{b\perp})] = \sqrt{T}\,\mathrm{vec}[\bar{b}'(\hat{\alpha}_{b\perp}-\alpha_{b\perp})] \xrightarrow{w} N(0,\Xi)$$

where $$\Xi = (\alpha_{b\perp}'\Omega\alpha_{b\perp}) \otimes (b'\Pi\Sigma_{xx.z}\Pi'b)^{-1}$$

This asymptotic covariance matrix can be consistently estimated by

$$(\hat{\alpha}_{b\perp}'\hat{\Omega}\hat{\alpha}_{b\perp}) \otimes (b'\hat{\Pi}M_{xx.z}\hat{\Pi}'b)^{-1}$$

PROOF We will first prove (5.4). Consider the distribution of a normalized version of α, $\alpha_b = \alpha(b'\alpha)^{-1}$. A first order expansion of $\hat{\alpha}_b$ around α gives

(5.5) $$\hat{\alpha}_b - \alpha_b = (I - \alpha_b b')(\hat{\alpha}-\alpha)(b'\alpha)^{-1} + O_p(\|\hat{\alpha}-\alpha\|^2)$$

and pre-multiplying by $\alpha_{b\perp}'$ one obtains

$$\alpha_{b\perp}'(\hat{\alpha}_b - \alpha_b) = \alpha_{b\perp}'(\hat{\alpha}-\alpha)(b'\alpha)^{-1} + O_p(T^{-1})$$

From lemma 2 the left-hand-side is equal to $-\alpha_b'(\hat{\alpha}_{b\perp}-\alpha_{b\perp})$, that is

(5.6) $$\alpha_b'(\hat{\alpha}_{b\perp}-\alpha_{b\perp}) = -(\alpha'b)^{-1}(\hat{\alpha}-\alpha)'\alpha_{b\perp} + O_p(T^{-1})$$

Applying the column stacking operator and multiplying by $\sqrt{T}$ one obtains

$$\sqrt{T}\,\mathrm{vec}\,(\alpha_b'(\hat{\alpha}_{b\perp}-\alpha_{b\perp})) = -\left(\alpha_{b\perp}' \otimes (\alpha'b)^{-1}\right)\sqrt{T}\,\mathrm{vec}[(\hat{\alpha}-\alpha)'] + O_p(T^{-1/2})$$

Thus from the previous expression and (5.2*c*) the asymptotic distribution (5.4) of $\hat{\alpha}_{b\perp}$ follows directly by Cramér's theorem.

Next consider $C = \beta_\perp(\alpha'_\perp\Psi\beta_\perp)^{-1}\alpha_\perp'$ which is seen to be invariant with respect of normalizations of $\beta_\perp$ and $\alpha_\perp$ of the type (4.3), that is $C = \beta_{c\perp}(\alpha_{b\perp}'\Psi\beta_{c\perp})^{-1}\alpha_{b\perp}'$ for any b, c. Substitution of the definition

[7] A similar derivation can also be employed to show that the limit distribution of $T\beta_c'(\hat{\beta}_{c\perp}-\beta_{c\perp})$ is mixed gaussian, although there is the additional complication that in one direction (the one of the linear trend) the distribution converges at the stronger rate $T^{-3/2}$.

(4.4) of $\beta_{c\perp}$ and $\alpha_{b\perp}$ in C makes clear that C is a function of $(\beta_c, \alpha_b, \Psi)$, $C = g(\beta_c, \alpha_b, \Psi)$. A first order expansion of C gives

$$(\hat{C} - C) = g_{\beta_c}(\hat{\beta}_c - \beta_c) + g_{\alpha_b}(\hat{\alpha}_b - \alpha_b) + g_{\Psi}(\hat{\Psi} - \Psi) +$$

$$+ O_p(\max\{\| \hat{\beta}_c - \beta_c \|^2, \| \hat{\alpha}_b - \alpha_b \|^2, \| \hat{\Psi} - \Psi \|^2\})$$

From (5.1) $\hat{\beta}_c - \beta_c = O_p(T^{-1})$, while $\hat{\alpha}_b - \alpha_b = O_p(T^{-1/2})$ and $\hat{\Psi} - \Psi = O_p(T^{-1/2})$ from (5.5) and (5.2) respectively. By the above superconsistency of $\hat{\beta}_c$ one can then consider β_c as fixed in $\hat{C}$; in fact, since α_b is function of α, one can write

(5.7) $$\sqrt{T}(\hat{C} - C) = g_{\alpha}(\sqrt{T}(\hat{\alpha} - \alpha)) + g_{\Psi}(\sqrt{T}(\hat{\Psi} - \Psi)) + o_p(1)$$

and again applying Cramér's theorem one obtains that $\sqrt{T}(\hat{C} - C)$ is asymptotically gaussian with mean zero. In order to obtain the asymptotic covariance matrix, the first order derivatives in (5.7) are needed. A first order expansion of C as a function of α_b gives

$$g_{\alpha_b}(\hat{\alpha}_b - \alpha_b) = -\beta_{\perp}(\alpha_{b\perp}'\Psi\beta_{\perp})^{-1}\bar{b}_{\perp}(\hat{\alpha}_b - \alpha_b)b'(I - \Psi C) + O_p(T^{-1}) =$$

$$= - C(\hat{\alpha}_b - \alpha_b)b'(I - \Psi C) + O_p(T^{-1})$$

where the second equality follows from (4.7). Substituting (5.5) it follows

$$g_{\alpha}(\hat{\alpha} - \alpha) = C(\hat{\alpha} - \alpha)(b'\alpha)^{-1}b'(\Psi C - I) + O_p(T^{-1})$$

Note that $\alpha_{\perp}'(\Psi C - I) = \alpha_{\perp}\Psi\beta_{\perp}(\alpha_{\perp}\Psi\beta_{\perp})^{-1}\alpha_{\perp}' - \alpha_{\perp}' = 0$ from which one has $\Psi C - I = (P_{\alpha} + P_{\alpha_{\perp}})(\Psi C - I) = P_{\alpha}(\Psi C - I)$ and therefore

(5.8) $$g_{\alpha}(\sqrt{T}(\hat{\alpha} - \alpha)) = C\sqrt{T}(\hat{\alpha} - \alpha)(b'\alpha)^{-1}b'\alpha\bar{\alpha}'(\Psi C - I) + O_p(T^{-1/2}) =$$

$$= C(\sqrt{T}(\hat{\alpha} - \alpha))\bar{\alpha}'(\Psi C - I) + o_p(1)$$

Let us now consider the first order derivatives with respect to Ψ; recall that the first order differential of a square invertible matrix A is $dA^{-1} = -A^{-1}dAA^{-1}$ so that

$$g_{\Psi}(\hat{\Psi} - \Psi) = - \beta_{\perp}(\alpha_{\perp}'\Psi\beta_{\perp})^{-1}\alpha_{\perp}'(\hat{\Psi} - \Psi)\beta_{\perp}(\alpha_{\perp}'\Psi\beta_{\perp})^{-1}\alpha_{\perp}' + O_p(T^{-1}) =$$

$$= - C(\hat{\Psi} - \Psi)C + O_p(T^{-1})$$

Remembering now that $\hat{\Psi}-\Psi=-(\sum_{i=1}^{k-1}\hat{\Gamma}_i-\Gamma_i)=-(\hat{\Gamma}-\Gamma)(i_{k-1}\otimes I_p)=-(\hat{\Phi}-\Phi)$, substituting and multiplying by $\sqrt{T}$ one obtains

(5.9a) $$g_{\Psi}(\sqrt{T}(\hat{\Psi}-\Psi))=C(\sqrt{T}(\hat{\Gamma}-\Gamma))(i_{k-1}\otimes I_p)C+o_p(1)$$

(5.9b) $$=C(\sqrt{T}(\hat{\Phi}-\Phi))C+o_p(1)$$

Combining (5.8) and (5.9a) the following expression results

(5.10a) $$\sqrt{T}(\hat{C}-C)=C(\sqrt{T}(\hat{\alpha}-\alpha))Q_1'+C(\sqrt{T}(\hat{\Gamma}-\Gamma))Q_2'+o_p(1)=$$

$$=C\sqrt{T}([\hat{\alpha},\hat{\Gamma}]-[\alpha,\Gamma])Q'+o_p(1)$$

where $Q_1\equiv(C'\Psi'-I)\bar{\alpha}$, $Q_2\equiv C'(i_{k-1}'\otimes I_p)$. Applying the column stacking operator one has

(5.11a) $$\sqrt{T}\,\mathrm{vec}(\hat{C}-C)=(Q\otimes C)\sqrt{T}\,\mathrm{vec}([\hat{\alpha},\hat{\Gamma}]-[\alpha,\Gamma])+o_p(1)$$

from which (5.3b) follows. Analogously from (5.8) and (5.9b) one derives

(5.10b) $$\sqrt{T}(\hat{C}-C)=C(\sqrt{T}(\hat{\alpha}-\alpha))Q_1'+C(\sqrt{T}(\hat{\Phi}-\Phi))C+o_p(1)=$$

$$=C\sqrt{T}([\hat{\alpha},\hat{\Phi}]-[\alpha,\Phi])Q^{*}{}'+o_p(1)$$

where $Q_1\equiv(C'\Psi'-I)\bar{\alpha}=(C'+C'\Phi'-I)\bar{\alpha}$; finally applying the column stacking operator

(5.11b) $$\sqrt{T}\,\mathrm{vec}(\hat{C}-C)=(Q^{*}\otimes C)\sqrt{T}\,\mathrm{vec}([\hat{\alpha},\hat{\Phi}]-[\alpha,\Phi])+o_p(1)$$

from which (5.3c) follows.

Q.E.D.

Note that theorem 4 is still valid under hypotheses on μ of the form $\mu=\alpha\beta_0$, since these affect only the limit distribution of $\hat{\beta}_c-\beta_c$, which does not contribute to the asymptotic covariance matrix (5.3) due to the superconsistency of $\hat{\beta}$. The modifications of the statistical calculations under $\mu=\alpha\beta_0$, that is the augmentation of x_t by 1 and the corresponding cancelling of 1 from the regressors in z_t or w_t, see Johansen (1991), do not affect the results either.

6 Wald tests on the common trends

The results of the previous section indicate the possibility to construct straightforward Wald type tests on C and on the normalized version of $\alpha_{\perp}$. It is the purpose of this section to discuss how to formulate appropriate tests with respect to $\alpha_\perp$. Linear restrictions will be considered first for simplicity. Consider therefore hypothesis of the form

(6.1) $$R\text{vec}(\alpha_{b\perp}) = q$$

where R is $m \times (p^2 - pr)$ and q is $m \times 1$. Since $\alpha_\perp$ has been normalized along the direction $b_\perp$, it is not sensible to test restrictions in the direction $b_\perp$. Note that in fact $\alpha_{b\perp} = \left(P_b + P_{b_\perp}\right)\alpha_{b\perp} = b\bar{b}'\alpha_{b\perp} + \bar{b}_\perp$ so that the left-hand-side of (6.1) can be rewritten as $R\text{vec}(\alpha_{b\perp}) = R(I_{p-r} \otimes b)\bar{b}\alpha_{b\perp} + R\text{vec}(\bar{b}_\perp) = R^*(\bar{b}'\alpha_{b\perp}) + R\text{vec}(\bar{b}_\perp)$; hypothesis (6.1) becomes then equivalent to

(6.2) $$R^*(\bar{b}'\alpha_{b\perp}) = q^*$$

where $R^* \equiv R(I_{p-r} \otimes b)$ is $m \times (pr - r^2)$ and $q^* \equiv q - R\text{vec}(\bar{b}_\perp)$ a $m \times p - r$. Note that R^* will be of full row rank m if the hypothesis (6.1) is well specified, that is if it does not pertain the already normalized coefficients. The reformulation (6.2) can then be used to check that the specified hypothesis concerns the normalization-unconstrained coefficients through a check on the rank of R^*.

As an illustration consider the choice of normalization $b = (I_r, 0)'$, $b_\perp = (0, I_{p-r})'$. The normalized version of $\alpha_\perp$ is therefore

(6.3) $$\alpha_{b\perp} = \begin{pmatrix} \alpha_1 \\ I_{p-r} \end{pmatrix}$$

and $\bar{b}' = b'$ selects α_1 from $\alpha_{b\perp}$, i.e. $\bar{b}'\alpha_{b\perp} = \alpha_1$. Reformulation (6.2) is pointing out in this example that questions relating the lower block of $\alpha_{b\perp}$ (the identity matrix I_{p-r}) are *not* statistical hypothesis. Note also that (6.2) provides the link between the hypothesis of interest and the result of the previous section. In fact the distribution of $\alpha_{b\perp}$ in (5.4) has been expressed as $\sqrt{T}\,\text{vec}[\alpha_b'(\hat{\alpha}_{b\perp} - \alpha_{b\perp})]$. From lemma 1.*e*) one has that $\alpha_b'(\hat{\alpha}_{b\perp} - \alpha_{b\perp}) = \bar{b}'(\hat{\alpha}_{b\perp} - \alpha_{b\perp})$, and thus it follows directly that under (6.2)

(6.4) $$T(R^*\text{vec}(\bar{b}'\hat{\alpha}_{b\perp}) - q^*)'[R^*\hat{\Xi}R^{*\prime}]^{-1}(R^*\text{vec}(\bar{b}'\hat{\alpha}_{b\perp}) - q^*) \xrightarrow{w} \chi^2(m)$$

A similar Wald type test can be constructed for a general smooth hypothesis of the form $\rho(\text{vec}(\bar{b}'\alpha_{b\perp})) = q^*$ simply substituting $\partial\rho(\text{vec}(\bar{b}'\hat{\alpha}_{b\perp}))/\partial\text{vec}(\bar{b}'\alpha_{b\perp})$ for R^* in (6.4).

One interesting special case of hypotheses of the form (6.1) (6.2) are exclusion type restrictions. A different way to express these type of restrictions is the following

$$R_1'\alpha_{b_\perp}R_2 = 0 \tag{6.5}$$

where R_1 is $p \times m_1$ and R_2 is $(p-r)\times m_2$. As in (6.2) we have to make sure that hypothesis (6.5) refers only to unnormalized coefficients, i.e. to α_1 in example (6.3). In order to do so R_1 has to lie in the span of b. In fact it is easy to see that were R_1 in the span of $b_\perp$, $R_1 = b_\perp d$, then $R_1'\alpha_{b_\perp}R_2 = d'b_\perp'\alpha_{b_\perp}R_2 = d'R_2$ which would be a known quantity. If R_1 had a non-null component in the span of $b_\perp$, that is $R_1 = b_\perp d + \bar{b}R_1{*}$, (6.5) would not express exclusion restrictions, as $R_1'\alpha_{b_\perp}R_2 = d'b_\perp'\alpha_{b_\perp}R_2 + R_1{*}'\bar{b}'\alpha_{b_\perp}R_2 = d'R_2 + R_1{*}'\bar{b}'\alpha_{b_\perp}R_2$ so that (6.5) would be equivalent to $R_1{*}'\bar{b}'\alpha_{b_\perp}R_2 = q$, where $q = -d'R_2$.

Therefore we assume that $R_1 \in span(b)$, or $R_1 = \bar{b}R_1{*}$; note that $R_1{*} = b'R_1$. Hypothesis (6.5) has thus equivalent representation

$$R_1{*}'(\bar{b}'\alpha_{b_\perp})R_2 = 0 \tag{6.6}$$

If m_1 or m_2 are greater than one, (6.6) is a joint hypothesis on the parameters in $\alpha_{b_\perp}$; it is then easy to transform (6.6) by applying the column stacking operator to obtain $(R_2' \otimes R_1{*}')\mathrm{vec}(\bar{b}'\alpha_{b_\perp}) = 0$, which is a special case of (6.2) with $R{*} = (R_2' \otimes R_1{*}')$ and $q{*} = 0$. Result (6.4) can therefore be applied with $m = m_1 m_2$.

If $m_1 = m_2 = 1$, (6.6) expresses a zero restriction on a scalar linear function of $\alpha_{b_\perp}$, and a simpler procedure in this case would be to derive t-ratio type statistics. $R_1{*}$ and R_2 will often be just selection column vectors e_h with all zero elements and a unit element in the h-th row, $R_1{*} \equiv e_i$ and $R_2 \equiv e_j$. Denoting with $(a)_{ij}$ the ij-th element of a matrix a, hypothesis (6.6) represent in this case exclusion restrictions of the type $(\bar{b}'\alpha_{b_\perp})_{ij} = 0$, or $(\alpha_1)_{ij} = 0$ in example (6.3). From (5.4) the corresponding asymptotic variance is $(\alpha_{b_\perp}'\Omega\alpha_{b_\perp})_{jj} \cdot (b'\Pi\Sigma_{xx.z}\Pi'b)^{-1}_{ii}$ which can be consistently estimated by $(\hat{\alpha}_{b_\perp}'\hat{\Omega}\hat{\alpha}_{b_\perp})_{jj} \cdot (b'\hat{\Pi}M_{xx.z}\hat{\Pi}'b)^{-1}_{ii}$. The corresponding t-ratio statistic is asymptotically standard normal distributed

$$\frac{\sqrt{T}(\bar{b}'\hat{\alpha}_{b_\perp})_{ij}}{\sqrt{(\hat{\alpha}_{b_\perp}'\hat{\Omega}\hat{\alpha}_{b_\perp})_{jj} \cdot (b'\hat{\Pi}M_{xx.z}\hat{\Pi}'b)^{-1}_{ii}}} \xrightarrow{w} N(0,1) \tag{6.7}$$

If R_1* and R_2 are not selection vectors, from (5.4) the asymptotic variance of the right-hand-side of (6.6) is seen to be $d \equiv R_2'\alpha_{b\perp}'\Omega\alpha_{b\perp}R_2 \cdot R_1^{*\prime}(b'\Pi\Sigma_{xx.z}\Pi'b)^{-1}R_1^*$ or $R_2'\alpha_{b\perp}'\Omega\alpha_{b\perp}R_2 \cdot R_1'b(b'\Pi\Sigma_{xx.z}\Pi'b)^{-1}b'R_1$ which is consistently estimated by $\hat{d} \equiv R_2'\hat{\alpha}_{b\perp}'\hat{\Omega}\hat{\alpha}_{b\perp}R_2 \cdot R_1'b(b'\hat{\Pi}M_{xx.z}\hat{\Pi}'b)^{-1}b'R_1$. The corresponding t-ratio statistic $\sqrt{T}R_1'b\bar{b}'\hat{\alpha}_{b\perp}R_2/\hat{d}$ has also a standard normal asymptotic distribution,

$$\sqrt{T}R_1'b\bar{b}'\hat{\alpha}_{b\perp}R_2/\hat{d} \overset{w}{\rightarrow} N(0,1)$$

7 Some remarks on Wald tests on the moving average impact matrix C

General linear hypotheses about C can be formulated as follows

(7.1) $$R\,\mathrm{vec}(C) = q$$

where R is $m \times p^2$ and q is $m \times 1$. The matrix C has rank $p - r$ and is therefore singular, i.e. some linear combinations of its elements will be exactly zero; this is reflected in the singularity if the asymptotic covariance matrix $\aleph$. Nevertheless if the asymptotic covariance matrix of the R linear combinations $R'(Q\Sigma_{vv.z2}^{-1}Q' \otimes C\Omega C')R$ is non-singular, under (7.1)

(7.2) $$T(R\ \mathrm{vec}(\hat{C}) - q)'(R\hat{\aleph}R')^{-1}(R\ \mathrm{vec}(\hat{C}) - q) \overset{w}{\rightarrow} \chi^2(m)$$

that is the Wald test statistic is asymptotically $\chi^2(m)$ distributed, with number of degrees of freedom m equal to the number of restrictions imposed by (7.1). As before one could consider general smooth hypotheses about C of the form $\rho(\mathrm{vec}(C)) = q$; result (7.2) would be still valid when substituting $\partial\rho(\mathrm{vec}(\hat{C}))/\partial\mathrm{vec}\,(C)$ for R.

Exclusion type restrictions are a special case of (7.1); they can be expressed as

(7.3) $$R_1'CR_2 = 0$$

where R_i are $p \times m_i$ vectors, $i = 1,2$. We will consider two special cases of (7.3).

1. $m_1 = m_2 = 1$, $(p - r > 1)$[8]. Hypothesis (7.3) expresses exclusion restrictions on a scalar linear function of C; R_i will typically be selection vectors, $R_1 \equiv e_i$, $R_2 \equiv e_j$, so that (7.3) corresponds to $C_{ij} = 0$. If no restrictions have been imposed on α and β in the estimation of the system, the

[8] For the case $p - r = 1$ hypothesis (7.3) is just the union of hypothesis of the form (7.6), (7.7) see the following.

asymptotic covariance matrix of (7.3) will be positive and will have the form $(Q\Sigma_{vv.z2}^{-1}Q')_{jj}\cdot(C\Omega C')_{ii}$ or $(Q^*\Sigma_{ss.w2}^{-1}Q^{*}')_{jj}\cdot(C\Omega C')_{ii}$ which can be consistently estimated by $(\hat{Q}M_{vv.z2}^{-1}\hat{Q}')_{jj}\cdot(\hat{C}\hat{\Omega}\hat{C}')_{ii}$. The corresponding t-ratio type statistic

$$(7.4)\qquad \frac{\sqrt{T}\,\hat{C}_{ij}}{\sqrt{(\hat{Q}M_{vv.z2}^{-1}\hat{Q}')_{jj}\cdot(\hat{C}\hat{\Omega}\hat{C}')_{ii}}}\xrightarrow{w}N(0,1)$$

is asymptotically standard normal distributed.

2. $R_1 = U, R_2 = U_\perp$. This type of hypothesis nests neutrality restrictions, see Mosconi and Giannini (1992) and reference therein. Let us partition C as follows

$$C\equiv\begin{pmatrix}C_{11} & C_{12}\\ C_{21} & C_{22}\end{pmatrix}$$

where C_{22} is $m_1\times m_1$; for the choice $U=(0,I_{m1})'$ (7.3) becomes

$$(7.5)\qquad (0,I_{m_1})\begin{pmatrix}C_{11} & C_{12}\\ C_{21} & C_{22}\end{pmatrix}\begin{pmatrix}I_{p-m_1}\\ 0\end{pmatrix}=0 \qquad \text{or} \qquad C_{21}=0$$

Such a restriction states that innovations in the lower block of m_1 equations do not contribute to the stochastic trends of the first block of $p-m_1$ variables[9]. It is easy to see that (7.5) can be translated in terms of (7.1) by applying the column stacking operator

$$(U_\perp'\otimes U')\mathrm{vec}(C)=0$$

For this special case of (7.1) $R=(U_\perp'\otimes U')$ and $q=0$. The corresponding asymptotic covariance matrix will be nonsingular if α and β are estimated unrestrictedly, and result (7.2) applies.

Several remarks have to be made concerning Wald type tests on C.

As the structure $C=\beta_\perp(\alpha_\perp'\Psi\beta_\perp)^{-1}\alpha_\perp'$ makes clear, the C matrix contains both information on $\alpha_\perp$, which spans the row space of C, and information about $\beta_\perp$, which spans its column space; it also contains additional information on the remaining autoregressive parameters through Ψ. When the hypothesis of interest really pertains either $\alpha_\perp$, or $\beta_\perp$ or the other parameters in Ψ, it seems a bad

[9] Note that innovations do not have a diagonal covariance matrix. Note also that, in general, the rank of C is not affected by the hypothesis. See Mosconi and Giannini (1992) for cases in which (7.5) corresponds to an analogous structure of the Π matrix.

idea to translate it in terms of C, which contains combined information about $\alpha_\perp$, $\beta_\perp$ and Ψ. It seems sensible, instead, to formulate hypothesis only with respect to the parameters of interest, and this in most cases leads to simpler inferences.

Note also that hypothesis about the spaces spanned by $\alpha_\perp$ and $\beta_\perp$ can be translated in terms of the space spanned by α and β. Take for instance the hypothesis $R_1'C = 0$, where R_1 is a $p \times 1$ vector, as e.g. $R_1 = (1, 0, \ldots, 0)'$. The hypothesis $R_1'C = 0$ is concerned with the column space of C, that is with the space spanned by $\beta_\perp$; more specifically it is stating that the vector $R_1 = (1, 0, \ldots, 0)'$ is orthogonal to the span of $\beta_\perp$, or equivalently that R_1 is in the span of β. Note that restricting the first row of C to zero really means that no I(1) component $\sum_{i=1}^{t} \varepsilon_i$ enters in the determination of the first variable in the system, which is just as saying that the first variable is stationary by itself, or that $R_1 = (1, 0, \ldots, 0)'$ is a cointegration vector.

From the above discussion it is easy to see that hypotheses of the form

$$
\begin{aligned}
&R_1'C = 0 \\
&CR_2 = 0
\end{aligned}
\tag{7.6}
$$

should indeed be reformulated as

$$
\begin{aligned}
&R_1'\beta_\perp = 0 \qquad \text{or} \qquad R_1 \in span(\beta) \\
&R_2'\alpha_\perp = 0 \qquad \text{or} \qquad R_2 \in span(\alpha)
\end{aligned}
\tag{7.7}
$$

Likelihood ratio tests and Wald type tests of hypotheses of the form (7.7) can be readily derived in the autoregressive representation of the process, and it seems much more sensible to address these issues in that context. Procedures to maximize the likelihood under such constraints are contained in Johansen (1991) and Johansen and Juselius (1990).

8 Conclusions

In this paper we have shown that inference about two orthogonal bases is indeed a single problem. Using this duality we have derived the asymptotic distribution of the moving average impact matrix C in autoregressive systems integrated of order one; in the derivation a natural by-product is the distribution of the linear combinations $\alpha_\perp$ of the process which are responsible for the common trends behavior of the system.

These asymptotic results lend themselves to the derivation of Wald type tests of general smooth hypothesis about C and $\alpha_\perp$; a word of caution is mandatory in this respect, as some relevant hypotheses can be better formulated with respect to single component spaces. A final remark seems in order to exploit the duality of hypothesis with respect to β and $\beta_\perp$ and to α and $\alpha_\perp$; often in fact either likelihood ratio tests or Wald tests are known or can be easily derived for either one of the bases, e.g. in this case for α and β.

Acknowledgments

This paper was written while visiting the Institute of Mathematical Statistics, University of Copenhagen. I am most indebted to Søren Johansen for discussions on this topic and suggestions on the presentation of the paper.

References

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

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

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

Johansen S. (1991) "Estimation and Hypothesis Testing of Cointegration vectors in Gaussian Vector Autoregressive Models", *Econometrica,* 59, pp. 1551-1580.

Johansen S. (1992) "A statistical Analysis of Cointegration for I(2) Variables", Institute of Mathematical Statistics, University of Copenhagen.

Johansen S. - Juselius K. (1990) "Maximum Likelihood Estimation and Inference on Cointegration - with Applications to the Demand of Money", *Oxford Bullettin of Economics and Statistics,* 52, pp. 169-210.

Lutkepohl H. - Reimers H. E. (1988) "Impulse response Analysis of Co-integrated Systems", University of Kiel.

Marzocchi W. - Mulargia F. - Paruolo P. (1991) "The Correlation of Geomagnetic Reversals and Mean Sea Level", forthcoming *Earth Planetary Science Letters.*

Mosconi R. - Giannini C. (1992) "Non-Causality in Cointegrated Systems: Representation, Estimation and Testing", to appear in *Oxford Bullettin of Economics and Statistis.*

Park J. (1990) "Disequilibrium Impulse Analysis", Aarhus University Economics Discussion Paper 1990-19.

Stock J. - Watson M. (1988) "Testing for common trends", *Journal of the American Statistical Association,* 83, pp. 1097-1107.

Warne A. (1990) "Vector Autoregression and Common Trends in Macro and Financial Economics", Stockholm School of Economics.

Appendix

Proof of lemma 1

Statements 1*a*) 1*b*) are obvious from the definition of the spaces and (4.3) and (4.4). Property 1.*c*) can be directly verified by substitution. In order to verify (4.5) one can simply use property 1.*a*) and 1.*b*); in order to obtain (4.6) consider

$$\beta_c = \left(P_c + P_{c_\perp}\right)\beta_c = \bar{c} + c_\perp \bar{c}_\perp{}'\beta_c = \bar{c} - c_\perp(\bar{c}_\perp{}'(I - \beta_c c'))\bar{c} = (I - c_\perp \beta_{c\perp}{}')\bar{c}$$

Finally consider property 1.*e*); substituting (4.4) in the left-hand-side of eq. (4.7) one obtains

$$\beta_{c\perp}{}'(\delta_c - \beta_c) = \bar{c}_\perp{}'(I - \beta_c c')(\delta_c - \beta_c) = \bar{c}_\perp{}'(\delta_c - \beta_c) - \bar{c}_\perp{}'\beta_c(c'\delta_c - c'\beta_c) = \bar{c}_\perp{}'(\delta_c - \beta_c)$$

from property 1.*b*). Eq. (4.8) follows from (4.7) and 1.*d*).

An alternative proof of lemma 2

A different route of proof is to show that (4.10) is true, and then derive (4.9) from (4.10). Take in fact $\hat{\beta}$, β, $\hat{\beta}_\perp$, $\beta_\perp$, which satisfy $\hat{\beta}'\hat{\beta}_\perp = 0$, $\beta'\beta_\perp = 0$, and consider the projection of $\hat{\beta}$ on the spaces $\mathcal{B}$ and $\mathcal{B}_\perp$, $\hat{\beta} = P_\beta \hat{\beta} + P_{\beta_\perp}\hat{\beta} = \beta h + \beta_\perp f$. Post-multiplying by h^{-1} one obtains $\hat{\beta}_{\bar{\beta}} = \beta + \beta_\perp f h^{-1}$or, $\tilde{\beta} - \beta = \beta_\perp d$. In order to solve for d, pre-multiply by $\beta^*_\perp{}' \equiv (\hat{\beta}_\perp)_{\beta_\perp}{}'$; one thus has $-\beta^*_\perp{}'\beta = \beta^*_\perp{}'\beta_\perp d$ or $d = -(\beta^*_\perp{}'\beta_\perp)^{-1}\beta^*_\perp{}'\beta$ and substituting back one finds

$$\tilde{\beta} - \beta = -\beta_\perp(\beta^*_\perp{}'\beta_\perp)^{-1}\beta^*_\perp{}'\beta$$

Note that *every* pair of normalizations of $\hat{\beta}_\perp$ and $\beta_\perp$ would satisfy the above relation as $\beta_\perp(\hat{\beta}_\perp{}'\beta_\perp)^{-1}\hat{\beta}_\perp{}'$ is invariant to normalization of both these spaces. Nevertheless for the specific choice (4.4) $c_\perp = \beta_\perp$ one also has $\beta^*_\perp{}'\beta_\perp = I_{p-r}$, so that one can rewrite the above relation as $\tilde{\beta} - \beta = -\beta_\perp(\beta^*_\perp - \bar{\beta}_\perp)'\beta$ or

$$\bar{\beta}_\perp{}'(\tilde{\beta} - \beta) = -(\beta^*_\perp - \bar{\beta}_\perp)'\beta \tag{4.10}$$

In order to obtain (4.9) from (4.10) post-multiply the latter equation by $(c'\tilde{\beta})^{-1} = (\bar{\beta}'\hat{\beta})(c'\hat{\beta})^{-1}$ and pre-multiply it by $(\bar{\beta}_\perp{}'c_\perp)^{-1}$; from lemma 1.*c*) one has

$$\beta_{c\perp}{}'\hat{\beta}_c = -(\bar{\beta}_\perp{}'c_\perp)^{-1}\beta^*_\perp{}'\beta(\bar{\beta}'\hat{\beta})(c'\hat{\beta})^{-1} \tag{A.1}$$

From definition (4.4) $\beta*_{\perp}' \equiv \bar{\beta}_{\perp}'(I - \hat{\beta}(\bar{\beta}'\hat{\beta})^{-1}\bar{\beta}')$ so that when post-multiplied by $\beta(c'\hat{\beta})^{-1} = \beta(\bar{\beta}'\hat{\beta})\,(c'\hat{\beta})^{-1}$ one obtains

$$\beta*_{\perp}'\beta(\bar{\beta}'\hat{\beta})\,(c'\hat{\beta})^{-1} = 0 - \bar{\beta}_{\perp}'\hat{\beta}_c = \bar{\beta}_{\perp}'(I - \hat{\beta}_c c')\beta_c = \bar{\beta}_{\perp}'\left(P_c + P_{c_{\perp}}\right)(I - \hat{\beta}_c c')\beta_c$$

$$= (\bar{\beta}_{\perp}'c_{\perp})\bar{c}_{\perp}'(I - \hat{\beta}_c c')\beta_c = (\bar{\beta}_{\perp}'c_{\perp})\hat{\beta}_{c\perp}'\beta_c$$

Reinserting the product back in equation (A.1) one has $\beta_{c\perp}'\hat{\beta}_c = -\hat{\beta}_{c\perp}'\beta_c$ or

(4.9) $$\beta_{c\perp}'(\hat{\beta}_c - \beta_c) = -(\hat{\beta}_{c\perp} - \beta_{c\perp})'\beta_c$$

Proof that for $p = 2$ (cfr. fig. 1)

$$\| \beta_c \| \cdot \| \hat{\beta}_{c\perp} - \beta_{c\perp} \| = \| \beta_{c\perp} \| \cdot \| \hat{\beta}_c - \beta_c \|$$

From (4.4) one has

(A.2) $$\| \beta_{c\perp} \|^2 = \bar{c}_{\perp}'(I + \beta_c(c'c)\beta_c')\bar{c}_{\perp} = (c_{\perp}'c_{\perp})^{-1} + (\bar{c}_{\perp}'\beta_c)^2(c'c)$$

where the second equality follows from the fact that $p = 2$, $r = 1$. Analogously

(A.3) $$\| \hat{\beta}_{c\perp} - \beta_{c\perp} \|^2 = \bar{c}_{\perp}'(\hat{\beta}_c - \beta_c)\,(c'c)\,(\hat{\beta}_c - \beta_c)'\bar{c}_{\perp} =$$

$$= (c_{\perp}'c_{\perp})^{-1}\,(c'c)\,(\hat{\beta}_c - \beta_c)'P_{c_{\perp}}(\hat{\beta}_c - \beta_c)$$

where again the last equality follows from the fact that $p = 2$, $r = 1$. Consider now

(A.4) $$\| \hat{\beta}_c - \beta_c \|^2 = (\hat{\beta}_c - \beta_c)'(\hat{\beta}_c - \beta_c) = (\hat{\beta}_c - \beta_c)'\left(P_c + P_{c_{\perp}}\right)(\hat{\beta}_c - \beta_c) =$$

$$= (\hat{\beta}_c - \beta_c)'P_{c_{\perp}}(\hat{\beta}_c - \beta_c)$$

and analogously

(A.5) $$\| \beta_c \|^2 = \beta_c'\beta_c = \beta_c'P_c\beta_c + \beta_c'P_{c_{\perp}}\beta_c = (c'c)^{-1} + (\bar{c}_{\perp}'\beta_c)^2(c_{\perp}'c_{\perp})$$

One can then verify that the product of (*A*.2) and (*A*.4) is equal to the product of (*A*.3) and (*A*.5).

PREPRINTS 1991

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: Determination of Cointegration Rank in the Presence of a Linear Trend.

No. 2 Johansen, Søren: A Statistical Analysis of Cointegration for I(2) Variables.

No. 3 Jacobsen, Martin: Homogeneous Gaussian Diffusions in Finite Dimensions.

No. 4 Johansen, Søren: Testing Weak Exogeneity and the Order of Cointegration in UK Money Demand Data.

No. 5 Johansen, Søren: An I(2) Cointegration Analysis of the Purchasing Power Parity between Australia and USA.

No. 6 Johansen, Søren: Estimating Systems of Trending Variables.

No. 7 Jacobsen, Martin and Keiding, Niels: Random Censoring and Coarsening at Random.

PREPRINTS 1992

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: The Role of the Constant Term in Cointegration Analysis of Nonstationary Variables.

No. 2 Paruolo, Paolo: Asymptotic Inference on the Moving Average Impact Matrix in Cointegrated I(1) VAR Systems.