Econometric Theory, **18**, 2002, 673–690. Printed in the United States of America.
DOI: 10.1017.S0266466602183058

ASYMPTOTIC INFERENCE ON THE MOVING AVERAGE IMPACT MATRIX IN COINTEGRATED *I*(2) VAR SYSTEMS

PAOLO PARUOLO
Università dell'Insubria

This paper provides asymptotic standard errors for the moving average (MA) impact matrix for the second differences of a vector autoregressive (VAR) process integrated of order 2, $I(2)$. Standard errors of the row space of the MA impact matrix are also provided; bases of this row space define the common $I(2)$ trends linear combinations. These standard errors are then used to formulate Wald-type tests. The MA impact matrix is shown to be linked to impact factors that measure the total effect of disequilibrium errors on the growth rate of the system. Most of the relevant limit distributions are Gaussian, and we report artificial regressions that can be used to calculate the estimators of the asymptotic variances. The use of the techniques proposed in the paper is illustrated on UK money data.

1. INTRODUCTION

Since its introduction, the concept of common trends has been central in the representation and in the interpretation of cointegrated systems. See Stock and Watson (1988) and references therein. In $I(1)$ systems the notion of common trends is linked to a reduced-rank impact matrix for the moving average (MA) representation of the first differences of the process. See, e.g., Paruolo (1997a) and discussion therein; see also Phillips (1998) for the relation of these concepts with impulse responses. This reduced-rank structure is the basis of the Beveridge–Nelson decomposition of a system into stationary and nonstationary components; see Proietti (1997) and Johansen (1999). The MA impact matrix is also one fundamental determinant of the long-run variance matrix of the system; see, e.g., Paruolo (1997b).

Although originally introduced in the context of $I(1)$ systems, the concept of common trends was extended in Johansen (1992) to $I(2)$ systems, where it is again associated with the rank deficiency of the impact matrix for the MA representation of the second differences of the process, indicated by C_2 in the fol-

Partial financial support from Italian MURST grants ex 60% is gratefully acknowledged. The paper was presented in Bertinoro, Forlì, Italy, January 27–29, 2000, where I received useful comments. I also thank two anonymous referees for useful comments. Address correspondence to: Paolo Paruolo, Dipartimento di Economia, Università dell'Insubria, via Ravasi 2, 21100 Varese, Italy; e-mail: paolo.paruolo@uninsubria.it.

© 2002 Cambridge University Press 0266-4666/02 $9.50

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

lowing. Just as in the $I(1)$ case it is easy to realize that the long-run variance of the second differences of the process is a quadratic function of C_2 (see Section 2.1, which follows).

Also in the $I(2)$ context the reduced-rank structure of C_2 is the basis for a Beveridge–Nelson type of decomposition that separates components of various order of integration; see Juselius (1994, 1995). In particular, indicating by α_2' a basis of the row space of C_2, the common $I(2)$ trends are the result of the double cumulation of α_2' times the innovation of the system. Thus C_2, and particularly α_2, plays a fundamental role in any common trend decomposition in $I(2)$ systems. The rows in α_2' are called the common $I(2)$ trends (I2CT) linear combinations.

In the spirit of Bedini and Mosconi (2000), measures of the total adjustment to given disequilibrium errors (impact factors [IF]) are shown to be functions of C_2 and its components. Moreover Proietti's (1997) decomposition into transitory and permanent components for the differences of an $I(2)$ process can be shown to depend on C_2; see also Johansen (1999) on the $I(1)$ case.

This paper considers inference on α_2, the I2CT linear combinations, C_2, the MA impact matrix for the second differences of an $I(2)$ vector autoregressive (VAR) process, and functions thereof. It is shown how these coefficients can be estimated by maximum likelihood (ML). Most of the relevant limit distributions are shown to be Gaussian; it is discussed how the asymptotic standard errors can be estimated via the use of artificial regressions. These expressions can then be used to formulate Wald-type tests on α_2 and C_2. Inference on the impact factor can be addressed similarly.

The paper follows a (quasi-)likelihood approach, in the spirit of Johansen (1996). Other approaches to inference in $I(2)$ systems have been proposed; they include Phillips and Chang (1994), Stock and Watson (1993), and Kitamura (1995).

The rest of the paper is organized as follows. Section 2 defines the parameters of interest; Section 3 derives the estimators; Section 4 reports the asymptotic results on the estimators (Section 4.1), describes artificial regressions that can be used to calculate the asymptotic standard errors (Section 4.2), and discusses the associated Wald tests (Section 4.3). Section 5 discusses some other special cases, whereas Section 6 reports an illustration of the techniques on UK money data. Section 7 reports extensions and conclusions. Proofs are placed in the Appendix.

In the following discussion $a := b$ and $b =: a$ indicate that a is defined by b; $(a : b)$ indicates the matrix obtained by horizontally concatenating a and b. For any full column rank matrices H, A, B, $sp(H)$ is the linear span of the columns of H, $\bar{H}$ indicates $H(H'H)^{-1}$, and $H_\perp$ indicates a basis of $sp(H)^\perp$, the orthogonal complement of $sp(H)$. Moreover $P_H := H\bar{H}'$, $H_{AB} := \bar{A}'H\bar{B}$, $H_{AB.C} := H_{AB} - H_{AC}H_{CC}^{-1}H_{CB}$, whereas $H_A := H(A'H)^{-1}$. Finally $(\cdot)_{ij}$ indicates the ijth element of the argument matrix, vec is the column stacking operator, $\otimes$ is the Kronecker product (i.e., $A \otimes B$ is the matrix with generic block $a_{ij}B$, where $A := [a_{ij}]$), and $\xrightarrow{w}$ indicates weak convergence.

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

2. REPRESENTATIONS

In this paper we consider kth-order vector autoregressive (VAR(k)) systems of the type

$$\Delta^2 X_t = \Pi X_{t-1} + \Gamma \Delta X_{t-1} + \sum_{i=1}^{k-2} \Upsilon_i \Delta^2 X_{t-i} + \mu d_t + \epsilon_t,$$

where X_t and ϵ_t are $p \times 1$ and ϵ_t is independent and identically distributed (i.i.d.) $N(0, \Omega)$, d_t is a $q \times 1$ vector of deterministic terms. In the following discussion we assume $k \geq 2$; several particular cases arise when $k = 2$ and certain restrictions hold; these cases are presented in Section 5.

In the rest of this section we recall the common trends (CT) representation in Section 2.1 and an equilibrium correction (EC) formulation of the system in Section 2.2. The CT representation shows the central role played by the MA impact matrix C_2 in the definition of common $I(2)$ trends. The EC formulation is the one that is used in the estimation of the parameters of the system and hence also of C_2. Finally some impact coefficients of disequilibrium errors on the system are derived; it is shown how these coefficients depend on C_2.

2.1. Common Trends Representation

We first list some assumptions. Let $\Upsilon := \sum_{i=1}^{k-2} \Upsilon_i$.

$I(2)$ Assumption.

1. Every root z of the characteristic polynomial of X_t satisfies $z = 1$ or $|z| > 1$.
2. $\Pi = \alpha\beta'$, where α and β are $p \times r$ matrices of full rank $r < p$.
3. $P_{\alpha_\perp} \Gamma P_{\beta_\perp} = \alpha_1 \beta_1'$, where α_1 and β_1 are $p \times s$ matrices of full rank $s < p - r$.
4. $\alpha_2' \theta \beta_2$ has full rank $p - r - s$, where $\alpha_2 = (\alpha : \alpha_1)_\perp$, $\beta_2 = (\beta : \beta_1)_\perp$, and θ is defined as

$$\theta := \Gamma \bar{\beta} \bar{\alpha}' \Gamma + I - \Upsilon. \tag{1}$$

Johansen's $I(2)$ representation theorem (see Johansen, 1992 or 1996, Theorem 4.6) establishes under 1 that necessary and sufficient conditions for X_t to be integrated of order 2 are the conditions listed in 2, 3, and 4. In the following discussion "$I(2)$ assumption" and "$I(2)$ conditions" are used as synonyms.

Under the $I(2)$ assumption X_t has a CT representation

$$X_t = C_2 \sum_{s=1}^{t} \sum_{i=1}^{s} (\epsilon_i + \mu d_i) + C_1 \sum_{i=1}^{t} (\epsilon_i + \mu d_i) + C_0(L)(\epsilon_t + \mu d_t) + m_0 + m_1 t, \tag{2}$$

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

where m_0, m_1 depend on initial values, and

$$C_2 = \beta_2(\alpha_2'\theta\beta_2)^{-1}\alpha_2' \tag{3}$$

is the matrix of loadings of the $I(2)$ trends into the system. This matrix can be decomposed in the product of $\beta_2(\alpha_2'\theta\beta_2)^{-1}$ and α_2'; the linear combinations $\alpha_2'\epsilon_i$ are the innovations that form the common $I(2)$ trends. In the following discussion α_2 is called "the I2CT linear combinations."

Representation (2) is parallel to the one exploited in $I(1)$ systems in the CT Beveridge–Nelson decomposition (see Stock and Watson, 1988). From (2) it is also simple to deduce that the long-run variance of $\Delta^2 X_t$ equals $C_2\Omega C_2'$, a quadratic function of C_2. Inference on $C_2\Omega C_2'$ may be addressed using the results of this paper along the lines in Paruolo (1997b), to which we refer for details.

Note that α_2 and β_2 can be substituted within C_2 with any bases of $sp(\alpha_2)$ and $sp(\beta_2)$ without affecting C_2. In fact any other basis of $sp(\alpha_2)$ (respectively, $sp(\beta_2)$) is of the form $\alpha_2 c_1$ (respectively, $\beta_2 c_2$) with c_i square and nonsingular, and the factors c_i cancel within (3). In other words C_2 is invariant with respect to the choice of bases of $sp(\alpha_2)$ and $sp(\beta_2)$. In particular it can be expressed as $C_2 = \bar{\beta}_2(\alpha_2'\theta\bar{\beta}_2)^{-1}\alpha_2'$; this formulation turns out to be useful in equation (4), which follows.

The CT representation (2) can be used to decompose the growth rate of X_t into transitory and permanent components along the lines of Proietti (1997). In fact one can define $\Delta X_t^{\text{perm}} := C_2\theta\Delta X_t$ and $\Delta X_t^{\text{trans}} := (I - C_2\theta)\Delta X_t$ (see Johansen, 1999). It is simple to verify, in fact, that $C_2\theta C_2 = C_2$, so that ΔX_t^{perm} still contains the common trends, whereas $(I - C_2\theta) = (I - C_2\theta)P_\tau$, where $\tau := (\beta : \beta_1)$, because $(I - C_2\theta)\bar{\beta}_2 = \bar{\beta}_2 - \bar{\beta}_2 = 0$. This shows that $\Delta X_t^{\text{trans}} = (I - C_2\theta)\bar{\tau}\tau'\Delta X_t$ does not contain CT. Note that again this transitory-permanent decomposition is based on C_2.

It is useful for later development to express (3) in terms of the EC formulation of Paruolo and Rahbek (1999), which is reported in the next subsection and summarized in Tables 1 and 2. One has

$$\theta\bar{\beta}_2 = \zeta_1\delta + (I - \Upsilon)\bar{\beta}_2, \tag{4}$$

where $\bar{\beta}_2$ is used in place of β_2, conformably with $C_2 = \bar{\beta}_2(\alpha_2'\theta\bar{\beta}_2)^{-1}\alpha_2'$; note also that $\Upsilon = 0$ if $k = 2$. Equations (3) and (4) show how C_2 depends on α_2, ζ_1, Υ, and the cointegration parameters $\gamma_1 := (\beta, \beta_1, \delta)$.

2.2. Equilibrium Correction Form and Impact Factors

In this section we first report an EC representation that is used in the estimation of the system. Other EC representations may be defined; see the discussion in Paruolo and Rahbek (1999). We also give a brief reference to impact factors in $I(2)$ systems; a systematic exposition of the latter is given in Omtzigt and Paruolo (2002).

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

TABLE 1. Parameter definitions; $n := p(k-2)+q$

Parameter	Definition	Matrix dimensions
δ	$\bar{\alpha}'\Gamma\bar{\beta}_2$	$r \times (p-r-s)$
ζ_1	$\Gamma\bar{\beta}$	$p \times r$
ζ_2	$\Gamma\bar{\beta}_1$	$p \times s$
ς	$(\alpha : \zeta_2)$	$p \times (r+s)$
Υ	$\sum_{i=1}^{k-2}\Upsilon_i$	$p \times p$
τ	$(\beta : \beta_1)$	$p \times (r+s)$
γ_1	(τ, δ)	This is a set
γ_2	$(\varsigma : \zeta_1 : \Psi)$	$p \times (2r+s+n)$

Under the $I(2)$ conditions the linear combinations $Y_t := \beta' X_{t-1} + \delta\beta_2'\Delta X_{t-1}$, $N_t := \beta'\Delta X_{t-1}$ and $\beta_1'\Delta X_{t-1}$ are stationary, i.e., β, β_1, and δ are cointegration coefficients. See Johansen (1996, Theorem 4.6).[1] The system allows the following EC representation (see Paruolo and Rahbek, 1999, equation (2.11)):

$$\Delta^2 X_t = \alpha(\beta' X_{t-1} + \delta\beta_2'\Delta X_{t-1}) + (\zeta_1 : \zeta_2)(\beta : \beta_1)'\Delta X_{t-1} + \Psi U_t + \epsilon_t.$$

We now introduce some notation that simplifies inference on α_2. Let $\varsigma := (\alpha : \zeta_2)$; it can be noted that $sp(\varsigma) = sp(\alpha, \alpha_1)$ as a result of the definition of $\zeta_2 := \Gamma\bar{\beta}_1$; see Table 1.[2] This shows that $\alpha_2 = \varsigma_\perp$ and suggests rewriting the EC representation as follows:

$$Z_t = (\varsigma : \zeta_1 : \Psi)(J_t' : N_t' : U_t')' + \epsilon_t, \tag{5}$$

where $Z_t := \Delta^2 X_t$ is the vector of dependent variables, $U_t := (\Delta^2 X_{t-1}', \ldots, \Delta^2 X_{t-k+2}', d_t')'$ contains the lagged second differences and deterministics,

TABLE 2. Variable definitions; $n := p(k-2)+q$

Variables	Definition	Vector dimensions
Z_t	$\Delta^2 X_t$	$p \times 1$
V_t	$\Delta^2 X_{t-1}$	$p \times 1$
U_t	$(\Delta^2 X_{t-1}' : \ldots : \Delta^2 X_{t-k+2}' : d_t')'$	$n \times 1$
W_t	$(\Delta^3 X_{t-1}' : \ldots : \Delta^3 X_{t-k+3}' : d_t')'$	$(n-p) \times 1$
Y_t	$\beta' X_{t-1} + \delta\beta_2'\Delta X_{t-1}$	$r \times 1$
J_t	$(Y_t' : \Delta X_{t-1}'\beta_1)'$	$(r+s) \times 1$
N_t	$\beta'\Delta X_{t-1}$	$r \times 1$
F_t	$(J_t' : N_t' : V_t')'$	$(p+2r+s) \times 1$
G_t	$(J_t' : N_t' : U_t')'$	$(n+2r+s) \times 1$

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

$J_t := (Y_t' : \Delta X_{t-1}' \beta_1)'$ is $(r+s) \times 1$ and contains the EC terms with coefficient $\varsigma := (\alpha : \zeta_2)$, and $N_t := \beta' \Delta X_{t-1}$ of dimension $r \times 1$ contains the remaining EC terms. Note that the definitions of J_t and N_t involve the cointegration parameters β, β_1, and δ and that the timing of variables appearing in the EC terms $\beta' X_{t-1} + \delta \beta_2' \Delta X_{t-1}$ and $(\beta : \beta_1)' \Delta X_{t-1}$ could be shifted, just as in $I(1)$ systems.[3] However, any time shift would not affect the results of the paper.

The MA impact matrix C_2 is also involved in the definition of impact factor (IF). In Proposition 1, which follows, we give an example of impact factor for system (5). Let a_t and b_t be variables belonging to a stationary VAR process, with companion form $\widetilde{X}_t = \tilde{A} \widetilde{X}_{t-1} + \tilde{\varepsilon}_t$. The best forecasts under quadratic loss are given by the conditional expectation $E(\widetilde{X}_{t+h} | \widetilde{X}_t)$. Bedini and Mosconi (2000) suggest measuring the total impact of a_0 on $\sum_{i=1}^{t} b_i$ by

$$\mathrm{IF}_{a,b} := \sum_{i=1}^{\infty} \frac{\partial E(b_i | \widetilde{X}_0)}{\partial a_0'} = \sum_{i=1}^{\infty} \frac{\partial E(b_{t+i} | \widetilde{X}_t)}{\partial a_t'}, \tag{6}$$

where the equality follows from the stationarity of $\widetilde{X}_t$. The last expression gives a forecasting interpretation to $\mathrm{IF}_{a,b}$, which measures the cumulated effect on the forecasts at time t of b_{t+i} due to changes in a_t.

In the present context system (5) can be rewritten as a stationary process for an appropriate choice of state space vector $\widetilde{X}_t$; in particular one can choose the components $a_t := Y_t$ and $b_t := \Delta^2 X_t$ of $\widetilde{X}_t$ in (6). For this choice, the following proposition shows that $\mathrm{IF}_{a,b}$ is a linear function of C_2. For further reference on impact factors and a proof of Proposition 1, refer to Omtzigt and Paruolo (2002).

PROPOSITION 1. *Under the* I(2) *conditions*

$$\psi := \mathrm{IF}_{Y, \Delta^2 X} = -C_2 \zeta_1.$$

Taking the cumulation implicit in definition (6) into account, ψ measures the total impact of the disequilibrium error Y_0 on the expected value of the growth rate of the system, $\Delta X_t - \Delta X_0$.

3. ESTIMATION

Consider the EC representation (5); the parameters can be partitioned into three subsets, γ_1, γ_2, and Ω, where $\gamma_1 := (\beta, \beta_1, \delta)$ and $\gamma_2 := (\varsigma : \zeta_1 : \Psi)$. Here γ_1 contains the cointegrating parameters, while γ_2 contains the "short-run-dynamics" parameters. In this section both γ_1 and γ_2 are assumed to be unrestricted; i.e., the corresponding parameter space for $(\gamma_1, \gamma_2, \Omega)$ is $\mathbb{R}^m \times \mathbb{R}^{p(2r+s+n)} \times \mathbb{S}_p$, for an appropriate dimension m, where $\mathbb{S}_p$ is the set of square symmetric positive definite matrices of order p. Section 7 extends the present results to the restricted case, i.e., where restrictions are imposed on γ_1 and/or γ_2. We observe that for fixed values of γ_1, J_t and N_t are observable regressors and the ML es-

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

timator of γ_2, indicated by $\hat{\gamma}_2(\gamma_1)$, can be obtained by regression. Setting $G_t := (J_t' : N_t' : U_t')'$ one finds

$$(\hat{\varsigma}(\gamma_1) : \hat{\xi}_1(\gamma_1) : \hat{\Psi}(\gamma_1)) = M_{ZG} M_{GG}^{-1},$$

where $M_{ab} := T^{-1} \sum_{t=1}^{T} a_t b_t'$. In the following we will consider either ML estimators for the (multi-)cointegration parameters γ_1 (see Johansen, 1997) or the efficient 2SI2 estimator of Johansen (1995). Both estimators of γ_1 are superconsistent and have the same limit distribution (see Paruolo, 2000); both of them are indicated by $\hat{\gamma}_1$. In the following discussion we indicate by $\hat{\gamma}_2$ both the ML and 2SI2 estimators of γ_2, $\hat{\gamma}_2 := \hat{\gamma}_2(\hat{\gamma}_1)$; in both cases $\hat{\gamma}_2$ presents the same limit distribution.[4] In the following discussion we also let $\tau := (\beta : \beta_1)$ and indicate by $\hat{\epsilon}_t := Z_t - \hat{\gamma}_2 G_t$ the estimated least squares residuals obtained in regression (5) when γ_1 is substituted by $\hat{\gamma}_1$ within G_t. This also defines the estimator $\hat{\Upsilon}$ of Υ, $\hat{\Upsilon} := \sum_{i=1}^{k-2} \hat{\Upsilon}_i$.

To obtain an identified estimator of α_2 we assume that a known matrix a of the same order of α_2 exists such that $a'\alpha_2$ is of full rank; we then define

$$\alpha_{2a} := \alpha_2 (a'\alpha_2)^{-1}$$

as, e.g., in Johansen (1996, chap. 5.2). We consider the corresponding estimator of α_2,

$$\hat{\alpha}_{2a} := \hat{\varsigma}_\perp (a'\hat{\varsigma}_\perp)^{-1}. \tag{7}$$

We observe that $C_2 = \bar{\beta}_2 (\alpha_2' \theta \bar{\beta}_2)^{-1} \alpha_2'$ is an already estimable function of the parameters in the model and does not need any normalization. We consider the estimators of C_2 and ψ obtained by substituting $\hat{\gamma}_1$, $\hat{\varsigma}$, $\hat{\xi}_1$, and $\hat{\Upsilon}$ in expression (3), i.e.,

$$\hat{C}_2 = \hat{\bar{\beta}}_2 (\hat{\varsigma}_\perp' (\hat{\xi}_1 \hat{\delta} + (I - \hat{\Upsilon}) \hat{\bar{\beta}}_2))^{-1} \hat{\varsigma}_\perp', \qquad \hat{\psi} = -\hat{C}_2 \hat{\xi}_1,$$

where $\hat{\beta}_2$ is the identified estimator of $\beta_2 := \tau_\perp$ used in the estimation of δ.[5] The special VAR(2) case, $k = 2$, with the true parameter value $\delta = 0$ is treated separately in Section 5; in that section a constrained estimator of C_2 is considered, which incorporates the restriction $\delta = 0$. This constrained estimator is still based on the unrestricted estimator $\hat{\gamma}_1$, $\hat{\gamma}_2$ described earlier. Extensions to more generally constrained estimators of γ_1 and γ_2 are treated in Section 7.[6]

4. ASYMPTOTICS

This section reports the asymptotic analysis referred to the unconstrained estimators $\hat{\gamma}_1$ and $\hat{\gamma}_2$. We work under the following hypothesis.

Assumption $\mathcal{C}$ (correct specification). The $I(2)$ conditions hold; the integration indices r and s and the lag length k are correctly specified.

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

In the following discussion the asymptotic variances of a generic asymptotically Gaussian estimator $\hat{b}$ are indicated by $\Lambda(\hat{b})$. If this variance matrix has a Kronecker product form, it is indicated by $\Lambda(\hat{b}) = \Xi(\hat{b}) \otimes \kappa(\hat{b})$. The corresponding (ML) estimator is indicated by $\hat{\Lambda}(\hat{b}) = \hat{\Xi}(\hat{b}) \otimes \hat{\kappa}(\hat{b})$.

In the following discussion it is convenient to separate Υ from within Ψ. We define $F_t := K_1' G_t = (J_t' : N_t' : V_t')'$, $W_t := K_2' G_t = (\Delta^3 X_{t-1}' : \ldots : \Delta^3 X_{t-k+3}' : d_t')'$, where

$$K_1' := (I_\ell \quad 0_{\ell \times n}) \qquad K_2' := \begin{pmatrix} I_p & -I_p & & & \\ \cdots & \cdots & \cdots & & \\ & & I_p & -I_p & 0_{p\times q} \\ & & & & I_q \end{pmatrix},$$

$\ell := 2r + s + p$, and $K := (K_1 : K_2)$ is a square nonsingular matrix. It is easy to check that this transformation of variables corresponds to the reparameterization

$$\Psi U_t = \Upsilon V_t + \phi W_t,$$

which isolates Υ within Ψ. This piece of notation is reported in Table 2, and it is used also in artificial regressions in Section 4.2.

4.1. Standard Errors

Let Σ be the second moment matrix of F_t corrected for W_t and let Σ_{xy} indicate a subblock of Σ corresponding to the generic variables x_t and y_t. Also let $S_{xy} := M_{xy} - M_{xW} M_{WW}^{-1} M_{Wy}$.

THEOREM 2. *Under Assumption C*

$$T^{1/2}(\mathrm{vec}(\bar{a}_\perp'(\hat{\alpha}_{2a} - \alpha_{2a}))) \xrightarrow{w} N(0, \Xi(\bar{a}_\perp' \hat{\alpha}_{2a}) \otimes \kappa(\bar{a}_\perp' \hat{\alpha}_{2a})),$$

where $\Xi(\bar{a}_\perp' \hat{\alpha}_{2a}) = \alpha_{2a}' \Omega \alpha_{2a}$ *and* $\kappa(\bar{a}_\perp' \hat{\alpha}_{2a}) = (\varsigma' a_\perp)^{-1} \Sigma_{JJ.(N,V)}^{-1} (a_\perp' \varsigma)^{-1}$. *The asymptotic variance matrix can be consistently estimated by*

$$\hat{\Lambda}(\bar{a}_\perp' \hat{\alpha}_{2a}) = \hat{\alpha}_{2a}' \hat{\Omega} \hat{\alpha}_{2a} \otimes (a_\perp' \hat{\varsigma} S_{JJ.(N,V)} \hat{\varsigma}' a_\perp)^{-1}. \tag{8}$$

Moreover $T^{1/2}\mathrm{vec}(\hat{C}_2 - C_2) \xrightarrow{w} N(0, \Xi(\hat{C}_2) \otimes \kappa(\hat{C}_2))$ *where* $\Xi(\hat{C}_2) = Q' \Sigma_{FF}^{-1} Q$ *and* $\kappa(\hat{C}_2) = C_2 \Omega C_2'$ *with*

$$Q' := ((C_2' \theta' - I)\bar{\varsigma} : -C_2' \beta_2 \delta' : C_2'), \tag{9}$$

which can be consistently estimated by

$$\hat{\Lambda}(\hat{C}_2) = \hat{Q}' S_{FF}^{-1} \hat{Q} \otimes \hat{C}_2 \hat{\Omega} \hat{C}_2', \tag{10}$$

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

where $\hat{Q}$ is defined as in (9) with parameters substituted by the corresponding estimators. If

(i) $k > 2$ or (ii) $k = 2$ and δ is of full column rank **(11)**

then Q is a full column rank matrix.

Analogously $T^{1/2}\mathrm{vec}(\hat{\psi} - \psi) \xrightarrow{w} N(0, \Xi(\hat{\psi}) \otimes \kappa(\hat{\psi}))$, *where* $\Xi(\hat{\psi}) = L'\Sigma_{FF}^{-1}L$, $\kappa(\hat{\psi}) = \kappa(\hat{C}_2)$, *and*

$$L' := -(\zeta_1'(C_2'\theta' - I_p)\bar{\varsigma} : (I_r - \zeta_1' C_2'\beta_2\delta') : \zeta_1' C_2'). \quad \textbf{(12)}$$

Here L is of full column rank provided $\alpha_2'(I - \Upsilon)\beta_2$ is of full rank.

Several remarks follow from Theorem 2.

Remark (i). The estimator of α_2 is a function of $\hat{\gamma}_1$ through $\hat{\gamma}_2(\gamma_1)$. Because $\hat{\gamma}_1$ is superconsistent (see Paruolo, 2000), one can substitute γ_1 in place of $\hat{\gamma}_1$ in the derivations. For γ_1 fixed at the true value, α_2 is a function of the regression coefficient to stationary variables, which is $T^{1/2}$ consistent and asymptotically normal under correct specification.

Remark (ii). The same arguments mentioned in remark (i) are valid for the estimator of C_2, which, however, is singular by construction. This is reflected in the singularity of the asymptotic variance matrix of C_2. In fact choosing the linear combination $R_1 \in sp(\tau)$ one has that $T^{1/2}R_1'(\hat{C}_2 - C_2)$ has asymptotically null variance. In other words, the econometrician has to make sure that he or she is not testing for cointegration ($R_1 \in sp(\tau)$) when making inference on C_2. To account for the singularity of the asymptotic distribution of C_2, in the following section we consider a decomposition of hypotheses on it in the spirit of Paruolo (1997b).

Remark (iii). The asymptotic standard errors in Theorem 2 require correct specification of the integration indices r and s. They provide the basis for Wald tests on α_2 in the course of a London School of Economics–type modeling strategy, where tests are used to reduce model dimensions sequentially. In this spirit, the integration indices are determined first, and further model reductions corresponding to hypotheses on α_2 and C_2 are sought next. One of the referees pointed out that an alternative approach is the one of constructing inferential procedures that do not require prespecification of r and s, in the spirit of Phillips and Chang (1994).

Remark (iv). It may be observed that under the conditions (11) the matrix $\Xi(\hat{C}_2)$ is of full rank; thus $T^{1/2}e_i'(\hat{C}_2 - C_2)R_2$ has asymptotically positive variance for any vector R_2 if $e_i \notin sp(\tau)$ despite the fact that $C_2R_2 = 0$ if $R_2 \in sp(\varsigma)$. This phenomenon is due to the fact that the first-order asymptotic variance $\hat{C}_2$ receives a positive contribution from the estimation of ζ_1 (and of Υ if $k > 2$) in addition to the one associated with ς.

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

4.2. Artificial Regressions

In this section we present artificial regressions that can be used to calculate the variance matrices reported in Theorem 2 of the previous section. Recall from Section 2 that $\hat{\epsilon}_t := Z_t - (\hat{\gamma}_2 : \hat{\phi})(F_t' : W_t')'$ denotes the estimated least residuals obtained in regression (5). We consider the following artificial regressions:

$$\hat{\alpha}_{2a}'\hat{\epsilon}_t = A_1(a_\perp'\hat{\varsigma}J_t) + A_2(N_t' : U_t')' + \text{error}_t, \tag{13}$$

$$\hat{C}_2\hat{\epsilon}_t = A_1(\hat{\bar{H}}'F_t) + A_2(F_t'\hat{H}_\perp : W_t')' + \text{error}_t, \tag{14}$$

where $H = Q$ or L in (14), $\hat{H}_\perp$ is any basis of $sp(\hat{H})^\perp$, and A_1 and A_2 are artificial parameters. Let m_1 be the number of regressors in (13) and m_2 the number of regressors in (14). The following proposition summarizes the results.

PROPOSITION 3. *In order to calculate* $\hat{\Lambda}(\bar{a}_\perp'\hat{\alpha}_{2a})$ *compute the artificial regression (13); the* OLS *estimated covariance matrix corresponding to* $\text{vec}(A_1')$ *multiplied by* $(T - m_1)/T$ *is equal to the estimated covariance matrix* $\hat{\Lambda}(\bar{a}_\perp'\hat{\alpha}_{2a})$ *defined in (8).*

Let the conditions (11) hold; in order to calculate $\hat{\Lambda}(\hat{C}_2)$ *compute the artificial regression (14) with* $H := Q$*; the* OLS *estimated covariance matrix corresponding to* $\text{vec}(A_1)$ *multiplied by* $(T - m_2)/T$ *is equal to the estimated covariance matrix* $\hat{\Lambda}(\hat{C}_2)$ *defined in (10). The same artificial regression (14) produces* $\hat{\Lambda}(\hat{\psi})$ *when* $H := L$.

It may be observed that in case conditions (11) do not hold, then Q is of reduced rank $q < p$ and the attempt to run artificial regression (14) results in a computational error, because in this case the regressor's second moment matrix is singular and cannot be inverted. Therefore an OLS routine diagnostic of "collinearity" when applied to (14) could signal the empirical violation of condition (11). Note that this can take place only if $k = 2$. The case when (11) is violated is treated in Section 5.

4.3. Wald Tests

Following Paruolo (1997a) we consider hypotheses on C_2 of the form

$$H_0 : R_1'C_2R_2 = 0, \tag{15}$$

where R_i has n_i columns and is assumed to be of full column rank, $i = 1, 2$. Hypothesis H_0 in (15) is decomposed as $H_0 = \cap_{i=1}^{n_1} \cap_{j=1}^{n_2} {}^{ij}H_0$ where ${}^{ij}H_0$: $R_{1i}'C_2R_{2j} = 0$, $R_1 =: (R_{11} : \ldots : R_{1n_1})$, and $R_2 =: (R_{21} : \ldots : R_{2n_2})$. Each hypothesis ${}^{ij}H_0$ is a unidimensional hypothesis because it regards one linear combination of elements in C_2. The hypothesis ${}^{ij}H_0$ is further decomposed as ${}^{ij}H_0 = {}^{\cdot j}H_{01} \cup {}^{i\cdot}H_{02} \cup {}^{ij}H_{03}$ where

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

$^{\cdot j}H_{01} : R_{2j} \in sp(\alpha : \alpha_1)$

$^{i\cdot}H_{02} : R_{1i} \in sp(\tau)$

$^{ij}H_{03} : R'_{1i} C_2 R_{2j} = 0,\ R_{2j} \notin sp(\alpha : \alpha_1),\ R_{1i} \notin sp(\tau).$

For definitions of parameters see Tables 1 and 2. The first two hypotheses concern $sp(\alpha : \alpha_1)$ and $sp(\beta : \beta_1)$ and are not treated here; see Johansen (2000), Paruolo and Rahbek (1999), and references therein. In this paper we concentrate on the hypothesis $^{ij}H_{03}$, which assumes that $R_{2j} \notin sp(\varsigma), R_{1i} \notin sp(\tau)$, and we consider Wald-type statistics of the type

$$t(R'_{1i} C_2 R_{2j}) := \frac{T^{1/2} R'_{1i} \hat{C}_2 R_{2j}}{(R'_{2j} \hat{\Xi}(\hat{C}) R_{2j} \cdot R'_{1i} \hat{\kappa}(\hat{C}) R_{1i})^{1/2}}$$

(see Theorem 2). Each test on $^{(\cdot)}H_{0h}$ can be recombined as in Paruolo (1997a) to give a test of (15) with chosen size.

We also consider hypotheses on α_2 of the type

$$R' \mathrm{vec}(\bar{a}'_\perp \alpha_{2a}) = R_0, \tag{16}$$

where R has n columns; for this hypothesis we define the Wald statistics

$$J(\bar{a}'_\perp \alpha_{2a}) := T(R' \mathrm{vec}(\bar{a}'_\perp \hat{\alpha}_{2a}) - R_0)'(R' \hat{\Lambda}(\bar{a}'_\perp \hat{\alpha}_{2a}) R)^{-1} (R' \mathrm{vec}(\bar{a}'_\perp \hat{\alpha}_{2a}) - R_0).$$

The special case of $R = (e_j \otimes e_i)$, $R_0 = 0$, where e_i is a unit vector with all zeros and 1 in position i, corresponds to the hypothesis $(\bar{a}'_\perp \hat{\alpha}_{2a})_{ij} = 0$; in this case one can also consider the t-statistic

$$t((\bar{a}'_\perp \alpha_{2a})_{ij}) := \frac{T^{1/2} (\bar{a}'_\perp \hat{\alpha}_{2a})_{ij}}{((\hat{\Xi}(\bar{a}'_\perp \hat{\alpha}_{2a}))_{jj} \cdot (\hat{\kappa}(\bar{a}'_\perp \hat{\alpha}_{2a}))_{ii})^{1/2}}.$$

COROLLARY 4. *Let Assumption C hold. Under* $^{ij}H_{03}$ *one has* $t(R'_{1i} C_2 R_{2j}) \xrightarrow{w} N(0,1)$ *and under (16) one has* $J(\bar{a}'_\perp \alpha_{2a}) \xrightarrow{w} \chi^2(n)$, $t((\bar{a}'_\perp \alpha_{2a})_{ij}) \xrightarrow{w} N(0,1)$.

5. SOME SPECIAL CASES

In this section we discuss a special case for $k = 2$ that requires a special treatment. This case corresponds to $\delta = 0$, where equation (4) implies $\theta \beta_2 = \beta_2 := \tau_\perp$, showing that C_2 is just a function of ς and τ and does not depend on ζ and Υ. In the rest of this section we concentrate on this case, where $\alpha'_2 \theta \beta_2 = \alpha'_2 \beta_2 := \varsigma'_\perp \tau_\perp$, which is assumed to be of full rank by the $I(2)$ conditions. Here C_2 takes the form

$$C_2 = \tau_\perp (\varsigma'_\perp \tau_\perp)^{-1} \varsigma'_\perp = I - \varsigma(\tau' \varsigma)^{-1} \tau'.$$

We consider the estimator of C_2 obtained substituting $\hat{\varsigma}$ and $\hat{\tau}$ within the previous expression for C_2, where $\hat{\varsigma}$ and $\hat{\tau}$ represent the unconstrained estimators of the previous sections. The following theorem considers the hypothesis $^{ij}H_{03}$ about C_2.

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

THEOREM 5. *Let Assumption C and* ${}^{ij}H_{03}$ *hold; if* $k = 2$ *and* $\alpha_2'\theta\beta_2 = \alpha_2'\beta_2 := \varsigma_\perp'\tau_\perp$, *then the following cases apply to the estimator* $\hat{C}_2 = I - \hat{\varsigma}(\hat{\tau}'\hat{\varsigma})^{-1}\hat{\tau}'$:

1. *if* $R_{2j} \notin sp(\tau_\perp)$ *then*

$$T^{1/2}R_{1i}'(\hat{C}_2 - C_2)R_{2j} = -R_{1i}'C_2(T^{1/2}(\hat{\varsigma} - \varsigma))(\tau'\varsigma)^{-1}\tau'R_{2j} + o_p(1);$$

2. *if* $R_{2j} \in sp(\tau_\perp)$ *and* $R_{1i} \notin sp(\varsigma_\perp)$ *then*

$$TR_{1i}'(\hat{C}_2 - C_2)R_{2j} = -R_{1i}'\varsigma(\tau'\varsigma)^{-1}(T(\hat{\tau} - \tau)')C_2R_{2j} + o_p(1);$$

3. *if* $R_{2j} \in sp(\tau_\perp)$ *and* $R_{1i} \in sp(\varsigma_\perp)$ *then*

$$T^{3/2}R_{1i}'(\hat{C}_2 - C_2)R_{2j} = -R_{1i}'C_2(T^{1/2}(\hat{\varsigma} - \varsigma))(T(\hat{\tau} - \tau)')C_2R_{2j} + o_p(1).$$

The results in Theorem 5 can be used to derive Wald tests as in Section 4.3; e.g., for the first case, one can immediately derive that, provided $R_{2j} \notin sp(\tau_\perp)$,

$$T^{1/2}R_{1i}'(\hat{C}_2 - C_2)R_{2j} \xrightarrow{w} N(0, R_{2j}'\tau(\tau'\varsigma\Sigma_{JJ.(N,V)}\varsigma'\tau)^{-1}\tau'R_{2j} \cdot R_{1i}'C_2\Omega C_2'R_{1i})$$

The asymptotic variance is positive and can be consistently estimated by substituting estimators in place of parameters and S_{FF} in place of Σ. Analogous derivation can be applied to the other cases, making use of the results in Johansen (1997) and Paruolo (2000).

We observe that in case 3 one finds non-Gaussian limit distributions, as the asymptotic expression involves the product of a mixed normal and a Gaussian distribution; see Paruolo (1997a, Section 7) for the construction of Wald tests in this situation.

6. AN APPLICATION

To illustrate the preceding procedures we considered the data on UK money, income, prices, and interest rates analyzed in Doornik, Hendry, and Nielsen (1998) and Rahbek, Kongsted, and Jørgensen (1999), among others. The quarterly seasonally adjusted data cover the period 1963Q1 to 1989Q2 and include the log of nominal money (M1), indicated with m, the log of total final expenditure at 1985 prices, *inc*, the log of the implicit price deflator, p, and a measure of the opportunity cost of holding money given by R^*, the difference between the 3-month local authority interest rate and the learning-adjusted retail sight-deposit interest rate. The data consist of the time series of $X_t := (m_t : p_t : inc_t : R_t^*)'$; following previous analyses, the lag length has been set equal to $k = 5$ and the period 1964Q3 to 1989Q2 has been selected as the estimation period, for a total of 100 effective observations.

On the basis of tests on the integration indices, Paruolo (1996) and Rahbek et al. (1999) selected $r = 1$, $s = 1$, $p - r - s = 2$; moreover they could not reject the model with the constant, $d_t = 1$, restricted not to generate quadratic trends in the data, i.e., $\mu = (\alpha : \alpha_1)\rho$.

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

TABLE 3. 2SI2 estimates of of β, β_1, and β_2

	β	β_1	β_2		δ	
m	1	−0.13	1	0	10.97	−4.61
p	−1	0.22	0	1		
inc	−1	1	0.16	−0.24		
R^*	7.00	0.19	−0.12	0.11		

Within such a model, the same authors find that the hypothesis $\beta = (1 : -1 : -1 : *)'$ cannot be rejected by the data, where $*$ indicates an unrestricted entry; we thus selected $r = 1$, $s = 1$, $p - r - s = 2$ as integration indices and imposed the restriction $\beta = (1 : -1 : -1 : *)'$ via the 2SI2 estimator. Table 3 reports the 2SI2 estimates of β, β_1, and β_2 and δ properly normalized. The time series of the estimated disequilibrium errors are plotted in Figure 1.

The estimates of α_2 and C_2, along with *t*-ratios in parentheses, are reported in Table 4. Let e_i be a unit vector with all zeros and 1 in position i. It can be noted that no coefficient in the third column of C_2 is significant at conventional significance levels, suggesting possibly that $e_3'\alpha_2 = 0$, i.e., that all I2CT coefficients on the innovations of the income variable may be zero. A test of this hypothesis can be conducted through the $J(\bar{a}_\perp'\alpha_{2a})$ statistic, choosing $a := (e_4 : e_2)$ and $R := (e_2 : e_4)$. This test statistic equals 4.377, which compared with a $\chi^2(2)$ degrees of freedom gives a *p*-value of 0.112. Thus this hypothesis cannot be rejected at conventional significance levels.

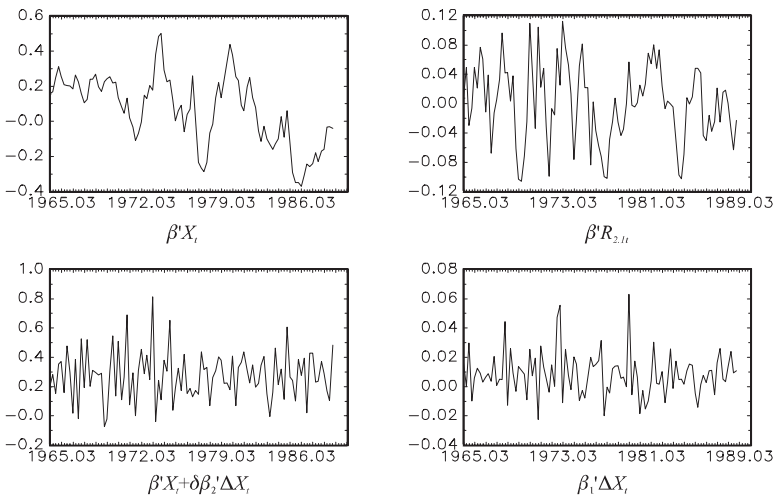


FIGURE 1. Estimated disequilibrium errors; $R_{2.1t}$ indicates the residuals of X_{t-1} regressed on ΔX_{t-1}, $\Delta^2 X_{t-1}$, $\Delta^3 X_{t-i}$, $i = 1, \ldots, k - 3$, and the constant.

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

TABLE 4. 2SI2 estimates of C_2 and α_2

Variables	α_2		C_2				$\psi := -C_2\zeta_1$
m	0.260	**0.169**	−0.077	**0.346**	−0.055	**−0.520**	**0.089**
	(1.854)	(2.107)	(−1.012)	(2.609)	(−0.727)	(−5.199)	(5.041)
p	0	1	**0.143**	**0.572**	0.131	0.178	−0.017
			(2.089)	(4.765)	(1.917)	(1.971)	(−1.076)
inc	0.214	0.163	−0.047	**−0.084**	−0.041	**−0.125**	**0.018**
	(1.471)	(1.953)	(−1.948)	(−2.002)	(−1.688)	(−3.946)	(3.248)
R^*	1	0	0.025	0.020	0.021	**0.082**	**−0.013**
			(1.784)	(0.835)	(1.504)	(4.474)	(−3.891)

t-ratios are in parentheses; entries with *t*-ratio greater than 2 in absolute value are in boldface.

The coefficients in α_2 have been normalized on the innovations of interest rates R_t^* and prices p_t. Although the coefficients in the first linear combination are only marginally significant, the coefficients to the innovation in money *m* and income *inc* are significant in the second I2CT linear combination. This suggests that one I2CT could consist of the innovations of the interest rate R^* alone, whereas the second one could contain the innovations on the other variables.

A joint Wald test of the type $J(\bar{a}_\perp'\alpha_{2a})$ that the unnormalized coefficients of the first column in α_{2a} are zero gives a $\chi^2(2)$ test statistic of 5.820, with a *p*-value of 0.055 that is marginally insignificant. Finally the same test applied to the unnormalized coefficients of the second column in α_{2a} gives a $\chi^2(2)$ test statistic of 8.585, with a *p*-value of 0.0137 that is significant at the 5% level. This suggests that one I2CT could indeed consist of the innovations interest rates R^* alone, whereas the second one could contain the innovations on the other variables. A test that all coefficients except 0.169 in α_{2a} are null gives a $\chi^2(3)$ test statistics of 8.0331, with a *p*-value of 0.0453, marginally rejecting the hypothesis.

The estimates of the impact factors $-C_2\zeta_1$, along with *t*-ratios in parentheses, are also reported in Table 4. It can be noted that all the coefficients are significant, except for the impact factor on prices. A disequilibrium in the long-run money demand thus seems to imply a positive impact on the growth rate of *m* and of *inc*, although determining a negative effect on the growth rate of R_t^*.[7]

7. EXTENSIONS AND CONCLUSIONS

In the previous sections we have derived the limit distributions of the estimators of C_2, α_2, and ψ under no restrictions on the parameters γ_1 and γ_2. We now consider restrictions on these parameters in turn.

In the case where the cointegration parameters γ_1 satisfy smooth restrictions and are estimated under the correctly specified constraints, none of the formulas in Section 4 are affected. In fact, as a result of the superconsistency of $\hat{\gamma}_1$, the standard errors in Section 4 reflect only the variability induced by $\hat{\gamma}_2$. Re-

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

garding the results in Section 5, all the asymptotic expansions remain unaffected, even though, e.g., the limit distribution of $\beta_2'(\hat{\tau} - \tau)$ appearing in case 3 of Theorem 5 would differ from the unrestricted case.

Also the short-run parameters γ_2 could satisfy certain constraints. Under correct specification, the restricted estimator would present a different (and smaller) asymptotic variance matrix. Substituting the restricted estimator $\hat{\gamma}_2$ in the calculation of $\hat{C}_2$, $\hat{\alpha}_2$, and $\hat{\psi}$ would provide more efficient estimators. In this case the standard errors of Theorem 2 would be incorrect. However, all the asymptotic expansions used in the proof of Theorem 2 would continue to hold and could be easily used to find the correct standard errors.

We illustrate the necessary changes by considering $\hat{C}_2$ under constraints on $(\varsigma : \zeta_1 : \Upsilon)$. Let $(\tilde{\varsigma} : \tilde{\zeta}_1 : \tilde{\Upsilon})$ be a constrained estimator of $(\varsigma : \zeta_1 : \Upsilon)$ with asymptotic variance Θ and let $\tilde{C}_2$ be the corresponding estimator of C_2. It is straightforward to see from the proof of Theorem 2 that the corresponding asymptotic variance of $\tilde{C}_2$ would be $(Q' \otimes C_2)\Theta(Q \otimes C_2')$. Analogous considerations apply to $\hat{C}_2$, $\hat{\alpha}_2$, and $\hat{\psi}$ under restrictions on γ_2.

The results of this paper can also be applied to derive the asymptotic variances of other impact factors. Again the method of proof would be the same as the one applied here.

NOTES

1. Recall that β_2 is any basis of $sp(\beta : \beta_1)^{\perp}$, i.e., it is a function of β, β_1.

2. In fact by inserting $I = P_{\alpha} + P_{\alpha_{\perp}}$ in front of ζ_2 one obtains $\zeta_2 = \alpha\Gamma_{\alpha\beta_1} + \alpha_1$. See Paruolo and Rahbek (1999, equation (2.10)).

3. For more on time shifts within this EC formulation see Omtzigt and Paruolo (2002).

4. If $\hat{\gamma}_1$ is the ML estimator the resulting estimator $\hat{\gamma}_2$ is also a ML estimator. Any other superconsistent estimator of γ_1 would also imply the same limit distributions for α_2 and C_2.

5. Again one can note that $\hat{\gamma}_1$ is the ML estimator of γ_1; then $\hat{C}_2$ is the ML estimator of C_2. Constrained estimators of C_2 are also considered in Section 5 for special cases.

6. One referee pointed out that in the $I(1)$ case the limit distribution of the estimator of $\alpha_{\perp}$ can be derived directly from the one of canonical covariates. Similarly in the $I(2)$ case an estimator of α_2 can be obtained in the second stage of the 2SI2 analysis. This estimator of α_2 is different from the one considered in the paper, which is asymptotically efficient due to the superconsistency of the cointegrating parameters. The alternative estimator is not studied here because of space constraints.

7. It must be remembered that R^* is the difference between two interest rates; however, the last result is difficult to interpret from an economic point of view.

REFERENCES

Bedini, C. & R. Mosconi (2000) New Tools for the Dynamic Analysis of Cointegrated VAR Models. Mimeo, Department of Economics and Production, Politecnico di Milano, Italy.

Doornik, J., D. Hendry, & B. Nielsen (1998) Inference in cointegrating models: UK M1 revisited. *Journal of Economic Surveys* 12, 533–540.

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

Johansen, S. (1995) A statistical analysis of cointegration for $I(2)$ variables. *Econometric Theory* 11, 25–59.

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

Johansen, S. (1996) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, second printing. Oxford: Oxford University Press.
Johansen, S. (1997) A likelihood analysis of the $I(2)$ model. *Scandinavian Journal of Statistics* 24, 433–462.
Johansen, S. (1999) Common Trends in the Cointegrated VAR Models. Mimeo, Department of Economics, European University Institute, Fiesole (Fi), Italy.
Johansen, S. (2000) Testing Hypotheses in the $I(2)$ Model. Mimeo, Department of Economics, European University Institute, Fiesole (Fi), Italy.
Juselius, K. (1994) On the duality between long-run relations and common trends in the $I(1)$ and $I(2)$ case. An application to aggregate money holdings. *Econometric Reviews* 13, 151–178.
Juselius, K. (1995) Do purchasing power parity and uncovered interest rate parity hold in the long run? An example of likelihood inference in a multivariate time series model. *Journal of Econometrics* 69, 211–240.
Kitamura, Y. (1995) Estimation of cointegrated systems with $I(2)$ processes. *Econometric Theory* 11, 1–24.
Omtzigt, P. & P. Paruolo (2002) Impact Factors. Working paper in Economics 2002/4, Università dell'Insubria (Va), Italy.
Paruolo, P. (1996) On the determination of integration indices in $I(2)$ systems. *Journal of Econometrics* 72, 313–356.
Paruolo, P. (1997a) Asymptotic inference on the moving average impact matrix in cointegrated $I(1)$ VAR systems. *Econometric Theory* 13, 79–118.
Paruolo, P. (1997b) Standard errors for the long-run variance matrix. *Econometric Theory* 13, Problem 97.1.1, 305; *Econometric Theory* 14, Solution 97.1.1, 152–153.
Paruolo, P. (2000) Asymptotic efficiency of the two stage estimator in $I(2)$ systems. *Econometric Theory* 16, 524–550.
Paruolo, P. & A. Rahbek (1999) Weak exogeneity in $I(2)$ VAR systems. *Journal of Econometrics* 93, 281–308.
Phillips, P.C.B. (1998) Impulse response and forecast error variance asymptotics in nonstationary VARs. *Journal of Econometrics* 83, 21–56.
Phillips, P.C.B. & Y. Chang (1994) Fully modified least squares in $I(2)$ regression. *Econometric Theory* 10, Problem 94.5.1, 967.
Proietti, T. (1997) Short run dynamics in cointegrated systems. *Oxford Bulletin of Economics and Statistics* 59, 405–422.
Rahbek, A., H.C. Kongsted, & C. Jørgensen (1999) Trend-stationarity in the $I(2)$ cointegration model. *Journal of Econometrics* 90, 265–289.
Srivastava, M.S. & C.G. Khatri (1979) *An Introduction to Multivariate Statistics*. New York: Elsevier North-Holland.
Stock, J. & Watson, M. (1988) Testing for common trends. *Journal of the American Statistical Association* 83, 1097–1107.
Stock, J. & Watson M. (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.

APPENDIX

Proof of Theorem 2. The results of the theorem follow by Cramér's theorem. In particular by the results (B.2) and (B.3) in Paruolo (1997a) one has

$$\bar{a}_\perp'(\hat{\alpha}_{2a} - \alpha_{2a}) = \varsigma_{a_\perp}'(\hat{\alpha}_{2a} - \alpha_{2a}) = -(\hat{\varsigma}_{a_\perp} - \varsigma_{a_\perp})'\bar{a}. \quad \textbf{(A.1)}$$

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

A first-order expansion of $\hat{\varsigma}_{a_\perp}$ around $\varsigma_{a_\perp}$ gives $(\hat{\varsigma}_{a_\perp} - \varsigma_{a_\perp}) = (I - \varsigma_{a_\perp} a'_\perp)(\hat{\varsigma} - \varsigma) \times (a'\varsigma)^{-1} + O_p(\|\hat{\varsigma} - \varsigma\|^2)$, which substituted within (A.1) gives, by (D.2) in Paruolo (1997a),

$$T^{1/2} \bar{a}'_\perp (\hat{\alpha}_{2a} - \alpha_{2a}) = -(\varsigma' a_\perp)^{-1} T^{1/2} (\hat{\varsigma} - \varsigma)' \alpha_{2a} + o_p(1).$$

From, e.g., Theorem 4.3 in Paruolo (2000), we know that $\hat{\varsigma}$ is asymptotically Gaussian,

$$T^{1/2} \mathrm{vec}(\hat{\varsigma} - \varsigma) \xrightarrow{w} N(0, \Sigma^{-1}_{JJ.(N,V)} \otimes \Omega).$$

Hence one finds

$$T^{1/2} \mathrm{vec}(\bar{a}'_\perp (\hat{\alpha}_{2a} - \alpha_{2a})) \xrightarrow{w} N(0, \alpha'_{2a} \Omega \alpha_{2a} \otimes (a'_\perp \varsigma \Sigma_{JJ.(N,V)} \varsigma' a_\perp)^{-1}),$$

where the asymptotic variance can be consistently estimated by

$$\hat{\Lambda}(\bar{a}'_\perp \hat{\alpha}'_{2a}) = \hat{\alpha}'_{2a} \hat{\Omega} \hat{\alpha}_{2a} \otimes (a'_\perp \hat{\varsigma} S_{JJ.(N,V)} \hat{\varsigma}' a_\perp)^{-1},$$

given the consistency of $\hat{\alpha}'_2$, $\hat{\Omega}$, and $\hat{\varsigma}$. We next turn to C_2; as a result of the superconsistency of the estimators of γ_1, the parameters γ_1 can be considered fixed in C_2. The results of the theorem follow by Cramér's theorem, again applying the results (B.2) and (B.3) in Paruolo (1997a). In fact a first-order expansion gives

$$\begin{aligned} T^{1/2}(\hat{C}_2 - C_2) = {} & C_2 T^{1/2} (\hat{\varsigma} - \varsigma) \bar{\varsigma}' (\theta C_2 - I) - C_2 T^{1/2} (\hat{\zeta}_1 - \zeta_1) \delta \beta'_2 C_2 \\ & + C_2 T^{1/2} (\hat{\Upsilon} - \Upsilon) C_2 + o_p(1) \end{aligned}$$

(see Paruolo, 1997a, Theorems 5.5 and 7.1), from which the results on C_2 follow. We next want to show that Q is of full column rank under (11). Consider the matrix $B := (\varsigma : \theta \bar{\beta}_2)$. Here B is of full rank because, premultiplying by the full rank square matrix $A' := (\bar{\varsigma} : \alpha_2)'$, one finds

$$A'B = \begin{pmatrix} I_{r+s} & \bar{\varsigma}' \theta \bar{\beta}_2 \\ 0 & \alpha'_2 \theta \bar{\beta}_2 \end{pmatrix},$$

which is nonsingular under the $I(2)$ conditions. Thus $\mathrm{rank}(Q') = \mathrm{rank}(B'Q')$, where

$$B'Q' = \begin{pmatrix} I_{r+s} & 0 & 0 \\ 0 & -\bar{\beta}'_2 \theta' C'_2 \beta_2 \delta' & \bar{\beta}'_2 \theta' C'_2 \end{pmatrix} = \begin{pmatrix} I_{r+s} & 0 & 0 \\ 0 & -\delta' & \bar{\beta}'_2 \end{pmatrix} \tag{A.2}$$

because $C_2 \theta \bar{\beta}_2 = \bar{\beta}_2$ from the definition of C_2. When $k > 2$ the matrix $B'Q'$ is of full rank because $\bar{\beta}'_2$ is of full row rank. When $k = 2$ the last block of columns $(0 : \bar{\beta}_2)'$ in (A.2) is missing because $\Upsilon = 0$; if $k = 2$ and if δ is of full column rank then $B'Q' = \mathrm{diag}(I_{r+s}, -\delta')$ is of full rank.

We next consider the estimator of $\psi := -C_2 \zeta_1$ and observe that $T^{1/2}(\hat{\psi} - \psi) = -T^{1/2}(\hat{C}_2 - C_2)\zeta_1 - C_2 T^{1/2}(\hat{\zeta}_1 - \zeta_1) + o_p(1)$; vectorizing one finds the stated result. Let $\Phi := I - \Upsilon$. We next want to show that L is of full rank if $\Phi_{\alpha_2 \beta_2}$ is of full rank by showing that the second block in L, equal to $I_r - \delta \beta'_2 C_2 \zeta_1 = I_r - \Gamma_{\alpha \beta_2} \theta^{-1}_{\alpha_2 \beta_2} \Gamma_{\alpha_2 \beta}$, is of full rank. Consider the matrix

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press

$$A := \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} := \begin{pmatrix} I_r & -\Gamma_{\alpha\beta_2} \\ \Gamma_{\alpha_2\beta} & \Phi_{\alpha_2\beta_2} \end{pmatrix}.$$

Note that $A_{11} = I_r$ and $A_{22.1} := A_{22} - A_{21}A_{11}^{-1}A_{12} = \Gamma_{\alpha_2\beta}\Gamma_{\alpha\beta_2} + \Phi_{\alpha_2\beta_2} = \theta_{\alpha_2\beta_2}$ are nonsingular under the $I(2)$ conditions and hence the matrix A is of full rank. By the appropriate partitioned-inverse-matrix formula the first block on the main diagonal of A^{-1} is $A_{11}^{-1} + A_{11}^{-1}A_{12}A_{22.1}^{-1}A_{21}A_{11}^{-1} = I_r - \Gamma_{\alpha\beta_2}\theta_{\alpha_2\beta_2}^{-1}\Gamma_{\alpha_2\beta}$. Because $\Phi_{\alpha_2\beta_2}$ is assumed of full rank, this block is equal to $A_{11.2}^{-1} = (I_r + \Gamma_{\alpha\beta_2}\Phi_{\alpha_2\beta_2}^{-1}\Gamma_{\alpha_2\beta})^{-1}$ (see, e.g., Srivastava and Khatri, 1979, Corollary 1.4.2) and is thus invertible. This completes the proof. ■

Proof of Proposition 3. The artificial regressions are of the form

$$B\hat{\epsilon}_t = A_1 x_{1t} + A_2 x_{2t} + \text{error}_t,$$

where x_{1t} and x_{2t} are elements of the regressors (F_t, W_t) and $B = \hat{\alpha}_{2a}, \hat{C}_2$. We observe that, because $\hat{\epsilon}_t$ is the residual of a regression of $\Delta^2 X_t$ on F_t, W_t, the estimated coefficients A_1 and A_2 are all equal to 0 by construction, such that the artificial regression residuals $B\hat{\epsilon}_t - \hat{A}_1 x_{1t} - \hat{A}_2 x_{2t}$ are equal to $B\hat{\epsilon}_t$.

The results for $\bar{a}_\perp' \alpha_{2a}$ then follow from the standard formula of the estimated covariance matrix. For the case of the artificial regression (14) observe that $Q'D^{-1}Q = D_{QQ.Q_\perp}^{-1} := (D_{QQ} - D_{QQ_\perp}D_{Q_\perp Q_\perp}^{-1}D_{Q_\perp Q})^{-1}$, where $D := S_{FF}$, before applying the preceding argument. ■

Proof of Corollary 4. The results follow from Theorem 2. Note that for the case of the t-test statistics on C_2 under the ${}^{ij}H_{03}$ hypothesis the relevant asymptotic variance is positive. ■

https://doi.org/10.1017/S0266466602183058 Published online by Cambridge University Press