UvA-DARE (Digital Academic Repository)

Efficient inference on cointegration parameters in structural error correction models

Boswijk, H.P.

DOI

[10.1016/0304-4076\(94\)01665-M](https://doi.org/10.1016/0304-4076(94)01665-M)

Publication date

1995

Published in

Journal of Econometrics

[Link to publication](#)

Citation for published version (APA):

Boswijk, H. P. (1995). Efficient inference on cointegration parameters in structural error correction models. *Journal of Econometrics*, 69(1), 133-158. [https://doi.org/10.1016/0304-4076\(94\)01665-M](https://doi.org/10.1016/0304-4076(94)01665-M)

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Materiaal : Abx PPN : 840121202,541891855
Titel : Journal of econometrics
Auteur : Aigner, D.J. (ed.)
Deel / Supplem. :
Corporatie : Externe Database :
Jaar / Editie : 1973 Extern Nummer :
Uitgave : Amsterdam North-Holland
Serie / Sectie :
Pag-ISSN / ISBN : 0304-4076

840121202 ; P 11.961 ; ; 1973 V1 - ~eH2323950~eV~c

Jaar	: 1995-00-00	Datum Indienen	: 15-04-2005 18:32
Volume	: 69	Datum Plaatsing	: 15-04-2005 18:32
Aflevering	: 1	Datum Rappel	: 13-05-2005
Leenvorm	: KOPIE	Particulier	: N
Leveringswijze	: E	Geplaatst bij	: 0008
Coöperatiecode	: R	Indiener	: 0004/9999
Aanvrager	: 0004/9998	Eindgebruiker	: 041631433
Aanvragerident.	: UVA KEUR (UB GRONINGEN)	Aanvraagident.	:
Auteur	: Boswijk, H.Peter (ed.)		
Artikel	: Efficient inference on cointegration parameters in structural error cor		
Bladzijden	: 133-158		
Bron	:		
Opmerking	: ARNO:7576		

Indiener	: 0004/9999	Stuur rekening	: N
Aanvrager	: 0004/9998	Eindgebruiker	: 041631433
Aanvragerident.	: UVA KEUR (UB GRONINGEN)	Aanvraagident.	:

Afleveradres Post UvA Keur
UB Groningen

Broerstraat 4
9700 AN Groningen

NL

Fax
E-mail m.s.van.delden@rug.nl
Ftp
Ariel
Telefoon 050-3635057

Faktuuraadres Clearing House

[1] origineel gestuurd
[2] kopie gestuurd
[3] overige

[4] nog niet aanwezig
[5] niet aanwezig
[6] niet beschikbaar

[7] uitgeleend
[8] wordt niet uitgeleend
[9] bibliografisch onjuist
[0] bij de binder

Aantal eenheden : 26
Aanvraagnummer : A078352355

ELSEVIER

Journal of Econometrics 69 (1995) 133–158

JOURNAL OF
Econometrics

Efficient inference on cointegration parameters in structural error correction models

H. Peter Boswijk

Department of Actuarial Science and Econometrics, University of Amsterdam, 1018 WB Amsterdam, The Netherlands

Abstract

This paper proposes inferential procedures for error correction models in structural form. Particular attention is paid to the issues of exogeneity of conditioning variables and identification of cointegration parameters as well as short-run parameters. The model leads to two classes of estimators and associated test statistics, depending on the exogeneity status of the conditioning variables. A Monte Carlo experiment shows how their asymptotic properties are reflected in finite sample behaviour.

Key words: Cointegration; Error correction models; Exogeneity; Identification; Structural models

JEL classification: C32

1. Introduction

Since the introduction of the notion of cointegration by Granger (1981) and Engle and Granger (1987), the literature on this topic may be categorized into three separate groups. The first approach focuses on *cointegrating regressions* or *regressions with integrated processes*. Starting from the super-consistency result proved by Krämer (1986), Phillips and Durlauf (1986), and Stock (1987), it is argued that the short-run dynamics may be neglected for consistent estimation or treated nonparametrically to obtain efficient inference. This approach was

This paper was presented at the International Conference on Bayesian and Classical Econometric Modeling of Time Series, Marseille, 19–20 June 1992. Thanks are due to Neil Ericsson, Jan Kiviet, Teun Kloek, Søren Johansen, Katarina Juselius, Michel Lubrano, and two anonymous referees for helpful comments. Financial support from the Tinbergen Institute and the Royal Netherlands Academy of Arts and Sciences is gratefully acknowledged.

developed by Phillips and his co-authors; see Park and Phillips (1988, 1989), Park (1992), Phillips and Hansen (1990), Phillips (1991), Phillips and Loretan (1991), and Saikkonen (1991). Generalizations of this approach to simultaneous cointegrating regressions are considered by Park (1990) and Saikkonen (1993).

The second approach, developed by Johansen, starts from a *vector autoregressive error correction model* (VAR-ECM) with Gaussian innovations to obtain maximum likelihood inference (see Johansen, 1988, 1991, 1995; Johansen and Juselius, 1990, 1992, 1994). One of the most apparent differences with the previous approach is the joint parametric treatment of the short-run dynamics and the long-run relations. Another difference is that the determination of the number of long-run relations is an integral part of the analysis.

A third line of research is based on error correction models that are conditional upon a set of variables (hence *conditional error correction models*; see Boswijk, 1990; Kremers et al., 1992; Johansen, 1992). In general such conditional models seem to relate naturally to economic theory, which often stipulates a (large) number of exogenous variables in addition to the endogenous variables. The link with economic theory can be analyzed even more conveniently in error correction models in structural form, i.e., transformations of the conditional (reduced form) model such that the model is formulated in terms of parameters that allow an economic interpretation. These shall be referred to as *structural error correction models* (SECMs); cf. Boswijk (1992), Davidson and Hall (1991), and Urbain (1992). The formulation of an SECM requires that the *identification* of cointegration parameters as well as the short-run parameters be analyzed. Moreover, because some variables are conditioned upon, the notion of *weak exogeneity* plays an important role in this analysis.

In this paper we present efficient estimation procedures of cointegration parameters in structural error correction models. In contrast with previous studies, we shall not restrict ourselves to single-equation models. In addition, both models with and without weak exogeneity will be considered. For the latter case, fully iterated as well as two-step estimators are suggested. Since the procedures are based on the likelihood analysis of a reparametrized VAR-ECM, they can be seen as a reformulation of Johansen's (1992) analysis for the weak exogeneity case and of Johansen's (1991) analysis for the general case. In particular, the asymptotic properties of the estimators and test statistics logically follow from Johansen's results. Therefore, the focus of this paper is on model formulation and parametrization. In specific, the formulation chosen here will be shown to facilitate imposing (over-) identifying restrictions, both on the cointegration parameters and on the short-run structure.

The plan of the paper is as follows. In Section 2, we derive the structural error correction model from the VAR-ECM via conditioning and reparametrization. In the third section, we propose the estimators and test statistics and state their asymptotic properties. Section 4 contains a small Monte Carlo investigation of

the finite sample behaviour of the statistics, and in Section 5 the results are discussed.

2. The model

2.1. The VAR-ECM

Consider the vector autoregressive (VAR) model of order p of an n -vector time series $\{x_t, t = 1, \dots, T\}$ with starting values $\{x_{1-p}, \dots, x_0\}$:

$$x_t = \sum_{j=1}^p \Phi_j x_{t-j} + \varepsilon_t, \quad t = 1, \dots, T, \quad (1)$$

where $\{\varepsilon_t, t = 1, \dots, T\}$ is an independent $N(0, \Sigma)$ sequence, and where Φ_j , $j = 1, \dots, p$, and Σ are $n \times n$ parameter matrices, with Σ positive definite. For notational convenience, no deterministic variables such as a constant or linear trend term are included. Defining

$$A = -I_n + \sum_{i=1}^p \Phi_i, \quad A_j = -\sum_{i=j+1}^p \Phi_i, \quad j = 1, \dots, p-1, \quad (2)$$

the error correction form of (1) (VAR-ECM) is obtained:

$$\Delta x_t = A x_{t-1} + \sum_{j=1}^{p-1} A_j \Delta x_{t-j} + \varepsilon_t, \quad t = 1, \dots, T. \quad (3)$$

Assumption 1 (Cointegration). In (3), $\text{rank}(A) = r$, $0 < r < n$, and the characteristic equation

$$\phi(\zeta) = \left| I_n - \sum_{j=1}^p \Phi_j \zeta^j \right| = \left| (1 - \zeta) I_n - A \zeta - (1 - \zeta) \sum_{j=1}^{p-1} A_j \zeta^j \right| = 0 \quad (4)$$

has $(n - r)$ roots ζ equal to one and all other roots outside the unit circle.

The rank reduction of A implies that this matrix can be expressed as $A = \alpha\beta'$, with α and β $n \times r$ matrices of full column rank, so that (3) becomes

$$\Delta x_t = \alpha\beta' x_{t-1} + \sum_{j=1}^{p-1} A_j \Delta x_{t-j} + \varepsilon_t, \quad t = 1, \dots, T. \quad (5)$$

As follows from Johansen's (1991) version of Granger's representation theorem (see also Boswijk, 1992, Theorem 2.2), the assumption on the characteristic roots implies that x_t is integrated of order 1, whereas the r linear combinations $\beta' x_t$ are stationary. That is, x_t is cointegrated of order (1,1), see Engle and Granger

(1987), with cointegrating rank r ; the columns of the matrix β are cointegrating vectors, and α contains the adjustment (error correction) parameters. Johansen (1988, 1991) provides a likelihood analysis of the VAR-ECM, consisting of a sequential testing procedure to determine the cointegrating rank r , the maximum likelihood estimators of α and β , and likelihood ratio tests for restrictions on α and β .

In the VAR-ECM all variables in x_t are allowed to be determined simultaneously. However, economic models often entail a distinction between endogenous and exogenous variables. The latter set of variables are considered as fixed, i.e., they are conditioned upon. This will be considered next.

2.2. Conditioning and weak exogeneity

Let x_t be partitioned as $x_t = (y_t', z_t')'$, where the g -vector y_t contains the endogenous variables and the k -vector z_t consists of conditioning variables ($n = g + k$). Let $\{\mathcal{F}_t, t = 0, 1, \dots\}$ be an increasing sequence of σ -fields such that $\mathcal{F}_t$ equals the σ -field generated by $\{x_t, x_{t-1}, \dots, x_{1-p}\}$. Similarly, let $\mathcal{F}_{t-1}^z = \sigma(z_t, x_{t-1}, \dots, x_{1-p})$. Because $\{\varepsilon_t\}$ is assumed to be an independent $N(0, \Sigma)$ sequence, we have

$$\Delta x_t | \mathcal{F}_{t-1} \sim N(\mu_t, \Sigma), \quad t = 1, \dots, T, \quad (6)$$

where

$$\mu_t = \alpha \beta' x_{t-1} + \sum_{j=1}^{p-1} A_j \Delta x_{t-j}, \quad t = 1, \dots, T. \quad (7)$$

Letting $\mu_t, \alpha, A_j, j = 1, \dots, p-1$, and Σ be partitioned conformably with x_t , (6) implies

$$\begin{aligned} \Delta y_t | \mathcal{F}_{t-1}^z &\sim N(\mu_{1.2,t}, \Sigma_{11.2}), \\ \Delta z_t | \mathcal{F}_{t-1} &\sim N(\mu_{2t}, \Sigma_{22}), \end{aligned} \quad t = 1, \dots, T. \quad (8)$$

where

$$\begin{aligned} \mu_{1.2,t} &= \mu_{1t} + \Sigma_{12} \Sigma_{22}^{-1} (\Delta z_t - \mu_{2t}), \\ \Sigma_{11.2} &= \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}. \end{aligned} \quad (9)$$

After some rearranging of terms, this yields the conditional error correction model,

$$\Delta y_t = \Pi_0 \Delta z_t + \alpha_{1.2} \beta' x_{t-1} + \sum_{j=1}^{p-1} A_{1.2,j} \Delta x_{t-j} + \varepsilon_{1.2,t}, \quad t = 1, \dots, T, \quad (10)$$

where

$$\begin{aligned} \Pi_0 &= \Sigma_{12} \Sigma_{22}^{-1}, & \alpha_{1,2} &= \alpha_1 - \Pi_0 \alpha_2, \\ A_{1,2,j} &= A_{1j} - \Pi_0 A_{2j}, & j &= 1, \dots, p-1, \end{aligned} \quad (11)$$

and $\varepsilon_{1,2t} = \varepsilon_{1t} - \Pi_0 \varepsilon_{2t}$, a Gaussian innovation relative to $\mathcal{F}_t^z$ with covariance matrix $\Sigma_{11,2}$. The marginal error correction model consists of the last k equations of (5):

$$\Delta z_t = \alpha_2 \beta' x_{t-1} + \sum_{j=1}^{p-1} A_{2j} \Delta x_{t-j} + \varepsilon_{2t}, \quad t = 1, \dots, T. \quad (12)$$

Because the conditional and marginal models constitute a reparametrization of the VAR-ECM, the maximum likelihood (ML) estimator of β may be obtained either from (5) or from joint estimation of (10) and (12). However, in some circumstances the ML estimator of β may be obtained from the conditional model only. Engle et al. (1983) define a vector of conditioning variables z_t to be weakly exogenous for the parameters of interest, if (i) the parameters of the conditional and marginal models are variation free, and (ii) the parameters of interest are a function of the conditional model parameters only. Comparing (10) and (12), we observe that in general, β is contained in the marginal model parameters as well as the conditional model parameters, leading to a violation of the variation-free requirement. Johansen (1992, Theorem 1) shows that $\alpha_2 = 0$ is a necessary and sufficient condition for z_t to be weakly exogenous for α and β : if there is no error correction mechanism in the marginal model, then all relevant information on α and β is contained in the conditional model. Thus in that case efficient inference on α and β can be conducted by analyzing (10) alone.¹

Johansen (1992) analyzes inference on cointegration based on reduced rank regressions in the conditional (or *partial*) model (10). The analysis in this paper differs from Johansen's because we shall consider (over-)identified cointegrating vectors. Moreover, our methods will be based on an error correction model in structural form, i.e., on a reparametrization of (10). In the next subsection we consider identification of the long-run (cointegration) parameters, as well as of the short-run structure.

2.3. Identification and the structural form

Consider the error correction term in the VAR-ECM, $\alpha \beta' x_{t-1}$. For all possible transformations $(\alpha^*, \beta^*) = (\alpha C', \beta C^{-1})$ with C nonsingular, it is the case

¹ A referee has pointed out that if a subset of the cointegrating vectors is considered of interest, then only the corresponding part of α_1 is required to vanish for weak exogeneity. For example, the conditional and marginal model may contain completely separate sets of cointegrating vectors, in which case z_t is weakly exogenous for the cointegrating vectors entering the conditional model.

that $\alpha^*\beta^{*'} = \alpha C'(C')^{-1}\beta' = \alpha\beta'$. Therefore, without any further restrictions, α and β are not identified. A number of different approaches to resolve this identification problem have been proposed in the literature.

Let β^* denote a matrix of cointegrating vectors under some normalization. In the cointegrating regression literature (see, e.g., Phillips and Durlauf, 1986; Park and Phillips, 1988, 1989; Phillips, 1991), β^* is normalized with respect to the first r components of x_t , i.e., $\beta^{*'} = [I_r; -\Pi]$. Defining $u_t \equiv \beta^{*'}x_t$ and partitioning $x'_t = (x'_{1t}, x'_{2t})$ conformably with β^* , we have

$$x_{1t} = \Pi x_{2t} + u_t, \quad t = 1, \dots, T. \quad (13)$$

Johansen (1991) considers a related class of identifying restrictions of the form $c'\beta^* = I_r$, where c is a known $n \times r$ matrix. Choosing $c' = [I_r; 0]$ leads to the same parametrization $\beta^{*'} = [I_r; -\Pi]$, but other choices may be considered as well. The normalized matrix β^* is obtained from an arbitrary $n \times r$ matrix β of cointegrating vectors as $\beta^* = \beta(c'\beta)^{-1}$. Therefore, this normalization requires that $\text{rank}(c'\beta) = r$, which can be seen as a rank condition for identification. Boswijk (1993b) proposes an identifiability test for a such a rank condition. Under this condition (and with $c' = [I_r; 0]$), there are no linear combinations of the columns of β with zeros in the first r rows, i.e., there are no cointegrating relationships involving x_{2t} only. The system (13) can be seen as a reduced form system, which may be reparametrized into a structural form. Park (1990) and Saikkonen (1993) analyze identification of and asymptotic inference on such simultaneous cointegrating regression equations.

In this paper we consider a more general class of identifying restrictions. Following Johansen (1995) and Johansen and Juselius (1992, 1994), we can formulate these restrictions as

$$\beta_i = H_i \varphi_i, \quad i = 1, \dots, r, \quad (14)$$

where H_i are known $n \times s_i$ matrices of full column rank, and φ_i are $s_i \times 1$ parameter vectors which are subsequently normalized with respect to one of their components. Letting R_i denote $l_i \times n$ matrices ($l_i = n - s_i$) of full row rank such that $R_i H_i = 0$, (14) is equivalent to

$$R_i \beta_i = 0, \quad i = 1, \dots, r. \quad (15)$$

The rank condition for identification is:

Assumption 2. The matrices R_i , $i = 1, \dots, r$, and the matrix β of cointegrating vectors satisfy

$$\text{rank}(R_i \beta) = r - 1, \quad i = 1, \dots, r. \quad (16)$$

As discussed by Johansen (1995), (16) involves unknown parameters and thus cannot be established without knowledge of β . However, Johansen provides an operational criterion for a set of restrictions to be *generically identifying*, i.e., identifying for all β except for a set with measure zero. Boswijk (1993b) proposes a test for the hypothesis that β in fact lies in this measure-zero set, and hence that β_t is not identified.

Next, we consider a structural reformulation of the conditional error correction model. In a recent paper, Spanos (1990) provides a reinterpretation of the classical simultaneous equations model that unifies information on the statistical properties of the data with theory information. In Spanos' analysis, the model in structural form is viewed as derived from the joint density of the observables, represented here by the VAR-ECM, by conditioning [to obtain the reduced form system such as (10)] and reparametrization.

Therefore, let the system (10) be premultiplied by a particular nonsingular $g \times g$ matrix Γ_0 :

$$\Gamma_0 \Delta y_t = \lambda \beta' x_{t-1} + B_0 \Delta z_t + \sum_{j=1}^{p-1} (\Gamma_j \Delta y_{t-j} + B_j \Delta z_{t-j}) + v_t, \quad (17)$$

where

$$\lambda = \Gamma_0 \alpha_{1,2}, \quad B_0 = \Gamma_0 \Pi_0, \quad [\Gamma_j; B_j] = \Gamma_0 A_{1,2,j}, \quad j = 1, \dots, p-1, \quad (18)$$

and where $\{v_t\} = \{\Gamma_0 \varepsilon_{1,2,t}\}$ is a Gaussian innovation process relative to $\mathcal{F}_t^z$ with covariance matrix $\Omega = \Gamma_0 \Sigma_{11,2} \Gamma_0'$. Here λ is a $g \times r$ matrix of error correction coefficients, and Γ_j and B_j , $j = 0, \dots, p-1$, are parameter matrices of order $g \times g$ and $g \times k$, respectively. Notice that because the disturbances v_t are orthogonal to the conditioning variables z_t , weak exogeneity of z_t for the parameters of (17) holds if z_t is weakly exogenous for the parameters of the conditional model (10), i.e., if $\alpha_2 = 0$. Urbain (1992) considers the possibility that the orthogonality condition is violated, so that weak exogeneity may fail either because $\alpha_2 \neq 0$ or because (17) does not represent a reparametrized conditional model.

Although β is identified as indicated above, identification of the remaining parameters in (17) requires further restrictions. The system can be normalized by

$$\Gamma_{0,ii} = 1, \quad i = 1, \dots, g. \quad (19)$$

Additional restrictions are to be obtained from economic theory. However, quite often economic theory will only be informative about the long-run relationships. A possible solution to this problem is to start with a just-identified model that can be given an economic interpretation, and subsequently restrict the model according to some statistical model selection procedure. Two approaches are worth mentioning explicitly.

The first class of identified structures is obtained by setting $\Gamma_0 = I_g$. This of course renders the structural form identical to the conditional ECM (10), so that the conditional ECM parameters are considered to be of interest themselves. In this case overidentifying restrictions are simply restrictions on the conditional model. A second class is obtained by restricting Γ_0 to be a triangular matrix and Ω to be a diagonal matrix. These restrictions yield a so-called recursive system. The main characteristic of this formulation is that it imposes an ordering in the contemporaneous effects (or *causal chain*). Because these restrictions lead to a factorization of the joint likelihood function of the model into the product of the likelihood functions of the separate equations, these equations can be estimated separately (conditional upon knowledge, or an estimate, of β), which solves the identification problem and simplifies the model specification process considerably. An example of this approach to cointegration modelling is given in Johansen and Juselius (1994).

In this paper, an alternative approach to the identification of (17) is adopted, which requires the following assumption:

Assumption 3. In (17) it is the case that (i) $g = r$, and (ii) $\text{rank}(\lambda \beta_y') = r$, where $\beta' = [\beta_y' : \beta_z']$ with β_y an $r \times r$ matrix.

Thus we require that the cointegrating rank is known, that it equals the number of endogenous variables in the conditional model, and that the conditional (and hence structural) ECM is stable, in the sense that its characteristic equation does not have any unit roots. An estimate of r can be obtained from Johansen's cointegration tests in the VAR-ECM; conditional upon this estimate, one can choose r endogenous variables, associated with the r long-run relationships, and test for stability of the system using Boswijk's (1994) test. Note that (ii) limits the choice of the matrices H_i in (14).

Under these conditions, we may restrict λ to be a diagonal matrix. This implies that restrictions on the i th long-run relation (15) are also imposed on the i th structural error correction equation. Letting δ_i denote the i th vector of $\beta \lambda'$, and hence the coefficient vector of x_{t-1} in the i th equation, we have

$$R_i \delta_i = R_i \beta_i \lambda_i = 0. \quad (20)$$

Because of (20), the rank condition (16) is sufficient for identification of the i th equation of the system (17), provided that Assumption 3 holds, so that $\lambda_i \neq 0$, $i = 1, \dots, g$. This follows from the fact that under these conditions, there is no linear combination of the remaining $(g - 1)$ equations that satisfies (20).

The main motivation for considering a diagonal error correction matrix is that it provides a simple interpretation of individual equations from the structural system. Here a structural equation is identified by its long-run properties. For example, the distinguishing feature of a money demand error correction equation is that its long-run solution can be interpreted as a money

demand cointegrating relation. Moreover, this formulation simplifies testing the stability of the system (see Boswijk, 1994) and inference on the long-run parameters. Notice that the possibility that each endogenous variable is affected by all disequilibria is not excluded: the reduced form system (10) will have no restrictions imposed on its error correction matrix $\alpha_{1,2} = \Gamma_0^{-1} \lambda$. The chosen structure is related to Bergstrom's (1976) approach, who considers continuous time models where each of the differential equations involves only one long-run relationship.

It can be argued that a large class of structural models is excluded by Assumption 3 and the diagonality restriction. If Assumption 3 does not hold, e.g., because the number of endogenous variables of interest exceeds the cointegrating rank, then the methods proposed in this paper cannot be applied. In that case Johansen's (1992) methods can be used to obtain an estimate of the cointegration space, or, if identifying restrictions are considered, the model may be estimated by ML under the relevant nonlinear cross-equation restrictions. If, on the other hand, Assumption 3 is satisfied but λ is not diagonal, then the model may always be transformed into a system with diagonal error correction matrix, so that the methods derived below can still be applied. However, in those cases this system does not represent the genuine structural form, and the term 'pseudo-reduced form' seems more appropriate.

The SECM (17) has been presented here as a statistical reduction from the VAR (1), so that all properties of the former stem from the latter, i.e., the VAR model encompasses the structural model. On the other hand, one of the motivations for considering structural models is that they purport to describe behavioural relations, which remain constant even if the process generating the exogenous variables changes. In this light, the VAR model may be seen as reduced from the structural model, augmented with the marginal system (12). If the structural model does indeed remain constant when the marginal system changes, then it can explain the nonconstancy in the reduced form and thus encompasses the VAR; cf. Hendry and Mizon (1993).

3. Estimation and hypothesis testing

In this section a number of estimation and testing procedures for the long-run parameters is proposed. In the first subsection the estimators are based on the structural model only. It is well-known (cf. Johansen, 1992) that if the conditioning variables are not weakly exogenous for the long-run parameters, the asymptotic distribution is not mixed normal, making the construction of tests cumbersome. Therefore, in the second subsection a Lagrange multiplier test for weak exogeneity is derived and an iterative procedure is proposed that takes into account both the structural and the marginal model.

3.1. Conditional model estimation

Consider first the case with only one long-run relationship ($g = 1$), so that in (17), $\Gamma_0 = 1$. Note that the conditional ECM (10) and the structural ECM (17) now coincide. Express the model more concisely as follows:

$$\Delta y_t = \delta' x_{t-1} + \pi' w_t + v_t, \quad t = 1, \dots, T, \quad (21)$$

where $\delta = \lambda\beta$, where w_t contains all regressors other than the error correction term, and where π is the corresponding parameter vector. Restrictions may be imposed on the short-run dynamics by redefining w_t and π appropriately; e.g., the maximal lag length may differ across variables. Assume that a sample $\{x_t, t = 1 - p, \dots, T\}$ of observations is available and define the vectors and matrices $\Delta y = (\Delta y_1, \dots, \Delta y_T)'$, $X_{-1} = (x_0, \dots, x_{T-1})'$, and $W = (w_1, \dots, w_T)'$. If we assume that the starting values are fixed, that the disturbances are Gaussian innovations, and that the conditioning variables are weakly exogenous, then the ML estimator of (δ', π') in (21) equals the ordinary least squares (OLS) estimator

$$\begin{bmatrix} \hat{\delta} \\ \hat{\pi} \end{bmatrix} = \begin{bmatrix} X_{-1}' X_{-1} & X_{-1}' W \\ W' X_{-1} & W' W \end{bmatrix}^{-1} \begin{bmatrix} X_{-1}' \Delta y \\ W' \Delta y \end{bmatrix}. \quad (22)$$

Its covariance matrix is estimated by

$$\hat{V} \begin{bmatrix} \hat{\delta} \\ \hat{\pi} \end{bmatrix} = \hat{\omega}^2 \begin{bmatrix} X_{-1}' X_{-1} & X_{-1}' W \\ W' X_{-1} & W' W \end{bmatrix}^{-1}, \quad (23)$$

where $\hat{\omega}^2$ is the least squares variance estimator.

Because there is only one cointegrating vector, a normalization of β is sufficient to identify λ and β . Under Assumption 3, the first component of β differs from zero, so that β may be normalized as $(1, -\theta')$, where θ represents the long-run effect of z on y . The normalization implies

$$\delta = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} \lambda \\ -\lambda\theta \end{bmatrix}, \quad (24)$$

so that $-\theta$ is defined by δ_2/δ_1 , and the ML estimators of λ and θ are given by $\hat{\lambda} = \hat{\delta}_1$ and

$$\hat{\theta} = -\hat{\delta}_2/\hat{\delta}_1. \quad (25)$$

The covariance matrix of $\hat{\theta}$ is estimated by (see, e.g., Bårdsen, 1989)

$$\hat{V}[\hat{\theta}] = [\hat{\theta} : I_k] \hat{V}[\hat{\delta}] [\hat{\theta} : I_k]' / \hat{\lambda}^2. \quad (26)$$

Consider a set of linear restrictions on θ :

$$R_2 \theta = r_1, \quad (27)$$

where R_2 is a known $h \times k$ matrix of full row rank and r_1 is a known h -vector. This hypothesis corresponds to

$$R\delta = r_1\delta_1 + R_2\delta_2 = \lambda[r_1 - R_2\theta] = 0, \quad (28)$$

where $R = [r_1, R_2]$ is an $h \times n$ matrix. Because δ is proportional to β , this may be expressed analogously to (14) as

$$\delta = H\varphi, \quad (29)$$

where H is an $n \times s$ matrix ($s = n - h$) such that $RH = 0$, and φ is an s -vector of unrestricted parameters. The restricted ML estimator of (δ', π') is obtained from the least-squares regression

$$\Delta y_t = \varphi' H' x_{t-1} + \pi' w_t + v_t, \quad t = 1, \dots, T, \quad (30)$$

with $\hat{\delta} = H\hat{\varphi}$. The likelihood ratio test statistic for the null hypothesis (27) is given by

$$LR = T \ln(RSS_0/RSS_1), \quad (31)$$

where RSS_1 is the residual sum of squares from the unrestricted regression (21) and RSS_0 is the restricted residual sum of squares from (30). As is well-known, there is a one-to-one correspondence between the LR statistic and the Wald statistic

$$Wald = (R\hat{\delta})' [R\hat{V}[\hat{\delta}]R']^{-1} (R\hat{\delta}) = T \frac{RSS_0 - RSS_1}{RSS_1}. \quad (32)$$

An alternative version of the Wald test statistic may be constructed from $(R_2\hat{\theta} - r_1)$, using the estimated covariance matrix of $\hat{\theta}$. However, as shown in Boswijk (1993a), the statistics (31) or (32) should be favoured over this alternative Wald statistic, because the likelihood ratio test is invariant to reformulations of the null hypothesis, whereas the Wald test is not. Moreover, that study shows that the finite sample behaviour of a Wald test based on (26) is adversely influenced by the lack of finite moments of $\hat{\theta}$. This is caused by the fact that $\hat{\theta}$ is defined as a ratio, cf. (25), where the denominator has positive density at the origin.

Consider now the case where $g > 1$. Let ΔY_i denote the $T \times (g - 1)$ matrix containing observations on all endogenous variables other than Δy_{it} . Next, let $W_i = [\Delta Y_i; W_{i2}]$ contain all regressors in the i th equation other than the error correction term, with coefficient vector π_i . Notice that W_{i2} is allowed to vary with i , so that overidentifying restrictions may be imposed on the short-run

dynamics. Let δ_i denote the i th column of $\beta\lambda'$, the coefficient vector of x_{t-1} in the i th equation. The identifying restrictions are incorporated via $\delta_i = H_i\varphi_i$, cf. (14). Thus the i th equation, in matrix notation, is

$$\Delta y_t = X_{-1}H_i\varphi_i + W_i\pi_i + v_i. \quad (33)$$

These equations are stacked into the system

$$\Delta y = (I_g \otimes X_{-1})H\varphi + W\pi + v, \quad (34)$$

where

$$\Delta y = \begin{bmatrix} \Delta y_1 \\ \vdots \\ \Delta y_g \end{bmatrix}, \quad H = \begin{bmatrix} H_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & H_g \end{bmatrix}, \quad W = \begin{bmatrix} W_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_g \end{bmatrix}, \quad (35)$$

and where $\varphi = (\varphi'_1, \dots, \varphi'_g)'$, $\pi = (\pi'_1, \dots, \pi'_g)'$, and $v = (v'_1, \dots, v'_g)'$.

This system is a standard linear simultaneous equations model with no cross-equation restrictions, so that a number of familiar estimators are available. A survey of these estimators is given by Hendry (1976) and Hausman (1983). Hendry derives a general expression for a class of estimators of the parameters of a simultaneous equations model, the so-called *estimator generating equation*. For any estimate $(\hat{\varphi}', \hat{\pi}')$, let $\hat{\Omega}$ denote the estimated covariance matrix of v_i , let $\Delta \hat{Y}_i$ denote the matrix of fitted values of ΔY_i , and define $\hat{W}_i = [\Delta \hat{Y}_i; W_{i2}]$ and $\hat{W} = \text{diag}(\hat{W}_1, \dots, \hat{W}_g)$. Then we have

$$\begin{bmatrix} \hat{\varphi} \\ \hat{\pi} \end{bmatrix} = \{[X_{-1}H; \hat{W}]'(\hat{\Omega}^{-1} \otimes I_T)[X_{-1}H; W]\}^{-1} \\ \times [X_{-1}H; \hat{W}]'(\hat{\Omega}^{-1} \otimes I_T)\Delta y. \quad (36)$$

If $(\hat{\varphi}', \hat{\pi}')$, $\hat{\Omega}$, and $\hat{W}$ are mutually consistent, then these are full information maximum likelihood (FIML) estimators. Hendry shows that a number of alternative estimation procedures are obtained from other choices for $\hat{\Omega}$ and $\hat{W}$. For example, if $\hat{\Omega} = I_g$ and $\hat{W}$ is obtained from the unrestricted reduced form, then (36) yields the two-stage least squares (2SLS) estimator. If the same choice for $\hat{W}$ is combined with $\hat{\Omega}$, estimated from the 2SLS residuals, then the three-stage least squares (3SLS) estimator results. If the system is just identified, then all procedures yield identical estimates. Even with overidentification, FIML and 3SLS are asymptotically equivalent and efficient. The estimated covariance matrix of the FIML/3SLS estimator equals

$$\hat{V} \begin{bmatrix} \hat{\varphi} \\ \hat{\pi} \end{bmatrix} = \{[X_{-1}H; \hat{W}]'(\hat{\Omega}^{-1} \otimes I_T)[X_{-1}H; \hat{W}]\}^{-1}. \quad (37)$$

Analogous to the single-equation case, the cointegrating vector can be normalized with respect to its first element, leading to the vector of estimated long-run effects:

$$\hat{\theta}_i = -\hat{\phi}_{i2}/\hat{\phi}_{i1}. \quad (38)$$

Mutatis mutandis, the covariance matrix of $\hat{\theta}_i$ can be obtained from (26). The likelihood ratio test statistic for the overidentifying restrictions is given by

$$LR = T \ln \frac{|\hat{\Gamma}_0^{-1} \hat{\Omega} \hat{\Gamma}_0^{-1}|}{|\hat{\Sigma}_{11.2}|}, \quad (39)$$

where $\hat{\Sigma}_{11.2}$ is the least squares covariance matrix estimator from the unrestricted reduced form, and $\hat{\Gamma}_0$ and $\hat{\Omega}$ are the FIML estimators of Γ_0 and Ω .

Using Johansen's (1992) more general results, Boswijk (1992) shows that under Assumptions 1–3, the conditional model estimators are consistent as $T \rightarrow \infty$, and in particular that the long-run parameter estimators $\hat{\theta}$ are super-consistent, i.e., $(\hat{\theta} - \theta) = O_p(T^{-1})$, regardless of the question whether or not the weak exogeneity restriction $\alpha_2 = 0$ holds. However, the estimators are only asymptotically efficient and mixed normally distributed if $\alpha_2 = 0$. In that case, the test statistics satisfy, as $T \rightarrow \infty$ and under the null hypothesis,

$$Wald, LR \xrightarrow{d} \chi^2(h), \quad (40)$$

where h is the number of overidentifying restrictions. If $\alpha_2 \neq 0$, then the asymptotic distribution of the estimators and test statistics can be expressed as a function of a vector Brownian motion process, and depends on nuisance parameters. Therefore, in that case an estimate of the long-run parameters should be based on the full system, which is considered next.

3.2. Full system estimation and the LM test for weak exogeneity

Consider first the single-equation case, and define $\gamma = \alpha_2/\lambda$, so that $\alpha_2 \beta' x_{t-1} = \gamma \delta' x_{t-1}$. Let $\Delta z = \text{vec}(\Delta Z) = \text{vec}(\Delta z_1, \dots, \Delta z_T)'$ denote the $Tk \times 1$ vector of observations on Δz_t , whereas $\varepsilon_2 = \text{vec}(E_2) = \text{vec}(\varepsilon_{21}, \dots, \varepsilon_{2T})'$. Next, let V denote the $T \times k$ matrix of regressors in the marginal model, other than the error correction variables (assumed to be the same across equations), and let ψ be its coefficient vector. Then the system, in vectorized notation, reads

$$\begin{aligned} \Delta y &= X_{-1} \delta + W \pi + v, \\ \Delta z &= (I_k \otimes X_{-1} \delta) \gamma + (I_k \otimes V) \psi + \varepsilon_2 \\ &= (\gamma \otimes X_{-1}) \delta + (I_k \otimes V) \psi + \varepsilon_2. \end{aligned} \quad (41)$$

Assume that the starting values are given and that $(v_t, \varepsilon'_{2t})'$ is normally distributed with covariance matrix $\text{diag}(\omega^2, \Sigma_{22})$. If we keep δ, γ, ω^2 , and Σ_{22} fixed, it is easily seen that the ML estimators of π and ψ are

$$\begin{aligned}\hat{\pi} &= (W'W)^{-1}W'(\Delta y - X_{-1}\delta), \\ \hat{\psi} &= [I_k \otimes (V'V)^{-1}V'](\Delta z - (\gamma \otimes X_{-1})\delta),\end{aligned}\quad (42)$$

the least squares estimators. Defining the familiar projection matrix $M_W = I_T - W(W'W)^{-1}W'$ (and similarly M_V), this leads to the concentrated log-likelihood (up to a constant term)

$$\begin{aligned}\mathcal{L} &= -\frac{T}{2} \ln \omega^2 - \frac{1}{2} \omega^{-2} (\Delta y - X_{-1}\delta)' M_W (\Delta y - X_{-1}\delta) \\ &\quad - \frac{T}{2} \ln |\Sigma_{22}| - \frac{1}{2} (\Delta z - (\gamma \otimes X_{-1})\delta)' [\Sigma_{22}^{-1} \otimes M_V] (\Delta z - (\gamma \otimes X_{-1})\delta).\end{aligned}\quad (43)$$

Because the information matrix satisfies the usual block diagonality restriction with respect to $(\delta, \gamma, \pi, \psi)$, ω^2 , and Σ_{22} , the derivatives with respect to the variance/covariance parameters are not given; as expected, they lead to $\hat{\omega}_2 = (\hat{v}'\hat{v})/T$ and $\hat{\Sigma}_{22} = (\hat{E}'_2\hat{E}_2)/T$. The derivatives of the log-likelihood with respect to δ and γ are

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \delta} &= \omega^{-2} X'_{-1} M_W (\Delta y - X_{-1}\delta) + [\gamma' \Sigma_{22}^{-1} \otimes X'_{-1} M_V] (\Delta z - (\gamma \otimes X_{-1})\delta), \\ \frac{\partial \mathcal{L}}{\partial \gamma} &= [\Sigma_{22}^{-1} \otimes \delta' X'_{-1} M_V] (\Delta z - (I_k \otimes X_{-1})\delta) \gamma.\end{aligned}\quad (44)$$

The corresponding Hessian matrix is given by

$$Q = \begin{bmatrix} Q_{\delta\delta} & Q_{\delta\gamma} \\ Q'_{\delta\gamma} & Q_{\gamma\gamma} \end{bmatrix}, \quad (45)$$

where

$$\begin{aligned}Q_{\delta\delta} &= -\omega^{-2} X'_{-1} M_W X_{-1} - (\gamma' \Sigma_{22}^{-1} \gamma) X'_{-1} M_V X_{-1}, \\ Q_{\delta\gamma} &= -\gamma' \Sigma_{22}^{-1} \otimes X'_{-1} M_V u_{-1} + X'_{-1} M_V E_2 \Sigma_{22}^{-1}, \\ Q_{\gamma\gamma} &= -\Sigma_{22}^{-1} \otimes u'_{-1} M_V u_{-1},\end{aligned}\quad (46)$$

with $u_{-1} = X_{-1}\delta$. The second term of $Q_{\delta\gamma}$ is obtained from (44) via

$$\begin{aligned}[\gamma' \Sigma_{22}^{-1} \otimes X'_{-1} M_V] (\Delta z - (\gamma \otimes X_{-1})\delta) &= [\gamma' \Sigma_{22}^{-1} \otimes X'_{-1} M_V] \text{vec } E_2 \\ &= \text{vec } X'_{-1} M_V E_2 \Sigma_{22}^{-1} \gamma.\end{aligned}\quad (47)$$

Because the ML residuals satisfy $M_V \hat{E}_2 = 0$, this term will vanish from the estimated Hessian.

From (44)–(46) an expression for the Lagrange multiplier (LM) test statistic for the weak exogeneity hypothesis $\alpha_2 = 0$ (hence $\gamma = 0$) is readily obtained. Under the null hypothesis, the conditional and marginal model can be estimated separately, leading to the conditional model estimators $\hat{\delta}$ and $\hat{\omega}^2$ and the marginal model estimator $\hat{\Sigma}_{22} = (\Delta Z' M_V \Delta Z) / T$. The components of the concentrated score vector, evaluated at $(\delta, \gamma, \omega^2, \Sigma_{22}) = (\hat{\delta}, 0, \hat{\omega}^2, \hat{\Sigma}_{22})$, are

$$\frac{\partial \mathcal{L}(\hat{\delta}, 0, \hat{\omega}^2, \hat{\Sigma}_{22})}{\partial \delta} = 0, \quad \frac{\partial \mathcal{L}(\hat{\delta}, 0, \hat{\omega}^2, \hat{\Sigma}_{22})}{\partial \gamma} = [\hat{\Sigma}_{22} \otimes \hat{u}'_{-1} M_V] \Delta z, \quad (48)$$

whereas the estimated Hessian equals

$$\hat{Q} = \begin{bmatrix} -\hat{\omega}^{-2} X'_{-1} M_W X_{-1} & 0 \\ 0 & -\hat{\Sigma}_{22}^{-1} \otimes \hat{u}'_{-1} M_V \hat{u}_{-1} \end{bmatrix}, \quad (49)$$

so that the LM statistic is given by

$$\begin{aligned} LM &= \Delta z' [\hat{\Sigma}_{22}^{-1} \otimes M_V \hat{u}_{-1} [\hat{u}'_{-1} M_V \hat{u}_{-1}]^{-1} \hat{u}'_{-1} M_V] \Delta z \\ &= \hat{\gamma}' [\hat{\Sigma}_{22} \otimes (\hat{u}'_{-1} M_V \hat{u}_{-1})^{-1}]^{-1} \hat{\gamma}, \end{aligned} \quad (50)$$

where $\hat{\gamma}$ is the multivariate least squares estimate in

$$\Delta z = (I_k \otimes \hat{u}_{-1}) \gamma + (I_k \otimes V) \psi + \varepsilon_z. \quad (51)$$

Hence LM is the variable addition statistic for $\hat{u}_{-1}$ in (51). This test (or its F -version) was proposed by Johansen (1992) as a weak exogeneity test; see also Boswijk (1992) for a more extensive derivation of (50) as an LM statistic. Using the superconsistency of the long-run parameter estimator $\hat{\theta}$, it follows that LM has an asymptotic χ^2 null distribution.

Making use of the partitioned inverse of the Hessian, Newton–Raphson leads to the following revision formula:

$$\begin{aligned} \tilde{\delta} &= \hat{\delta} + \left[X'_{-1} \left(\hat{\omega}^{-2} M_W + (\hat{\gamma}' \hat{\Sigma}_{22}^{-1} \hat{\gamma}) M_{V;\hat{u}_{-1}} \right) X_{-1} \right]^{-1} \\ &\quad \times X'_{-1} \left(\hat{\omega}^{-2} M_W \hat{v} + [\hat{\gamma}' \hat{\Sigma}_{22}^{-1} \otimes M_{V;\hat{u}_{-1}}] \hat{e}_2 \right). \end{aligned} \quad (52)$$

Here $\tilde{\delta}$ denotes the revised estimator, and a hat indicates evaluated at the previous estimate. The following iterative scheme emerges:

1. Estimate δ , conditional upon estimates of ω^2 , Σ_{22} , and γ , from (52).
2. Estimate ω^2 , Σ_{22} , and γ , conditional upon an estimate of δ , by least squares regressions of $(\Delta y - X_{-1} \delta)$ on W and of ΔZ on $X_{-1} \delta$ and V .

In order to obtain ML estimates, steps 1 and 2 should be repeated until convergence. Alternatively, re-estimation of δ may be performed only once. Because the conditional model estimator $\hat{\delta}$ is (super-)consistent, irrespective of γ , the estimators of γ and Σ_{22} in step 2 will be consistent if δ is replaced by $\hat{\delta}$. Therefore, substituting these estimates in (52) leads to an asymptotically efficient estimator. A similar two-step method to obtain efficiency is proposed by Engle and Yoo (1991).

The covariance matrix of $\hat{\delta}$ is estimated by

$$\tilde{V}[\hat{\delta}] = \left[X'_{-1} \left(\tilde{\omega}^{-2} M_W + (\tilde{\gamma}' \tilde{\Sigma}_{22}^{-1} \tilde{\gamma}) M_{V;\hat{\delta}_{-1}} \right) X_{-1} \right]^{-1}, \quad (53)$$

where a tilde indicates evaluated at the ML estimates. This matrix can be used to construct a Wald test for $R\delta = 0$. Using the parametrization $\delta = H\varphi$, restrictions of this form can also be imposed and tested by replacing X_{-1} and δ in the above formulae by $X_{-1}H$ and φ , respectively, and comparing the unrestricted and restricted log-likelihood (note that this requires full optimization of the likelihood).

For the simultaneous equations case ($g > 1$), we shall not give the full derivations of the ML estimator but rather provide the results, combining the derivations leading to (36) and (52). Define the $k \times g$ matrix $\gamma = \alpha_2 \lambda^{-1}$, and let U_{-1} be the $T \times g$ matrix of observations on $X_{-1}H_i\varphi_i$, $i = 1, \dots, g$, so that $X_{-1}\beta\alpha_2' = U_{-1}\gamma'$. Moreover, note that $\text{vec } U_{-1} = [I_g \otimes X_{-1}]H\varphi$, so that

$$\text{vec } U_{-1}\gamma' = [\gamma \otimes I_T] \text{vec } U_{-1} = [\gamma \otimes X_{-1}]H\varphi. \quad (54)$$

Hence the marginal model may now be expressed as

$$\begin{aligned} \Delta z &= (I_k \otimes U_{-1})\text{vec}(\gamma') + (I_k \otimes V)\psi + \varepsilon_2 \\ &= (\gamma \otimes X_{-1})H\varphi + (I_k \otimes V)\psi + \varepsilon_2. \end{aligned} \quad (55)$$

Starting from the FIML estimator $\hat{\varphi}$, obtained from the conditional model, the parameter matrices γ and Σ_{22} are estimated consistently using multivariate least squares in (55), substituting $\hat{\varphi}$ for φ . Again the LM test for weak exogeneity is a test for the significance of γ in this regression; see Johansen (1992) and Boswijk (1992). Defining $\hat{W}$ as before, and letting $P = (\hat{\Omega}^{-1/2} \otimes I_T)$ denote the GLS transformation matrix, the following revision formula for the ML estimator of φ results:

$$\begin{aligned} \tilde{\varphi} &= \hat{\varphi} + [H'(I_g \otimes X'_{-1})(P'M_{P\hat{W}}P + (\hat{\gamma}' \hat{\Sigma}_{22}^{-1} \hat{\gamma}) \otimes M_{V;\hat{\delta}_{-1}})(I_g \otimes X_{-1})H]^{-1} \\ &\quad \times H'(I_g \otimes X'_{-1})(P'M_{P\hat{W}}\hat{\varphi} + [\hat{\gamma}' \hat{\Sigma}_{22}^{-1} \otimes M_{V;\hat{\delta}_{-1}}]\hat{\varepsilon}_2). \end{aligned} \quad (56)$$

An iterative procedure can now be defined entirely analogous to the single-equation case. Note that at each iteration $\hat{W}$ as well as the parameter estimates should be updated.

A consistent estimator of the asymptotic covariance matrix of $\tilde{\varphi}$ is provided by the first right-hand-side matrix in (56). As before, Wald tests for overidentifying restrictions can be constructed using this covariance matrix. A likelihood ratio test for the overidentifying restrictions is obtained by comparing the maximized log-likelihood

$$\mathcal{L} = -\frac{T}{2} \ln |\tilde{\Sigma}| = -\frac{T}{2} (\ln |\tilde{\Gamma}_0^{-1} \tilde{\Omega} \tilde{\Gamma}_0^{-1}| + \ln |\tilde{\Sigma}_{22}|), \quad (57)$$

with and without the restrictions imposed. Finally, the parameters θ are estimated by (25) with $\hat{\delta}$ replaced by $\tilde{\delta}$ if $g = 1$, and *mutatis mutandis* by (38) if $g > 1$; their covariance matrix is provided by the analogue of (26). Because the statistics are based on the likelihood analysis of the full (reparametrized) system, the long-run parameter estimators will be superconsistent and asymptotically efficient and mixed normal, so that the LR test statistics have an asymptotic χ^2 null distribution (see Boswijk, 1992).

The estimation procedures proposed in this subsection can be summarized as follows. First, the cointegrating vectors are estimated on the basis of the conditional model. These are used to compute the estimated disequilibrium errors. The LM test for weak exogeneity is obtained as a variable addition test of these disequilibrium errors in the marginal model. If this test statistic is larger than its χ^2 critical value, then the cointegrating vectors are re-estimated on the basis of the full system, where the score and Hessian contributions of the conditional and marginal model are combined.

4. A Monte Carlo experiment

In this section the performance of the methods proposed in this paper are analyzed in a small-scale Monte Carlo study. In Phillips and Hansen (1990) and Phillips and Loretan (1991), a number of different estimation and testing procedures are compared; their data generating mechanism is

$$y_t = \theta z_t + u_{1t}, \quad z_t = z_{t-1} + u_{2t}, \quad t = 1, \dots, T, \quad (58)$$

where

$$\begin{aligned} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} &= \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} + \begin{bmatrix} 0.3 & -0.4 \\ \gamma & 0.6 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{bmatrix}, \\ \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} &\sim \text{IN} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \end{aligned} \quad (59)$$

and where $T = 50$, and $\theta = 2$. They consider a number of different parameter values for γ and ρ ; however, their results suggest that the main differences arise

from the choice of the parameter γ . The role of this parameter can be seen by reparametrizing (58) and (59) in a conditional and marginal ARMAX model:

$$\begin{aligned}\Delta y_t &= -(1.6 + \rho\gamma)(y_{t-1} - \theta z_{t-1}) + (\theta + \rho)\Delta z_t \\ &\quad - (0.4 - 0.3\rho)\Delta z_{t-1} + c(L)v_t, \\ \Delta z_t &= \gamma(y_{t-1} - \theta z_{t-1}) - 0.3\Delta z_{t-1} + c(L)\varepsilon_{2t},\end{aligned}\tag{60}$$

where $c(L) = 1 + 0.9L + (0.18 + 0.4\gamma)L^2$ and $v_t = \varepsilon_{1t} - \rho\varepsilon_{2t}$. It can be seen that γ equals the error correction coefficient in the marginal model and thus determines whether or not z_t is weakly (and in this case strongly) exogenous for θ . Therefore ρ is fixed at -0.8 (as in Phillips and Hansen, 1990), and for γ two values are considered: $\gamma \in \{0.8, 0.0\}$. The results are based on 30,000 replications.

Three estimators of θ and test statistics for $\theta = 2$ are based on an autoregressive approximation of (60), with a constant added to the regressors and $p = 3$. The first estimator (ECM) is the long-run parameter estimator (25), and the first t -statistic is a t -version of (32), with a degrees of freedom correction allowing for eight estimated parameters. The second estimator and t -statistic (2ST) are based on (52) and (53), after one iteration, where $\hat{\omega}^2$ and $\hat{\Sigma}_{22}$ are corrected for the number of estimated parameters. Finally Johansen's ML estimator is analyzed, together with a degrees of freedom corrected LR statistic, 'studentized' by taking the square root and multiplying by $\text{sign}(\hat{\theta} - 2)$. Note that these should give the same results as the iterative procedure suggested in Section 3.2, provided that it is iterated until convergence.

The results are displayed in Figs. 1–4 and Tables 1 and 2. First the case is considered where $\gamma = 0.8$ (no weak exogeneity). Fig. 1 contains the estimated densities of the three estimators in deviation from θ . These densities are estimated using Bierens' (1983) sample moments integrating normal kernel estimator, with a window width of 0.025 for Figs. 1 and 3, and a window width of 0.25 for Figs. 2 and 4.

The conditional model estimator ECM appears to be skewed and biased. This may be explained by its asymptotic distribution, which in the absence of weak exogeneity is a mixture of a unit root type distribution and a mixed Gaussian distribution (cf. Park and Phillips, 1988). The distribution of Johansen's ML estimator appears to be symmetric and unbiased. The density of the two-step estimator (2ST) lies between the other two densities. Although there is still some bias and skewness left, the correction in the second step seems to be rather successful. An additional property of the 2ST and ML estimators is that the shape of the densities is quite similar to a Cauchy density, and in particular that their tails seem to die out slowly. This may be explained by the lack of finite moments (causing extreme observations) of these estimators, because they are defined as a ratio, see (25), the denominator of which has nonzero density at the origin (see also Phillips, 1994). Note that for the same reason, the ECM estimator has no finite moments.

Fig. 1. Distribution of $\hat{\theta} - \theta$, $\gamma = 0.8$. Fig. 2. Distribution of t -ratios, $\gamma = 0.8$.

Given this lack of moments, it can be expected that the Monte Carlo sample mean and root mean square error provide a poor measure of location and dispersion, respectively. Nevertheless, these statistics are reported in Table 1, because they may provide an indication of the relative magnitude of the fat tail

Table 1
Mean, RMSE, and quantiles of $\hat{\theta} - \theta$

		Mean	RMSE	Median	$Q_{0.1}$	$Q_{0.9}$
$\gamma = 0.8$	ECM	– 0.092	0.030	– 0.067	– 0.246	0.036
	2ST	– 0.052	2.919	– 0.019	– 0.172	0.095
	ML	0.036	7.065	0.004	– 0.133	0.166
$\gamma = 0.0$	ECM	– 0.000	0.040	0.000	– 0.045	0.044
	2ST	– 0.002	0.045	– 0.002	– 0.051	0.045
	ML	– 0.004	0.644	– 0.002	– 0.055	0.047

problem. For a finite number of Monte Carlo replications, any sample moment will be finite, even if the corresponding population moment does not exist. If the tails of a particular distribution have little probability mass relative to another distribution, then extreme observations will occur relatively infrequently, so that the sample moments will be smaller. In order to provide a more reliable assessment of the bias, dispersion, and skewness of the distributions, the median and 10% and 90% quantiles are given.

From the upper three rows of Table 1 it can be observed that ECM has the largest bias and the ML estimator has the smallest bias, both in terms of mean and median. The sample RMSE suggests that the ML estimator suffers most from fat tails, and the ECM estimator least. The quantiles indicate that the conditional model estimator has the largest skewness, and that there are no large differences in dispersion.

In Fig. 2 the estimated densities of the t -ratios corresponding to the ECM, 2ST, and ML estimators are depicted and compared with the standard normal density. The ECM t -ratio seems to perform best as far as the shape of the density is concerned, but the problem with this statistic lies in its noncentrality. Again the 2ST and ML procedures have similar properties; the mean of these densities is much better (although there is some bias left in 2ST), but their variance is too large. A possible explanation for this is that the lag length of 3 is insufficient to capture the dynamics in the system; in addition, the asymptotic approximations may be relatively poor for the nonlinear procedures. The same picture emerges from Table 2: the rejection rate for all methods is more than twice the nominal size of 5%. This is caused by the wrong location for the ECM method (and, to a lesser extent, for the 2ST method) and by too much dispersion for the full system methods.

The results for the case when $\gamma = 0$ can be summarized quite briefly. In this case all methods are asymptotically equivalent, and this is reflected in Figs. 3 and 4 and Tables 1 and 2, bottom rows. All estimators are almost unbiased, the ECM estimator even up to three significant digits. This estimator is also most

Table 2
Mean, standard error, and rejection rates of t -ratios

		Mean	S.E.	$t < -1.96$	$t > 1.96$	$ t > 1.96$
$\gamma = 0.8$	ECM	-0.827	1.029	0.133	0.004	0.137
	2ST	-0.314	1.247	0.097	0.027	0.124
	ML	0.053	1.215	0.050	0.057	0.107
$\gamma = 0.0$	ECM	-0.005	1.061	0.034	0.032	0.065
	2ST	-0.081	1.169	0.055	0.040	0.095
	ML	-0.093	1.124	0.049	0.033	0.082

Fig. 3. Distribution of $\hat{\theta} - \theta$, $\gamma = 1$

concentrated around the true value, which is a finite sample efficiency gain from imposing a correct restriction (i.e., $\gamma = 0$). The ECM method is also superior with respect to the rejection frequency and distributional shape of the t -ratio; the 2ST method performs worst, but is very similar to the ML method in this case.

The Monte Carlo results can be summarized as follows. Firstly, the asymptotic problems with the single-equation methods in the absence of weak exogeneity, viz. skewness, bias, and nonsimilarity of test statistics, are clearly reflected in the behaviour of the ECM method. Secondly, the two-step method suggested in Section 3.2 seems to be moderately successful in correcting for these problems. This may, however, be influenced by the small number of parameters, so that it is advised to fully iterate this procedure or (in this case equivalently)

Fig. 4. Distribution of t -ratios, $\gamma = 0.0$.

use Johansen's ML estimator in practical applications. Thirdly, the ML estimator seems to suffer from fat tails more than the other estimators. Fourthly, there is a gain in imposing the weak exogeneity restriction when this is valid. Finally, the autoregressive approximations appear to perform quite reasonably, although there is some overrejection in the case of no weak exogeneity.

Because the same Monte Carlo DGP is used by Phillips and Hansen (1990) and Phillips and Loretan (1991), the results in this section may be directly compared with theirs. They consider the following four classes of estimators: (i) OLS, the least squares static regression estimator, (ii) FM (fully modified), a nonparametric correction to OLS, (iii) H (referring to Hendry), the least squares estimate in the augmented regression

$$y_t = \theta z_t + \sum_{j=1}^l \gamma_j \Delta y_{t-j} + \sum_{j=0}^m \beta_j \Delta z_{t-j} + \varepsilon_t, \quad (61)$$

(this estimator is labelled 'ECM' in Phillips and Hansen, 1990), and finally (iv) PL (Phillips–Loretan), the nonlinear least squares estimator in

$$y_t = \theta z_t + \sum_{j=1}^l \gamma_j (y_{t-j} - \theta z_{t-j}) + \sum_{j=0}^m \beta_j \Delta z_{t-j} + \varepsilon_t. \quad (62)$$

For the latter two estimators, models with *leads* of Δz_t are also considered. It can easily be shown that (62) with $l = p$ and $m = p - 1$ is observationally equivalent with the single-equation error correction model, and that (61) with $l = m = p - 1$ is a restricted version of this model, with $\lambda = 1$. As this restriction

seems to be quite implausible in practice, and is never imposed in Hendry's work (see, e.g., Ericsson et al., 1990, for a recent survey of Hendry's methodology) it seems that the PL estimator (without leads of Δz_t) deserves the label 'ECM' or 'Hendry' more than the H estimator.

Concentrating on the t -ratios in Phillips and Loretan (1991),² it appears that when $\gamma = 0.8$, the ML and 2ST methods appear to perform slightly better than the FM and PL estimators (the latter with leaded terms), and much better than the remaining procedures. If $\gamma = 0$, the ECM method clearly outperforms any of the four methods listed above, and this also seems to be the case for the 2ST and ML procedures. Part of this better performance will be due to the degrees of freedom corrections. In conclusion, the methods proposed in this paper work extremely well in this Monte Carlo experiment when there is weak exogeneity, and the suggested two-step procedure works reasonably well when the exogeneity assumption does not hold.

5. Discussion

In this paper we have studied the analysis of cointegrating relations in structural error correction models. The traditional analysis of long-run multipliers in conditional error correction models is extended in two directions, viz. simultaneous cointegrating relations and systems where the explanatory variables are not weakly exogenous. In a small Monte Carlo experiment, the asymptotic properties have been shown relevant in explaining the finite sample behaviour of the statistics.

In order to obtain the results, a set of rather stringent assumptions was required, viz. that the number of cointegrating relationships is known and equal to the number of endogenous variables, and that the conditional model is stable. As indicated in Section 2.3, these assumptions can be checked as follows. The cointegrating rank can only be consistently determined in the VAR model; hence any analysis should start with a sequence of Johansen's likelihood ratio tests to estimate r . The choice of the endogenous variables may then be guided by the outcome of these tests. Next, the stability of the conditional system can be tested using Boswijk's (1994) test. Whether in practice one is willing to take the second step (choosing g equal to r) is an open question; if not, the inferential procedures proposed in this paper cannot be directly applied. If Assumption 3 does hold, then the diagonality of the error correction matrix does not further limit the analysis; any system of simultaneous equations can be reformulated into this form, although the presence of overidentifying restrictions on the short-run structure will complicate the analysis.

²In their generating process, $\rho = -0.85$ instead of -0.8 .

Given the growing availability of asymptotically equivalent, mixed normal, and efficient estimators, the question emerges how to choose between these various procedures. One approach to this problem may be to base the choice on small sample performance, as we have done in this paper. Another consideration however is the purpose of the analysis: if we only want to estimate the long-run parameters and test hypotheses on them, we might use the semiparametric static regression-based procedures. If, on the other hand, the aim is to obtain a structural model, in which both the short- and the long-run parameters are of interest, there is a clear case in favour of the methods proposed in this paper. An additional problem with the semiparametric methods is that there is no way to check whether the corrections are sufficient; in other words, there is no analogue of the diagnostics that are available for autoregressive-type models.

Because our model is a reparametrization of the VAR-ECM, the question of the advantage of our methods compared to Johansen's procedure remains. First, it is clear that our methods require more assumptions, notably Assumption 3. Moreover, there may be larger computational costs associated with the estimators proposed in this paper, because they sometimes require iterations, whereas the VAR procedure only requires the solution of a small-scale eigenvalue problem. However, this disadvantage diminishes if overidentified systems are considered, because in that case the VAR procedure also requires iterations; see Johansen (1995) and Johansen and Juselius (1992, 1994). Moreover, the VAR model does not allow for a restricted short-run dynamic structure, whereas our methods do. In any case, the two procedures are complementary even if our assumptions are satisfied, because the VAR analysis is required to test for the number of unit roots in the system.

References

- Bårdsen, G., 1989, Estimation of long-run coefficients in error correction models, *Oxford Bulletin of Economics and Statistics* 51, 345–350.
- Bergstrom, A.R., ed., 1976, *Statistical inference in continuous time economic models* (North-Holland, Amsterdam).
- Bierens, H.J., 1983, Sample moments integrating normal kernel estimators of multivariate density and regression functions, *Sankhya B* 45, 160–192.
- Boswijk, H.P., 1990, *On the scope of conditional dynamic modelling of cointegrated variables* (with comment), *Tinbergen Institute Research Bulletin* 2, 97–111.
- Boswijk, H.P., 1992, *Cointegration, identification and exogeneity: Inference in structural error correction models* (Thesis Publishers, Amsterdam).
- Boswijk, H.P., 1993a, On the formulation of Wald tests on long-run parameters, *Oxford Bulletin of Economics and Statistics* 55, 137–144.
- Boswijk, H.P., 1993b, *Testing stability and identifiability of long-run equilibria*, Discussion paper (University of Amsterdam, Amsterdam).
- Boswijk, H.P., 1994, Testing for an unstable root in conditional and structural error correction models, *Journal of Econometrics* 63, 37–60.

- Davidson, J.E.H. and S.G. Hall, 1991, Cointegration in recursive systems, *Economic Journal* 101, 239–251.
- Engle, R.F. and C.W.J. Granger, 1987, Cointegration and error correction: Representation, estimation, and testing, *Econometrica* 55, 251–276.
- Engle, R.F. and B.S. Yoo, 1991, Cointegrated economic time series: A survey with new results, in: R.F. Engle and C.W.J. Granger, eds., *Modelling long-run economic relationships: Readings in cointegration* (Oxford University Press, Oxford).
- Engle, R.F., D.F. Hendry, and J.-F. Richard, 1983, Exogeneity, *Econometrica* 51, 277–304.
- Ericsson, N.R., J. Campos, and H.-A. Tran, 1990, PC-GIVE and David Hendry's econometric methodology, *Revista de Econometria* 10, 7–117.
- Granger, C.W.J., 1981, Some properties of time series data and their use in econometric model specification, *Journal of Econometrics* 16, 121–130.
- Hausman, J.A., 1983, Specification and estimation of simultaneous equations models, in: Z. Griliches and M. Intriligator, eds., *Handbook of econometrics* (North-Holland, Amsterdam).
- Hendry, D.F., 1976, The structure of simultaneous equations estimators, *Journal of Econometrics* 4, 51–88.
- Hendry, D.F. and G.E. Mizon, 1993, Evaluating dynamic econometric models by encompassing the VAR, in: P.C.B. Phillips, ed., *Models, methods and applications of econometrics: Essays in honor of A.R. Bergstrom* (Basil Blackwell, Cambridge, MA).
- Johansen, S., 1988, Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* 12, 231–254.
- Johansen, S., 1991, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* 59, 1551–1580.
- Johansen, S., 1992, Cointegration in partial systems and the efficiency of single-equation analysis, *Journal of Econometrics* 52, 389–402.
- Johansen, S., 1995, Identifying restrictions of linear equations – With applications to simultaneous equations and cointegration, *Journal of Econometrics*, this issue.
- Johansen, S. and K. Juselius, 1990, Maximum likelihood estimation and inference on cointegration – With applications to the demand for money, *Oxford Bulletin of Economics and Statistics* 52, 169–210.
- Johansen, S. and K. Juselius, 1992, Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK, *Journal of Econometrics* 53, 211–244.
- Johansen, S. and K. Juselius, 1994, Identification of the long-run and the short-run structure: An application to the ISLM model, *Journal of Econometrics* 63, 7–36.
- Krämer, W., 1986, Least squares regression when the independent variable follows an ARIMA process, *Journal of the American Statistical Association* 81, 150–154.
- Kremers, J.J.M., N.R. Ericsson, and J.J. Dolado, 1992, The power of cointegration tests, *Oxford Bulletin of Economics and Statistics* 54, 325–348.
- Park, J.Y., 1990, Maximum likelihood estimation of simultaneous cointegrated models, Discussion paper (University of Aarhus, Aarhus).
- Park, J.Y., 1992, Canonical cointegrating regressions, *Econometrica* 60, 119–143.
- Park, J.Y. and P.C.B. Phillips, 1988, Statistical inference in regressions with integrated processes: Part 1, *Econometric Theory* 4, 468–497.
- Park, J.Y. and P.C.B. Phillips, 1989, Statistical inference in regressions with integrated processes: Part 2, *Econometric Theory* 5, 95–131.
- Phillips, P.C.B., 1991, Optimal inference in cointegrated systems, *Econometrica* 59, 283–306.
- Phillips, P.C.B., 1994, Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models, *Econometrica* 62, 73–93.
- Phillips, P.C.B. and S.N. Durlauf, 1986, Multiple time series regression with integrated processes, *Review of Economic Studies* 53, 473–495.

- Phillips, P.C.B. and B.E. Hansen, 1990, Statistical inference in instrumental variable regression with $I(1)$ processes, *Review of Economic Studies* 57, 99–125.
- Phillips, P.C.B. and M. Loretan, 1991, Estimating long-run economic equilibria, *Review of Economic Studies* 58, 407–436.
- Saikkonen, P., 1991, Asymptotically efficient estimation of cointegration regressions, *Econometric Theory* 7, 1–21.
- Saikkonen, P., 1993, Estimation of cointegration vectors with linear restrictions, *Econometric Theory* 9, 19–35.
- Spanos, A., 1990, The simultaneous-equations model revisited: statistical adequacy and identification, *Journal of Econometrics* 44, 87–105.
- Stock, J.H., 1987, Asymptotic properties of least squares estimators of cointegrating vectors, *Econometrica* 55, 1035–1056.
- Urbain, J.-P., 1992, On weak exogeneity in error correction models, *Oxford Bulletin of Economics and Statistics* 54, 187–207.