Edinburgh School of Economics

Discussion Paper Series

Number 44

Long-Run Structural Modelling

Mohammad Hashem Pesaran (Trinity College, Cambridge)
Yongcheol Shin (University of Edinburgh)

Date
April 1999

Published by
School of Economics
University of Edinburgh
30 -31 Buccleuch Place
Edinburgh EH8 9JT
+44 (0)131 650 8361
http://www.ed.ac.uk/schools-departments/economics

THE UNIVERSITY *of* EDINBURGH

Long-Run Structural Modelling*

M. Hashem Pesaran
Trinity College, Cambridge

Yongcheol Shin
Department of Economics, University of Edinburgh

April 1999

Abstract

The paper develops a general framework for identification, estimation, and hypothesis testing in cointegrated systems when the cointegrating coefficients are subject to (possibly) non-linear and cross-equation restrictions, obtained from economic theory or other relevant *a priori* information. It provides a proof of the consistency of the maximum likelihood (ML) estimators, establishes the relative rates of convergence of the ML estimators of the short-run and the long-run parameters, and derives their asymptotic distribution; thus generalizing the results already available in the literature for the linear case. The paper also develops tests of the over-identifying (possibly) non-linear restrictions on the cointegrating vectors. The estimation and hypothesis testing procedures are applied to an Almost Ideal Demand System estimated on U.K. quarterly observations. Unlike many other studies of consumer demand this application does not treat relative prices and real per capita expenditures as exogenously given.

JEL Classifications: C1, C3, D1, E1.
Key Words: Cointegration, identification, testing non-linear restrictions, consistency, asymptotic distribution, Almost Ideal Demand Systems.

*We are grateful to Michael Binder, Cheng Hsiao, Kees Jan van Garderen, Soren Johansen, Kevin Lee, Adrian Pagan, Bahram Pesaran, and in particular to Peter Boswijk, Peter Robinson, Richard Smith, and an anonymous referee for most helpful comments. Partial financial support from the ESRC (grant No. R000233608) and the Isaac Newton Trust of Trinity College, Cambridge, are gratefully acknowledged.

1 Introduction

One of the main problems in the econometric analysis of cointegrated systems is that in general the cointegrating relations are identified only up to a non-singular linear transformation. In the simple case of $r = 1$ (r being the number of cointegrating relations), the one restriction needed to identify the cointegrating relation can be viewed as "normalizing" and applied to the coefficient of any one of the integrated variables that enter the cointegrating relation. However, when $r > 1$, the number of such "normalizing" or "scaling" restrictions is equal to r and needs to be supplemented with further $r^2 - r$ restrictions. As with the identification of a simultaneous equations system *à la* Cowles Commission, these additional restrictions should be obtained from *a priori* knowledge of the long-run equilibrium relations, often characterized by economic theory, (see Pesaran , 1997).

There are two main system approaches currently available for estimating cointegrating relations: Johansen's (1988, 1991) autoregressive vector error correction model and Phillips' (1991, 1995) triangular vector error correction model. Both approaches deal with the identification problem implicitly by imposing r^2 arbitrary restrictions, which seem to have been adopted for their mathematical convenience rather than their appropriateness on economic grounds. For a comprehensive review of the cointegration literature see Watson (1994).

More recently, Johansen (1995) has developed an eigenvalue routine for testing linear homogeneous restrictions imposed on one cointegrating vector at a time, implicitly assuming that the unrestricted part of the cointegrating space is exactly-identified. But he does not allow for non-linear restrictions or restrictions across different cointegrating vectors. Using Phillips' triangular framework, Saikkonen (1993a) has considered estimation of cointegrating vectors subject to linear restrictions, and develops tests of the validity of the restrictions. His procedure, however, requires *a priori* decomposition of m integrated variables in the system into r and $m - r$ subsets, such that the variables in the latter are not cointegrated, as in Phillips.

In this paper we consider the problem of identification, estimation, and hypothesis testing in cointegrated systems subject to general non-linear restrictions on the cointegrating vectors. We explicitly deal with the identification problem and derive rank and order conditions for identification of the cointegrating vectors, allowing for parametric restrictions across the cointegrating relations as well as for restrictions on individual cointegrating vectors. Our approach emphasizes the use of economic theory and does not require the *a priori* decomposition of the system variables as in Phillips and Saikkonen.

In general, it is taken for granted that the maximum likelihood (ML) estimators in a cointegrated vector autoregressive (VAR) model are consistent, but to our knowledge no general proof of the consistency of the ML estimators of the cointegrating vectors is available in the literature. The difficulty lies in the fact that the average log-likelihood function does not have a finite limit when the underlying variables are trended, and standard proofs of consistency and asymptotic normality of the ML estimators are therefore not applicable. This problem has been addressed in Saikkonen (1995) in the context of a relatively simple model, where he provides a proof of the consistency of the ML estimators of the long-run parameters *conditional* on the true values of the short-run parameters and *vice versa*, and establishes the asymptotic normality of the ML estimators. Building on Saikkonen's (1993b,

1995) work, this paper provides a formal proof of the consistency and super-consistency of the ML estimators of short-run and long-run parameters, respectively, allowing for general *non-linear* restrictions on the cointegrating coefficients. It further establishes stochastic equicontinuity conditions for the weak convergence of the sample information matrix and derives the asymptotic distribution of the ML estimators. Finally, it establishes the validity of the standard χ^2 tests for testing general non-linear over-identifying restrictions on the cointegrating vectors.

The estimation and testing procedures are then applied to an Almost Ideal Demand System estimated for three non-durable expenditure categories using U.K. quarterly observations over the period 1955(1) - 1993(2). This application provides an example where economic theory predicts cross-equation restrictions on the long-run relations.

The plan of the paper is as follows: Section 2 develops the ML theory for the analysis of the cointegrated systems subject to general non-linear restrictions on the cointegrating vectors. Section 2.2 deals with the identification problem. Section 2.4 provides the proof of consistency of the ML estimators and establishes their relative rates of convergence. Section 2.5 derives the asymptotic distribution of the ML estimators. Section 3 gives the asymptotic theory relevant to testing the over-identifying restrictions on cointegrating vectors. Section 4 presents the empirical application and Section 5 offers some concluding remarks. Some of the mathematical derivations and proofs are provided in the Appendix.

The following notation will be used throughout: The symbol $\Rightarrow$ signifies weak convergence in probability measure, $\overset{a}{\sim}$ asymptotic equality in distributions, MN mixture normal, $I(d)$ an integrated variable of order d, $\mathrm{Tr}(\cdot)$ the trace of a matrix, $vec(\cdot)$ columns of a matrix stacked into a column vector, $vech(\cdot)$ elements on and below the main diagonal of a symmetric matrix stacked into a column vector, $\mathbf{I}_m$ an identity matrix of order m, $diag[\cdot]$ a diagonal matrix, and $\|\mathbf{A}\| = (Tr(\mathbf{A}\mathbf{A}'))^{1/2}$ the Euclidean norm of $\mathbf{A}$.

2 Maximum Likelihood Analysis of the Vector Error Correction Model

2.1 VAR Approach to Cointegration Analysis

Consider the finite order VAR representation in an $m \times 1$ vector of $I(1)$ variables, $\mathbf{x}_t$:

$$\mathbf{A}_0\mathbf{x}_t = \mathbf{b}_0 + \mathbf{b}_1 t + \mathbf{A}_1\mathbf{x}_{t-1} + ... + \mathbf{A}_p\mathbf{x}_{t-p} + \boldsymbol{\zeta}_t,\ t = 1, 2, ...T, \tag{2.1}$$

where $\mathbf{b}_0$ and $\mathbf{b}_1$ are $m \times 1$ vectors of unknown coefficients, $\mathbf{A}_i,\ i = 0, 1, ..., p$, are $m \times m$ matrices of unknown parameters, $\mathbf{A}_0$ is non-singular, and $\boldsymbol{\zeta}_t$ is an $m \times 1$ vector of disturbances. For cointegration analysis it is convenient to rewrite (2.1) as

$$\mathbf{A}_0\Delta\mathbf{x}_t = \mathbf{b}_0 + \mathbf{b}_1 t + \sum_{i=1}^{p-1} \Psi_i \Delta\mathbf{x}_{t-i} - \mathbf{A}(1)\mathbf{x}_{t-1} + \boldsymbol{\zeta}_t,\ t = 1, 2, ...T, \tag{2.2}$$

where $\Psi_i = -\sum_{j=i+1}^{p} \mathbf{A}_j$, and $\mathbf{A}(1) = \mathbf{A}_0 - \sum_{i=1}^{p} \mathbf{A}_i$. The equilibrium properties of (2.2) are characterized by the rank of $\mathbf{A}(1)$. If $\mathbf{A}(1)$ is of rank r $(0 < r < m)$, then $\mathbf{A}(1)$ can be expressed as

$$\mathbf{A}(1) = \boldsymbol{\alpha}_* \boldsymbol{\beta}',$$

where $\boldsymbol{\alpha}_*$ and $\boldsymbol{\beta}$ are $m \times r$ matrices of full column rank, and $\boldsymbol{\beta}'\mathbf{x}_t$ gives the r linear combinations of $\mathbf{x}_t$ that are cointegrated.

The two forms of the model given by (2.1) and (2.2) highlight the two types of identification problem that are present in structural modelling with $I(1)$ variables. The first is the traditional identification problem and involves the identification of the contemporaneous coefficients, $\mathbf{A}_0$, and the short-run dynamic coefficients, $\mathbf{A}_1, ..., \mathbf{A}_p$. The second concerns the identification of the long-run coefficients, $\boldsymbol{\beta}$, which arises only when the $\mathbf{x}_t$'s are $I(1)$. As the above derivations make clear, the identification of the coefficients in $\mathbf{A}_j,\ j = 0, 1, ..., p$, does not necessarily provide information on those of $\boldsymbol{\beta}$, and knowledge of $\boldsymbol{\beta}$ does not provide information with which to identify the short-run dynamics. For example, without further *a priori* restrictions, the cointegrating vectors of the model are only identified up to a non-singular linear transformation, since for any non-singular $r \times r$ matrix, $\mathbf{Q}$, $\check{\boldsymbol{\alpha}}_* = \boldsymbol{\alpha}_* \mathbf{Q}'^{-1}$ and $\check{\boldsymbol{\beta}} = \boldsymbol{\beta}\mathbf{Q}$ give the same value of $\mathbf{A}(1)$, and therefore $(\check{\boldsymbol{\alpha}}_*, \check{\boldsymbol{\beta}})$ and $(\boldsymbol{\alpha}_*, \boldsymbol{\beta})$ cannot be distinguished using data alone. Assumptions on the endogeneity and exogeneity of the variables in the system may provide some information which may help to identify the parameters of interest, but in the case of $I(1)$ variables such distinctions are not essential.

The focus of this paper is on long-run structural modelling. It considers the problem of identification and estimation of the long-run coefficients, $\boldsymbol{\beta}$, and assumes that the short-run coefficients, $\mathbf{A}_0, \mathbf{A}_1, ..., \mathbf{A}_p$, are unrestricted. Consequently, we pre-multiply (2.2) by $\mathbf{A}_0^{-1}$, and work with the autoregressive vector error correction (VEC) model,

$$\Delta \mathbf{x}_t = \mathbf{a}_0 + \mathbf{a}_1 t + \sum_{i=1}^{p-1} \Gamma_i \Delta \mathbf{x}_{t-i} - \Pi \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t,\ t = 1, 2, ...T, \tag{2.3}$$

where $\mathbf{a}_0 = \mathbf{A}_0^{-1}\mathbf{b}_0$, $\mathbf{a}_1 = \mathbf{A}_0^{-1}\mathbf{b}_1$, $\Gamma_i = \mathbf{A}_0^{-1}\Psi_i$, $\Pi = \mathbf{A}_0^{-1}\mathbf{A}(1)$ and $\boldsymbol{\varepsilon}_t = \mathbf{A}_0^{-1}\boldsymbol{\zeta}_t$. Notice that

$$\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}', \tag{2.4}$$

where $\boldsymbol{\alpha} = \mathbf{A}_0^{-1}\boldsymbol{\alpha}_*$. We assume that the order of the VAR model is known, and that the initial values, $\mathbf{x}_0, \mathbf{x}_{-1}, ..., \mathbf{x}_{-p+1}$, are given.

We develop a general ML theory for the analysis of cointegrated systems subject to non-linear restrictions on the cointegrating relations in the context of the following VAR version of the model (2.3):

$$\Phi(L)\mathbf{x}_t = \mathbf{a}_0 + \Phi(1)\mathbf{c}\ t + \boldsymbol{\varepsilon}_t,\ t = 1, 2, ..., T, \tag{2.5}$$

where $\Phi(L) = \mathbf{I}_m - \sum_{i=1}^{p} \Phi_i L^i$, $\Phi_i = \mathbf{A}_0^{-1}\mathbf{A}_i$, for $i = 1, 2, ..., p$, $\mathbf{c}$ is an $m \times 1$ vector of unknown coefficients and the trend coefficients, $\mathbf{a}_1 = \Phi(1)\mathbf{c}$, are appropriately restricted so that the deterministic component of $\mathbf{x}_t$ is linearly trended for all values of r.[1]

To ensure that $\mathbf{x}_t$ are at most $I(1)$ and to rule out the possibility of explosive or seasonal unit roots we assume:

[1]See, for example, Pesaran, Shin and Smith (1998).

Assumption 2.1 *All the roots of the determinantal equation,* $|\mathbf{I}_m - \Phi_1 z - \cdots - \Phi_p z^p| = 0$, *are either on or outside the unit circle.*

Assumption 2.2 $\boldsymbol{\alpha}'_{\perp}\Gamma(1)\boldsymbol{\beta}_{\perp}$ *has full rank, where* $\Gamma(1) = \mathbf{I}_m - \sum_{i=1}^{p-1}\Gamma_i$, *and* $\boldsymbol{\alpha}_{\perp}$ *and* $\boldsymbol{\beta}_{\perp}$ *are* $m \times (m-r)$ *matrices of full column rank such that* $\boldsymbol{\alpha}'\boldsymbol{\alpha}_{\perp} = \mathbf{0}$ *and* $\boldsymbol{\beta}'\boldsymbol{\beta}_{\perp} = \mathbf{0}$.

Under the above assumptions we have[2]

$$\Delta \mathbf{x}_t = \boldsymbol{\mu} + \mathbf{C}(L)\boldsymbol{\varepsilon}_t, \tag{2.6}$$

where

$$\mathbf{C}(L) = \sum_{i=0}^{\infty} \mathbf{C}_i L^i = \mathbf{C}(1) + (1-L)\mathbf{C}^*(L),\ \mathbf{C}_0 = \mathbf{I}_m, \tag{2.7}$$

$$\mathbf{C}^*(L) = \sum_{i=0}^{\infty} \mathbf{C}_i^* L^i,\ \ \boldsymbol{\mu} = \mathbf{C}(1)\mathbf{a}_0 + \mathbf{c}, \tag{2.8}$$

and[3]

$$\mathbf{C}(1)\Phi(1) = \mathbf{C}(1)\Pi = \mathbf{0},\ \mathbf{C}^*(1)\Pi = \mathbf{I}_m. \tag{2.9}$$

Solving for $\mathbf{x}_t$, we now have

$$\mathbf{x}_t = \mathbf{x}_0 + \boldsymbol{\mu}t + \mathbf{C}(1)\mathbf{s}_t + \mathbf{C}^*(L)\boldsymbol{\varepsilon}_t,\ t = 1, 2, ..., T, \tag{2.10}$$

where $\mathbf{s}_t = \sum_{j=1}^{t}\boldsymbol{\varepsilon}_j$. The condition for cointegration is given by[4]

$$\mathbf{C}'(1)\boldsymbol{\beta} = \mathbf{0}. \tag{2.11}$$

Finally, pre-multiplying (2.10) by $\boldsymbol{\beta}'$ we have

$$\boldsymbol{\beta}'\mathbf{x}_t = \boldsymbol{\beta}'\mathbf{x}_0 + (\boldsymbol{\beta}'\mathbf{c})t + \sum_{i=0}^{\infty}\boldsymbol{\beta}'\mathbf{B}_i\boldsymbol{\varepsilon}_{t-i}, \tag{2.12}$$

where $\mathbf{B}_i = \sum_{j=0}^{i}\mathbf{C}_j$. It is clear that the cointegrating relations $\boldsymbol{\beta}'\mathbf{x}_t$ will contain r deterministic trends, characterized by the $r \times 1$ vector, $\boldsymbol{\beta}'\mathbf{c}$.

[2] See Johansen (1991, Theorem 4.1, p. 1559) and Johansen (1995, Theorem 4.2, p. 49).

[3] The coefficient matrices $\mathbf{C}_i^*$ satisfy the recursions, $\mathbf{C}_i^* = \mathbf{C}_{i-1}^*\Phi_1 + \mathbf{C}_{i-2}^*\Phi_2 + ... + \mathbf{C}_{i-p}^*\Phi_p$, for $i = 1, 2, ...,$ with $\mathbf{C}_0^* = \mathbf{I}_m - \mathbf{C}(1)$ and $\mathbf{C}_i^* = \mathbf{0}$, $i < 0$. Summing these relations across $i = 0, 1, 2, ...,$ it then follows that $\mathbf{C}^*(1)\Pi = \mathbf{I}_m$.

[4] See Engle and Granger (1987).

2.2 Identification of the Long Run Parameters: Rank and Order Conditions

When Π is of full rank, m, Π and the other parameters of the model are identified under fairly general conditions, and can be consistently estimated by OLS. See, for example, Lütkepohl (1991). However, if the rank of Π is $r < m$, then Π is subject to $(m - r)^2$ non-linear restrictions, and therefore determined uniquely in terms of the $(2mr - r^2)$ underlying unknown parameters.

We shall assume that $\boldsymbol{\alpha}$ is unrestricted and has full column rank and concentrate on the case where the identifying restrictions are imposed only on $\boldsymbol{\beta}$. We suppose that the $mr \times 1$ vector $\boldsymbol{\theta} = vec(\boldsymbol{\beta})$ satisfies the non-linear restrictions,

$$\boldsymbol{\theta} = \mathbf{f}\left(\boldsymbol{\phi}\right), \tag{2.13}$$

where $\boldsymbol{\phi}$ is an $s \times 1$ vector of unknown parameters. In particular we assume:

Assumption 2.3 *$\boldsymbol{\kappa} = vec(\boldsymbol{\alpha}') \in \Upsilon_\kappa$ and $\boldsymbol{\phi} \in \Upsilon_\phi$ where Υ_κ and Υ_ϕ are compact subsets of $\mathbb{R}^{mr}$ and $\mathbb{R}^s$, respectively, with their true values, $\boldsymbol{\kappa}_0$ and $\boldsymbol{\phi}_0$, being interior points of Υ_κ and Υ_ϕ. $\boldsymbol{\alpha}$ has the full column rank r for all $\boldsymbol{\kappa} \in \Upsilon_\kappa$, and the mapping $\mathbf{f}$, defined by (2.13), is continuously differentiable such that $\mathbf{F}(\boldsymbol{\phi}) = \partial \mathbf{f}(\boldsymbol{\phi})/\partial \boldsymbol{\phi}'$ has the full column rank, s, for all $\boldsymbol{\phi} \in \Upsilon_\phi$.*

A necessary and sufficient condition for identification of the long-run coefficients can be derived using (2.11). Denoting the true value of $\mathbf{C}(1)$ by $\mathbf{C}_0(1)$, it must be the case that

$$\mathbf{C}_0'(1)\boldsymbol{\beta}(\boldsymbol{\phi}) = \mathbf{0} \text{ if and only if } \boldsymbol{\phi} = \boldsymbol{\phi}_0. \tag{2.14}$$

Vectorizing the left hand side of (2.14), and using the mean-value expansion of $\boldsymbol{\theta} = \mathbf{f}\left(\boldsymbol{\phi}\right)$ around $\boldsymbol{\phi}_0$, we have

$$vec[\mathbf{C}_0'(1)\boldsymbol{\beta}(\boldsymbol{\phi})] = [\mathbf{I}_r \otimes \mathbf{C}_0'(1)]\,\mathbf{f}\left(\boldsymbol{\phi}\right) = [\mathbf{I}_r \otimes \mathbf{C}_0'(1)]\,\mathbf{F}(\bar{\boldsymbol{\phi}})\left(\boldsymbol{\phi} - \boldsymbol{\phi}_0\right), \tag{2.15}$$

where the (i, j) element of $\mathbf{F}(\bar{\boldsymbol{\phi}})$ is evaluated at $(\bar{\phi}_i, \bar{\phi}_j)$, and $\bar{\phi}_i$ is a convex combination of ϕ_{0i} and ϕ_i. For (2.14) to hold, we must therefore have

$$\text{Rank}\left\{[\mathbf{I}_r \otimes \mathbf{C}_0'(1)]\,\mathbf{F}(\bar{\boldsymbol{\phi}})\right\} = s \leq mr - r^2, \text{ for all } \bar{\phi}_i, \bar{\phi}_j \in \Upsilon_\phi. \tag{2.16}$$

The above identification condition is difficult to use in practice, but is needed in our proof of the consistency of the ML estimators. (See Section 2.4 below.) Restrictions on the cointegrating vectors, $\boldsymbol{\theta} = vec(\boldsymbol{\beta})$, are usually specified directly rather than indirectly through $\mathbf{f}\left(\boldsymbol{\phi}\right)$. It is therefore useful to consider the problem of identification and testing of over-identifying restrictions when $\boldsymbol{\theta}$ is subject to the following k general non-linear restrictions:

$$\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}, \tag{2.17}$$

where $\boldsymbol{\theta} \in \Theta$, $\Theta \subset \mathbb{R}^{mr}$, $\mathbf{h}(\boldsymbol{\theta}) = (h_1(\boldsymbol{\theta}), h_2(\boldsymbol{\theta}), ..., h_k(\boldsymbol{\theta}))'$, and $h_i(\boldsymbol{\theta})$, $i = 1, 2, ..., k$, is a known continuously differentiable scalar function of $\boldsymbol{\theta}$. Let $\boldsymbol{\theta}_0 = \mathbf{f}(\boldsymbol{\phi}_0)$ be the true value of $\boldsymbol{\theta}$, and assume that $\mathbf{H}(\boldsymbol{\theta}) = \partial\mathbf{h}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ has the full rank $k\,(\leq mr)$ for all $\boldsymbol{\theta} \in \Upsilon_\theta$, where $\Upsilon_\theta \ = \Theta$ $\cap(\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0})$. (See also Assumption 3.2 below.)

The analysis of identification of the cointegrating vectors can now be approached noting that $\Pi_0 = \boldsymbol{\alpha}_0\boldsymbol{\beta}_0' = \boldsymbol{\alpha}_0\mathbf{Q}'^{-1}\mathbf{Q}'\boldsymbol{\beta}_0' = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\mathbf{Q}$ is any arbitrary $r \times r$ non-singular matrix. Vectorizing $\boldsymbol{\beta} = \boldsymbol{\beta}_0\mathbf{Q}$ we obtain

$$\boldsymbol{\theta} = (\mathbf{I}_r \otimes \boldsymbol{\beta}_0)vec(\mathbf{Q}). \tag{2.18}$$

Consider now the mean value expansion of $\mathbf{h}(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$,

$$\mathbf{h}(\boldsymbol{\theta}) = \mathbf{h}(\boldsymbol{\theta}_0) + \mathbf{H}(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta} - \boldsymbol{\theta}_0), \tag{2.19}$$

where the (i,j) element of $\mathbf{H}(\bar{\boldsymbol{\theta}})$ is evaluated at $(\bar{\theta}_i, \bar{\theta}_j)$, and $\bar{\theta}_i$ is a convex combination of θ_{0i} and θ_i. Under (2.17) we have

$$\mathbf{H}(\bar{\boldsymbol{\theta}})\boldsymbol{\theta} = \mathbf{b}(\bar{\boldsymbol{\theta}}, \boldsymbol{\theta}_0), \tag{2.20}$$

where $\mathbf{b}(\bar{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = \mathbf{H}(\bar{\boldsymbol{\theta}})\boldsymbol{\theta}_0 - \mathbf{h}(\boldsymbol{\theta}_0) \neq \mathbf{0}$. Substituting for $\boldsymbol{\theta}$ from (2.18) in (2.20) yields

$$\mathbf{H}(\bar{\boldsymbol{\theta}})(\mathbf{I}_r \otimes \boldsymbol{\beta}_0)vec(\mathbf{Q}) = \mathbf{b}(\bar{\boldsymbol{\theta}}, \boldsymbol{\theta}_0), \tag{2.21}$$

and a unique solution exists for $vec(\mathbf{Q})$ if and only if

$$\text{Rank}\left\{\mathbf{H}(\bar{\boldsymbol{\theta}})(\mathbf{I}_r \otimes \boldsymbol{\beta}_0)\right\} = r^2, \text{ for all } \bar{\theta}_i, \bar{\theta}_j \in \Upsilon_\theta. \tag{2.22}$$

This condition can be viewed as the dual of the rank condition (2.16). A necessary condition for (2.22) to hold is given by the order condition, $k \geq r^2$. Since $s + k = mr$, this order condition is equivalent to the order condition given by (2.16).

2.3 The Log-Likelihood Function

Writing the VEC model, (2.3), in matrix notation, we have:

$$\Delta\mathbf{X} = \mathbf{Z}\mathbf{A} + \mathbf{Y}\Gamma - \mathbf{X}_{-1}\Pi' + \mathbf{E}, \tag{2.23}$$

where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)'$, $\boldsymbol{\tau} = (1,1,...,1)'$, $\mathbf{t} = (1,2,...,T)'$, $\mathbf{Z} = (\boldsymbol{\tau}, \mathbf{t})$, $\mathbf{Y} = (\Delta\mathbf{X}_{-1}, \Delta\mathbf{X}_{-2}, \ldots, \Delta\mathbf{X}_{-p+1})$, $\mathbf{E} = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, ..., \boldsymbol{\varepsilon}_T)'$, $\mathbf{A} = (\mathbf{a}_0, \mathbf{a}_1)'$, and $\Gamma = (\Gamma_1, \Gamma_2, ..., \Gamma_{p-1})'$ are $2 \times m$ and $m(p-1) \times m$ matrices of unknown coefficients, respectively. Conditional on the initial values, $(\mathbf{x}_{-p+1}, ..., \mathbf{x}_0)$, and assuming that $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \Omega)$, the log-likelihood function of the model is given by

$$\ell_T(\mathbf{a}, \boldsymbol{\varphi}) = \frac{-mT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Omega| - \frac{1}{2}\text{Tr}\left[\Omega^{-1}(\Delta\mathbf{X} - \mathbf{Z}\mathbf{A} - \mathbf{Y}\Gamma + \mathbf{X}_{-1}\Pi')'(\Delta\mathbf{X} - \mathbf{Z}\mathbf{A} - \mathbf{Y}\Gamma + \mathbf{X}_{-1}\Pi')\right], \tag{2.24}$$

where $\mathbf{a} = vec(\mathbf{A})$, $\boldsymbol{\varphi} = (\boldsymbol{\gamma}', \boldsymbol{\kappa}', \boldsymbol{\phi}', \boldsymbol{\omega}')'$, with $\boldsymbol{\gamma} = vec(\Gamma)$, $\boldsymbol{\kappa} = vec(\boldsymbol{\alpha}')$, and $\boldsymbol{\omega} = vech(\Omega)$.

2.4 Consistency of the ML Estimators of the Cointegrating Vectors Subject to General Non-Linear Restrictions

In general, it is taken for granted that the ML estimators in a cointegrated VAR model are consistent, but to our knowledge no general proof of the consistency of the ML estimators of the cointegrating vectors is available in the literature.[5] In this subsection we provide a general proof of the consistency of the ML estimators which is valid irrespective of the trending or cointegrating properties of $\mathbf{x}_t$.[6] Without loss of generality we shall work with the following concentrated log-likelihood function (the coefficients on deterministic components of the model being concentrated out):

$$\ell_T(\boldsymbol{\varphi}) = \frac{-mT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Omega| - \tfrac{1}{2}\mathrm{Tr}\left[\Omega^{-1}(\Delta\mathbf{X} - \mathbf{Y}\Gamma + \mathbf{X}_{-1}\Pi')'\mathbf{M}(\Delta\mathbf{X} - \mathbf{Y}\Gamma + \mathbf{X}_{-1}\Pi')\right], \tag{2.25}$$

where $\mathbf{M} = \mathbf{I}_T - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is a $T \times T$ idempotent matrix. We now provide a proof of the consistency of the ML estimators under the following additional assumptions:

Assumption 2.4 *The $m \times 1$ vector of errors, $\boldsymbol{\varepsilon}_t$, is such that*
(a) $E(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}) = \mathbf{0}$ and $Var(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}) = \Omega$, where $\mathcal{F}_{t-1} = (\mathbf{x}_{t-1,}\mathbf{x}_{t-2}, \mathbf{x}_{t-3}, ...)$ is a non-decreasing information set, and Ω is a positive definite symmetric matrix;
(b) $\sup_t E\left(\|\boldsymbol{\varepsilon}_t\|^j\right) < \infty$ for some $j > 2$.[7]

Assumption 2.5 *$\boldsymbol{\varphi} \in \Upsilon_\varphi$, where $\Upsilon_\varphi = \Upsilon_\gamma \times \Upsilon_\kappa \times \Upsilon_\phi \times \Upsilon_\omega$ is a compact subset of $\mathbb{R}^{h_\varphi}$ with $h_\varphi = m^2(p-1)+mr+s+\frac{1}{2}m(m+1)$. The true value of $\boldsymbol{\varphi}$, denoted by $\boldsymbol{\varphi}_0 = (\boldsymbol{\gamma}_0', \boldsymbol{\kappa}_0', \boldsymbol{\phi}_0', \boldsymbol{\omega}_0')'$, is an interior point of Υ_φ.*

Partition $\boldsymbol{\varphi} = (\boldsymbol{\phi}', \boldsymbol{\rho}')'$ into the long-run parameters, $\boldsymbol{\phi}$, and the short-run parameters, $\boldsymbol{\rho} = (\boldsymbol{\gamma}', \boldsymbol{\kappa}', \boldsymbol{\omega}')'$. Let $\hat{\boldsymbol{\varphi}}$ be the ML estimator of $\boldsymbol{\varphi}$. As noted by Saikkonen (1993b, 1995), proving the consistency of $\hat{\boldsymbol{\varphi}}$ is complicated in models with unit roots due to the fact that the ML estimators of the short-run parameters, $\hat{\boldsymbol{\rho}} = (\hat{\boldsymbol{\gamma}}', \hat{\boldsymbol{\kappa}}', \hat{\boldsymbol{\omega}}')'$, and those of the long-run parameters, $\hat{\boldsymbol{\phi}}$, converge to their true values at different rates. In the context of a relatively simple model, Saikkonen (1995, Section 5.3) provides a proof of consistency of the ML estimators of the long-run parameters *conditional* on the true values of the short-run parameters and *vice versa*.

In this subsection we will directly examine the convergence properties of $T^{-1}\left[\ell_T(\boldsymbol{\varphi}) - \ell_T(\boldsymbol{\varphi}_0)\right]$ and show that

$$\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0 = o_p(1), \text{ and } \hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0 = o_p(T^{-1/2}). \tag{2.26}$$

[5] As noted in the introduction, the difficulty lies in the fact that the average log-likelihood function does not have a finite limit when $\mathbf{x}_t$ is trended, and hence the usual proof of the consistency of the ML estimators along the lines set out, for example, in Davidson and MacKinnon (1993, Section 8.4) will not be applicable. A proof of the consistency of the least squares estimator of an exactly identified cointegrating vector is given by Stock (1987).

[6] This approach can be readily applied to the analysis of models with more complicated trends, or to cases where the nature of the trend depends on the values of one or more unknown parameters of the model.

[7] For a discussion of this assumption in the VEC models see, for example, Pesaran, Shin and Smith (1998).

Using (2.25), it is easily seen that

$$T^{-1}\left[\ell_T(\boldsymbol{\varphi}_0)-\ell_T(\boldsymbol{\varphi})\right]=\tfrac{1}{2}\left(\mathcal{A}_T+\mathcal{B}_T\right), \tag{2.27}$$

where

$$\mathcal{A}_T=-m\ln\left|\Omega^{-1}\Omega_0\right|-m-\mathrm{Tr}\left\{(\Omega_0^{-1}-\Omega^{-1})T^{-1}\mathbf{E}'\mathbf{M}\mathbf{E}\right\}, \tag{2.28}$$

and

$$\mathcal{B}_T=\mathrm{Tr}\left[\Omega^{-1}\left\{T^{-1}(\Delta\mathbf{X}-\mathbf{Z}\mathbf{A}-\mathbf{Y}\Gamma+\mathbf{X}_{-1}\Pi')'(\Delta\mathbf{X}-\mathbf{Z}\mathbf{A}-\mathbf{Y}\Gamma+\mathbf{X}_{-1}\Pi')-T^{-1}\mathbf{E}'\mathbf{M}\mathbf{E}\right\}\right]. \tag{2.29}$$

Using (2.23), under the true parameter values we have

$$\Delta\mathbf{X}=\mathbf{Z}\mathbf{A}_0+\mathbf{Y}\Gamma_0-\mathbf{X}_{-1}\Pi_0'+\mathbf{E}.$$

Also

$$\Pi-\Pi_0=\boldsymbol{\alpha}\boldsymbol{\beta}'-\boldsymbol{\alpha}_0\boldsymbol{\beta}_0'=\left(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\right)\boldsymbol{\beta}_0'+\boldsymbol{\alpha}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)',$$

where for notational convenience we are denoting $\boldsymbol{\beta}(\boldsymbol{\phi})$ and $\boldsymbol{\beta}(\boldsymbol{\phi}_0)$ by $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$, respectively. Using these results in (2.29) and noting that $\mathrm{Tr}(\mathbf{ABCD})=(vec\mathbf{D})'\left(\mathbf{A}\otimes\mathbf{C}'\right)vec\mathbf{B}'$,[8] then after some algebra we obtain:

$$\begin{aligned}\mathcal{B}_T &= (\gamma-\gamma_0)'\left(\Omega^{-1}\otimes\frac{\mathbf{Y}'\mathbf{M}\mathbf{Y}}{T}\right)(\gamma-\gamma_0)+(\boldsymbol{\kappa}-\boldsymbol{\kappa}_0)'\left(\Omega^{-1}\otimes\frac{\boldsymbol{\beta}_0'\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1}\boldsymbol{\beta}_0}{T}\right)(\boldsymbol{\kappa}-\boldsymbol{\kappa}_0)\\
&\quad+(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\left(\boldsymbol{\alpha}'\Omega^{-1}\boldsymbol{\alpha}\otimes\frac{\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1}}{T}\right)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)-2(\gamma-\gamma_0)'\left(\Omega^{-1}\otimes\frac{\mathbf{Y}'\mathbf{M}\mathbf{X}_{-1}\boldsymbol{\beta}_0}{T}\right)(\boldsymbol{\kappa}-\boldsymbol{\kappa}_0)\\
&\quad-2(\gamma-\gamma_0)'\left(\Omega^{-1}\boldsymbol{\alpha}\otimes\frac{\mathbf{Y}'\mathbf{M}\mathbf{X}_{-1}}{T}\right)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)+2(\boldsymbol{\kappa}-\boldsymbol{\kappa}_0)'\left(\Omega^{-1}\boldsymbol{\alpha}\otimes\frac{\boldsymbol{\beta}_0'\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1}}{T}\right)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\\
&\quad-2(\gamma-\gamma_0)'\left(\Omega^{-1}\otimes\frac{\mathbf{Y}'\mathbf{M}\mathbf{E}}{T}\right)vec(\mathbf{I}_m)+2(\boldsymbol{\kappa}-\boldsymbol{\kappa}_0)'\left(\Omega^{-1}\otimes\frac{\boldsymbol{\beta}_0'\mathbf{X}_{-1}'\mathbf{M}\mathbf{E}}{T}\right)vec(\mathbf{I}_m)\\
&\quad+2(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\left[\boldsymbol{\alpha}'\Omega^{-1}\otimes\frac{\mathbf{X}_{-1}'\mathbf{M}\mathbf{E}}{T}\right]vec(\mathbf{I}_m).\end{aligned} \tag{2.30}$$

Define the open balls, $B(\boldsymbol{\gamma}_0,\delta_\gamma)=\{\boldsymbol{\gamma}\in\Upsilon_\gamma:\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|<\delta_\gamma\}$, $B(\boldsymbol{\kappa}_0,\delta_\kappa)=\{\boldsymbol{\kappa}\in\Upsilon_\kappa:\|\boldsymbol{\kappa}-\boldsymbol{\kappa}_0\|<\delta_\kappa\}$, $B(\boldsymbol{\phi}_0,\delta_\phi)=\{\boldsymbol{\phi}\in\Upsilon_\phi:\|\boldsymbol{\phi}-\boldsymbol{\phi}_0\|<\delta_\phi\}$, and $B(\boldsymbol{\omega}_0,\delta_\omega)=\{\boldsymbol{\omega}\in\Upsilon_\omega:\|\boldsymbol{\omega}-\boldsymbol{\omega}_0\|<\delta_\omega\}$, and their complements, $\overline{B}(\boldsymbol{\gamma}_0,\delta_\gamma)=\{\boldsymbol{\gamma}\in\Upsilon_\gamma:\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|\geq\delta_\gamma\}$, $\overline{B}(\boldsymbol{\kappa}_0,\delta_\kappa)=\{\boldsymbol{\kappa}\in\Upsilon_\kappa:\|\boldsymbol{\kappa}-\boldsymbol{\kappa}_0\|\geq\delta_\kappa\}$, $\overline{B}(\boldsymbol{\phi}_0,\delta_\phi)=\{\boldsymbol{\phi}\in\Upsilon_\phi:\|\boldsymbol{\phi}-\boldsymbol{\phi}_0\|\geq\delta_\phi\}$, and $\overline{B}(\boldsymbol{\omega}_0,\delta_\omega)=\{\boldsymbol{\omega}\in\Upsilon_\omega:\|\boldsymbol{\omega}-\boldsymbol{\omega}_0\|\geq\delta_\omega\}$, respectively.

First, to prove the consistency of the ML estimator of the long-run parameters $\boldsymbol{\phi}$ (namely $\hat{\boldsymbol{\phi}}$), it is sufficient to show that for all values of $\boldsymbol{\rho}\in\Upsilon_\rho=\Upsilon_\gamma\times\Upsilon_\kappa\times\Upsilon_\omega$ and for every $\delta_\phi>0$ (see, for example, Saikkonen,1995, p.903),

$$\lim_{T\to\infty}\Pr\left\{\inf_{\varphi\in\overline{B}(\phi_0,\delta_\phi)\times\Upsilon_\rho}T^{-1}\left[\ell_T(\boldsymbol{\varphi}_0)-\ell_T(\boldsymbol{\varphi})\right]>0\right\}=1. \tag{2.31}$$

[8] See, for example, Magnus and Neudecker (1988, p. 31).

Using (A.7), (A.10) and (A.11) in Appendix , and under Assumption 2.5, it is easily seen that except for the third term in (2.30), all other terms of $T^{-1}\left[\ell_T(\boldsymbol{\varphi}_0)-\ell_T(\boldsymbol{\varphi})\right]$ are at most $O_p(1)$. Therefore,

$$2T^{-1}\left[\ell_T(\boldsymbol{\varphi}_0)-\ell_T(\boldsymbol{\varphi})\right]=T\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)'\left[\boldsymbol{\alpha}'\Omega^{-1}\boldsymbol{\alpha}\otimes\mathbf{C}_0(1)\frac{\mathbf{S}_{-1}'\mathbf{M}\mathbf{S}_{-1}}{T^2}\mathbf{C}_0(1)'\right]\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)+O_p(1), \tag{2.32}$$

where $\mathbf{S}_{-1}=(\mathbf{s}_0,\mathbf{s}_1,...,\mathbf{s}_{T-1})'$, $\mathbf{s}_t=\mathbf{s}_{t-1}+\boldsymbol{\varepsilon}_t$, $t=1,2,..$, with $\mathbf{s}_0=\mathbf{0}$. Hence upon using (2.15)

$$2T^{-1}\left[\ell_T(\boldsymbol{\varphi}_0)-\ell_T(\boldsymbol{\varphi})\right]=T(\boldsymbol{\phi}-\boldsymbol{\phi}_0)'\mathbf{Q}_{T,\phi\phi}(\boldsymbol{\phi}-\boldsymbol{\phi}_0)+O_p(1),$$

where

$$\mathbf{Q}_{T,\phi\phi}=\left\{\left[\mathbf{I}_r\otimes\mathbf{C}_0(1)'\right]\mathbf{F}(\bar{\boldsymbol{\phi}})\right\}'\left\{\boldsymbol{\alpha}'\Omega^{-1}\boldsymbol{\alpha}\otimes\frac{\mathbf{S}_{-1}'\mathbf{M}\mathbf{S}_{-1}}{T^2}\right\}\left\{\left[\mathbf{I}_r\otimes\mathbf{C}_0(1)'\right]\mathbf{F}(\bar{\boldsymbol{\phi}})\right\}, \tag{2.33}$$

and the (i,j) element of $\mathbf{F}(\bar{\boldsymbol{\phi}})$ is evaluated at $(\bar{\phi}_i,\bar{\phi}_j)$, and $\bar{\phi}_i$ is a convex combination of ϕ_{0i} and ϕ_i. By the rank condition (2.16), $\left[\mathbf{I}_r\otimes\mathbf{C}_0(1)'\right]\mathbf{F}(\bar{\boldsymbol{\phi}})$ has the full column rank s ($\leq mr-r^2$), and $T^{-2}\mathbf{S}_{-1}'\mathbf{M}\mathbf{S}_{-1}$ weakly converges to the positive definite (with probability 1) matrix $\mathbf{Q}_{SS}$ defined by (A.9) (see also (A.8)), and by assumption $\boldsymbol{\alpha}'\Omega^{-1}\boldsymbol{\alpha}$ is an $r\times r$ positive definite matrix for all values of $\boldsymbol{\kappa}$ and $\boldsymbol{\omega}$ in Υ_ρ. Hence, $\mathbf{Q}_{T,\phi\phi}$ also weakly converges (with probability 1) to the positive definite matrix $\mathbf{Q}_{\phi\phi}$ defined by

$$\mathbf{Q}_{\phi\phi}=\left\{\left[\mathbf{I}_r\otimes\mathbf{C}_0(1)'\right]\mathbf{F}(\bar{\boldsymbol{\phi}})\right\}'\left\{\boldsymbol{\alpha}'\Omega^{-1}\boldsymbol{\alpha}\otimes\mathbf{Q}_{SS}\right\}\left\{\left[\mathbf{I}_r\otimes\mathbf{C}_0(1)'\right]\mathbf{F}(\bar{\boldsymbol{\phi}})\right\}. \tag{2.34}$$

Therefore,

$$\inf_{\boldsymbol{\varphi}\in\overline{B}(\boldsymbol{\phi}_0,\delta_\phi)\times\Upsilon_\rho}T^{-1}\left[\ell_T(\boldsymbol{\varphi}_0)-\ell_T(\boldsymbol{\varphi})\right]\geq T\delta_\phi^2\lambda_{\min}(\mathbf{Q}_{T,\phi\phi})+O_p(1), \tag{2.35}$$

where $\lambda_{\min}(\mathbf{A})$ denotes the minimum eigenvalue of matrix $\mathbf{A}$. As $T\to\infty$, $\lambda_{\min}(\mathbf{Q}_{T,\phi\phi})$ weakly converges to $\lambda_{\min}(\mathbf{Q}_{\phi\phi})>0$, and the right hand side of (2.35) will increase without bounds with probability 1. This establishes the consistency of $\hat{\boldsymbol{\phi}}$, and also shows that the presence of stationary regressors does not affect the consistency of the long-run parameters.

Next, we prove both the super-consistency of $\hat{\boldsymbol{\phi}}$, and the consistency of $\hat{\boldsymbol{\rho}}$ simultaneously. Since the consistency of $\hat{\boldsymbol{\phi}}$ has already been established, we now focus on values of $\boldsymbol{\phi}$ that are sufficiently close to $\boldsymbol{\phi}_0$. Formally we define[9]

$$\boldsymbol{\phi}=\boldsymbol{\phi}_0+T^{-1/2}\mathbf{d}, \tag{2.36}$$

where we take $\mathbf{d}$ to be an $s\times 1$ vector of fixed constants defined on a compact set.[10] Accordingly, following Saikkonen (1995) we define the open shrinking ball $N_T(\boldsymbol{\phi}_0,\delta_d)=\{\boldsymbol{\phi}\in\Upsilon_\phi:$

[9]In general, the consistency of $\hat{\boldsymbol{\phi}}$ implies that $\boldsymbol{\phi}-\boldsymbol{\phi}_0=T^{-\delta}\mathbf{d}$ with $0<\delta<1$. The choice of $\delta=\frac{1}{2}$ is made ensuring that all the decomposed terms of the average log-likelihood can be of the same order of magnitude at most. Notice that the order of consistency of $\hat{\boldsymbol{\phi}}$ is determined by the rate at which this ball shrinks to zero.

[10]The case where elements of $\mathbf{d}$ are allowed to increase without bound is covered in the proof of (2.31).

$T^{\frac{1}{2}} \|\boldsymbol{\phi}-\boldsymbol{\phi}_0\| < \delta_d\}$ and its complement $\overline{N}_T(\boldsymbol{\phi}_0, \delta_d) = \{\boldsymbol{\phi} \in \Upsilon_\phi : T^{\frac{1}{2}} \|\boldsymbol{\phi}-\boldsymbol{\phi}_0\| \geq \delta_d\}$, and note that on $\overline{N}_T(\boldsymbol{\theta}_0, \delta_d)$ we also have $\|\mathbf{d}\| \geq \delta_d$. Let $C(\boldsymbol{\varphi}_0, \delta_d, \delta_\rho) = \cup_{\delta_d,\delta_\rho} \left(\overline{N}_T(\boldsymbol{\phi}_0, \delta_d) \times \overline{B}(\boldsymbol{\rho}_0, \delta_\rho)\right)$, where $\overline{B}(\boldsymbol{\rho}_0, \delta_\rho) = \{\boldsymbol{\rho} \in \Upsilon_\rho : \|\boldsymbol{\rho}-\boldsymbol{\rho}_0\| \geq \delta_\rho\}$, and the union is taken over all values of δ_d and δ_ρ such that $\delta_\varphi = (\delta_d^2+\delta_\rho^2)^{1/2}$ and $\delta_\rho = (\delta_\gamma^2+\delta_\kappa^2+\delta_\omega^2)^{1/2}$. We then prove that for every $\delta_\varphi > 0$,

$$\lim_{T\to\infty} \Pr\left\{ \inf_{\boldsymbol{\varphi}\in C(\boldsymbol{\varphi}_0,\delta_d,\delta_\rho)} T^{-1}\left[\ell_T(\boldsymbol{\varphi}_0) - \ell_T(\boldsymbol{\varphi})\right] > 0 \right\} = 1. \tag{2.37}$$

Using (2.15) we first rewrite (2.30) compactly as

$$\mathcal{B}_T = (\boldsymbol{\eta}-\boldsymbol{\eta}_0)'\mathbf{Q}_{1T}(\boldsymbol{\eta}-\boldsymbol{\eta}_0) + 2(\boldsymbol{\eta}-\boldsymbol{\eta}_0)'\mathbf{Q}_{2T}\boldsymbol{\nu}, \tag{2.38}$$

where $\boldsymbol{\eta}-\boldsymbol{\eta}_0 = \left[(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)', (\boldsymbol{\kappa}-\boldsymbol{\kappa}_0)', (\boldsymbol{\phi}-\boldsymbol{\phi}_0)'\right]'$, $\boldsymbol{\nu} = \left[(vec(\mathbf{I}_m))', (vec(\mathbf{I}_m))', (vec(\mathbf{I}_m))'\right]'$,

$$\mathbf{Q}_{1T} = \begin{bmatrix} \Omega^{-1}\otimes\frac{\mathbf{Y}'\mathbf{MY}}{T} & \Omega^{-1}\otimes\frac{\mathbf{Y}'\mathbf{MX}_{-1}\beta_0}{T} & \left(\Omega^{-1}\boldsymbol{\alpha}\otimes\frac{\mathbf{Y}'\mathbf{MX}_{-1}}{T}\right)\mathbf{F}(\bar{\boldsymbol{\phi}}) \\ \Omega^{-1}\otimes\frac{\beta_0'\mathbf{X}_{-1}'\mathbf{MY}}{T} & \Omega^{-1}\otimes\frac{\beta_0'\mathbf{X}_{-1}'\mathbf{MX}_{-1}\beta_0}{T} & \left(\Omega^{-1}\boldsymbol{\alpha}\otimes\frac{\beta_0'\mathbf{X}_{-1}'\mathbf{MX}_{-1}}{T}\right)\mathbf{F}(\bar{\boldsymbol{\phi}}) \\ \mathbf{F}(\bar{\boldsymbol{\phi}})'\left(\boldsymbol{\alpha}'\Omega^{-1}\otimes\frac{\mathbf{X}_{-1}'\mathbf{MY}}{T}\right) & \mathbf{F}(\bar{\boldsymbol{\phi}})'\left(\boldsymbol{\alpha}'\Omega^{-1}\otimes\frac{\mathbf{X}_{-1}'\mathbf{MX}_{-1}\beta_0}{T}\right) & \mathbf{F}(\bar{\boldsymbol{\phi}})'\left(\boldsymbol{\alpha}'\Omega^{-1}\boldsymbol{\alpha}\otimes\frac{\mathbf{X}_{-1}'\mathbf{MX}_{-1}}{T}\right)\mathbf{F}(\bar{\boldsymbol{\phi}}) \end{bmatrix},$$

and

$$\mathbf{Q}_{2T} = \begin{bmatrix} \Omega^{-1}\otimes\frac{\mathbf{Y}'\mathbf{ME}}{T} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Omega^{-1}\otimes\frac{\beta_0'\mathbf{X}_{-1}'\mathbf{ME}}{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{F}(\bar{\boldsymbol{\phi}})'\left(\boldsymbol{\alpha}'\Omega^{-1}\otimes\frac{\mathbf{X}_{-1}'\mathbf{ME}}{T}\right) \end{bmatrix}.$$

Using (2.38) in (2.27), we note that

$$2\inf T^{-1}\left[\ell_T(\boldsymbol{\varphi}_0)-\ell_T(\boldsymbol{\varphi})\right] \geq \inf(\mathcal{A}_T) + \inf\left[(\boldsymbol{\eta}-\boldsymbol{\eta}_0)'\mathbf{Q}_{1T}(\boldsymbol{\eta}-\boldsymbol{\eta}_0)\right] + 2\inf\left[(\boldsymbol{\eta}-\boldsymbol{\eta}_0)'\mathbf{Q}_{2T}\boldsymbol{\nu}\right], \tag{2.39}$$

where all the inf operations are taken over the set $\boldsymbol{\varphi} \in C(\boldsymbol{\varphi}_0, \delta_d, \delta_\rho)$. Defining $\mathbf{K}_T = diag(\mathbf{I}_{m^2(p-1)}, \mathbf{I}_{mr}, T^{1/2}\mathbf{I}_s)$, then

$$(\boldsymbol{\eta}-\boldsymbol{\eta}_0)'\mathbf{Q}_{1T}(\boldsymbol{\eta}-\boldsymbol{\eta}_0) = \left[\mathbf{K}_T(\boldsymbol{\eta}-\boldsymbol{\eta}_0)\right]'\left(\mathbf{K}_T^{-1}\mathbf{Q}_{1T}\mathbf{K}_T^{-1}\right)\left[\mathbf{K}_T(\boldsymbol{\eta}-\boldsymbol{\eta}_0)\right], \tag{2.40}$$

and

$$(\boldsymbol{\eta}-\boldsymbol{\eta}_0)'\mathbf{Q}_{2T}\boldsymbol{\nu} = \left[\mathbf{K}_T(\boldsymbol{\eta}-\boldsymbol{\eta}_0)\right]'\mathbf{K}_T^{-1}\mathbf{Q}_{2T}, \tag{2.41}$$

where for $\boldsymbol{\varphi} \in C(\boldsymbol{\varphi}_0, \delta_d, \delta_\rho)$, $\|\mathbf{K}_T(\boldsymbol{\eta}-\boldsymbol{\eta}_0)\| \geq \delta_\vartheta$ with $\delta_\vartheta = (\delta_\gamma^2+\delta_\kappa^2+\delta_d^2)^{1/2}$. Using (A.7), (A.10), and (A.11), it is then easily seen that

$$\mathbf{K}_T^{-1}\mathbf{Q}_{1T}\mathbf{K}_T^{-1} = \mathcal{J}_{T,\eta\eta} + o_p(1),\ \mathbf{K}_T^{-1}\mathbf{Q}_{2T} = o_p(1), \tag{2.42}$$

where

$$\mathcal{J}_{T,\eta\eta} = \begin{bmatrix} \Omega^{-1}\otimes\frac{\mathbf{Y}'\mathbf{MY}}{T} & \Omega^{-1}\otimes\frac{\mathbf{Y}'\mathbf{MX}_{-1}\boldsymbol{\beta}_0}{T} & \mathbf{0} \\ \Omega^{-1}\otimes\frac{\boldsymbol{\beta}_0'\mathbf{X}_{-1}'\mathbf{MY}}{T} & \Omega^{-1}\otimes\frac{\boldsymbol{\beta}_0'\mathbf{X}_{-1}'\mathbf{MX}_{-1}\boldsymbol{\beta}_0}{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_{T,\phi\phi} \end{bmatrix},$$

and $\mathbf{Q}_{T,\phi\phi}$ is defined by (2.33). Using (2.42) and recalling that $\mathbf{d}$ and $\boldsymbol{\rho}$ are defined on compact sets, it follows that $(\boldsymbol{\eta}-\boldsymbol{\eta}_0)'\mathbf{Q}_{2T}\boldsymbol{\nu} = o_p(1)$, and therefore,

$$2\inf T^{-1}\left[\ell_T(\boldsymbol{\varphi}_0)-\ell_T(\boldsymbol{\varphi})\right] \geq \inf(\mathcal{A}_T) + \inf\left\{\left[\mathbf{K}_T(\boldsymbol{\eta}-\boldsymbol{\eta}_0)\right]'\mathcal{J}_{T,\eta\eta}\left[\mathbf{K}_T(\boldsymbol{\eta}-\boldsymbol{\eta}_0)\right]\right\} + o_p(1), \tag{2.43}$$

where as before all the inf operations are taken over the set $\boldsymbol{\varphi} \in C\left(\boldsymbol{\varphi}_0, \delta_d, \delta_\rho\right)$.

Consider $\mathcal{A}_T$ defined by (2.28), which can be rewritten as

$$\begin{aligned}\mathcal{A}_T &= -m\ln\left|\Omega^{-1}\Omega_0\right| - m + \operatorname{Tr}\left(\Omega^{-1}\Omega_0\right) - \operatorname{Tr}\left\{\left(\Omega_0^{-1}-\Omega^{-1}\right)\left(T^{-1}\mathbf{E}'\mathbf{M}\mathbf{E}-\Omega_0\right)\right\}, \\ &= \sum_{i=1}^{m}(\lambda_i - 1 - \ln\lambda_i) - \operatorname{Tr}\left\{\left(\Omega_0^{-1}-\Omega^{-1}\right)\left(T^{-1}\mathbf{E}'\mathbf{M}\mathbf{E}-\Omega_0\right)\right\}, \end{aligned} \tag{2.44}$$

where $\lambda_i > 0,\ i = 1, 2, ..., m$, denote the eigenvalues of $\Omega^{-1}\Omega_0$. Since $T^{-1}\mathbf{E}'\mathbf{M}\mathbf{E}$ uniformly converges to Ω_0, the second term in (2.44) uniformly converges to 0. Notice also that $\lambda_i - 1 - \ln\lambda_i$ attains its unique minimum at $\lambda_i = 1$, and is strictly positive for all feasible values of λ_i not equal to unity. When $\lambda_i = 1$ for all i, we must have $\Omega = \Omega_0$. Therefore

$$\inf(\mathcal{A}_T) > 0 \Leftrightarrow \delta_\omega > 0. \tag{2.45}$$

For the second term in (2.43), we have[11]

$$\inf_{\boldsymbol{\varphi}\in C(\boldsymbol{\varphi}_0,\delta_d,\delta_\rho)}\left\{\left\{\left[\mathbf{K}_T(\boldsymbol{\eta}-\boldsymbol{\eta}_0)\right]'\mathcal{J}_{T,\eta\eta}\left[\mathbf{K}_T(\boldsymbol{\eta}-\boldsymbol{\eta}_0)\right]\right\}\right\} \geq \delta_\vartheta^2\lambda_{\min}(\mathcal{J}_{T,\eta\eta}). \tag{2.46}$$

As $T\to\infty$, $\lambda_{\min}(\mathcal{J}_{T,\eta\eta})$ converges weakly to $\lambda_{\min}(\mathcal{J}_{\eta\eta}) > 0$, which is the smallest eigenvalue of the positive definite (with probability 1) matrix $\mathcal{J}_{\eta\eta}$ given by

$$\mathcal{J}_{\eta\eta} = \begin{bmatrix} \Omega^{-1}\otimes\mathbf{Q}_{yy} & \Omega^{-1}\otimes\mathbf{Q}_{y\beta_0} & \mathbf{0} \\ \Omega^{-1}\otimes\mathbf{Q}_{y\beta_0} & \Omega^{-1}\otimes\mathbf{Q}_{\beta_0\beta_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_{\phi\phi} \end{bmatrix},$$

where $\mathbf{Q}_{yy}$, $\mathbf{Q}_{y\beta_0}$ and $\mathbf{Q}_{\beta_0\beta_0}$ are defined by (A.12). Using (2.45) and (2.46) in (2.43), and recalling that $\delta_\varphi = (\delta_\vartheta^2 + \delta_\omega^2)^{1/2}$ we obtain (2.37) for $\delta_\vartheta > 0$ and/or $\delta_\omega > 0$. This establishes the desired result given by (2.26), which we summarize:

Theorem 2.1 *Under Assumptions 2.1 to 2.5, and the identification condition (2.16), the ML estimator of $\boldsymbol{\varphi}$, obtained from the VEC model (2.23), is consistent. Furthermore, the ML estimator of the long-run parameters is super-consistent such that $\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0 = o_p(T^{-1/2})$.*

2.5 Asymptotic Distribution of the ML Estimators

Noting that under Assumption 2.4, $\Delta\mathbf{x}_{t-1}, ..., \Delta\mathbf{x}_{t-p+1}$, and $\boldsymbol{\beta}_0'\mathbf{x}_{t-1}$ are distributed independently of $\boldsymbol{\varepsilon}_t$, and using the results in Sections A.1 and A.2 of the Appendix, it is readily seen that

$$T^{-\frac{1}{2}}\frac{\partial\ell_T(\boldsymbol{\varphi}_0)}{\partial\boldsymbol{\gamma}} = (\Omega_0^{-1}\otimes\mathbf{I}_{m(p-1)})vec(T^{-\frac{1}{2}}\mathbf{Y}'\mathbf{M}\mathbf{E}) \overset{a}{\sim} N(\mathbf{0}, \Omega_0^{-1}\otimes\mathbf{Q}_{yy}),$$

[11] Recall that $\|\mathbf{K}_T(\boldsymbol{\eta}-\boldsymbol{\eta}_0)\| \geq \delta_\vartheta$, where $\delta_\vartheta = (\delta_\gamma^2 + \delta_\kappa^2 + \delta_d^2)^{1/2}$.

$$T^{-\frac{1}{2}}\frac{\partial \ell_T(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\kappa}} = (\Omega_0^{-1}\otimes \mathbf{I}_r)vec(T^{-\frac{1}{2}}\boldsymbol{\beta}_0'\mathbf{X}_{-1}'\mathbf{ME}) \overset{a}{\sim} N(\mathbf{0}, \Omega_0^{-1}\otimes \mathbf{Q}_{\beta_0\beta_0}),$$

and

$$T^{-\frac{1}{2}}\frac{\partial \ell_T(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\omega}} = \frac{1}{2}\mathbf{D}_m'(\Omega_0^{-1}\otimes \Omega_0^{-1})\mathbf{D}_m vech[T^{-\frac{1}{2}}(\mathbf{E}'\mathbf{ME}-T\Omega_0)] \overset{a}{\sim} N\left(0, \tfrac{1}{2}\mathbf{D}_m'\left[\Omega_0^{-1}\otimes \Omega_0^{-1}\right]\mathbf{D}_m\right).$$

where $\mathbf{Q}_{yy}$ and $\mathbf{Q}_{\beta_0\beta_0}$ are defined by (A.12) and $\mathbf{D}_m$ is an $m\times \frac{1}{2}m(m+1)$ duplication matrix.

The asymptotic distribution of $T^{-1}\partial \ell_T(\boldsymbol{\varphi}_0)/\partial \boldsymbol{\phi}$ is more complicated and involves functionals of Brownian motions. From (A.1), we have

$$T^{-1}\frac{\partial \ell_T(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\phi}} = \mathbf{F}'(\boldsymbol{\phi}_0)(\boldsymbol{\alpha}_0'\Omega_0^{-1}\otimes \mathbf{I}_m)vec\left(T^{-1}\mathbf{X}_{-1}'\mathbf{ME}\right),$$

which upon using (A.10) yields

$$T^{-1}\frac{\partial \ell_T(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\phi}} = \mathbf{F}'(\boldsymbol{\phi}_0)vec\left[\int_0^1 \mathbf{v}_1(a)d\mathbf{v}_2'(a)\right] + o_p(1), \tag{2.47}$$

where $\mathbf{v}_1(a) = \mathbf{C}_0(1)\mathbf{w}^*(a)$, $\mathbf{v}_2(a) = \boldsymbol{\alpha}_0'\Omega_0^{-1}\mathbf{w}(a)$, and $\mathbf{w}(a)$ and $\mathbf{w}^*(a)$, $a\in[0,1]$, are the standard and the demeaned and detrended Brownian motions, respectively (see Appendix A.2). But, noting from (2.9) that $\mathbf{C}_0(1)\boldsymbol{\alpha}_0\boldsymbol{\beta}_0' = \mathbf{0}$ and Rank$(\boldsymbol{\beta}_0) = r$, then

$$E[\mathbf{v}_1(a)\mathbf{v}_2'(a)] = \mathbf{C}_0(1)E\left[\mathbf{w}^*(a)\mathbf{w}'(a)\right]\Omega_0^{-1}\boldsymbol{\alpha}_0 = \mathbf{C}_0(1)\boldsymbol{\alpha}_0 = \mathbf{0}.$$

Hence, $\mathbf{v}_1(a)$ and $\mathbf{v}_2(a)$ are independently distributed, and the locally asymptotically mixed normal (LAMN) theory of Jeganathan (1982) is directly applicable to (2.47). (See also Phillips,1991, p. 289). Therefore,

$$T^{-1}\frac{\partial \ell_T(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\phi}} \overset{a}{\sim} MN\left\{\mathbf{0}, \Im_{\phi\phi}(\boldsymbol{\varphi}_0)\right\}, \tag{2.48}$$

where

$$\Im_{\phi\phi}(\boldsymbol{\varphi}_0) = \mathbf{F}'(\boldsymbol{\phi}_0)\left[\boldsymbol{\alpha}_0'\Omega_0^{-1}\boldsymbol{\alpha}_0\otimes \mathbf{C}_0(1)\mathbf{Q}_{SS}\mathbf{C}_0'(1)\right]\mathbf{F}(\boldsymbol{\phi}_0), \tag{2.49}$$

and $\mathbf{Q}_{SS}$, defined by (A.9), is a positive definite matrix with probability 1. Similarly,

$$\mathbf{D}_T\left\{\frac{-\partial^2 \ell_T(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'}\right\}\mathbf{D}_T \Rightarrow \Im(\boldsymbol{\varphi}_0), \tag{2.50}$$

where $\mathbf{D}_T = diag\left(T^{-\frac{1}{2}}\mathbf{I}_{m^2(p-1)}, T^{-\frac{1}{2}}\mathbf{I}_{mr}, T^{-1}\mathbf{I}_s, T^{-\frac{1}{2}}\mathbf{I}_{m(m+1)/2}\right)$,

$$\Im(\boldsymbol{\varphi}_0) = \begin{bmatrix} \Omega_0^{-1}\otimes \mathbf{Q}_{yy} & \Omega_0^{-1}\otimes \mathbf{Q}_{y\beta_0} & \mathbf{0} & \mathbf{0} \\ \Omega_0^{-1}\otimes \mathbf{Q}_{y\beta_0}' & \Omega_0^{-1}\otimes \mathbf{Q}_{\beta_0\beta_0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Im_{\phi\phi}(\boldsymbol{\varphi}_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{2}\mathbf{D}_m'(\Omega_0^{-1}\otimes \Omega_0^{-1})\mathbf{D}_m \end{bmatrix}, \tag{2.51}$$

and $\mathbf{Q}_{yy}$, $\mathbf{Q}_{y\beta_0}$ and $\mathbf{Q}_{\beta_0\beta_0}$ are defined in (A.12).

Combining the above results, and making use of the results in Sections A.1 and A.2 of the Appendix we have

Theorem 2.2 *In the context of the VEC model (2.23), and under Assumptions 2.1 to 2.5, and the identification condition (2.16),*

$$\mathbf{D}_T \left\{ \frac{\partial \ell_T(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}_0} \right\} \overset{a}{\sim} MN\left\{0, \mathfrak{I}(\boldsymbol{\varphi}_0)\right\}, \tag{2.52}$$

where $\mathfrak{I}(\boldsymbol{\varphi}_0)$, defined by (2.51), is a positive definite matrix with probability 1.

Consider the mean-value expansion of $\partial \ell_T(\hat{\boldsymbol{\varphi}})/\partial \boldsymbol{\varphi}$ around $\boldsymbol{\varphi}_0$:

$$\frac{\partial \ell_T(\hat{\boldsymbol{\varphi}})}{\partial \boldsymbol{\varphi}} = \frac{\partial \ell_T(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}} + \frac{\partial^2 \ell_T(\bar{\boldsymbol{\varphi}})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0),$$

where the (i, j) element of $\partial^2 \ell_T(\bar{\boldsymbol{\varphi}})/\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'$ is evaluated at $(\bar{\varphi}_i, \bar{\varphi}_j)$, and $\bar{\varphi}_i$ is a convex combination of φ_{i0} and $\hat{\varphi}_i$. Using the first-order conditions, $\partial \ell_T(\hat{\boldsymbol{\varphi}})/\partial \boldsymbol{\varphi} = \mathbf{0}$, we have:

$$\frac{\partial \ell_T(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}} = \left\{ -\frac{\partial^2 \ell_T(\bar{\boldsymbol{\varphi}})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'} \right\} (\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0). \tag{2.53}$$

Define

$$\mathfrak{I}_T(\bar{\boldsymbol{\varphi}}) = \mathbf{D}_T \frac{-\partial^2 \ell_T(\bar{\boldsymbol{\varphi}})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'} \mathbf{D}_T, \tag{2.54}$$

and write (2.53) as

$$\mathbf{D}_T \frac{\partial \ell_T(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}} = \mathfrak{I}_T(\bar{\boldsymbol{\varphi}}) \mathbf{D}_T^{-1} (\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0).$$

To derive the asymptotic distribution of $\mathbf{D}_T^{-1}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)$ it now remains to show that $\mathfrak{I}_T(\bar{\boldsymbol{\varphi}}) \Rightarrow \mathfrak{I}(\boldsymbol{\varphi}_0)$. Unlike the standard ML method, in the case of integrated and cointegrated systems the consistency of $\hat{\boldsymbol{\varphi}}$ is not sufficient to guarantee the weak convergence of $\mathfrak{I}_T(\bar{\boldsymbol{\varphi}})$ to $\mathfrak{I}(\boldsymbol{\varphi}_0)$, and additional conditions are needed. In particular, as shown by Saikkonen (1995, Proposition 3.2) $\mathfrak{I}_T(\bar{\boldsymbol{\varphi}}) \Rightarrow \mathfrak{I}(\boldsymbol{\varphi}_0)$, if $\mathbf{D}_T^{-1}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0) = o_p(1)$ and if the sample information matrix $\mathfrak{I}_T(\boldsymbol{\varphi})$ satisfies his stochastic equicontinuity condition.[12] The former condition is already established. See Theorem 2.1 and (2.26). The latter is proved in the Appendix (Section A.3), under the following assumption:

Assumption 2.6 *For $\boldsymbol{\phi}_* \in \Upsilon_\phi$, $\boldsymbol{\beta}(\boldsymbol{\phi})$ and $\mathbf{F}(\boldsymbol{\phi})$ satisfy the Lipschitz conditions:*

$$\|\boldsymbol{\beta}(\boldsymbol{\phi}_*) - \boldsymbol{\beta}(\boldsymbol{\phi})\| \leq c_\beta \|\boldsymbol{\phi}_* - \boldsymbol{\phi}\|, \tag{2.55}$$

$$\|\mathbf{F}(\boldsymbol{\phi}_*) - \mathbf{F}(\boldsymbol{\phi})\| \leq c_F \|\boldsymbol{\phi}_* - \boldsymbol{\phi}\|, \tag{2.56}$$

where c_β and c_F are positive constants.

[12] On the concept of stochastic equicontinuity and its use in establishing uniform convergence results in econometrics see Davidson (1994, pp. 335-340) and references cited therein; in particular Andrews (1987, 1992) and Pötscher and Prucha (1994).

These conditions impose a certain degree of smoothness on the non-linear dependence of $\boldsymbol{\beta}(\boldsymbol{\phi})$ and its derivatives, $\mathbf{F}(\boldsymbol{\phi})$, on $\boldsymbol{\phi}$, and are clearly satisfied when the restrictions on $\boldsymbol{\beta}$ are linear. The following theorem summarizes the main result on the asymptotic distribution of the ML estimators:

Theorem 2.3 *Consider the VEC model given by (2.23). Suppose that Assumptions 2.1 to 2.6, and the identification condition (2.16) hold. Then, the sample information matrix $\mathfrak{I}_T(\bar{\boldsymbol{\varphi}})$ defined by (2.54) weakly converges to $\mathfrak{I}(\boldsymbol{\varphi}_0)$, defined by (2.51), and the ML estimator of $\boldsymbol{\varphi}$, obtained subject to the general non-linear restrictions vec$(\boldsymbol{\beta}) = \boldsymbol{\theta} = \mathbf{f}(\boldsymbol{\phi})$, asymptotically has the mixture normal distribution:*

$$\mathbf{D}_T^{-1}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0) \overset{a}{\sim} MN\left\{\mathbf{0}, \mathfrak{I}^{-1}(\boldsymbol{\varphi}_0)\right\}. \tag{2.57}$$

Remark 2.1 *When vec$(\boldsymbol{\beta}) = \boldsymbol{\theta}$ is unrestricted, its associated information matrix, $\boldsymbol{\alpha}_0'\Omega_0^{-1}\boldsymbol{\alpha}_0 \otimes \mathbf{C}_0(1)\mathbf{Q}_{SS}\mathbf{C}_0'(1)$, is singular, having rank $mr - r^2$ with probability 1. Therefore, we need at least r^2 independent restrictions to identify $\boldsymbol{\theta}$. This represents a generalization of the result obtained by Rothenberg (1971) for the case where the underlying processes are stationary.*

Remark 2.2 *Given that $\mathfrak{I}(\boldsymbol{\varphi}_0)$ is block-diagonal, the ML estimators of the short-run parameters, $\hat{\boldsymbol{\rho}}$, are asymptotically distributed independently of $\hat{\boldsymbol{\phi}}$. Therefore, for large enough T, inferences on the short-run parameters can be carried out treating $\hat{\boldsymbol{\phi}}$ as if they were given. Thus, the results obtained in the literature for the case where the restrictions on $\boldsymbol{\theta}$ are linear extend readily to models with non-linear (over-identifying) restrictions.*

3 Testing Over-Identifying Restrictions on Cointegrating Vectors

In this section we consider the problem of testing over-identifying restrictions imposed directly on the cointegrating vectors. See also the discussion at the end of Section 2.2. Consider the following partition of $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$, the k $(> r^2)$ restrictions on $\boldsymbol{\theta}$ given by (2.17):

$$\mathbf{h}(\boldsymbol{\theta}) = [\mathbf{h}_A'(\boldsymbol{\theta}), \mathbf{h}_B'(\boldsymbol{\theta})] = \mathbf{0},$$

where $\mathbf{h}_A(\boldsymbol{\theta})$ and $\mathbf{h}_B(\boldsymbol{\theta})$ are $r^2 \times 1$ and $(k - r^2) \times 1$ vector functions, respectively. Without loss of generality, $\mathbf{h}_A(\boldsymbol{\theta}) = \mathbf{0}$ can be regarded as the r^2 just-identifying restrictions, and the remaining restrictions, $\mathbf{h}_B(\boldsymbol{\theta}) = \mathbf{0}$, then constitute the $k - r^2$ over-identifying restrictions.[13].

Let $\boldsymbol{\psi} = (\boldsymbol{\rho}', \boldsymbol{\theta}')'$, where $\boldsymbol{\rho} = (\boldsymbol{\gamma}', \boldsymbol{\kappa}', \boldsymbol{\omega}')'$ and $\boldsymbol{\theta}$ are the short-run and the long-run parameters, respectively, and consider the following assumptions that correspond to the Assumptions 2.3, 2.5 and 2.6 of the previous sections:

Assumption 3.1 $\boldsymbol{\theta} \in \Theta$ *where* $\Theta \subset \mathbb{R}^{mr}$ *and* $\mathbf{h}(\boldsymbol{\theta})$ *is a continuously differentiable function of* $\boldsymbol{\theta}$. *Under* $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$, $\boldsymbol{\theta} \in \Upsilon_\theta$, *where* Υ_θ *is a compact subset of* Θ, *and the* $k \times mr$ *Jacobian matrix,* $\mathbf{H}(\boldsymbol{\theta}) = \partial\mathbf{h}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$, *has full rank* $k \leq mr$ *for all* $\boldsymbol{\theta} \in \Upsilon_\theta$.

[13] It is easily seen that the test results are invariant to the way $\mathbf{h}(\boldsymbol{\theta})$ is decomposed into these two sets of restrictions.

Assumption 3.2 *$\boldsymbol{\psi} \in \Upsilon_\psi$, where $\Upsilon_\psi = \Upsilon_\rho \times \Upsilon_\theta$, is a compact subset of $\mathbb{R}^{h_\psi}$ with $h_\psi = m^2(p-1) + mr + \frac{1}{2}m(m+1) + mr$. The true value of $\boldsymbol{\psi}$, denoted by $\boldsymbol{\psi}_0 = (\boldsymbol{\rho}_0', \boldsymbol{\theta}_0')'$, is an interior point of Υ_ψ.*

Assumption 3.3 *For $\boldsymbol{\theta}_* \in \Upsilon_\theta$, $\mathbf{h}(\boldsymbol{\theta})$ and $\mathbf{H}(\boldsymbol{\theta})$ satisfy the Lipschitz conditions:*

$$\|\mathbf{h}(\boldsymbol{\theta}_*) - \mathbf{h}(\boldsymbol{\theta})\| \leq c_h \|\boldsymbol{\theta}_* - \boldsymbol{\theta}\|, \tag{3.1}$$

$$\|\mathbf{H}(\boldsymbol{\theta}_*) - \mathbf{H}(\boldsymbol{\theta})\| \leq c_H \|\boldsymbol{\theta}_* - \boldsymbol{\theta}\|, \tag{3.2}$$

where c_h and c_H are positive constants.

Using similar results as in Section 2.5, we have

$$\mathbf{D}_{\psi T} \frac{\partial \ell_T(\boldsymbol{\psi}_0)}{\partial \boldsymbol{\psi}} \equiv \begin{bmatrix} \mathbf{d}(\boldsymbol{\rho}_0) \\ \mathbf{d}(\boldsymbol{\theta}_0) \end{bmatrix} \overset{a}{\sim} MN\left\{\mathbf{0}, \Im(\boldsymbol{\psi}_0)\right\}, \tag{3.3}$$

where $\mathbf{D}_{\psi T} = diag\left(T^{-\frac{1}{2}}\mathbf{I}_{h_\rho}, T^{-1}\mathbf{I}_{mr}\right)$, $h_\rho = m^2(p-1)+mr+m(m+1)/2$, $\mathbf{d}(\boldsymbol{\rho}_0) = T^{-\frac{1}{2}}\frac{\partial \ell_T(\boldsymbol{\psi}_0)}{\partial \boldsymbol{\rho}}$, $\mathbf{d}(\boldsymbol{\theta}_0) = T^{-1}\frac{\partial \ell_T(\boldsymbol{\psi}_0)}{\partial \boldsymbol{\theta}}$, and $\Im(\boldsymbol{\psi}_0)$ is defined by

$$\mathbf{D}_{\psi T}\left\{-\frac{\partial^2 \ell_T(\boldsymbol{\psi}_0)}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'}\right\}\mathbf{D}_{\psi T} \Rightarrow \Im(\boldsymbol{\psi}_0) = \begin{bmatrix} \Im_{\rho\rho}(\boldsymbol{\psi}_0) & \mathbf{0} \\ \mathbf{0} & \Im_{\theta\theta}(\boldsymbol{\psi}_0) \end{bmatrix}. \tag{3.4}$$

Note that $\Im_{\rho\rho}(\boldsymbol{\psi}_0)$ is a positive definite matrix, but

$$\Im_{\theta\theta}(\boldsymbol{\psi}_0) = \boldsymbol{\alpha}_0'\Omega_0^{-1}\boldsymbol{\alpha}_0 \otimes \mathbf{C}_0(1)\mathbf{Q}_{SS}\mathbf{C}_0'(1) \tag{3.5}$$

is singular, having rank $mr - r^2$ with probability 1.[14]

Let $\tilde{\boldsymbol{\psi}} = (\tilde{\boldsymbol{\rho}}', \tilde{\boldsymbol{\theta}}')'$ and $\hat{\boldsymbol{\psi}} = (\hat{\boldsymbol{\rho}}', \hat{\boldsymbol{\theta}}')'$ be the ML estimators of $\boldsymbol{\psi}$ obtained subject to the r^2 exactly-identifying restrictions (say, $\mathbf{h}_A(\boldsymbol{\theta}) = \mathbf{0}$), and subject to the k restrictions, $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$, respectively. Then, the $k - r^2$ over-identifying restrictions on $\boldsymbol{\theta}$ can be tested using the log-likelihood ratio statistic

$$LR_T = 2\left\{\ell_T(\tilde{\boldsymbol{\psi}}) - \ell_T(\hat{\boldsymbol{\psi}})\right\}, \tag{3.6}$$

where $\ell_T(\hat{\boldsymbol{\psi}})$ and $\ell_T(\tilde{\boldsymbol{\psi}})$ represent the maximized values of the log-likelihood function obtained under $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$ and $\mathbf{h}_A(\boldsymbol{\theta}) = \mathbf{0}$, respectively.

Under Assumptions 2.1, 2.2, 2.4, and 3.1 through 3.3, and using a similar line of reasoning as in Sections 2.4 and 2.5, it can be shown that $\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0 = o_p(1)$ and $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = o_p(T^{-\frac{1}{2}})$, and the sample information matrix, $\Im_T(\boldsymbol{\psi}) = \mathbf{D}_{\psi T}\left\{-\frac{\partial^2 \ell_T(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'}\right\}\mathbf{D}_{\psi T}$, also satisfies Saikkonen's (1995) stochastic equicontinuity condition SE_o. Therefore, we have

[14] For a general analysis of ML estimation in the case of singular information matrices see Silvey (1959, Section 6) and Breusch (1986). However, their analysis is not directly applicable to models with unit roots.

Theorem 3.1 *Under Assumptions 2.1, 2.2, 2.4, and 3.1 through 3.3,*

$$\sqrt{T}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0) \overset{a}{\sim} MN\{\mathbf{0}, \mathfrak{I}_{\rho\rho}^{-1}(\boldsymbol{\psi}_0)\},\ T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \overset{a}{\sim} MN\{\mathbf{0}, \mathbf{V}_{\theta\theta}(\boldsymbol{\psi}_0)\}, \tag{3.7}$$

where

$$\mathbf{V}_{\theta\theta}(\boldsymbol{\psi}_0) = \mathbf{J}_{\theta\theta}^{-1}(\boldsymbol{\psi}_0) - \mathbf{J}_{\theta\theta}^{-1}(\boldsymbol{\psi}_0)\mathbf{H}'(\boldsymbol{\theta}_0)\{\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{J}_{\theta\theta}^{-1}(\boldsymbol{\psi}_0)\mathbf{H}'(\boldsymbol{\theta}_0)\}^{-1}\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{J}_{\theta\theta}^{-1}(\boldsymbol{\psi}_0), \tag{3.8}$$

is a singular random matrix having rank $mr - k$ with probability 1, and

$$\mathbf{J}_{\theta\theta}(\boldsymbol{\psi}_0) = \mathfrak{I}_{\theta\theta}(\boldsymbol{\psi}_0) + \mathbf{H}_A'(\boldsymbol{\theta}_0)\mathbf{H}_A(\boldsymbol{\theta}_0), \tag{3.9}$$

is a positive definite matrix.

Proof. See Section A.4 in the Appendix .

Theorem 3.2 *Under Assumptions 2.1, 2.2, 2.4, and 3.1 through 3.3, the log-likelihood ratio statistic for testing $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$, defined in (3.6), is asymptotically distributed as a χ^2 variate with $k - r^2$ degrees of freedom.*

Proof. See Section A.5 in the Appendix.

Remark 3.1 *The Wald statistic (W) for testing the $k - r^2$ over-identifying restrictions, $\mathbf{h}_B(\boldsymbol{\theta}) = 0$, is given by*

$$W = T^2\mathbf{h}_B'(\tilde{\boldsymbol{\theta}})[\mathbf{H}_B(\tilde{\boldsymbol{\theta}})\mathbf{V}_{\theta\theta}^A(\tilde{\boldsymbol{\psi}})\mathbf{H}_B'(\tilde{\boldsymbol{\theta}})]^{-1}\mathbf{h}_B(\tilde{\boldsymbol{\theta}}), \tag{3.10}$$

where $\mathbf{H}_B(\tilde{\boldsymbol{\theta}}) = \partial\mathbf{h}_B(\tilde{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}'$. Since

$$\mathbf{h}_B(\tilde{\boldsymbol{\theta}}) = \mathbf{H}_B(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(1),$$

hence using (A.24),

$$T\mathbf{h}_B(\tilde{\boldsymbol{\theta}}) \overset{a}{\sim} MN\{\mathbf{0}, \mathbf{H}_B(\boldsymbol{\theta}_0)\mathbf{V}_{\theta\theta}^A(\boldsymbol{\psi}_0)\mathbf{H}_B'(\boldsymbol{\theta}_0)\}.$$

Therefore, $W \overset{a}{\sim} \chi^2_{k-r^2}$. Finally, the Lagrange Multiplier statistic (LM) for testing the over-identifying restrictions can be written as

$$LM = \hat{\boldsymbol{\lambda}}'\{\mathbf{V}_{\lambda\lambda}(\hat{\boldsymbol{\psi}})\}^{-1}\hat{\boldsymbol{\lambda}}, \tag{3.11}$$

where $\hat{\boldsymbol{\lambda}}$ is the ML estimator of the Lagrange multipliers $\boldsymbol{\lambda}$ obtained under $\mathbf{h}(\boldsymbol{\theta}) = 0$ (see (A.15)), and $\mathbf{V}_{\lambda\lambda}(\hat{\boldsymbol{\psi}})$ is defined by (A.23). Then, using similar methods as used in the proof of Lemma 6 in Silvey (1959), it can be shown that $LM \overset{a}{\sim} \chi^2_{k-r^2}$.

4 An Empirical Application: Long-Run Estimates of Consumer Demand Equations for the UK

In this section we apply the long-run structural modelling techniques to Almost Ideal Demand Systems (AIDS) estimated for three non-durable expenditure categories using the UK quarterly observations over the period 1956(1)-1993(2). (The available observations before 1956 were used to create the necessary lagged variables). This application provides a good example where economic theory provides strong testable restrictions (such as homogeneity and symmetry) on the long-run equilibrium relations. The symmetry restrictions are of particular interest, since they provide an example of cross-equation restrictions.

Under the AIDS model of Deaton and Muellbauer (1980), the expenditure share of the i-th commodity group, w_{it}, is determined in the long run by

$$w_{it} = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln P_{jt} + \delta_i \ln(Y_t/P_t),\ i = 1, 2, ..., n,\ t = 1, 2, ..., T, \tag{4.1}$$

where P_{jt} is the price deflator of the commodity group j, Y_t is the per capita expenditure on all n commodities, and P_t is a general price index which we approximate using the Stone formula:[15] $\ln P_t = \sum_{j=1}^{n} w_{j0} \ln P_{jt}$, where w_{j0} refer to budget shares in the base year.

Consumer theory imposes the following restrictions on the parameters of the share equations:

- Adding-up restrictions: $\sum_{i=1}^{n} \alpha_i = 1,\ \sum_{i=1}^{n} \gamma_{ij} = 0,\ \sum_{i=1}^{n} \delta_i = 0.$
- Homogeneity restrictions: $\sum_{j=1}^{n} \gamma_{ij} = 0.$
- Symmetry restrictions: $\gamma_{ij} = \gamma_{ji}.$

The adding-up restrictions are not testable, and are imposed indirectly by first estimating the $n-1$ share equations, and then estimating the parameters of the remaining equation from the adding-up restrictions. In system estimation of the share equations the results are invariant to the choice of the $n-1$ commodities included in the analysis. Although there have been a number of attempts to deal with the dynamics of the AIDS model, these analyses invariably consider rather restricted set-ups, and all treat prices and real per capita expenditure as exogenously given.[16]

The long-run structural modelling approach of this paper considers the share equations, (4.1), as the long-run equilibrium relations of a VAR model in the $2n$ variables,

[15] The exact expression for $\ln P_t$ is given by (see Deaton and Muellbauer, 1980):

$$\ln P_t = \alpha_0 + \sum_{k=1}^{n} \alpha_k \ln P_{kt} + \frac{1}{2} \sum_{k=1}^{n} \sum_{j=1}^{n} \gamma_{kj} \ln P_{kt} \ln P_{jt}.$$

Its use in our work will give rise to a non-linear VAR model, the analysis of which is outside the scope of this paper. For an empirical application of Stone's approximation in a static AIDS model see Pashardes (1993).

[16] The most general dynamic model used is by Anderson and Blundell (1983), which is a VAR(1) in budget shares, and is estimated assuming exogenously given prices and per capita real expenditures.

$\mathbf{x}_t = (w_{1t}, w_{2t}, ..., w_{n-1,t}, \ln P_{1t},..., \ln P_{nt}, \ln(Y_t/P_t))$. This approach has two important advantages. Firstly, apart from the order of the VAR, it does not impose any arbitrary restrictions not supported by *a priori* theory on the short-run dynamics. Secondly, it allows for any possible interdependencies that may exist among the budget shares, prices, and the real per capita expenditure. This approach has, however, one important limitation: due to its highly data-intensive nature, only demand systems with a few commodity groups can be analyzed in a satisfactory manner. Here we estimate a three-commodity system on the UK quarterly seasonally adjusted data over the period 1956(1)-1993(2). The three commodity groups are (1) food, drink and tobacco; (2) services (including rents and rates); and (3) energy and other non-durables.[17]

Since the analysis of the cointegrated VAR model pre-assumes $\mathbf{x}_t$ to be I(1), we computed augmented Dickey-Fuller (1979) and Phillips-Perron (1988) statistics for the three budget shares (w_{1t}, w_{2t}, w_{3t}), the price variables ($\ln P_{1t}$, $\ln P_{2t}$, $\ln P_{3t}$), and the per capita real expenditure variable, $\ln(Y_t/P_t)$. The results are summarized in Table 1, and show that for none of the variables it is possible to reject the unit root hypothesis at the 95 percent level.[18]

Table 1 here

Consumer theory predicts that there should be two cointegrating relations among the six variables, w_{1t}, w_{2t}, $\ln P_{1t}$, $\ln P_{2t}$, $\ln P_{3t}$, and $\ln(Y_t/P_t)$. To test this hypothesis, in what follows, we consider a VAR(4) model with restricted intercepts and no trends to ensure that there exist steady state values for the budget shares both under the null and the alternative hypotheses. In this case, we have

$$\Delta \mathbf{x}_t = \sum_{i=1}^{3} \Gamma_i \Delta \mathbf{x}_{t-i} - \boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{x}^*_{t-1} + \boldsymbol{\varepsilon}_t, \tag{4.2}$$

where $\mathbf{x}^*_{t-1} = (\mathbf{x}'_{t-1}, 1)'$ is an $(m+1) \times 1$ vector, and $\boldsymbol{\beta}$ is an $(m+1) \times r$ matrix. The last row of $\boldsymbol{\beta}$ gives the steady state values of the budget shares. Using (4.2) we computed the log-likelihood Trace and Maximum eigenvalue statistics over the period 1956(1)-1993(2). The test results are summarized in Table 2.

Table 2 here

[17] Consumer expenditures at current and constant 1990 prices for the three commodity groups were taken from Central Statistical Office's (CSO) quarterly Macroeconomic Database. Quarterly observations on population were obtained by interpolation of annual population figures taken from the CSO Annual Database. Price indices of individual commodity groups were obtained as implicit deflators of relevant expenditure categories. The general price index was approximated by the Stone index. All the data were converted into indices with base equal to 1 in 1990. This ensures that the estimates of the α's in (4.1) are close to the budget shares in the base year.

[18] We also computed unit root statistics for all the variables not including trends in the underlying regressions, but could not reject the unit root hypothesis in any case. Since budget shares are bounded between zero and unity, it may be argued that the non-rejection of the unit root hypothesis is due to the relatively small sample used and the lack of power of unit root tests. Nevertheless, it seems reasonable to proceed assuming that the budget shares can be approximated as unit-root processes (see also Chambers and Nowman, 1994).

At the five percent significance level, the Trace statistic (λ_{trace}) suggests two cointegrating vectors, while the Maximum eigenvalue statistic ($\lambda_{\max}$) does not reject the hypothesis that there is only one cointegrating vector among the six variables. At the ten percent level, neither statistic rejects the hypothesis that there are two cointegrating vectors.[19]

Given the fact that the evidence against theory's prediction is rather weak we proceed assuming that $r = 2$. Denote the corresponding cointegrating vectors on w_{1t}, w_{2t}, $\ln P_{1t}$, $\ln P_{2t}$, $\ln P_{3t}$, $\ln(Y_t/P_t)$ and the intercept by $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{21}, \beta_{31}, \beta_{41}, \beta_{51}, \beta_{61}, \beta_{71})'$ and $\boldsymbol{\beta}_2 = (\beta_{12}, \beta_{22}, \beta_{32}, \beta_{42}, \beta_{52}, \beta_{62}, \beta_{72})'$, respectively. The exact (theory) identifying restrictions implicit in the specification of the share equations in the AIDS model are given by

$$H_E : \left\{ \begin{array}{cc} \beta_{11} = -1, & \beta_{12} = 0 \\ \beta_{21} = 0, & \beta_{22} = -1 \end{array} \right\},$$

and the exactly identified estimates of the two cointegrating vectors are[20]

$\tilde{\boldsymbol{\beta}}'_E =$

−1	0	.2734 (.0372)	−.2215 (.0235)	−.0516 (.0433)	−.1761 (.0303)	.2866 (.0027)
0	−1	−.1672 (.0637)	.0536 (.0395)	.1030 (.0754)	.3218 (.0521)	.5196 (.0044)

with the maximized value of the log-likelihood function being 3404.5, where the asymptotic standard errors are given in brackets. The estimates in the last column of $\tilde{\boldsymbol{\beta}}'_E$ correspond to the steady-state budget shares for the first two expenditure categories, namely "food, drink, and tobacco" and "services and rent".

Next, we provide tests of the homogeneity and symmetry restrictions, taking $\tilde{\boldsymbol{\beta}}_E$ as the appropriate exactly identified estimates. Estimation of the cointegrating relations subject to the homogeneity restrictions (namely, $\beta_{31} + \beta_{41} + \beta_{51} = 0$, and $\beta_{32} + \beta_{42} + \beta_{52} = 0$) yielded the following results:

$\tilde{\boldsymbol{\beta}}'_H =$

−1	0	.2722 (.0166)	−.2218 (.0208)	−.0504 (.0216)	−.1753 (.0164)	.2866 (.0028)
0	−1	−.1160 (.0289)	.0721 (.0358)	.0439 (.0378)	.2825 (.0281)	.5191 (.0048)

with the maximized value of the log-likelihood function being 3404.0. Therefore, the log-likelihood ratio statistic for testing the homogeneity hypothesis is equal to $2(3404.5 - 3304.0) = 1.0$, which is well below the 95 percent critical value of the Chi-Squared test with 2 degrees of freedom.

Turning to the symmetry hypothesis, the relevant restriction is the cross-equation restriction $\beta_{41} = \beta_{32}$. The estimates of the cointegrating vectors subject to the homogeneity

[19] We obtained similar results when we estimated lower order VAR models. We did not consider models of order higher than 4 on grounds of data limitations.

[20] The cointegrating vectors subject to the exactly- and over-identifying restrictions are estimated using a generalized Newton-Raphson algorithm. For details of the numerical algorithm see Pesaran and Shin (1994) and Pesaran and Pesaran (1997).

and symmetry restrictions are as follows:[21]

$$\tilde{\boldsymbol{\beta}}'_{HS} = \begin{array}{|c|c|c|c|c|c|c|}\hline -1 & 0 & \begin{array}{l}.2565\\(.0399)\end{array} & \begin{array}{l}-.2150\\(.0281)\end{array} & \begin{array}{l}-.0415\\(.0411)\end{array} & \begin{array}{l}-.1771\\(.0222)\end{array} & \begin{array}{l}.2848\\(.0066)\end{array} \\ \hline 0 & -1 & \begin{array}{l}-.2150\\(.0281)\end{array} & \begin{array}{l}.0977\\(.1111)\end{array} & \begin{array}{l}.1173\\(.1295)\end{array} & \begin{array}{l}.2837\\(.0955)\end{array} & \begin{array}{l}.5072\\(.0188)\end{array} \\ \hline \end{array},$$

with the maximized log-likelihood value of 3402.8. The LR statistic for testing this joint hypothesis is equal to 3.37, which is well below the 95 percent critical value of the Chi-Squared test with three degrees of freedom, and does not support a rejection of this joint hypothesis.

Finally, the estimates of the price and income elasticities for the specification that imposes both the homogeneity and symmetry restrictions are presented in Table 3.

Table 3 *here*

The income elasticities all have the correct signs and plausible magnitudes. The estimated price elasticities are generally reasonable, except for the own price elasticity of the "food, drink and tobacco" category which is slightly positive but statistically insignificant. Finally, the estimates of the restricted intercepts, given in the last row of $\hat{\boldsymbol{\beta}}_{HS}$, namely .285 and .507, for the w_1 and w_2 share equations match closely the average budget shares of .286 and .519 in the base year (1990).

5 Concluding Remarks

We have argued that in cointegrated VAR models where there is more than one cointegrating relation, the statistical approach to identification of the long run cointegrating relations advanced in the literature is not satisfactory, and as far as the interpretation of the results and their use in policy analysis is concerned, can be misleading. Identification of the long-run relations requires *a priori* information, in the form predicted by economic theory, market arbitrage conditions or institutional characteristics. When there are r cointegrating relations, there must be at least r independent *a priori* restrictions (including one normalization restriction) on each of the r cointegrating relations. This paper provides a general theory for the identification of the cointegrating vectors when they are subject to non-linear parametric restrictions. It gives a rigorous proof of the consistency of the ML estimators, establishes the relative rates of convergence of the ML estimators of the short-run and the long-run coefficients, and derives their asymptotic distribution; thus providing a formal proof for many of the results routinely used in the literature. The empirical application in the paper also shows that the econometric and computational methods advanced in the paper are readily applicable to a wide variety of applied economic problems.

[21] Using the adding-up condition the third (cointegrating) share equation is given by

$$\widehat{w}_{3t} = 0.2080 - .0415 \ln P_{1t} + 0.1173 \ln P_{2t} - .0758 \ln P_{3t} - 0.1066 \ln(Y_t/P_t).$$

Table 1: Unit Root Test Results over 1956(1)-1993(2)

	ADF(p)*					PP(ℓ)+				
Variables	0	1	2	3	4	1	2	3	4	5
w_1	-2.18	-1.91	-1.83	-1.90	-1.85	-1.81	-1.69	-1.73	-1.82	-2.00
w_2	-1.24	-1.10	-1.25	-1.26	-1.22	-1.05	-1.16	-1.21	-1.31	-1.37
w_3	-1.76	-1.21	-1.07	-0.99	-1.03	-1.01	-.077	-0.65	-0.60	-0.36
$\ln P_1$	-2.68	-2.12	-2.10	-2.18	-2.33	-2.58	-2.57	-2.60	-2.65	-2.70
$\ln P_2$	-1.93	-1.71	-1.74	-1.75	-2.09	-1.97	-2.07	-2.15	-2.26	-2.35
$\ln P_3$	-1.91	-1.79	-1.96	-2.16	-2.03	-2.02	-2.15	-2.27	-2.37	-2.46
$\ln(Y/P)$	-1.60	-1.54	-1.89	-1.94	-2.45	-1.53	-1.85	-1.99	-2.23	-2.41

* The ADF (augmented Dickey-Fuller) statistics are computed using the ADF(p), ($p = 0, 1, 2, 3, 4$) regressions containing intercepts and linear trends. + The PP (Phillips-Perron) statistics are computed using the AR(1) regression containing an intercept and a linear trend, where the Bartlett window is used in computing the long-run variance of the residual, and ℓ denotes the lag truncation parameter, $\ell = 1, 2, 3, 4, 5$. The 95% critical value for both statistics is -3.44.

Table 2: Johansen's Cointegration Rank Test Statistics for the AID System Applied to UK Non-Durable Consumption Expenditures over 1956(1)-1993(2)*

H_0	Eigenvalues	λ_{trace}		λ_{max}	
$r = 0$	.2404	119.61	[102.56]	41.24	[40.53]
$r = 1$	.1938	78.38	[75.98]	32.31	[34.40]
$r = 2$	.1074	46.06	[53.48]	17.04	[28.27]
$r = 3$	.0892	29.03	[34.87]	14.01	[22.04]
$r = 4$	.0615	15.02	[20.18]	9.52	[15.87]
$r = 5$	.0360	5.50	[9.16]	5.50	[9.16]

* λ_{trace} and $\lambda_{\max}$ are the trace and the maximum eigenvalue statistics, respectively. r is the number of cointegrating relations. These values are estimated using the VAR(4) model with restricted intercepts and no trends in the six variables, w_1, w_2, $\ln P_1$, $\ln P_2$, $\ln P_3$, and $\ln(Y/P)$. The values in $[\cdot]$ are the 95% critical values.

Table 3: Own and Cross Price Elasticities and Income Elasticities of Main Three Non-Durable Expediture Categories in the UK over 1956(1)-1993(2)*

	Price Elasticities			Income Elasticities
	Food	Services	Others	
Food	.0413 (.1384)	$-.4228$ (.1255)	$-.0216$ (0.1300)	.4042 (.0745)
Services	$-.5932$ (.0980)	-1.0902 (.3044)	.1211 (.2264)	1.5623 (.1892)
Others	$-.0495$ (.3221)	.8638 (.9020)	-1.2760 (.6766)	.4617 (.5609)

* Elasticities are estimated using the VAR(4) model with restricted intercepts and no trends, imposing homogeneity and symmetry restrictions, at the 1990 budget shares. Asymptotic standard errors are given in brackets.

Appendix A: Mathematical Derivations and Proofs

A.1 Derivation of First and Second Derivatives of the Log-Likelihood Function

First- and second-order total differentials of $\ell_T(\boldsymbol{\varphi})$, defined by (2.25), are given by

$$\begin{aligned} d\ell_T(\boldsymbol{\varphi}) &= \frac{1}{2}\text{Tr }\left\{\Omega^{-1}(\mathbf{E}'\mathbf{ME}-T\Omega)\Omega^{-1}(d\Omega)\right\} \\ &\quad -\text{Tr }\left\{\Omega^{-1}\mathbf{E}'\mathbf{M}[-\mathbf{Y}(d\Gamma)+\mathbf{X}_{-1}(d\boldsymbol{\beta})\boldsymbol{\alpha}'+\mathbf{X}_{-1}\boldsymbol{\beta}(d\boldsymbol{\alpha}')]\right\}, \end{aligned}$$

$$\begin{aligned} d^2\ell_T(\boldsymbol{\varphi}) &= \text{Tr}\left\{\Omega^{-1}(d\Omega)\Omega^{-1}\left(\mathbf{E}'\mathbf{ME}-T\Omega\right)\Omega^{-1}(d\Omega)\right\}-\frac{T}{2}\text{Tr}\left[\Omega^{-1}(d\Omega)\Omega^{-1}(d\Omega)\right] \\ &\quad +\text{Tr}\left\{\Omega^{-1}\left[-\mathbf{Y}(d\Gamma)+\mathbf{X}_{-1}(d\boldsymbol{\beta})\boldsymbol{\alpha}'+\mathbf{X}_{-1}\boldsymbol{\beta}(d\boldsymbol{\alpha}')\right]'\mathbf{ME}\Omega^{-1}(d\Omega)\right\} \\ &\quad +\text{Tr}\left\{\Omega^{-1}(d\Omega)\Omega^{-1}\mathbf{E}'\mathbf{M}\left[-\mathbf{Y}(d\Gamma)+\mathbf{X}_{-1}(d\boldsymbol{\beta})\boldsymbol{\alpha}'+\mathbf{X}_{-1}\boldsymbol{\beta}(d\boldsymbol{\alpha}')\right]\right\} \\ &\quad -\text{Tr}\Omega^{-1}\left\{(d\Gamma)'(\mathbf{Y}'\mathbf{MY})(d\Gamma)-(d\Gamma)'(\mathbf{Y}'\mathbf{MX}_{-1}\boldsymbol{\beta})(d\boldsymbol{\alpha}')-(d\Gamma)'(\mathbf{Y}'\mathbf{MX}_{-1})(d\boldsymbol{\beta})\boldsymbol{\alpha}'\right\} \\ &\quad -\text{Tr}\Omega^{-1}\left\{-(d\boldsymbol{\alpha})\left(\boldsymbol{\beta}'\mathbf{X}_{-1}'\mathbf{MY}\right)(d\Gamma)+(d\boldsymbol{\alpha})(\boldsymbol{\beta}'\mathbf{X}_{-1}'\mathbf{MX}_{-1}\boldsymbol{\beta})(d\boldsymbol{\alpha}')+(d\boldsymbol{\alpha})\left(\boldsymbol{\beta}'\mathbf{X}_{-1}'\mathbf{MX}_{-1}\right)(d\boldsymbol{\beta})\boldsymbol{\alpha}'\right\} \\ &\quad -\text{Tr}\Omega^{-1}\left\{-\boldsymbol{\alpha}(d\boldsymbol{\beta}')(\mathbf{X}_{-1}'\mathbf{MY})(d\Gamma)+\boldsymbol{\alpha}(d\boldsymbol{\beta}')(\mathbf{X}_{-1}'\mathbf{MX}_{-1}\boldsymbol{\beta})(d\boldsymbol{\alpha}')+\boldsymbol{\alpha}(d\boldsymbol{\beta}')(\mathbf{X}_{-1}'\mathbf{MX}_{-1})(d\boldsymbol{\beta})\boldsymbol{\alpha}'\right\}. \end{aligned}$$

Using Theorem 3 on p. 31 and Theorem 1 on p. 192 in Magnus and Neudecker (1988), and also noting that

$$dvec(\boldsymbol{\beta})=\mathbf{F}(\boldsymbol{\phi})d\boldsymbol{\phi},\; dvec(\Omega)=\mathbf{D}_m\boldsymbol{\omega},\; vec(\mathbf{E}'\mathbf{ME}-T\Omega)=\mathbf{D}_m vech(\mathbf{E}'\mathbf{ME}-T\Omega),$$

where $\mathbf{D}_m$ is an $m\times\frac{1}{2}m(m+1)$ duplication matrix, then the first- and the second-partial differentials of $\ell_T(\boldsymbol{\varphi})$ are given by

$$\frac{\partial\ell_T(\boldsymbol{\varphi})}{\partial\boldsymbol{\varphi}}=\begin{bmatrix}\partial\ell_T(\boldsymbol{\varphi})/\partial\boldsymbol{\gamma}\\ \partial\ell_T(\boldsymbol{\varphi})/\partial\boldsymbol{\kappa}\\ \partial\ell_T(\boldsymbol{\varphi})/\partial\boldsymbol{\phi}\\ \partial\ell_T(\boldsymbol{\varphi})/\partial\boldsymbol{\omega}\end{bmatrix}=\begin{bmatrix}\left(\Omega^{-1}\otimes\mathbf{I}_{m(p-1)}\right)vec(\mathbf{Y}'\mathbf{ME})\\ \left(\Omega^{-1}\otimes\mathbf{I}_r\right)vec(\boldsymbol{\beta}'\mathbf{X}_{-1}'\mathbf{ME})\\ \mathbf{F}'(\boldsymbol{\phi})(\boldsymbol{\alpha}'\Omega^{-1}\otimes\mathbf{I}_m)\;vec(\mathbf{X}_{-1}'\mathbf{ME})\\ \frac{1}{2}\mathbf{D}_m'(\Omega^{-1}\otimes\Omega^{-1})\mathbf{D}_m\;vech(\mathbf{E}'\mathbf{ME}-T\Omega)\end{bmatrix}, \tag{A.1}$$

$$-\frac{\partial^2\ell_T(\boldsymbol{\varphi})}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'}=\begin{bmatrix}\Lambda_{11} & \Lambda_{12}\mathbf{D}_m\\ \mathbf{D}_m'\Lambda_{12}' & \mathbf{D}_m'\Lambda_{22}\mathbf{D}_m\end{bmatrix}, \tag{A.2}$$

where Λ_{11} is the symmetric matrix

$$\Lambda_{11}=\begin{bmatrix}\Omega^{-1}\otimes\mathbf{Y}'\mathbf{MY} & \Omega^{-1}\otimes\mathbf{Y}'\mathbf{MX}_{-1}\boldsymbol{\beta} & (\Omega^{-1}\boldsymbol{\alpha}\otimes\mathbf{Y}'\mathbf{MX}_{-1})\mathbf{F}(\boldsymbol{\phi})\\ \cdot & \Omega^{-1}\otimes\boldsymbol{\beta}'\mathbf{X}_{-1}'\mathbf{MX}_{-1}\boldsymbol{\beta} & (\Omega^{-1}\boldsymbol{\alpha}\otimes\boldsymbol{\beta}'\mathbf{X}_{-1}'\mathbf{MX}_{-1})\mathbf{F}(\boldsymbol{\phi})\\ \cdot & \cdot & \mathbf{F}'(\boldsymbol{\phi})(\boldsymbol{\alpha}'\Omega^{-1}\boldsymbol{\alpha}\otimes\mathbf{X}_{-1}'\mathbf{MX}_{-1})\mathbf{F}(\boldsymbol{\phi})\end{bmatrix},$$

$$\Lambda_{12}=\begin{bmatrix}\Omega^{-1}\otimes\mathbf{Y}'\mathbf{ME}\Omega^{-1}\\ \Omega^{-1}\otimes\boldsymbol{\beta}\mathbf{X}_{-1}'\mathbf{ME}\Omega^{-1}\\ (\Omega^{-1}\otimes\mathbf{X}_{-1}'\mathbf{ME}\Omega^{-1})\mathbf{F}(\boldsymbol{\phi})\end{bmatrix},$$

$$\Lambda_{22}=\frac{T}{2}\left\{(\Omega^{-1}\otimes\Omega^{-1})+(\Omega^{-1}\otimes\Omega^{-1})(T^{-1}\mathbf{E}'\mathbf{ME}-\Omega)\Omega^{-1}\right\}.$$

A.2 Derivation of Probability Limits Involving Unit Root Processes

Under Assumption 2.4 and using the multivariate invariance principle (see Phillips and Durlauf,1986)

$$T^{-\frac{1}{2}}\mathbf{s}_{[Ta]}\Rightarrow\mathbf{w}(a),\; a\in[0,1], \tag{A.5}$$

where $\mathbf{s}_{[Ta]} = \sum_{j=1}^{[Ta]} \boldsymbol{\varepsilon}_j$, $[Ta]$ is the smallest integer part of Ta, and $\mathbf{w}(a)$ is an $m \times 1$ vector of Brownian motion with the covariance matrix, Ω_0. Next, applying the continuous mapping theorem,

$$T^{-\frac{3}{2}} \sum_{t=1}^{T} \mathbf{s}_t \Rightarrow \int_0^1 \mathbf{w}(a)da, \ T^{-2} \sum_{t=1}^{T} \mathbf{s}_t \mathbf{s}_t' \Rightarrow \int_0^1 \mathbf{w}(a)\mathbf{w}'(a)da. \tag{A.6}$$

From (2.10) we have (see also (2.7))

$$\mathbf{X}_{-1} = \boldsymbol{\tau}(\mathbf{x}_0 - \boldsymbol{\mu})' + \mathbf{t}\boldsymbol{\mu}' + \mathbf{S}_{-1}\mathbf{C}_0(1)' + \sum_{i=0}^{\infty} \mathbf{E}_{-i-1}\mathbf{C}_{0i}^{*\prime},$$

where $\mathbf{S}_{-1} = (\mathbf{s}_0, \mathbf{s}_1, ..., \mathbf{s}_{T-1})'$ and $\mathbf{E}_{-i} = (\boldsymbol{\varepsilon}_{1-i}, \boldsymbol{\varepsilon}_{2-i}, ..., \boldsymbol{\varepsilon}_{T-i})'$, $i = 1, 2, ...$. Then,

$$\begin{aligned} \mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1} &= \mathbf{C}_0(1)\mathbf{S}_{-1}'\mathbf{M}\mathbf{S}_{-1}\mathbf{C}_0'(1) + \sum_{i=0}^{\infty} \mathbf{C}_0(1)\mathbf{S}_{-1}'\mathbf{M}\mathbf{E}_{-i-1}\mathbf{C}_{0i}^{*\prime} \\ &+ \sum_{i=0}^{\infty} \mathbf{C}_{0i}^{*}\mathbf{E}_{-i-1}'\mathbf{M}\mathbf{S}_{-1}\mathbf{C}_0'(1) + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \mathbf{C}_{0i}^{*}\mathbf{E}_{-i-1}'\mathbf{M}\mathbf{E}_{-j-1}\mathbf{C}_{0j}^{*\prime}, \end{aligned}$$

where $\mathbf{M} = \mathbf{I}_T - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ (see also (2.23)). Then, it is easily seen that

$$T^{-2}\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1} = T^{-2}\mathbf{C}_0(1)\mathbf{S}_{-1}'\mathbf{M}\mathbf{S}_{-1}\mathbf{C}_0'(1) + O_p(T^{-1}), \ T^{-1}\mathbf{X}_{-1}'\mathbf{M}\mathbf{E} = T^{-1}\mathbf{C}_0(1)\mathbf{S}_{-1}'\mathbf{M}\mathbf{E} + o_p(1). \tag{A.7}$$

Defining $\mathbf{D}_Z = diag\left[T^{-\frac{1}{2}}, T^{-\frac{3}{2}}\right]$, and using (A.5) and (A.6), as $T \to \infty$,

$$T^{-1}\mathbf{S}_{-1}'\mathbf{Z}\mathbf{D}_Z \Rightarrow \left[\int_0^1 \mathbf{w}(a)da, \int_0^1 a\mathbf{w}(a)da\right],$$

$$T^{-2}\mathbf{S}_{-1}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{S}_{-1} \Rightarrow \left[\int_0^1 \mathbf{w}(a)da, \int_0^1 a\mathbf{w}(a)da\right] \begin{bmatrix} 4 & -6 \\ -6 & 12 \end{bmatrix} \begin{bmatrix} \int_0^1 \mathbf{w}'(a)da \\ \int_0^1 a\mathbf{w}'(a)da \end{bmatrix},$$

$$T^{-2}\mathbf{S}_{-1}'\mathbf{M}\mathbf{S}_{-1} \Rightarrow \mathbf{Q}_{SS}, \ T^{-1}\mathbf{S}_{-1}'\mathbf{M}\mathbf{E} \Rightarrow \int_0^1 \mathbf{w}^*(a)d\mathbf{w}'(a), \tag{A.8}$$

where

$$\mathbf{Q}_{SS} = \int_0^1 \mathbf{w}^*(a)\mathbf{w}^{*\prime}(a)da \tag{A.9}$$

is a positive definite random matrix with probability 1 (see Phillips, 1991), and $\mathbf{w}^*(a) = \mathbf{w}(a) + (6a-4)\int_0^1 \mathbf{w}(a)da + (-12a+6)\int_0^1 a\mathbf{w}(a)da$ is an $m \times 1$ vector of demeaned and detrended Brownian motion with the covariance matrix, Ω_0. Using the above results in (A.7) , as $T \to \infty$,

$$T^{-2}\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1} \Rightarrow \mathbf{C}_0(1)\mathbf{Q}_{SS}\mathbf{C}_0'(1), \ T^{-1}\mathbf{X}_{-1}'\mathbf{M}\mathbf{E} \Rightarrow \mathbf{C}_0(1)\left\{\int_0^1 \mathbf{w}^*(a)d\mathbf{w}'(a)\right\}. \tag{A.10}$$

Finally, noting that $\Delta\mathbf{Y}$ and $\boldsymbol{\beta}_0'\mathbf{X}_{-1}$ are stationary and uncorrelated with $\mathbf{E}$, it is also easily verified that

$$T^{-1}\mathbf{Y}'\mathbf{M}\mathbf{E} = o_p(1), \ T^{-1}\boldsymbol{\beta}_0\mathbf{X}_{-1}'\mathbf{M}\mathbf{E} = o_p(1), T^{-1}\mathbf{Y}'\mathbf{M}\mathbf{X}_{-1} = O_p(1), \ T^{-1}\boldsymbol{\beta}_0'\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1} = O_p(1). \tag{A.11}$$

Furthermore, the following probability limits can be shown to exist:

$$\mathbf{Q}_{yy} = P\lim_{T\to\infty} \frac{\mathbf{Y}'\mathbf{M}\mathbf{Y}}{T}, \ \mathbf{Q}_{y\beta_0} = P\lim_{T\to\infty} \frac{\mathbf{Y}'\mathbf{M}\mathbf{X}_{-1}\boldsymbol{\beta}_0}{T}, \ \mathbf{Q}_{\beta_0\beta_0} = P\lim_{T\to\infty} \frac{\boldsymbol{\beta}_0'\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1}\boldsymbol{\beta}_0}{T}. \tag{A.12}$$

A.3 Stochastic Equicontinuity of $\Im_T(\boldsymbol{\varphi})$

Let $\boldsymbol{\rho} = (\boldsymbol{\gamma}', \boldsymbol{\kappa}', \boldsymbol{\omega}')'$ and define the open ball, $B(\boldsymbol{\rho}, \delta) = \{\boldsymbol{\rho}_* \in \Upsilon_\rho : \|\boldsymbol{\rho}_* - \boldsymbol{\rho}\| < \delta\}$, with $\Upsilon_\rho = \Upsilon_\gamma \times \Upsilon_\kappa \times \Upsilon_\omega$, and the open shrinking ball, $N_T(\boldsymbol{\phi}, \delta) = \{\boldsymbol{\phi}_* \in \Upsilon_\phi : T^{1/2}\|\boldsymbol{\phi}_* - \boldsymbol{\phi}\| < \delta\}$, where δ is a positive real number.[22] The sample information matrix, $\Im_T(\boldsymbol{\varphi})$, is given by

$$\Im_T(\boldsymbol{\varphi}) = \mathbf{D}_T \frac{-\partial^2 \ell_T(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'} \mathbf{D}_T, \tag{A.13}$$

where $\mathbf{D}_T = T^{-\frac{1}{2}} diag\left(\mathbf{I}_{m^2(p-1)}, \mathbf{I}_{mr}, T^{-\frac{1}{2}}\mathbf{I}_s, \mathbf{I}_{m(m+1)/2}\right)$ and $-\partial^2 \ell_T(\boldsymbol{\varphi})/\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'$ is defined by (A.2). $\Im_T(\boldsymbol{\varphi})$ satisfies Saikkonen's (1995) stochastic equicontinuity condition SE_0 if

$$\sup_{\boldsymbol{\varphi}_* \in B(\boldsymbol{\rho}_0, \delta) \times N_T(\boldsymbol{\phi}_0, \delta)} \|\Im_T(\boldsymbol{\varphi}_*) - \Im_T(\boldsymbol{\varphi}_0)\| = O_p(1). \tag{A.14}$$

In the case where the long-run coefficients $\boldsymbol{\phi}_0$ (or $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\phi}_0)$) are known the sample information matrix of the short-run coefficients $\boldsymbol{\rho}$ involve only stationary variables and the standard asymptotic theory is applicable. Therefore, in what follows we shall focus on establishing the condition SE_0 for those components of $\Im_T(\boldsymbol{\varphi})$ that involve the long-run coefficients, $\boldsymbol{\phi}$. These are $T^{-1}\mathbf{Y}'\mathbf{M}\mathbf{X}_{-1}\boldsymbol{\beta}$, $T^{-1}\boldsymbol{\beta}'\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1}\boldsymbol{\beta}$, $(\Omega^{-1}\boldsymbol{\alpha} \otimes T^{-3/2}\boldsymbol{\beta}'\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1})\mathbf{F}(\boldsymbol{\phi})$ and $\mathbf{F}'(\boldsymbol{\phi})(\boldsymbol{\alpha}'\Omega^{-1}\boldsymbol{\alpha} \otimes T^{-2}\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1})\mathbf{F}(\boldsymbol{\phi})$. Consider the first term. Denoting $\boldsymbol{\beta}(\boldsymbol{\phi}_*)$ and $\boldsymbol{\beta}(\boldsymbol{\phi}_0)$ by $\boldsymbol{\beta}_*$ and $\boldsymbol{\beta}_0$, respectively, and using the Lipschitz condition (2.55) we have

$$\sup_{\boldsymbol{\phi}_* \in N_T(\boldsymbol{\phi}_0, \delta)} \|T^{-1}\mathbf{Y}'\mathbf{M}\mathbf{X}_{-1}\boldsymbol{\beta}_* - T^{-1}\mathbf{Y}'\mathbf{M}\mathbf{X}_{-1}\boldsymbol{\beta}_0\| < c_\beta \delta \|T^{-3/2}\mathbf{Y}'\mathbf{M}\mathbf{X}\|.$$

But using (A.11), $T^{-3/2}\mathbf{Y}'\mathbf{M}\mathbf{X} = o_p(1)$, and hence $T^{-1}\mathbf{Y}'\mathbf{M}\mathbf{X}_{-1}\boldsymbol{\beta}$ clearly satisfies the SE_0 condition.

To prove the stochastic equicontinuity of $T^{-1}\boldsymbol{\beta}'\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1}\boldsymbol{\beta}$, let $\mathbf{Q}_{T,XX} = T^{-2}\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1}$ and note that

$$T(\boldsymbol{\beta}_*'\mathbf{Q}_{T,XX}\boldsymbol{\beta}_* - \boldsymbol{\beta}_0'\mathbf{Q}_{T,XX}\boldsymbol{\beta}_0) = T(\boldsymbol{\beta}_* - \boldsymbol{\beta}_0)'\mathbf{Q}_{T,XX}\boldsymbol{\beta}_0 + T\boldsymbol{\beta}_0'\mathbf{Q}_{T,XX}(\boldsymbol{\beta}_* - \boldsymbol{\beta}_0) + T(\boldsymbol{\beta}_* - \boldsymbol{\beta}_0)'\mathbf{Q}_{T,XX}(\boldsymbol{\beta}_* - \boldsymbol{\beta}_0).$$

Hence,

$$\sup_{\boldsymbol{\phi}_* \in N_T(\boldsymbol{\phi}_0, \delta)} \|T\boldsymbol{\beta}_*\mathbf{Q}_{T,XX}\boldsymbol{\beta}_* - T\boldsymbol{\beta}_0\mathbf{Q}_{T,XX}\boldsymbol{\beta}_0\| < 2c_\beta \delta \|T^{-3/2}\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1}\boldsymbol{\beta}_0\| + c_\beta^2 \delta^2 \|T^{-2}\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1}\|.$$

Again using (A.11), $T^{-3/2}\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1}\boldsymbol{\beta}_0 = o_p(1)$, and $\mathbf{Q}_{T,XX} = T^{-2}\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1} = O_p(1)$. Therefore, $T^{-1}\boldsymbol{\beta}'\mathbf{X}_{-1}'\mathbf{M}\mathbf{X}_{-1}\boldsymbol{\beta}$ satisfies the SE_0, in the sense that the above supremum can be made as small as desired by choosing a small enough value for δ.

Let $\Psi_* = (\Omega^{-1}\boldsymbol{\alpha} \otimes T^{1/2}\boldsymbol{\beta}_*'\mathbf{Q}_{T,XX})$, $\Psi_0 = (\Omega^{-1}\boldsymbol{\alpha} \otimes T^{1/2}\boldsymbol{\beta}_0'\mathbf{Q}_{T,XX})$, $\mathbf{F}_* = \mathbf{F}(\boldsymbol{\phi}_*)$, and $\mathbf{F}_0 = \mathbf{F}(\boldsymbol{\phi}_0)$. Then

$$\|\Psi_*\mathbf{F}_* - \Psi_0\mathbf{F}_0\| \le \|\Psi_* - \Psi_0\| \times \|\mathbf{F}_* - \mathbf{F}_0\| + \|\Psi_* - \Psi_0\| \times \|\mathbf{F}_0\| + \|\Psi_0\| \times \|\mathbf{F}_* - \mathbf{F}_0\|,$$

$$\|\Psi_* - \Psi_0\| = \|\Omega^{-1}\boldsymbol{\alpha}\| \times \|T^{1/2}(\boldsymbol{\beta}_* - \boldsymbol{\beta}_0)'\mathbf{Q}_{T,XX}\|,$$

and

$$\|\Psi_0\| = \|\Omega^{-1}\boldsymbol{\alpha}\| \times \|T^{1/2}\boldsymbol{\beta}_0'\mathbf{Q}_{T,XX}\|.$$

Hence

$$\sup_{\boldsymbol{\phi}_* \in N_T(\boldsymbol{\phi}_0, \delta)} \|\Psi_* - \Psi_0\| < c_\beta \delta \|\Omega^{-1}\boldsymbol{\alpha}\| \times \|\mathbf{Q}_{T,XX}\|.$$

Similarly, using (2.56)

$$\sup_{\boldsymbol{\phi}_* \in N_T(\boldsymbol{\phi}_0, \delta)} \|\mathbf{F}_* - \mathbf{F}_0\| < T^{-1/2} c_F \delta.$$

[22] Notice that in the construction of $B(\boldsymbol{\rho}, \delta)$ and $N_T(\boldsymbol{\phi}, \delta)$ the same δ is used as required by Saikkonen's (1995, p. 894) stochastic equicontinuity condition, SE_0.

Hence, under Assumption 2.6 we have

$$\sup_{\phi_*\in N_T(\phi_0,\delta)} \|\Psi_*\mathbf{F}_* - \Psi_0\mathbf{F}_0\| < \delta\|\Omega^{-1}\boldsymbol{\alpha}\| \left\{ c_\beta\|\mathbf{Q}_{T,XX}\| \times \|\mathbf{F}_0\| + c_F\|T^{-2}\mathbf{X}'_{-1}\mathbf{M}\mathbf{X}_{-1}\boldsymbol{\beta}_0\| + T^{-1/2}c_\beta c_F\delta\|\mathbf{Q}_{T,XX}\| \right\}.$$

Now using (A.11), and recalling that under our assumptions $\|\Omega^{-1}\boldsymbol{\alpha}\|$ and $\|\mathbf{F}_0\|$ are bounded, it follows that

$$\sup_{\phi_*\in N_T(\phi_0,\delta)} \|(\Omega^{-1}\boldsymbol{\alpha}\otimes T^{1/2}\boldsymbol{\beta}'_*\mathbf{Q}_{T,XX})\mathbf{F}(\boldsymbol{\phi}_*) - (\Omega^{-1}\boldsymbol{\alpha}\otimes T^{1/2}\boldsymbol{\beta}'_0\mathbf{Q}_{T,XX})\mathbf{F}(\boldsymbol{\phi}_0)\|$$

is bounded by an $O_p(1)$ variable and therefore $(\Omega^{-1}\boldsymbol{\alpha}\otimes T^{1/2}\boldsymbol{\beta}'\mathbf{Q}_{T,XX})\mathbf{F}(\boldsymbol{\phi})$ satisfies the SE_0 condition.

Finally, for the term $\mathbf{F}'(\boldsymbol{\phi})(\boldsymbol{\alpha}'\Omega^{-1}\boldsymbol{\alpha}\otimes T^{-2}\mathbf{X}'_{-1}\mathbf{M}\mathbf{X}_{-1})\mathbf{F}(\boldsymbol{\phi})$, we first note that

$$\begin{aligned}
& \mathbf{F}'(\boldsymbol{\phi}_*)(\mathbf{A}\otimes\mathbf{Q}_{T,XX})\mathbf{F}(\boldsymbol{\phi}_*) - \mathbf{F}'(\boldsymbol{\phi}_0)(\mathbf{A}\otimes\mathbf{Q}_{T,XX})\mathbf{F}(\boldsymbol{\phi}_0) \\
= & \left[\mathbf{F}(\boldsymbol{\phi}_*) - \mathbf{F}(\boldsymbol{\phi}_0)\right]'(\mathbf{A}\otimes\mathbf{Q}_{T,XX})\mathbf{F}(\boldsymbol{\phi}_0) + \mathbf{F}(\boldsymbol{\phi}_0)'(\mathbf{A}\otimes\mathbf{Q}_{T,XX})\left[\mathbf{F}(\boldsymbol{\phi}_*) - \mathbf{F}(\boldsymbol{\phi}_0)\right] \\
& + \left[\mathbf{F}(\boldsymbol{\phi}_*) - \mathbf{F}(\boldsymbol{\phi}_0)\right]'(\mathbf{A}\otimes\mathbf{Q}_{T,XX})\left[\mathbf{F}(\boldsymbol{\phi}_*) - \mathbf{F}(\boldsymbol{\phi}_0)\right],
\end{aligned}$$

where $\mathbf{A} = \boldsymbol{\alpha}'\Omega^{-1}\boldsymbol{\alpha}$ belongs to a compact set. Hence

$$\begin{aligned}
& \sup_{\phi_*\in N_T(\phi_0,\delta)} \|\mathbf{F}'(\boldsymbol{\phi}_*)(\mathbf{A}\otimes\mathbf{Q}_{T,XX})\mathbf{F}(\boldsymbol{\phi}_*) - \mathbf{F}'(\boldsymbol{\phi}_0)(\mathbf{A}\otimes\mathbf{Q}_{T,XX})\mathbf{F}(\boldsymbol{\phi}_0)\| \\
< \; & 2T^{-1/2}c_F\delta\|\mathbf{A}\otimes\mathbf{Q}_{T,XX}\| \times \|\mathbf{F}(\boldsymbol{\phi}_0)\| + T^{-1}c_F^2\delta^2\|\mathbf{A}\otimes\mathbf{Q}_{T,XX}\|.
\end{aligned}$$

It then readily follows that $\mathbf{F}'(\boldsymbol{\phi})(\boldsymbol{\alpha}'\Omega^{-1}\boldsymbol{\alpha}\otimes T^{-2}\mathbf{X}'_{-1}\mathbf{M}\mathbf{X}_{-1})\mathbf{F}(\boldsymbol{\phi})$ also satisfies the SE_0 condition.

A.4 Proof of Theorem 3.1

Using the Lagrangean function,

$$\Lambda(\boldsymbol{\psi},\boldsymbol{\lambda}) = \ell_T(\boldsymbol{\psi}) - T\boldsymbol{\lambda}'\mathbf{h}(\boldsymbol{\theta}), \tag{A.15}$$

where $\boldsymbol{\lambda}$ is a $k\times 1$ vector of the Lagrange multipliers, the constrained ML estimators, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\lambda}}$, satisfy the following first order-conditions:

$$\frac{\partial\ell_T(\hat{\boldsymbol{\psi}})}{\partial\boldsymbol{\rho}} = \mathbf{0},\; \frac{\partial\ell_T(\hat{\boldsymbol{\psi}})}{\partial\boldsymbol{\theta}} - T\mathbf{H}'(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\lambda}} = \mathbf{0},\; \mathbf{h}(\hat{\boldsymbol{\theta}}) = \mathbf{0}. \tag{A.16}$$

Using the mean-value expansion of $\partial\ell_T(\hat{\boldsymbol{\psi}})/\partial\boldsymbol{\psi}$ around $\boldsymbol{\psi}_0$, we have

$$\begin{bmatrix} \partial\ell_T(\hat{\boldsymbol{\psi}})/\partial\boldsymbol{\rho} \\ \partial\ell_T(\hat{\boldsymbol{\psi}})/\partial\boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \partial\ell_T(\boldsymbol{\psi}_0)/\partial\boldsymbol{\rho} \\ \partial\ell_T(\boldsymbol{\psi}_0)/\partial\boldsymbol{\theta} \end{bmatrix} - \begin{bmatrix} -\partial^2\ell_T(\bar{\boldsymbol{\psi}})/\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}' & -\partial^2\ell_T(\bar{\boldsymbol{\psi}})/\partial\boldsymbol{\rho}\partial\boldsymbol{\theta}' \\ -\partial^2\ell_T(\bar{\boldsymbol{\psi}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\rho}' & -\partial^2\ell_T(\bar{\boldsymbol{\psi}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' \end{bmatrix} \begin{bmatrix} (\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0) \\ (\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \end{bmatrix}, \tag{A.17}$$

where the (i,j) element of $(-\partial^2\ell_T(\bar{\boldsymbol{\psi}})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}')$ is evaluated at $(\bar{\psi}_i,\bar{\psi}_j)$, and $\bar{\psi}_i$ is a convex combination of $\hat{\psi}_i$ and ψ_{i0}. In view of the consistency results in Theorem 2.1 and the stochastic equicontinuity for the sample information matrix proved in Section A.3 above, we have

$$\Im_T(\bar{\boldsymbol{\psi}}) = \mathbf{D}_{\psi T}\frac{-\partial^2\ell_T(\bar{\boldsymbol{\psi}})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\mathbf{D}_{\psi T} \Rightarrow \Im(\boldsymbol{\psi}_0), \tag{A.18}$$

where $\Im(\boldsymbol{\psi}_0)$ is defined by (3.4). Similarly

$$\mathbf{H}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) = \mathbf{h}(\hat{\boldsymbol{\theta}}) - \mathbf{h}(\boldsymbol{\theta}_0) + o_p(1). \tag{A.19}$$

Notice that under the null hypothesis $\mathbf{h}(\boldsymbol{\theta}_0) = \mathbf{0}$. Using the first-order conditions (A.16) in (A.17) and (A.19), and then using (A.18) we obtain (after some algebra)

$$\begin{bmatrix} \Im_{\rho\rho}(\boldsymbol{\psi}_0) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Im_{\theta\theta}(\boldsymbol{\psi}_0) & \mathbf{H}'(\boldsymbol{\theta}_0) \\ \mathbf{0} & \mathbf{H}(\boldsymbol{\theta}_0) & \mathbf{0} \end{bmatrix} \begin{bmatrix} \sqrt{T}(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0) \\ T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \mathbf{d}(\boldsymbol{\rho}_0) \\ \mathbf{d}(\boldsymbol{\theta}_0) \\ \mathbf{0} \end{bmatrix} + o_p(1), \tag{A.20}$$

Because $\Im_{\theta\theta}(\boldsymbol{\psi}_0)$ is singular, a direct manipulation of (A.20) is not possible. However, this problem can be overcome following Silvey's (1959) approach. Since $\mathbf{H}(\boldsymbol{\theta}_0)T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=o_p(1)$, and therefore $\mathbf{H}_A(\boldsymbol{\theta}_0)T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=o_p(1)$, we can rewrite (A.20) as

$$\begin{bmatrix} \Im_{\rho\rho}(\boldsymbol{\psi}_0) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{\theta\theta}(\boldsymbol{\psi}_0) & \mathbf{H}'(\boldsymbol{\theta}_0) \\ \mathbf{0} & \mathbf{H}(\boldsymbol{\theta}_0) & \mathbf{0} \end{bmatrix} \begin{bmatrix} \sqrt{T}(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0) \\ T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \mathbf{d}(\boldsymbol{\rho}_0) \\ \mathbf{d}(\boldsymbol{\theta}_0) \\ \mathbf{0} \end{bmatrix} + o_p(1), \tag{A.21}$$

where $\mathbf{J}_{\theta\theta}(\boldsymbol{\psi}_0)=\Im_{\theta\theta}(\boldsymbol{\psi}_0)+\mathbf{H}_A'(\boldsymbol{\theta}_0)\mathbf{H}_A(\boldsymbol{\theta}_0)$ is a positive definite matrix with probability 1. Note that $\mathbf{H}_A(\boldsymbol{\theta}_0)$ has rank r^2, and the rank of $\Im_{\theta\theta}(\boldsymbol{\psi}_0)$ is equal to $mr-r^2$, with probability 1. Therefore,

$$\begin{bmatrix} \sqrt{T}(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0) \\ T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \Im_{\rho\rho}^{-1}(\boldsymbol{\psi}_0) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{\theta\theta}(\boldsymbol{\psi}_0) & \mathbf{V}_{\theta\lambda}(\boldsymbol{\psi}_0) \\ \mathbf{0} & \mathbf{V}_{\theta\lambda}'(\boldsymbol{\psi}_0) & \mathbf{V}_{\lambda\lambda}(\boldsymbol{\psi}_0) \end{bmatrix} \begin{bmatrix} \mathbf{d}(\boldsymbol{\rho}_0) \\ \mathbf{d}(\boldsymbol{\theta}_0) \\ \mathbf{0} \end{bmatrix} + o_p(1), \tag{A.22}$$

where $\mathbf{V}_{\theta\theta}(\boldsymbol{\psi}_0)$ is defined by (3.8) and

$$\mathbf{V}_{\theta\lambda}(\boldsymbol{\psi}_0)=\mathbf{J}_{\theta\theta}^{-1}(\boldsymbol{\psi}_0)\mathbf{H}'(\boldsymbol{\theta}_0)\left\{\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{J}_{\theta\theta}^{-1}(\boldsymbol{\psi}_0)\mathbf{H}'(\boldsymbol{\theta}_0)\right\}^{-1},\ \mathbf{V}_{\lambda\lambda}(\boldsymbol{\psi}_0)=-\left\{\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{J}_{\theta\theta}^{-1}(\boldsymbol{\psi}_0)\mathbf{H}'(\boldsymbol{\theta}_0)\right\}^{-1}. \tag{A.23}$$

Then, (3.7) readily follows from (3.3) and (3.4). The expression for the covariance matrix in (3.8) can also be easily obtained using the following results:

$$\mathbf{V}_{\theta\theta}(\boldsymbol{\psi}_0)\mathbf{H}'(\boldsymbol{\theta}_0)=\mathbf{0},\ \mathbf{V}_{\theta\theta}(\boldsymbol{\psi}_0)\Im_{\theta\theta}(\boldsymbol{\psi}_0)\mathbf{V}_{\theta\theta}(\boldsymbol{\psi}_0)=\mathbf{V}_{\theta\theta}(\boldsymbol{\psi}_0)\left[\mathbf{I}_{mr}-\mathbf{H}'(\boldsymbol{\theta}_0)\mathbf{V}_{\theta\lambda}(\boldsymbol{\psi}_0)\right]=\mathbf{V}_{\theta\theta}(\boldsymbol{\psi}_0),$$

which can be derived from inversion of the bordered matrix in (A.21). (For details see Appendix B of Pesaran and Shin,1994.) ■

A.5 Proof of Theorem 3.2

Using a similar procedure as in the proof of Theorem 3.1 above, we have

$$\sqrt{T}(\tilde{\boldsymbol{\rho}}-\boldsymbol{\rho}_0)\overset{a}{\sim}MN\left\{\mathbf{0},\Im_{\rho\rho}^{-1}(\boldsymbol{\psi}_0)\right\},\ T(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\overset{a}{\sim}MN\left\{\mathbf{0},\mathbf{V}_{\theta\theta}^{A}(\boldsymbol{\psi}_0)\right\}, \tag{A.24}$$

where

$$\mathbf{V}_{\theta\theta}^{A}(\boldsymbol{\psi}_0)=\mathbf{J}_{\theta\theta}^{-1}(\boldsymbol{\psi}_0)-\mathbf{J}_{\theta\theta}^{-1}(\boldsymbol{\psi}_0)\mathbf{H}_A'(\boldsymbol{\theta}_0)\{\mathbf{H}_A(\boldsymbol{\theta}_0)\mathbf{J}_{\theta\theta}^{-1}(\boldsymbol{\psi}_0)\mathbf{H}_A'(\boldsymbol{\theta}_0)\}^{-1}\mathbf{H}_A(\boldsymbol{\theta}_0)\mathbf{J}_{\theta\theta}^{-1}(\boldsymbol{\psi}_0), \tag{A.25}$$

is a random matrix, having rank $mr-r^2$ with probability 1.

Also in view of the consistency results in Theorem 2.1 and the stochastic equicontinuity results established in Section A.3, we are justified to write down the following Taylor series approximations of $\ell_T(\hat{\boldsymbol{\psi}})$ and $\ell_T(\tilde{\boldsymbol{\psi}})$ around $\boldsymbol{\psi}_0$:

$$\ell_T(\hat{\boldsymbol{\psi}})=\ell_T(\boldsymbol{\psi}_0)+\mathbf{d}'(\boldsymbol{\rho}_0)\sqrt{T}(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0)+\mathbf{d}'(\boldsymbol{\theta}_0)T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \tag{A.26}$$

$$-\frac{1}{2}\left\{\sqrt{T}(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0)'\Im_{\rho\rho}(\boldsymbol{\psi}_0)\sqrt{T}(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0)+T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\Im_{\theta\theta}(\boldsymbol{\psi}_0)T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\right\}+o_p(1),$$

and

$$\ell_T(\tilde{\boldsymbol{\psi}})=\ell_T(\boldsymbol{\psi}_0)+\mathbf{d}'(\boldsymbol{\rho}_0)\sqrt{T}(\tilde{\boldsymbol{\rho}}-\boldsymbol{\rho}_0)+\mathbf{d}'(\boldsymbol{\theta}_0)T(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \tag{A.27}$$

$$-\frac{1}{2}\left\{\sqrt{T}(\tilde{\boldsymbol{\rho}}-\boldsymbol{\rho}_0)'\Im_{\rho\rho}(\boldsymbol{\psi}_0)\sqrt{T}(\tilde{\boldsymbol{\rho}}-\boldsymbol{\rho}_0)+T(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\Im_{\theta\theta}(\boldsymbol{\psi}_0)T(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\right\}+o_p(1).$$

Using (3.7) in (A.26), and (A.24) in (A.27), substituting the results in (3.6), now yields

$$LR_T=\mathbf{d}'(\boldsymbol{\theta}_0)\left[\mathbf{V}_{\theta\theta}^{A}(\boldsymbol{\psi}_0)-\mathbf{V}_{\theta\theta}(\boldsymbol{\psi}_0)\right]\mathbf{d}(\boldsymbol{\theta}_0)+o_p(1). \tag{A.28}$$

Next, let $\mathbf{P} = \mathbf{I}_k - \mathbf{H}_*(\mathbf{H}_*'\mathbf{H}_*)^{-1}\mathbf{H}_*'$ and $\mathbf{P}_A = \mathbf{I}_k - \mathbf{H}_{A*}(\mathbf{H}_{A*}'\mathbf{H}_{A*})^{-1}\mathbf{H}_{A*}'$, where $\mathbf{H}_* = \mathbf{J}_{\theta\theta}^{-\frac{1}{2}}(\boldsymbol{\psi}_0)\mathbf{H}'(\boldsymbol{\theta}_0)$ and $\mathbf{H}_{A*} = \mathbf{J}_{\theta\theta}^{-\frac{1}{2}}(\boldsymbol{\psi}_0)\mathbf{H}_A'(\boldsymbol{\theta}_0)$. Then, using (3.8) and (A.25), we have

$$\mathbf{V}_{\theta\theta}(\boldsymbol{\psi}_0) = \mathbf{J}_{\theta\theta}^{-\frac{1}{2}}(\boldsymbol{\psi}_0)\mathbf{P}\mathbf{J}_{\theta\theta}^{-\frac{1}{2}}(\boldsymbol{\psi}_0),\ \mathbf{V}_{\theta\theta}^{A}(\boldsymbol{\psi}_0) = \mathbf{J}_{\theta\theta}^{-\frac{1}{2}}(\boldsymbol{\psi}_0)\mathbf{P}_A\mathbf{J}_{\theta\theta}^{-\frac{1}{2}}(\boldsymbol{\psi}_0).$$

Substituting these results in (A.28) we have

$$LR = \mathbf{u}'(\mathbf{P}_A - \mathbf{P})\mathbf{u} + o_p(1), \tag{A.29}$$

where $\mathbf{u} = \mathbf{J}_{\theta\theta}^{-\frac{1}{2}}(\boldsymbol{\psi}_0)\mathbf{d}(\theta_0)$. Using (3.3) and (2.54) it is then easily seen that $\mathbf{u} \overset{a}{\sim} N\{\mathbf{0}, \Sigma_u\}$, where $\Sigma_u = \mathbf{J}_{\theta\theta}^{-\frac{1}{2}}(\boldsymbol{\psi}_0)\Im_{\theta\theta}(\boldsymbol{\psi}_0)\mathbf{J}_{\theta\theta}^{-\frac{1}{2}}(\boldsymbol{\psi}_0) = \mathbf{I}_{mr} - \mathbf{H}_A'(\boldsymbol{\theta}_0)\mathbf{H}_A(\boldsymbol{\theta}_0)$. Notice that Σ_u is a non-stochastic matrix with rank $mr - r^2$. Then, by Theorem 9.21 of Rao and Mitra (1971, p. 171), it follows that the quadratic form, $\mathbf{u}'(\mathbf{P}_A - \mathbf{P})\mathbf{u}$, is χ^2 distributed with degrees of freedom equal to $\mathrm{Tr}[(\mathbf{P}_A - \mathbf{P})\Sigma_u]$ if and only if $\Sigma_u(\mathbf{P}_A - \mathbf{P})\Sigma_u(\mathbf{P}_A - \mathbf{P})\Sigma_u = \Sigma_u(\mathbf{P}_A - \mathbf{P})\Sigma_u$, or $[\Sigma_u(\mathbf{P}_A - \mathbf{P})]^3 = [\Sigma_u(\mathbf{P}_A - \mathbf{P})]^2$. Now note that

$$\Sigma_u(\mathbf{P}_A - \mathbf{P}) = \{\mathbf{I}_{mr} - \mathbf{H}_A'(\boldsymbol{\theta}_0)\mathbf{H}_A(\boldsymbol{\theta}_0)\}(\mathbf{P}_A - \mathbf{P})$$

$$= \mathbf{P}_A - \mathbf{P} - \mathbf{H}_A'(\boldsymbol{\theta}_0)\mathbf{H}_A(\boldsymbol{\theta}_0)\mathbf{P}_A + \mathbf{H}_A'(\boldsymbol{\theta}_0)\mathbf{H}_A(\boldsymbol{\theta}_0)\mathbf{P} = \mathbf{P}_A - \mathbf{P}.$$

Since $\mathbf{P}_A - \mathbf{P}$ is an idempotent matrix, we also have

$$[\Sigma_u(\mathbf{P} - \mathbf{P}_A)]^2 = [\Sigma_u(\mathbf{P} - \mathbf{P}_A)]^3 = \mathbf{P} - \mathbf{P}_A.$$

Therefore, the quadratic term in (A.29) is asymptotically χ^2 distributed with degrees of freedom equal to $\mathrm{Tr}[(\mathbf{P}_A - \mathbf{P})\Sigma_u] = \mathrm{Tr}(\mathbf{P}_A - \mathbf{P}) = k - r^2$. ■

References

[1] Aitchison, J. and S.D. Silvey (1958): "Maximum Likelihood Estimation of Parameters Subject to Restraints," *Annals of Mathematical Statistics*, 29, 813-828.

[2] Anderson, G. and R. Blundell (1983): "Testing Restrictions in a Flexible Dynamic Demand System: An Application to Consumer's Expenditure in Canada," *Review of Economic Studies*, 50, 397-410.

[3] Andrews, D.W.K. (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465-71.

[4] Andrews, D.W.K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 4, 458-467.

[5] Breusch, T.S. (1986): "Hypothesis Testing in Unidentified Models," *Review of Economic Studies*, 53, 635-651.

[6] Chambers, M.J. and K.B. Nowman (1994): "Forecasting with the Almost Ideal System," Discussion Paper Series No. 427, University of Essex.

[7] Davidson, J. (1994): *Stochastic Limit Theory*. Oxford: Oxford University Press.

[8] Davidson, R. and J.G. MacKinnon (1993): *Estimation and Inference in Econometrics*. Oxford: Oxford University Press.

[9] Deaton, A. and J. Muellbauer (1980): "An Almost Ideal Demand System," *American Economic Review*, 70, 312-336.

[10] Dickey, D.A. and W. A. Fuller (1979): "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 74, 427-431.

[11] Engle, R.F. and C.W.J. Granger (1987): "Cointegration and Error Correction Representation: Estimation and Testing," *Econometrica*, 55, 251-276.

[12] Jeganathan, P. (1982): "On the Asymptotic Theory of Estimation When the Limit of Log-Likelihood Ratios is Mixed Normal," *Sankya*, Series A, 44, 173-212.

[13] Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

[14] Johansen, S. (1991): "Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551-80.

[15] Johansen, S. (1995): "Identifying Restrictions of Linear Equations with Applications to Simultaneous Equations and Cointegration," *Journal of Econometrics*, 69, 111-132.

[16] Lütkepohl, H. (1991): *Introduction to Multiple Time Series Analysis.* Berlin: Springer Verlag.

[17] Magnus, J.R. and H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: John Wiley.

[18] Pashardes, P. (1993): "Bias in Estimating the Almost Ideal Demand System with the Stone Index Approximation," *Economic Journal*, 103, 908-915.

[19] Pesaran, M.H. (1997): "The Role of Economic Theory in Modelling the Long-Run," *The Economic Journal,* 107, 178-191.

[20] Pesaran, M.H. and B. Pesaran (1997): *Microfit 4.0: An Interactive Econometric Software Package.* Oxford: Oxford University Press.

[21] Pesaran, M.H. and Y. Shin (1994): "Long-Run Structural Modelling," *DAE Working Papers Series*, No. 9419, University of Cambridge.

[22] Pesaran, M.H., Y. Shin and R.J. Smith (1998): "Structural Analysis of Vector Autoregressive Models with Exogenous I(1) Variables," *DAE Working Papers Amalgamated Series* No. 9706, University of Cambridge. Revised November 1998. http://www.econ.cam.ac.uk/faculty/pesaran/

[23] Phillips, P.C.B. (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283-306.

[24] Phillips, P.C.B. (1995): "Fully Modified Least Squares and Vector Autoregressions," *Econometrica*, 63, 1023-1078.

[25] Phillips, P.C.B. and S. Durlauf (1986): "Multiple Time Series with Integrated Variables," *Review of Economic Studies,* 53, 473-496.

[26] Phillips, P.C.B. and P. Perron (1988): "Testing For Unit Roots in Time Series Regression," *Biometrika*, 75, 335-46.

[27] Pöstcher, B.M. and I.R. Prucha (1994): "Generic Uniform Convergence and Equicontinuity Concepts for Random Functions. An Exploration of the Basic Structure," *Journal of Econometrics*, 60, 23-63.

[28] Rao, C.R. and S.K. Mitra (1971): *Generalized Inverse of Matrices and their Applications*. New York: John Wiley.

[29] Rothenberg, T.J. (1971): "Identification in Parametric Models," *Econometrica*, 39, 577-592.

[30] Saikkonen, P. (1993a): "Estimation of Cointegration Vectors with Linear Restrictions," *Econometric Theory*, 9, 19-35.

[31] Saikkonen, P. (1993b): "Continuous Weak Convergence and Stochastic Equicontinuity Results for Integrated Processes with an Application to the Estimation of a Regression Model," *Econometric Theory,* 9, 155-188.

[32] Saikkonen, P. (1995): "Problems with the Asymptotic Theory of Maximum Likelihood Estimation in Integrated and Cointegrated Systems," *Econometric Theory,* 11, 888-911.

[33] Silvey, S.D. (1959): "The Lagrange Multiplier Test," *Annals of Mathematical Statistics*, 30, 389-407.

[34] Stock, J.H. (1987): "Asymptotic Properties of the Least Squares Estimators of the Cointegrating Vectors," *Econometrica*, 55, 1035-1056.

[35] Watson, M.W. (1994): "Vector Autoregression and Cointegration," in *Handbook of Econometrics*, Volume IV ed. by D. MacFadden and R.F. Engle. Amsterdam: North-Holland, 843-915.