Testing General Restrictions on the Cointegrating Space[1]

BY JURGEN A. DOORNIK
Nuffield College, Oxford OX1 1NF, UK

May 3, 1995

ABSTRACT

This paper considers identification and estimation of cointegration analysis under general restrictions on the cointegrating space. A solution is found to the identification problem and made operational. An algorithm for estimation is discussed. Some examples illustrate.

Key words: Cointegration; General restrictions; Identification; Numerical Rank; Singular value decomposition.

[1] All comments are welcome. Please direct them to Jurgen Doornik at econjad@vax.ox.ac.uk.

Testing General Restrictions on the Cointegrating Space

BY JURGEN A. DOORNIK
Nuffield College, Oxford OX1 1NF, UK

May 3, 1995

SUMMARY

This paper considers identification and estimation of cointegration analysis under general restrictions on the cointegrating space. A solution is found to the identification problem and made operational. An algorithm for estimation is discussed. Some examples illustrate.

Key words: Cointegration; General restrictions; Identification; Numerical Rank; Singular value decomposition.

1 INTRODUCTION

Consider the vector autoregressive (VAR) model

$$\boldsymbol{y}_t = \sum_{i=1}^{m} \boldsymbol{\pi}_i \boldsymbol{y}_{t-i} + \boldsymbol{v}_t, \boldsymbol{v}_t \sim \mathsf{IN}(\boldsymbol{0}, \boldsymbol{\Omega}), \tag{1}$$

consisting of n endogenous variables $\boldsymbol{y}_t$. In error correction form this is written as

$$\Delta \boldsymbol{y}_t = \boldsymbol{P}\boldsymbol{y}_{t-1} + \boldsymbol{v}_t, \tag{2}$$

setting $m = 1$ for simplicity. When $\boldsymbol{y}_t$ is integrated of order 1, I(1), then $\Delta \boldsymbol{y}_t$ is I(0) and $\boldsymbol{P}$ cannot be full rank. We may then write $\boldsymbol{P} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times p$ matrices of rank p, and $\boldsymbol{\beta}' \boldsymbol{y}_t$ must comprise p cointegrating I(0) relations inducing the restricted I(0) representation:

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha} \left(\boldsymbol{\beta}' \boldsymbol{y}_{t-1}\right) + \boldsymbol{v}_t. \tag{3}$$

Estimating the rank of $\boldsymbol{P}$ forms the cointegration analysis developed by Johansen (1988) (now well established, see e.g. Banerjee, Dolado, Galbraith and Hendry, 1993, or Johansen, 1995). The standard solution method for estimating the cointegrating rank involves a generalized eigenproblem.[1]

Conditional on a selected cointegrating rank, denoted by p here, it is of interest to test additional hypotheses about the cointegrating space. A range of tests of increasing complexity has been proposed by Johansen and Juselius. These are labelled $\mathsf{H}_a \ldots \mathsf{H}_e$ in the overview presented in Table 1. Note that the first two tests on $\boldsymbol{\beta}$ could easily be combined with restrictions on $\boldsymbol{\alpha}$ of type H_a.

It is important to know which constraints have explicit solutions (here: H_b and H_c, also when in conjunction with H_a). Moreover, it is useful to know that, where we need to use an iterative method, this will converge to the right solution, as is shown in Doornik (1995) for H_d and H_e. The drawback, however, is that each set requires separate implementation, while not exhausting all interesting hypotheses. Also, it is often more convenient to specify the restrictions directly on the elements of $\boldsymbol{\beta}$, rather than setting up

[1] Doornik (1995) shows that a numerical more stable solution avoids the Choleski decomposition.

Table 1. Summary of restrictions on cointegrating space.

	hypothesis	reference
H_a	$\boldsymbol{\alpha} = \boldsymbol{B}_1\boldsymbol{\theta}_1$	Johansen (1991), Johansen and Juselius (1990)
H_b	$\boldsymbol{\beta}_2 = \boldsymbol{H}_2\boldsymbol{\phi}_2$	Johansen (1988, 1991), Johansen and Juselius (1990)
H_c	$\boldsymbol{\beta}_3 = (\boldsymbol{H}_3 : \boldsymbol{\phi}_3)$	Johansen and Juselius (1992)
H_d	$\boldsymbol{\beta}_4 = (\boldsymbol{H}_4\boldsymbol{\varphi} : \boldsymbol{\psi})$	Johansen and Juselius (1992)
H_e	$\boldsymbol{\beta}_5 = (\boldsymbol{H}_1\boldsymbol{\varphi}_1 \ldots \boldsymbol{H}_p\boldsymbol{\varphi}_p)$	Johansen (1993), Johansen and Juselius (1994)
H_f	$\boldsymbol{\beta}_6 = \boldsymbol{f}_\beta(\boldsymbol{\theta}), \boldsymbol{\alpha} = \boldsymbol{f}_\alpha(\boldsymbol{\theta})$	Doornik and Hendry (1994), this paper

$\boldsymbol{H}$ matrices. In this paper, I discuss general restrictions (possibly non-linear) on the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ coefficients, labelled H_f. The method proposed here uses direct maximization of the concentrated likelihood function, based on an alternating quasi-Newton method, and has been successfully implemented in the computer program PcFiml (see Doornik and Hendry, 1994).

Ideally, a computer implementation would be as follows:

(1) Let the user specify restrictions of any type;
(2) Let the computer program analyse these restrictions, and select the most appropriate solution method.

In this implementation, the user need not have a thorough understanding of each test. The computer would recognize that the specified restrictions are, for example, of type H_b, and use the method appropriate for that type. The solution offered in PcFiml is only halfway towards the ideal implementation: users choose between an explicit method ($\mathsf{H}_a \cap \mathsf{H}_b$ or $\mathsf{H}_a \cap \mathsf{H}_c$), where they have to specify the $\boldsymbol{B}$ and $\boldsymbol{H}$ matrices, or the general method where constraints are expressed directly in terms of unconstrained parameters.

This paper addresses the two issues involved in testing general restrictions: how to estimate the restricted alpha and beta matrices, and how to verify identification of the selected specification. The latter is directly related to counting the number of free parameters to implement likelihood ratio tests. For $\mathsf{H}_b \ldots \mathsf{H}_e$ this can be derived directly, but for H_f it is more troublesome.

Table 2. Number of constraints under various restrictions of the cointegrating space.

	$\boldsymbol{\beta}$	$\boldsymbol{\alpha}$	restriction count
H_m	$\boldsymbol{\beta}$	$\boldsymbol{\alpha}$	0
H_a		$\boldsymbol{B}\boldsymbol{\theta}$	$(n-b)\,p$
H_b	$\boldsymbol{H}\boldsymbol{\phi}$	$\boldsymbol{\alpha}$	$p\,(n-s)$
$\mathsf{H}_b \cap \mathsf{H}_a$	$\boldsymbol{H}\boldsymbol{\phi}$	$\boldsymbol{B}\boldsymbol{\theta}$	$(n-b)\,p + p\,(n-s)$
H_c	$(\boldsymbol{H} : \boldsymbol{\phi})$	$\boldsymbol{\alpha}$	$s\,(n-p)$
$\mathsf{H}_c \cap \mathsf{H}_a$	$(\boldsymbol{H} : \boldsymbol{\phi})$	$\boldsymbol{B}\boldsymbol{\theta}$	$p\,(n-b) + s\,(n-p)$
H_d	$(\boldsymbol{H}\boldsymbol{\varphi} : \boldsymbol{\psi})$	$\boldsymbol{\alpha}$	$r_1\,(n-p+r_1-s)$
H_e	$(\boldsymbol{H}_1\boldsymbol{\varphi}_1 \ldots \boldsymbol{H}_p\boldsymbol{\varphi}_p)$	$\boldsymbol{\alpha}$	if identified: $\sum_{i=1}^{p}(n-p+1-s_i)$
H_f	$\boldsymbol{f}_\beta(\boldsymbol{\theta})$	$\boldsymbol{f}_\alpha(\boldsymbol{\theta})$	see §2

Table 3. Relevant matrix dimensions for various restrictions of the cointegrating space.

		dimensions		dimensions
H_m : unconstrained	$\boldsymbol{\alpha}$	$(n \times p)$	$\boldsymbol{\beta}$	$(n \times p)$
H_a	$\boldsymbol{\theta}$	$(b \times p), b \geq p$	$\boldsymbol{B}$	$(n \times b)$
H_b	$\boldsymbol{\phi}$	$(s \times p), s \geq p$	$\boldsymbol{H}$	$(n \times s)$
H_c	$\boldsymbol{\phi}$	$(n \times (p-s))$	$\boldsymbol{H}$	$(n \times s)$
H_d	$\boldsymbol{\varphi}$	$(s \times r_1), r_1 \leq s$	$\boldsymbol{\psi}$	$(n \times (p-r_1)), r_1 \leq p$
H_e	$\boldsymbol{\varphi}_i$	$(s_i \times 1)$	$\boldsymbol{H}$	$(n \times s_i)$

Table 2 gives the number of restrictions involved in the hypotheses considered by Johansen and Juselius. The relevant dimensions are given in Table 3. It is clear that restrictions on $\boldsymbol{\alpha}$ of the form $\boldsymbol{B}\boldsymbol{\theta}$ keep the identification burden fully on $\boldsymbol{\beta}$ (it does not restrict the $\boldsymbol{A}$ matrix in any way). The case H_e is treated in Johansen (1993) and Johansen and Juselius (1994), and the discussion is closely linked to the concept of identification. To find the degrees of freedom under H_e: check if the restrictions are identifying, if they are, then the degrees of freedom are computed as the sum of the overidentifying restrictions imposed on each cointegrating vector. If they are not identifying the situation is less simple. Moreover, constraints on $\boldsymbol{\alpha}$ (such as exclusion restrictions) may help to identify $\boldsymbol{\beta}$, thus making the situation more complex. So for H_f, I adopt a different approach.

Finally, I wish to point out that an intermediate form is considered by Boswijk (1994) (linear restrictions on $\mathrm{vec}\,\boldsymbol{\beta}$ with exclusion restrictions on $\mathrm{vec}\,\boldsymbol{\alpha}$).

2 Identification under general restrictions

Conditions for identification in the reduced rank vector error correction model under general restrictions will be somewhat weaker than those available for within-equation restrictions only. There, we may verify identification without reference to any parameter value, thus obtaining a global result (except for a set of measure zero). In the general case non-linear constraints are allowed and identification can only be checked locally for any parameter point. Under well-behaved constraints, however, the verification of identification at a random point in the restricted parameter space corresponds to global identification with probability one. The following proposition solves the identification problem under general parameter restrictions.

Proposition 1 (**Identification under general restrictions**): *A sufficient condition for the parameter value $\boldsymbol{\theta}_0$ in the model*

$$\boldsymbol{y}_t = \boldsymbol{P}(\boldsymbol{\theta})\, \boldsymbol{w}_t + \boldsymbol{v}_t, \quad \boldsymbol{P}(\boldsymbol{\theta}) = \boldsymbol{\alpha}(\boldsymbol{\theta})\boldsymbol{\beta}(\boldsymbol{\theta})', \quad \boldsymbol{v}_t \sim \mathsf{IN}(\boldsymbol{0}, \boldsymbol{\Omega}_v), \tag{4}$$

with

$$\mathrm{rank}(\boldsymbol{P}) = p, \quad \mathrm{rank}(\boldsymbol{\Omega}) = n_0, \quad \mathrm{rank}(\boldsymbol{W}) = n_1, \quad \boldsymbol{\Omega} \text{ is } n_0 \times n_0, \quad \boldsymbol{W} \text{ is } T \times n_1,$$

to be locally identified is that the $n_0 n_1 \times r$ matrix

$$\mathcal{J}' = \frac{\partial\, \mathrm{vec}\, \boldsymbol{P}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

has full column rank r at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$; with $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ respectively $n_0 \times p$, $n_1 \times p$ and $r \times 1$ matrices. It is also assumed that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are twice continuously differentiable functions of the coefficients $\boldsymbol{\theta}$.

PROOF. Identification of the statistical model (4) means that the joint probability distribution which generates data according to that model is uniquely determined by specifying the parameter values in the model. By requiring that $\boldsymbol{\Omega}$ and $\boldsymbol{W}$ are full rank, we can limit attention to the remaining parameters in the model. So identification requires that we can uniquely solve $\boldsymbol{P}'(\boldsymbol{\theta}_0) - \boldsymbol{P}'(\boldsymbol{\theta}) = \mathbf{0}$ (see Rothenberg, 1971, p.587 or Magnus and Neudecker, 1988, p.337). A sufficient condition for this is that the (transposed) Jacobian matrix $\partial \operatorname{vec} \boldsymbol{P}'(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ has full column rank when evaluated at $\boldsymbol{\theta}_0$.□

Application of Proposition 1 leads us to check the rank of

$$\mathcal{J}' = \frac{\partial \operatorname{vec} \boldsymbol{P}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = (\boldsymbol{I}_{n_0} \otimes \boldsymbol{\beta}(\boldsymbol{\theta}_0) : \boldsymbol{\alpha}(\boldsymbol{\theta}_0) \otimes \boldsymbol{I}_{n_1})\, \boldsymbol{J}', \tag{5}$$

where

$$\boldsymbol{J}' = \frac{\partial \left\{ \left[\operatorname{vec} \left(\boldsymbol{\alpha}(\boldsymbol{\theta}_0)' \right) \right]' : \left[\operatorname{vec} \boldsymbol{\beta}(\boldsymbol{\theta}_0) \right]' \right\}'}{\partial \boldsymbol{\theta}'}$$

is an $(n_0 p + p n_1) \times r$ matrix.

We can show that the approach taken by Boswijk (1994) (again applying results of Rothenberg, 1971) leads to the same condition. There, local identification of the value $\hat{\boldsymbol{\theta}}$ is obtained if the asymptotic information matrix:

$$\mathcal{I} = T \boldsymbol{J} \begin{pmatrix} \boldsymbol{I}_{n_0} \otimes \boldsymbol{\beta}' \\ \boldsymbol{\alpha}' \otimes \boldsymbol{I}_{n_1} \end{pmatrix} (\boldsymbol{\Omega}_v^{-1} \otimes \boldsymbol{S}_{11}) \begin{pmatrix} \boldsymbol{I}_{n_0} \otimes \boldsymbol{\beta} & \boldsymbol{\alpha} \otimes \boldsymbol{I}_{n_1} \end{pmatrix} \boldsymbol{J}', \tag{6}$$

is full rank; $\boldsymbol{S}_{11} = T^{-1} \boldsymbol{W}' \boldsymbol{W}$. Since the term in the middle is a positive definite $n_0 n_1 \times n_0 n_1$ matrix under the assumption that $\boldsymbol{W}$ and $\boldsymbol{\Omega}$ are full rank, it is sufficient to check that $\mathcal{J}'$ has rank r.

To make Proposition 1 operational, we need to be able to determine the rank of a matrix in a computer implementation. For this purpose we define the concept of numerical rank:

DEFINITION 1 (**Numerical rank**): *The rank of an $m \times n$ matrix $\boldsymbol{A}$, $m \geq n$, is determined by the number of singular values w_i satisfying:*

$$w_i > 10^4 \epsilon_m ||A||_\infty \equiv \epsilon_w ,$$

with ϵ_m the machine accuracy for double precision ($\approx 2 \times 10^{-16}$) and

$$||A||_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|.$$

The singular values can be found using the singular value decomposition. Deciding which singular values are small depends on the precision of the floating point arithmetic; the tolerance here is based on the IEEE standard for 8-byte floating point arithmetic, such as implemented on IBM compatible computers.

There is still a choice to be made, as the rank could be computed in three ways:

(1) using random values, uniform on $(0, 1)$, for $\boldsymbol{\theta}_0$ in equation (5);
(2) using the maximum likelihood estimates $\hat{\boldsymbol{\theta}}$ in (5);
(3) using the maximum likelihood estimates $\hat{\boldsymbol{\theta}}$ in the information matrix (6).

In numerical terms, verification of (5) is expected to be somewhat more reliable than (6), because the latter will be contaminated by the singular values of $\boldsymbol{\Omega}_v^{-1}$ and $\boldsymbol{S}_{11}$. Also, the use of a random parameter value comes closest to establishing global identification, because the probability of finding a $\boldsymbol{\theta}_0$ which

leads to a lower rank than $\hat{\boldsymbol{\theta}}$ is essentially zero. On the other hand, it is possible that the converged $\hat{\boldsymbol{\theta}}$ has a lower rank than $\boldsymbol{\theta}_0$, for example if some parameters go towards zero or infinity or if a restriction exacerbates a near singularity in the data (with the singularity causing local underidentification, rather than being caused by a generic lack of identification). This suggests that $\hat{\boldsymbol{\theta}}$ should not be used for generic identification, but that some information needs to be available for underidentification owing to singularity at the maximum.

The results may be summarized as follows.

ALGORITHM 1 (**Numerical identification under general restrictions**): *In a VAR of dimension n_0 and cointegrating rank p under general restrictions on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, define:*

s $\quad \operatorname{rank}\left[\mathcal{J}'\left(\boldsymbol{\theta}_0\right)\right]$, computed using Definition 1,
r $\quad$ the number of freely estimated parameters,

for a randomly chosen parameter value $\boldsymbol{\theta}_0$ and $\mathcal{J}$ given in (5). So s is the observed rank of $\mathcal{J}$, and r the dimension of $\boldsymbol{\theta}$. There is global identification with probability one if:

$$s = r,$$

that is, there are no zero singular values in $\mathcal{J}$. Following maximum likelihood estimation using Algorithm 2, which yields $\hat{\boldsymbol{\theta}}$, and when the model is identified, the 'variance' of the estimated parameters may be reported as

$$\widehat{\mathsf{V}\left[\hat{\boldsymbol{\theta}}\right]} = \frac{T}{T - r/n_0}\mathcal{I}^*\left(\hat{\boldsymbol{\theta}}\right)^{-1}, \quad r = s,$$

where $\mathcal{I}^*$ is given in (7) below, and r/n_0 uses integer division.[2] *Regardless of identification, the degrees of freedom of the* LR *test for restrictions on the cointegrating space is given by*

$$n_0 p + p n_1 - p^2 - s, \quad s = \operatorname{rank}\left[\mathcal{J}'\left(\boldsymbol{\theta}_0\right)\right],$$

when $n_1 - n_0$ parameters on non-modelled variables are restricted to lie in the cointegrating space.

$\mathcal{J}'$ is an $n_0 n_1 \times r$ matrix, and we can have $n_0 n_1 \geq r$ as well as $n_0 n_1 < r$ (e.g. when no restrictions are imposed at all: $r = n_0 p + p n_1$). In the latter case it is more efficient to compute the singular value decomposition of $\mathcal{J}$ instead. The LR test can be applied regardless of the rank of $\mathcal{J}'$, which we denote by s to distinguish it from the identification condition (there is identification if s equals r, but there are restrictions imposed if $s < n_0 p + p n_1 - p^2$).

As discussed in Johansen (1991) (also see Johansen, 1995), the distribution of $\hat{\boldsymbol{\beta}}$ is mixed normal, i.e. the variance matrix is stochastic. The discussion there also indicates that inference on $\hat{\boldsymbol{\beta}}$ may be done as if $\boldsymbol{\alpha}$ is known, and vice versa. In line with this, the 'variance' is computed as (i.e. setting the off-diagonal blocks to zero before applying the Jacobian matrix):

$$\mathcal{I}^* = (T - r/n)\,\boldsymbol{J}\begin{pmatrix} \boldsymbol{\Omega}_v^{-1} \otimes (\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\alpha}'\boldsymbol{\Omega}_v^{-1}\boldsymbol{\alpha} \otimes \boldsymbol{S}_{11} \end{pmatrix}\boldsymbol{J}', \tag{7}$$

where $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{J}$ are functions of $\hat{\boldsymbol{\theta}}$. If the inversion of $\mathcal{I}^*(\hat{\boldsymbol{\theta}})$ or $\mathcal{I}(\hat{\boldsymbol{\theta}})$ fails, there is a local singularity at $\hat{\boldsymbol{\theta}}$. Section 4 gives some examples, and illustrates that the random parameter value approach is the preferred method.

[2] The scale factor is in line with all reported variances in PcFiml by being scaled by $T - k$ rather than T.

3 ESTIMATION UNDER GENERAL RESTRICTIONS

Again assume a cointegrating rank p. The hypothesis is expressed as:

$$\mathsf{H}_f : \{\boldsymbol{\alpha} = \boldsymbol{f}_\alpha(\boldsymbol{\theta})\} \cap \{\boldsymbol{\beta} = \boldsymbol{f}_\beta(\boldsymbol{\theta})\}, \tag{8}$$

$\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are expressed as a function of the unrestricted elements $\boldsymbol{\theta}$ in $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$. This function may be non-linear, and even link $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Consider an example with $n = 4$, $p = 2$:

$$\boldsymbol{\alpha} = \begin{pmatrix} \varphi_0 & \varphi_1 \\ \varphi_2 & \varphi_3 \\ \varphi_4 & \varphi_5 \\ \varphi_6 & \varphi_7 \end{pmatrix}, \quad \boldsymbol{\beta}' = \begin{pmatrix} \varphi_8 & \varphi_9 & \varphi_{10} & \varphi_{11} \\ \varphi_{12} & \varphi_{13} & \varphi_{14} & \varphi_{15} \end{pmatrix}. \tag{9}$$

A possible restriction we may wish to test is $\varphi_6 = 0, \varphi_7 = 0, \varphi_{10} = \varphi_{11}, \varphi_{14} = \varphi_{15}$. Another could be $\varphi_6\varphi_{11} + \varphi_7\varphi_{15} = 0$ which could be expressed as $\varphi_6 = -\varphi_7\varphi_{15}/\varphi_{11}$; this type of restriction is considered by Hunter (1992) and Mosconi and Giannini (1992) (also see Toda and Phillips, 1991, who note the need for care in testing when some parameters vanish from the hypothesis when others are zero). In the first example we map from

$$\begin{pmatrix} \varphi_0 & \varphi_1 & \varphi_2 & \varphi_3 & \varphi_4 & \varphi_5 & \varphi_6 & \varphi_7 & : & \varphi_8 & \varphi_9 & \varphi_{10} & \varphi_{11} & \varphi_{12} & \varphi_{13} & \varphi_{14} & \varphi_{15} \end{pmatrix}$$

to

$$\begin{pmatrix} \varphi_0 & \varphi_1 & \varphi_2 & \varphi_3 & \varphi_4 & \varphi_5 & 0 & 0 & : & \varphi_8 & \varphi_9 & \varphi_{11} & \varphi_{11} & \varphi_{12} & \varphi_{13} & \varphi_{15} & \varphi_{15} \\ \theta_0 & \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 & & & : & \theta_6 & \theta_7 & & \theta_8 & \theta_9 & \theta_{10} & & \theta_{11} \end{pmatrix},$$

where $\boldsymbol{\theta}$ is the vector of parameters in the constrained likelihood, and $\boldsymbol{\varphi} = \left[(\operatorname{vec}\boldsymbol{\alpha}')' : (\operatorname{vec}\boldsymbol{\beta})'\right]'$ are the unconstrained parameters. The constrained parameters are partitioned according to alpha and beta: $\boldsymbol{\theta}' = \left(\boldsymbol{\theta}'_\alpha : \boldsymbol{\theta}'_\beta\right)$ which in the first example equals: $(\theta_0 \ldots \theta_5 : \theta_6 \ldots \theta_{11})$. The function $\boldsymbol{f}_\beta(\boldsymbol{\theta})$ maps from $\boldsymbol{\theta}$ to $\boldsymbol{\beta}$, whereas we shall write $\boldsymbol{f}_\beta^{-1}(\boldsymbol{\beta})$ to map from a $\boldsymbol{\beta}$ vector to $\boldsymbol{\theta}$ (i.e. just applying the constraints; only if no constraints are imposed will it hold that $\boldsymbol{f}_\beta\left(\boldsymbol{f}_\beta^{-1}(\boldsymbol{\beta})\right) = \boldsymbol{\beta}$).

The likelihood to maximize is (see the Appendix):

$$\ell_c(\boldsymbol{\theta}) \propto -\log\left|(\boldsymbol{I}_n : -\boldsymbol{\alpha}(\boldsymbol{\theta})\boldsymbol{\beta}(\boldsymbol{\theta})')\begin{pmatrix} \boldsymbol{S}_{00} & \boldsymbol{S}_{01} \\ \boldsymbol{S}_{10} & \boldsymbol{S}_{11} \end{pmatrix}\begin{pmatrix} \boldsymbol{I}_n \\ -\boldsymbol{\beta}(\boldsymbol{\theta})\boldsymbol{\alpha}(\boldsymbol{\theta})' \end{pmatrix}\right|, \tag{10}$$

where $\ell_c(\boldsymbol{\theta}) = \ell_c(\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta)$, writing $\boldsymbol{\alpha}(\boldsymbol{\theta}) = \boldsymbol{f}_\alpha(\boldsymbol{\theta})$ and $\boldsymbol{\beta}(\boldsymbol{\theta}) = \boldsymbol{f}_\beta(\boldsymbol{\theta})$. The estimation procedure under the constraints is implemented as an alternating variables method, where each step involves a non-linear maximization. The following algorithm describes the implementation in PcFiml:

ALGORITHM 2 (**Switching algorithm for general cointegration restrictions**):

(1) *determine starting values from the unrestricted parameters, assuming rank p;*
(2) *improve on the starting values using Newton's method;*
then choose one of the following:
(3) *direct maximization of the concentrated log-likelihood using BFGS;*
or:
(4) *switching, i.e. alternating maximization using BFGS:*
(α) *keep $\boldsymbol{\theta}_\beta$ fixed, and maximize (10) as a function of $\boldsymbol{\theta}_\alpha$;*
(β) *keep $\boldsymbol{\theta}_\alpha$ fixed, and maximize (10) as a function of $\boldsymbol{\theta}_\beta$;*

repeat steps (α) and (β) until convergence.

When moving from step (α) to (β) and back to (α), the likelihood will not go down (providing the restrictions do not link coefficients of $\boldsymbol{\alpha}$ to coefficients of $\boldsymbol{\beta}$). So this algorithm will converge to the global maximum as a consequence of the results in Sargan (1964) and Oberhofer and Kmenta (1974), provided a global maximum exists, there are no local maxima, and no numerical problems preventing us from reaching it (such as a failure to improve in the line search). BFGS is the Broyden-Fletcher-Goldfarb-Shanno method, see e.g. Fletcher (1987). The following remarks are relevant to Algorithm 2:

(1) Switching.
A major advantage of alternating (switching) between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is that the implicit identifying restrictions on $\boldsymbol{\beta}$, owing to the fact that $\boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\alpha}\boldsymbol{A}'^{-1}(\boldsymbol{\beta}\boldsymbol{A})'$ for non-singular $\boldsymbol{A}$, never have to be imposed. But when the constraints are not binding (do not restrict the long-run matrix: the constraints can be obtained by a certain $\boldsymbol{A}$ matrix), together with bad starting values, convergence can be relatively slow. However, in many cases (even without identification) we found that direct maximization worked well. There were a few instances where direct optimization led to offsetting rotation in $\boldsymbol{\theta}_\alpha$ and $\boldsymbol{\theta}_\beta$, taking the parameters into a numerically unstable area (ultimately leading to a failure to evaluate the log-likelihood). In those cases, switching did work well. Certain parameterizations (especially those linking coefficients of alpha and beta) may lead to very slow convergence. Here some manual intervention could be useful to try changing from switching to non-switching. Overall, I feel that switching should be the default, favouring robustness over speed.

(2) Starting values.
Since the main purpose is to restrict elements of $\boldsymbol{\beta}$, we shall only be concerned with starting values for beta. As pointed out under (1), some restrictions might not be binding, and could be accommodated by an appropriate $\boldsymbol{A}$ matrix. A simple approach is described here. Let $\hat{\boldsymbol{\alpha}}_{-1}$, $\hat{\boldsymbol{\beta}}_{-1}$ be the unconstrained standardized alpha and beta matrices. Write $\hat{\boldsymbol{b}}_{-1} = \boldsymbol{f}_\beta(\boldsymbol{f}_\beta^{-1}(\hat{\boldsymbol{\beta}}_{-1}))$ for the matrix resulting from applying the constraints to $\hat{\boldsymbol{\beta}}_{-1}$ and define $[\cdot]_r$ as dropping those rows which have no restrictions in them; if this yields less than p rows, than add rows back in, so that the $[\cdot]_r$ matrix is $q \times p$, with $q \geq p$. Then solve:

$$\left[\hat{\boldsymbol{\beta}}_{-1}\right]_r \boldsymbol{A} = \left[\hat{\boldsymbol{b}}_{-1}\right]_r ,$$

for $\boldsymbol{A}$ and use $\boldsymbol{\theta}_0' = \left[\boldsymbol{f}_\beta^{-1}(\hat{\boldsymbol{\beta}}_{-1}\hat{\boldsymbol{A}}) : \boldsymbol{f}_\alpha^{-1}(\hat{\boldsymbol{\alpha}}_{-1}\hat{\boldsymbol{A}}^{-1\prime})\right]$ to start.

(3) Newton steps.
Analogous to the hybrid Gauss Newton/BFGS method for solving non-linear least squares problems (Fletcher, 1987, p. 117) I found it helpful to start the iterating process with some (possibly none) Newton steps. Writing $\mathring{\boldsymbol{Q}} = \boldsymbol{J}\boldsymbol{Q}\boldsymbol{J}'$ and $\mathring{\boldsymbol{q}} = \boldsymbol{J}\boldsymbol{q}$ for the constrained Hessian matrix and score vector ($\boldsymbol{Q}$ and $\boldsymbol{q}$ are given below in (13), (14)), the Newton steps are performed on the whole vector of theta's:

$$\boldsymbol{\theta}_j = \boldsymbol{\theta}_{j-1} - \left(\mathring{\boldsymbol{Q}}_{j-1}\right)^+ \mathring{\boldsymbol{q}}_{j-1}.$$

Since $\mathring{\boldsymbol{Q}}$ could be singular, the generalized inverse based on the singular value decomposition is used. However, the steps are only taken if

$$\ell(\boldsymbol{\theta}_j) - \ell(\boldsymbol{\theta}_{j-1}) > \frac{1}{64}\left|\ell(\boldsymbol{\theta}_{j-1})\right| ,$$

i.e. only if the function increase is at least 1.5%. If any steps are taken, $-\left(\mathring{\boldsymbol{Q}}_j\right)^+$ forms the initial inverse Hessian; otherwise the identity matrix is used to initialize BFGS.

(4) Step (α) with unconstrained $\boldsymbol{\alpha}$ or step (β) with unconstrained $\boldsymbol{\beta}$ can be done without iteration:

$$\boldsymbol{\beta}' = \left(\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\Omega}}_v^{-1}\hat{\boldsymbol{\alpha}}\right)^{-1}\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\Omega}}_v^{-1}\boldsymbol{S}_{01}\left(\boldsymbol{S}_{11}\right)^{-1}, \tag{11}$$

see (13); and

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{S}_{01}\hat{\boldsymbol{\beta}}\left(\hat{\boldsymbol{\beta}}'\boldsymbol{S}_{11}\hat{\boldsymbol{\beta}}\right)^{-1}. \tag{12}$$

(5) In case the BFGS maximization procedure is used, it has been found advantageous to restart step (α) (or (β)) with the quasi-Newton approximation $\boldsymbol{K}$ to the inverse Hessian matrix at the end of the previous step (α) (or (β)). The alternative would be to reset $\boldsymbol{K}$ to the identity matrix before each maximization, but with a lot of alternating this would result in near linear behaviour.

(6) Convergence.
Convergence may be assumed after both steps have converged (with score vectors close to zero) and function values at the end of successive (β) steps are very close together.
When the restrictions are of the form H_e the alternating deflation method of Johansen (called switching method by him) will be more efficient because the function is concentrated with respect to $\boldsymbol{\alpha}$ while Algorithm 2 cycles over each $\boldsymbol{\beta}$ vector, rather than being conditional on a certain $\boldsymbol{\alpha}$ value for all $\boldsymbol{\beta}$ coefficients.

(7) The implementation of the algorithm in PcFiml allows two versions of the BFGS method to optimize the restricted likelihood:

   (a) BFGS with 'analytical' first derivatives;
   This method employs a mixture of analytical derivatives ($\partial\ell(\boldsymbol{\phi})/\partial\phi_i$) and numerical derivatives ($\partial\boldsymbol{\phi}/\partial\theta_i$). The first term is given in the appendix. This is the preferred method.

   (b) BFGS with numerical first derivatives.
   Uses numerical derivatives to compute $\partial\ell\left(\boldsymbol{\phi}\left(\boldsymbol{\theta}\right)\right)/\partial\theta_i$. The numerical scores are less accurate than analytical scores, (because they tend to zero as the iteration proceeds) and more costly to obtain.

4 SOME EXAMPLES

4.1 PcFiml tutorial

As a first application consider the tutorial example of Doornik and Hendry (1994, p. 86–87). There we have a four-dimensional VAR, with the trend restricted to the cointegrating space and a rank of three ($n_0 = 4, n_1 = 5, p = 3$). The restrictions considered are

$$\boldsymbol{\alpha} = \begin{pmatrix} * & 0 & 0 \\ 0 & * & * \\ 0 & * & * \\ 0 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{\beta}' = \begin{pmatrix} a & 0 & 0 & -a & * \\ 0 & b & -b & * & 0 \\ * & -c & c & 0 & 0 \end{pmatrix}.$$

The corresponding test statistic reported by PcFiml version 8.0 is $\chi^2(\approx 7) = 14.325\,[0.0457]$, with the p-value between square brackets. This, however, is incorrect, and the techniques of the previous section allow is to compute the correct degrees if freedom (as implemented in the newer PcFiml version 8.1).

Table 4 lists the singular values. The columns are from left to right: the proposed method of using random values in $\mathcal{J}'$; the maximum likelihood estimates in $\mathcal{J}'$; and finally the maximum likelihood estimates in the information matrix. Each of these suggest a matrix which has rank 8, the three zero singular values corresponding to an unidentified model (the zeros would disappear by normalizing $\boldsymbol{\beta}'$ along the diagonal). The unrestricted model has $12 + 15 - 9 = 18$ parameters, so that the observed rank 8 corresponds to 10

Table 4. Singular values of PcFiml tutorial example.

	$\mathcal{J}'(\boldsymbol{\theta}_0)$	$\mathcal{J}'(\hat{\boldsymbol{\theta}})$	$\mathcal{I}(\hat{\boldsymbol{\theta}})$
w_1	1.781	1.390	4.193×10^{7}
w_2	1.583	0.81034	6.646×10^{4}
w_3	1.409	0.80469	1.576×10^{4}
w_4	1.159	0.68091	1050.
w_5	1.061	0.66004	778.8
w_6	0.93190	0.19183	605.9
w_7	0.77074	0.17216	207.8
w_8	0.48373	0.073142	48.69
w_9	1.293×10^{-16}	4.312×10^{-17}	7.587×10^{-13}
w_{10}	1.936×10^{-17}	7.662×10^{-18}	4.589×10^{-12}
w_{11}	2.813×10^{-17}	6.420×10^{-19}	1.196×10^{-11}
ϵ_w	4.12795×10^{-12}	2.95779×10^{-12}	9.45369×10^{-5}

restrictions and not 7 as reported. In this case the correct test statistic is $\chi^2(10) = 14.325\ [0.159]$. Apparently, a zero in each of the cointegrating vector must be counted in addition to those in $\boldsymbol{\alpha}$. To see this, normalize $\boldsymbol{\beta}$: this is not a constraint as it can be achieved by rotating $\boldsymbol{\alpha}$, but now each vector is overidentified by one restriction. The tutorials considered a couple of variants. Setting the remaining element in the last column of $\boldsymbol{\beta}'$ to zero adds one further restriction. Reinstating this element, but setting the second and third diagonal element of $\boldsymbol{\alpha}$ to zero results in 12 degrees of freedom, still requiring normalization of each $\boldsymbol{\beta}$-vector for identification.

4.2 Australian data set

The second example uses the model presented in Johansen and Juselius (1994), hereafter referred to as JJ94, which is based on Australian data. The numerical issues which arise are very interesting and we shall discuss this model in considerable detail.

The endogenous variables in this model are the log of the M3 measure of the nominal money stock, the log of real GDP, the log of the GDP deflator, the three month commercial bill rate and the ten-year bond rate. These variables are denoted by $m3, y, p, r_3, r_{10}$.[3] In addition, there is a trend which is restricted to enter the cointegrating space. Finally, there is an unrestricted constant and centred seasonals, and an unrestricted step dummy which is zero up to 1983–4, and one thereafter. The dummy is introduced to capture the effects of floating exchange rate and financial deregulation.

The data are presented in Fig. 1. It is noticeable that both $m3$ and p closely resemble a linear trend. The seasonality is mainly confined to y, which also has a strong trend, and shows the break which the step dummy tries to capture.

The basic model is a VAR(2), which is estimated over the sample period 1976–1 to 1991–1 ($T = 61$). We have three implementations of the Johansen procedure at our disposal: CATS, PcFiml, and the numerically stable implementation in Ox (based on the singular value decomposition and described in Doornik (1995)). Each of these gives the same results for eigenvalues and eigenvectors of the unrestricted

[3] I wish to thank Katarina Juselius for kindly making the data available, and allowing me to use CATS version 4.2, see Hansen and Juselius (1994).

cointegration analysis.[4] The test output from PcFiml is summarized in Table 5. Although this suggests only one eigenvector, we adopt three as JJ94 do.

Table 5. Unrestricted cointegration analysis of Australian model of JJ94.

$\mathsf{H}_0 : \text{rank} = p$	(1)	(2)	95%	(3)	(4)	95%
$p = 0$	41.9^*	35.03	37.5	101.1^{**}	84.49	87.3
$p \leq 1$	26.65	22.28	31.5	59.15	49.45	63.0
$p \leq 2$	18.52	15.49	25.5	32.5	27.17	42.4
$p \leq 3$	11.92	9.965	19.0	13.98	11.69	25.3
$p \leq 4$	2.059	1.722	12.2	2.059	1.722	12.2

$(1) : \ -T \log(1-\mu)$ $\quad (2) : \ -(T-nm)\log(1-\mu)$

$(3) : \ -T \sum \log(1-\mu)$ $\quad (4) : \ -(T-nm)\sum \log(1-\mu)$

Table 6. Restrictions on cointegrating space.

	JJ94 table 3			JJ94 table 5			JJ94 table 6		
m	d	0	0	1	0	0	d	0	0
y	$*$	$*$	$*$	-1	0	0	1	0	0
p	$-d$	$*$	$*$	-1	0	$*$	$-d$	0	$*$
r_3	1	1	0	$*$	1	0	0	1	0
r_{10}	0	-1	1	$*$	-1	1	0	-1	1
t	$*$	$*$	$*$	$*$	0	$*$	$*$	0	$*$

Table 7. Singular values of Australian model, JJ94 Table 3.

	$\mathcal{J}'(\boldsymbol{\theta}_0)$	$\mathcal{J}'(\hat{\boldsymbol{\theta}})r_3, r_{10}$	$\mathcal{I}(\hat{\boldsymbol{\theta}})r_3, r_{10}$	$\mathcal{J}'(\hat{\boldsymbol{\theta}})R_3, R_{10}$	$\mathcal{I}(\hat{\boldsymbol{\theta}})R_3, R_{10}$
w_1	3.242	5.923	1.783×10^7	134.6	2.044×10^7
w_{22}	0.21929	0.11176	5.383	0.0015419	0.00029338
w_{23}	0.022014	0.039729	1.052	2.227×10^{-5}	0.00049829
w_{24}	0.011014	0.00069562	8.623×10^{-5}	1.158×10^{-6}	7.451×10^{-8}
ϵ_w	9.7×10^{-12}	1.6×10^{-11}	4.9×10^{-5}	5.3941×10^{-10}	7.28327×10^{-5}

Given the strong similarity of $m3$, p and the trend, there could be some concern about the ability to identify the cointegrating vectors. Estimation of the restriction on the cointegrating vectors presented in table 3 of JJ94 (see Table 6) bears this out. Using Algorithm 1, PcFiml is able to warn prior to estimation that the vectors are exactly identified, so that no estimation is required for the test, just a rotation. Proceeding with the estimation regardless leads to convergence.[5]

[4] The coefficients on r_3 and r_{10} are different from the results in Johansen and Juselius (1994), but this appears to be by a fixed scale factor of 1.12, and is of no concern.

[5] Somewhat before the maximum, namely at 1460.4386, where the maximum is 1460.4087. This, of course, is determined by the convergence tolerance, although attempting more precise estimation illustrates the numerical problem: the parameters change considerably, without much change in the function value.

Table 8. First beta vector, JJ94 Table 3.

		$m3$	y	p	R_3/r_3	R_{10}/r_{10}	$Trend$
PcFiml	R_3, R_{10}	1.062	103.7	−1.062	1	0	−0.72193
PcFiml	r_3, r_{10}	1.278	−5.459	−1.278	1	0	0.026125
CATS	R_3, R_{10}	21.750	−1.934	−21.750	1	0	−0.180
CATS	r_3, r_{10}	0.200	0.070	−0.200	1	0	−0.002
Ox	R_3, R_{10}	21.316	0.2892	−21.316	1	0	−0.19126
Ox	r_3, r_{10}	0.3677	−0.7872	−0.3677	1	0	0.002140

Investigation of the rank of the information matrix at the converged values, see Table 7, indicates that the model is close to reduced rank, even though it is exactly identified. Depending on the form of the restrictions, the information matrix singular values are scale variant. Here we standardize on the interest rates. If these are changed to percentages: $R_3 = 100 \times r_3$ and $R_{10} = 100 \times r_{10}$, the condition of the restricted information matrix will deteriorate, because the interest rate coefficients are fixed. The last two columns of the table have the numbers for the Jacobian and information matrix based rank condition at the converged values (the random parameter method is unaffected). Now the information matrix has reduced rank, and inverting it in PcFiml fails. Also the Jacobian at the maximum is much less clearly of full rank. The numerical instability is illustrated by the values found for the first beta vector (the second and third have roughly the same magnitude) from the various programs, given in Table 8. Care is required when interpreting the cointegrating vectors in this case. Fortunately, the test statistic for the imposed restrictions is much more stable, since an upper bound for the log-likelihood can be determined regardless of the underidentification.

The first cointegrating vector under the restrictions of JJ94 table 5 is also numerically unstable, but the test statistic is well determined by all programs at $\chi^2(5) = 22.48$. The results according to JJ94 table 6, are numerically stable: the results are unaffected by changing from r_3, r_{10} to R_3, R_{10}.

So, perhaps quite surprisingly, the numerical problems suggest using a more orthogonal data transformation along the lines of Hendry and Doornik (1994): replacing nominal money by real money ($m-p$), and using the inflation rate (Δp) instead of the GDP inflator, see Fig. 2. The results of this approach are reported in Doornik (1995).

5 Conclusion

Testing general cointegration restrictions provides an example where computation did lead theory for a period of time: estimation has been available in the computer program PcFiml (starting in 1992 with beta versions), but no reliable way to verify identification and count the degrees of freedom was available then. This paper solves the theoretical issues and provides the operational links required for practical application. Algorithm 1 makes it is possible to verify identification prior to estimation. The degrees of freedom (if any) of the test of overidentifying restrictions may now be presented automatically by a computer program, independent of whether the model is identified or not.

Acknowledgements

I wish to thank Bent Nielsen, David Hendry and Tom Rothenberg for helpful comments, and Katarina Juselius for making the Australian data set available to me. Financial support from the UK Economic and

Social Science Research Council is gratefully acknowledged.

REFERENCES

Banerjee, A., Dolado, J. J., Galbraith, J. W. and Hendry, D. F. (1993). *Co-integration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.

Boswijk, H. P. (1994). Identifiability of cointegrated systems. Mimeo, Department of Actuarial Sciences and Econometrics, University of Amsterdam.

Doornik, J. A. (1995). *Econometric Computing*. Oxford: University of Oxford. Ph.D Thesis.

Doornik, J. A. and Hendry, D. F. (1994). *PcFiml 8: An Interactive Program for Modelling Econometric Systems*. London: International Thomson Publishing.

Fletcher, R. (1987). *Practical Methods of Optimization* 2nd edn. New York: John Wiley & Sons.

Hansen, H. and Juselius, K. (1994). Manual to cointegration analysis of time series CATS in RATS. Discussion paper, Institute of Economics, University of Copenhagen.

Hendry, D. F. and Doornik, J. A. (1994). Modelling linear dynamic econometric systems. *Scottish Journal of Political Economy*, **41**, 1–33.

Hunter, J. (1992). Cointegrating exogeneity. *Economic Letters*, **34**, 33–35.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, **59**, 1551–1580.

Johansen, S. (1993). Identifying restrictions of linear equations. Discussion paper, Institute of Mathematical Statistics, University of Copenhagen. Forthcoming, Journal of Econometrics.

Johansen, S. (1995). *Likelihood based Inference on Cointegration in the Vector Autoregressive Model*: Oxford University Press. Forthcoming.

Johansen, S. and Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration – With application to the demand for money. *Oxford Bulletin of Economics and Statistics*, **52**, 169–210.

Johansen, S. and Juselius, K. (1992). Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK. *Journal of Econometrics*, **53**, 211–244.

Johansen, S. and Juselius, K. (1994). Identification of the long-run and the short-run structure. An application to the ISLM model. *Journal of Econometrics*, **63**, 7–36.

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

Mosconi, R. and Giannini, C. (1992). Non-causality in cointegrated systems: Representation, estimation and testing. *Oxford Bulletin of Economics and Statistics*, **54**, 399–417.

Oberhofer, W. and Kmenta, J. (1974). A general procedure for obtaining maximum likelihood estimates in generalized regression models. *Econometrica*, **42**, 579–590.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica*, **39**, 577–591.

Sargan, J. D. (1964). Wages and prices in the United Kingdom: A study in econometric methodology (with discussion). In Hart, P. E., Mills, G. and Whitaker, J. K. (eds.), *Econometric Analysis for National Economic Planning*, Vol. 16 of *Colston Papers*, pp. 25–63. London: Butterworth Co. Reprinted as pp. 275–314 in Hendry D. F. and Wallis K. F. (eds.) (1984). *Econometrics and Quantitative*

Economics. Oxford: Basil Blackwell, and as pp. 124–169 in Sargan J. D. (1988), *Contributions to Econometrics*, Vol. 1, Cambridge: Cambridge University Press.

Toda, H. and Phillips, P. C. B. (1991). Vector autoregressions and causality. Cowles Foundation discussion paper 977, Yale University.

APPENDIX

Start from the vector error-correction model

$$\Delta \boldsymbol{y}_t = \sum_{i=1}^{m-1} \boldsymbol{\delta}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{P}\boldsymbol{y}_{t-1} + \boldsymbol{v}_t, \boldsymbol{v}_t \sim \mathsf{IN}(\boldsymbol{0}, \boldsymbol{\Omega})$$

we obtain the concentrated system

$$\boldsymbol{R}_0' = \boldsymbol{P}\boldsymbol{R}_1' + \boldsymbol{V}'$$

by partialling out lagged differences and other unrestricted variables (such as e.g. a trend or seasonal dummies). The corresponding concentrated likelihood function is:

$$\ell_c(\boldsymbol{P}) = -\frac{T}{2}\log\left|T^{-1}\boldsymbol{V}'\boldsymbol{V}\right| = -\frac{T}{2}\log\left|T^{-1}(\boldsymbol{R}_0' - \boldsymbol{P}\boldsymbol{R}_1')(\boldsymbol{R}_0 - \boldsymbol{R}_1\boldsymbol{P}')\right|,$$

with $\boldsymbol{P} = \boldsymbol{P}(\boldsymbol{\theta}) = \boldsymbol{\alpha}(\boldsymbol{\theta})\boldsymbol{\beta}(\boldsymbol{\theta})'$ and p for the rank of the cointegrating space. Write

$$\boldsymbol{S}_{ij} = T^{-1}\boldsymbol{R}_i'\boldsymbol{R}_j \text{ for } i,j = 0,1.$$

The first derivatives, required for the maximization algorithm are for $\boldsymbol{\varphi}' = \left\{[\mathrm{vec}(\boldsymbol{\alpha}')]' : [\mathrm{vec}\,\boldsymbol{\beta}]'\right\}$:

$$\boldsymbol{q} = \frac{\partial \ell_c(\boldsymbol{\alpha}\boldsymbol{\beta}')}{\partial \boldsymbol{\varphi}} = T\begin{pmatrix} \mathrm{vec}\left\{\left[\boldsymbol{\Omega}_v^{-1}\left(\boldsymbol{S}_{01} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{S}_{11}\right)\boldsymbol{\beta}\right]'\right\} \\ \mathrm{vec}\left\{\left[\boldsymbol{\alpha}'\boldsymbol{\Omega}_v^{-1}\left(\boldsymbol{S}_{01} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{S}_{11}\right)\right]'\right\} \end{pmatrix}. \tag{13}$$

Asymptotic second derivatives are obtained by keeping $\boldsymbol{\Omega}_v$ fixed:

$$\begin{aligned} \mathrm{dd}\ell_c(\boldsymbol{P}) &= -\mathrm{tr}\left\{\boldsymbol{\Omega}_v^{-1}(\mathrm{d}\boldsymbol{P})\boldsymbol{R}_1\mathrm{d}\boldsymbol{P}'\right\} + \mathrm{tr}\left\{\boldsymbol{\Omega}_v^{-1}\boldsymbol{V}'\boldsymbol{R}_1\mathrm{dd}\boldsymbol{P}'\right\} \\ &= -\mathrm{tr}\left\{\boldsymbol{\Omega}_v^{-1}(\mathrm{d}\boldsymbol{P})\boldsymbol{R}_1\mathrm{d}\boldsymbol{P}'\right\}. \end{aligned}$$

The cross-derivative is found as follows. Analogous to Magnus and Neudecker (1988, p. 182) we can derive:

$$\begin{aligned} \mathrm{d}\,\mathrm{vec}(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) &= \left(\boldsymbol{I}\otimes\boldsymbol{X}'\right)\mathrm{d}\,\mathrm{vec}\,\boldsymbol{A}\boldsymbol{X}\boldsymbol{B} + (\boldsymbol{B}'\boldsymbol{X}'\boldsymbol{A}'\otimes\boldsymbol{I})\,\mathrm{d}\,\mathrm{vec}\,\boldsymbol{X}' \\ &= \left(\boldsymbol{B}'\otimes\boldsymbol{X}'\boldsymbol{A}\right)\mathrm{d}\,\mathrm{vec}\,\boldsymbol{X} + (\boldsymbol{B}'\boldsymbol{X}'\boldsymbol{A}'\otimes\boldsymbol{I})\,\boldsymbol{K}\,\mathrm{d}\,\mathrm{vec}\,\boldsymbol{X}, \end{aligned}$$

where $\boldsymbol{K}$ is the commutation matrix such that $\boldsymbol{K}\,\mathrm{vec}\,\boldsymbol{X} = \mathrm{vec}\,\boldsymbol{X}'$. Thus, we find for the cross-derivative:

$$\begin{aligned} &-\left(\boldsymbol{\Omega}_v^{-1}\boldsymbol{S}_{01}\otimes\boldsymbol{I}_p\right)\boldsymbol{K} + \left(\boldsymbol{\Omega}_v^{-1}\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{S}_{11}\otimes\boldsymbol{I}_p\right)\boldsymbol{K} + \left(\boldsymbol{\Omega}_v^{-1}\boldsymbol{\alpha}\otimes\boldsymbol{\beta}'\boldsymbol{S}_{11}\right) \\ &= \left(\boldsymbol{\Omega}_v^{-1}\boldsymbol{\alpha}\otimes\boldsymbol{\beta}'\boldsymbol{S}_{11}\right) - \left\{\boldsymbol{\Omega}_v^{-1}\left(\boldsymbol{S}_{01} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{S}_{11}\right)\otimes\boldsymbol{I}_p\right\}\boldsymbol{K} \\ &= \left(\boldsymbol{\Omega}_v^{-1}\boldsymbol{\alpha}\otimes\boldsymbol{\beta}'\boldsymbol{S}_{11}\right). \end{aligned}$$

The last step occurs because $\boldsymbol{S}_{01} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{S}_{11} = \boldsymbol{V}'\boldsymbol{R}_1$ vanishes asymptotically in the absence of misspecification.

So, for the asymptotic Hessian matrix of the CLF with respect to $\boldsymbol{\varphi}$ we obtain:

$$\begin{aligned} \boldsymbol{Q} &= \frac{\partial^2 \ell_c(\boldsymbol{\alpha}\boldsymbol{\beta}')}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'} = -T\begin{pmatrix} \boldsymbol{\Omega}_v^{-1}\otimes(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}) & \left(\boldsymbol{\Omega}_v^{-1}\boldsymbol{\alpha}\otimes\boldsymbol{\beta}'\boldsymbol{S}_{11}\right) \\ \boldsymbol{\alpha}'\boldsymbol{\Omega}_v^{-1}\otimes\boldsymbol{S}_{11}\boldsymbol{\beta} & \boldsymbol{\alpha}'\boldsymbol{\Omega}_v^{-1}\boldsymbol{\alpha}\otimes\boldsymbol{S}_{11} \end{pmatrix} \\ &= -T\begin{pmatrix} \boldsymbol{I}_{n_0}\otimes\boldsymbol{\beta}' \\ \boldsymbol{\alpha}'\otimes\boldsymbol{I}_{n_1} \end{pmatrix}\left(\boldsymbol{\Omega}_v^{-1}\otimes\boldsymbol{S}_{11}\right)\begin{pmatrix} \boldsymbol{I}_{n_0}\otimes\boldsymbol{\beta} & \boldsymbol{\alpha}\otimes\boldsymbol{I}_{n_1} \end{pmatrix}. \end{aligned} \tag{14}$$

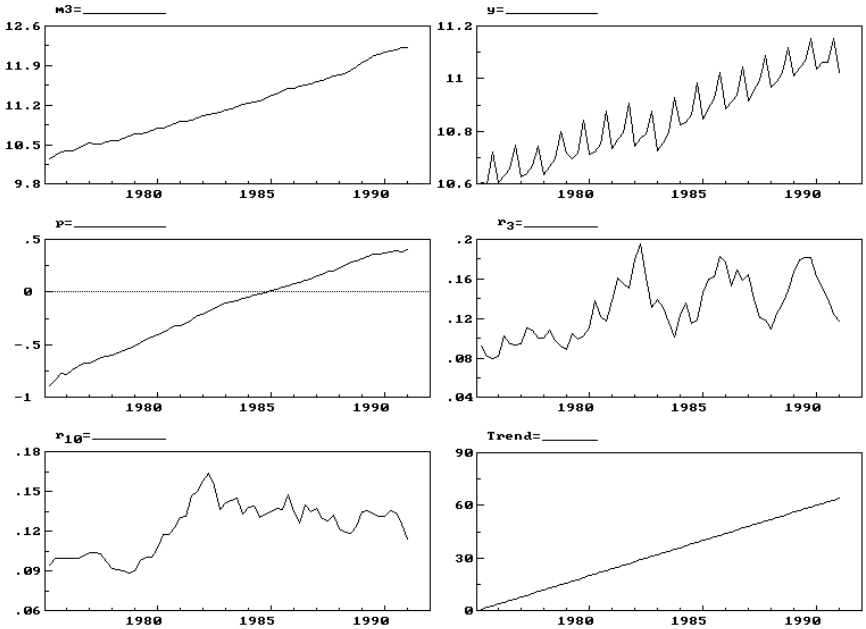

Fig. 1. Basic variables in Australian model.

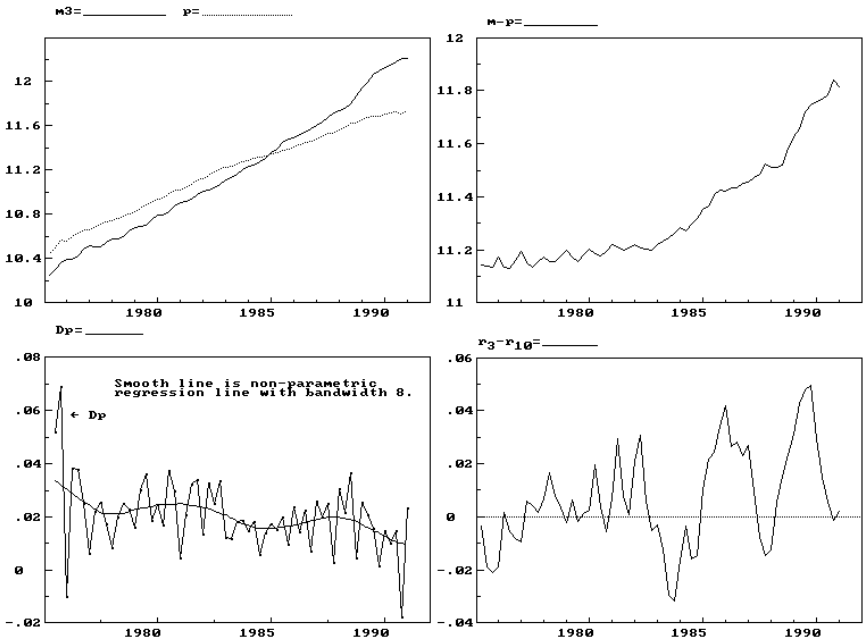

Fig. 2. Transformations of variables in Australian model.