EUROPEAN UNIVERSITY INSTITUTE
*Department of Economics*



# Essays on Cointegration Analysis

Pieter Omtzigt

*Thesis submitted for assessment with a view to obtaining the degree of Doctor of the European University Institute*

**Florence**
December 2003

**EUROPEAN UNIVERSITY INSTITUTE**
**Department of Economics**

# Essays on Cointegration Analysis

Pieter Omtzigt



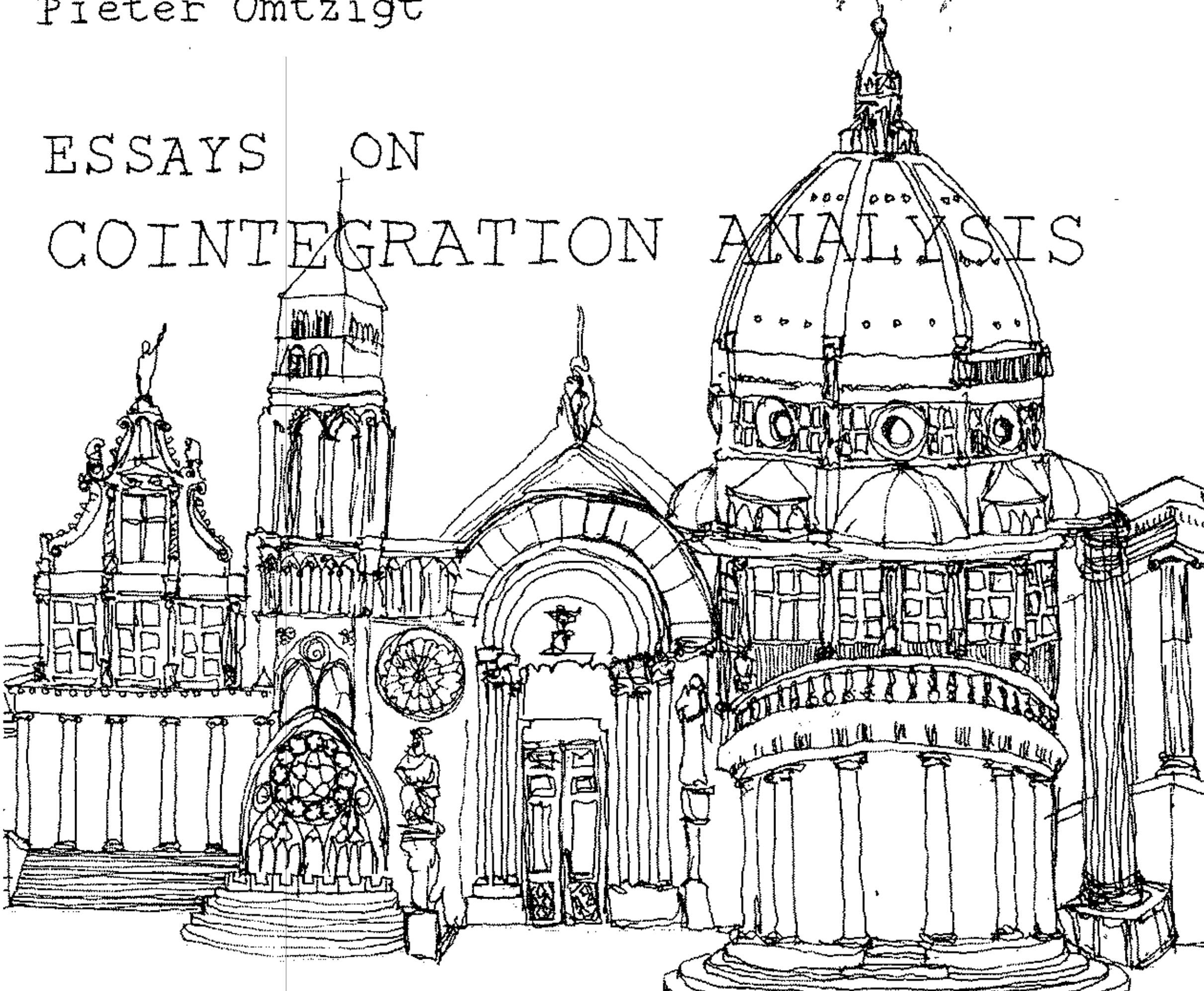
The Thesis Committee consists of:

Alvaro Escribano, Carlos III, Madrid
Søren Johansen, Supervisor, University of Copenhagen
Oliver Linton, London School of Economics
Helmut Lütkepohl, EUI

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

Pieter Omtzigt

# ESSAYS ON COINTEGRATION ANALYSIS



Amsterdam-Florence 2003

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

# Essays on cointegration analysis

PIETER OMTZIGT

EUROPEAN UNIVERSITY INSTITUTE
*Current address*: University of Amsterdam, Faculty of Economics and Econometrics, Roetersstraat 11, 1018WB Amsterdam, The Netherlands
*E-mail address*: P.H.Omtzigt@uva.nl
*URL*: http://www.fee.uva.nl/ke/Uva-Econometrics

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

# Contents





Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

# List of Figures





Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

# List of Tables





Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

# Acknowledgements

Without many people, this thesis would never have been. First and foremost Søren Johansen and Katarina Juselius have not only helped me enormously and stimulated my thinking on economic models, small sample corrections and automated model selection, but they have also been very generous throughout the seven years it took to write this thesis. Many thanks are due to Søren for being able to explain econometrics over and again: it all looks so straightforward the way he brings it, until I get home. And thanks to Katarina for showing how econometric models can be applied fruitfully.

The comments of the three other thesis committee members, Alvaro Escribano, Oliver Linton and Helmut Lütkepohl (beside Søren Johansen) have been most helpful for improving parts of this thesis.

The influence of other peoples ideas on my work is difficult to overestimate: the small sample corrections and then particularly the long Bartlett corrections have greatly benefited from input from and discussions with Søren and Bent Nielsen. After a course in Copenhagen by Joel Horowitz on bootstrapping, Stefano Fachin really got me involved in the difficult practice of bootstrapping. Together with Stefano I wrote chapter 4.

The papers on automated model selection owe a large deal to ideas from Katarina on one hand and the research group, lead by David Hendry in Oxford on the other. Discussions at the Sveriges Riksbanken on automated model selection (and bootstrapping) with Anders Vredin and Tor Jacobson have also considerably influenced the ideas, if only by giving me a large Swedish data set, which I did not manage to crack properly yet.

During my time in Varese, Paolo Paruolo and I together wrote chapter 6 on impact factors: with his precision and deep mathematic knowledge he has greatly helped me in other parts of the thesis as well.

Katarina Juselius guided me in the difficult analysis of the Dutch data. Peter Boswijk was kind in pointing me at the right directions for data and institutions.

For the I(2) models and software, discussions with Søren Johansen, Anders Rahbek, Hans Christian Kongsted and Heino Bohn Nielsen were extremely helpful in shaping the package Me2, included in this thesis.

Other work on pensions with Carolyn Dudek, on seasonal cointegration with Gianluca Cubadda and on MCMC with Antonietta Mira is not included here. Yet thanks for understanding so well I was doing other things at the same time, be it my thesis, be it politics. In my relatively short time in Amsterdam, it was great to work on Frank Kleibergens project.

In different places I met a large number of marvellous people. This list is not here for excluding anybody, but for the research part, Chiara Osbat, Massimiliano Marcellino, Juan Rojas (Florence) and Sophocles Mavroeidis (Amsterdam) have always been great friends and cheered me on at just the right times.

Then in seven Italian years, it was fantastic to live in beautiful villages with great people. Before going to the EUI, a year in Rome, then four years in Pian di Mugnone and two years in Varese. In all these three places and Verona I have just met so many friends, I keep struggling to stay in touch, be it with the group in Pian/le Caldine, the 'Amici del Sidamo' in Varese/Lombardy at large or the many in Verona. Going back to the Netherlands was a second cultural shock.



Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

This thesis might well have been on pensions, had I not changed topic after a few years. It may be ironic that this research, which started under the supervision of Ramon Marimon, is not only progressing, but has become part of my current work. The research was instrumental for getting in touch with politics in the Netherlands in general and the CDA in particular: Marnix van Rij, Nicolien van Vroonhoven-Kok, Alex Krijger and Jan-Peter Balkenende all kindly contributed with their discussions on the future of pensions in the Netherlands and Europe.

The beautiful illustrations of the places I lived in over the years, were drawn by Nina Mathijsen.

During my studies it has always been great to have the support of my parents, my twin brother Robbert and younger brother Dirk-Jan, who has now taken up econometrics as well.

And then, more than seven years ago in Spanish class at the EUI - one of my side tracks in research - I met April. It took over six years to get in touch again, but thanks so much for just being there.

CHAPTER 1

# Introduction

Essays in cointegration analysis never was the working title of the work in progress for the last seven years. I started this project with the aim of doing applied research in economics and econometrics. Hence the last chapter of this thesis, Money demand in the Netherlands, was the first chapter written. Yet that very first version, written in Florence in 1998, bears little resemblance to the present version, included in this thesis. The only substantial agreement with the first version are the data used: the short term interest rates were collected from private banks in the Netherlands: as they increasingly offered above money-market interest rates to retail investors, those official interest rates were not relevant for retail investors and even small firms.

That interesting problem - the irrelevance of money market rates to money demand - was thus solved rather quickly. Yet the other five chapters all evolved from practical problems, I ran into, during the cointegration analysis of the Dutch data set. I shall thus very briefly describe the cointegrated VAR models and then discuss in turn the problems, these chapters deal with.

Each chapter is stand-alone, in that it can be read without having to read any of the others first: this also means that short general overviews of the methodology are presented in each one of them and not repeated here. This introduction just points at the general problems tackled in the papers.

For each of the chapters, Matlab programs are available to replicate the results. These are included on a CD, which is part of this thesis.

Two programs have been developed into stand-alone packages: *me2* for maximum likelihood estimation of I(2) models and *datamine* for the automatic identification and restriction of the cointegration space.

Both of them are used in the replication of the results. The replication notes are thus a good starter for using these programs.

## 1.1. The cointegrated VAR model

**1.1.1. The I(2) model.** One representation of the $p$-dimensional I(2) model (Johansen, 1992) with 2 lags is given by:

$$\Delta^2 X_t = \Pi X_{t-1} - \Gamma \Delta X_{t-1} + \mu_0 + \mu_1 t + \varepsilon_t \tag{1.1}$$

where $\varepsilon_t$ is distributed normally with mean zero and variance-covariance matrix $\Omega$ and $t = 1, \ldots, T$.

Define the characteristic polynomial of this process as:

$$F(\lambda) = \lambda^2 I - (\Pi + 2I - \Gamma)\lambda - (\Gamma - I) \tag{1.2}$$

and let $\lambda_1, \ldots, \lambda_{2p}$ be the roots of $|F(\lambda)| = 0$

The following assumptions apply:

**I(2) a:** $\Pi = \alpha\beta'$ where $\alpha$ and $\beta$ are $p \times r$ matrices of full column rank. $r < p$

**I(2) b:** $2p - 2r - s$ roots $\lambda$ of the characteristic polynomial (1.2) equal one $\lambda = 1$. The other $2r + s$ roots are smaller than one in absolute value $|\lambda| < 1$. Let $\{\lambda_i^*\}, i = 1, \ldots, 2r + s$ indicate the roots of the second group. It then follows that $\alpha'_\perp \Gamma \beta_\perp = \xi\eta'$ where $\xi$ and $\eta$ are full rank matrices of dimension $(p - r) \times s$. Another equivalent way of stating



Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

this result is $\bar{\alpha}_\perp \alpha'_\perp \Gamma \bar{\beta}_\perp \beta'_\perp = \alpha_1 \beta'_1$ where $\alpha_1$ and $\beta_1$ are full rank matrices of dimension $p \times s$, $s < p - r$.

**I(2) c:** $\alpha'_2 \Theta \beta_2$ where $\Theta = \Gamma \bar{\beta} \bar{\alpha}' \Gamma + I$, $\alpha_2 = (\alpha : \alpha_1)_\perp$ and $\beta_2 = (\beta : \beta_1)_\perp$ is a matrix of full rank $(p - r - s)$.

**1.1.2. The I(1) model.** The $p$-dimensional vector autoregressive model with 2 lags can be represented in its reduced form as:

$$\Delta X_t = \alpha \beta' X_{t-1} + \Gamma \Delta X_{t-1} + \mu_0 + \mu_1 t + \varepsilon_t \tag{1.3}$$

where $\varepsilon_t$ are distributed $N(0, \Omega)$ and $t = 1, \ldots, T$.

Define the characteristic polynomial of this process as:

$$F(\lambda) = \lambda^2 I - (\Pi + I + \Gamma)\lambda + \Gamma \tag{1.4}$$

and let $\lambda_1, \ldots, \lambda_{2p}$ be the roots of $|F(\lambda)| = 0$

The assumptions, which assure that the model is I(1) are:

**I(1) a:** $p - r$ roots $\lambda$ of the characteristic polynomial (1.4) equal one: $\lambda = 1$. The other $p + r$ roots are smaller than one in absolute value $|\lambda| < 1$. Let $\{\lambda_i^*\}, i = 1, \ldots, p + r$ indicate the roots of the second group.

**I(1) b:** $\alpha$ and $\beta$ are full rank $p \times r$ matrices, $r < p$.

## 1.2. The chapters

**1.2.1. Chapter 2: Automatic identification of simultaneous equation models.** One set of frequently applied restrictions on the cointegration space $\beta$ in equation (1.3) is:

$$\beta = (H_1 \varphi_1, \ldots, H_r \varphi_r) \tag{1.5}$$

Johansen (1995a) proposes a way to maximize the likelihood function, if these restrictions are generically identifying. However when the restrictions are not generically identifying, general optimization algorithms are available. In chapter 2, I propose an algorithm to add non-binding restrictions, which render any set of non-generically identifying restrictions, identifying.

The method is presented in a simultaneous equation model setting, with within equation restrictions, but the application is in terms of the just described cointegration model. On Australian money demand data, the algorithm with identification finds a higher maximum than the same algorithm without identification.

**1.2.2. Chapter 3: Automatic identification and restriction of the cointegration space.** It takes a considerable amount of time to test a large number of hypotheses of the kind (1.5). In chapter 3 I automate the process of restricting the cointegration space $\beta$. The applied econometrician only has to specify which column vectors could make up the various matrices $H$. The algorithm then combines them and finds the model with the largest number of overidentifying restrictions.

A Monte Carlo study and application to UK money demand data illustrate the use of the algorithm.

**1.2.3. Chapter 4: Bootstrapping and Bartlett corrections in the cointegrated VAR.** The small sample properties of tests on $\beta$ are known to be poor. Two alternative approaches have been proposed in the literature: bootstrapping, for which Gredenhoff and Jacobson (2001) propose a procedure and a Bartlett corrections, derived by Johansen (2000a). In this thesis chapter, which is a co-authored article with Stefano Fachin, we compare these two methods and a refinement to bootstrapping: the feasible double bootstrap, proposed by Davidson and MacKinnon (2000).

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

We find that in our Monte Carlo DGP none of these three procedures has much power, if they are based on the estimates under the null hypothesis. We thus propose to base all these methods on the estimates under the alternative hypothesis.

**1.2.4. Chapter 5: A Bartlett correction in stationary autoregressive models.** In the longest chapter I derive a Bartlett correction in the following multivariate model:

$$Y_t = AX_t + \eta_{2t}$$

where

$$X_t = Q(L)\eta_{t-1} = Q_0\eta_{t-1} + Q_1\eta_{t-2} + Q_2\eta_{t-3} + \ldots$$
$$\eta_t = \left[\begin{array}{cc} \eta_{1t}' & \eta_{2t}' \end{array}\right]' \sim MIIDN(0, \Omega)$$

under the assumption that $Q(L)$ is an exponentially decreasing polynomial

The hypothesis, for which I derive the Bartlett correction is $\mathcal{H} : A = A_0$, both when $var\ (\eta_t)$ is known and when it is unknown.

Applications of this Bartlett correction include a test for no autocorrelation, a null hypothesis in a multivariate autoregressive process and two tests for 'no long-run level feedback' in the I(1) model. A Monte Carlo study for each of these examples shows the usefulness of the derived correction

**1.2.5. Chapter 6: Impact Factors.** In this chapter, which is co-authored with Paolo Paruolo, we derive explicit expressions for the long-run effect of a one-time perturbation to a stationary VAR, the I(1) model and the I(2) model.

A one-off perturbation in an I(2) model may have an effect on the growth level of the individual variables.

An application to mark-up on prices in Australia illustrates the use of Impact Factors.

**1.2.6. Chapter 7: Money demand in the Netherlands.** In this chapter I model money demand in the twenty years prior to the introduction of the Euro. Applying most of the techniques in the previous chapters, I find a stationary money demand equation over the period under study, as well as an IS-curve. The data set has been specially compiled for this study.

**1.2.7. The CD with software.** All programs used in this thesis have been put on a CD, such that the results are replicable in Matlab. Two particular parts of the software have been developed further and can be used as applications:

1.2.7.1. *Me2: full maximum likelihood in I(2) models.* The Dutch data in chapter 7 are found to be I(2). To determine the rank, transform them to I(1) by means of a so-called nominal-to-real transformation and to calculate the impact factors, I wrote a complete package, named, me2. It executes a large amount of tests and writes its output straight into latex files. For windows-users with Matlab 6.x or higher and the statistics toolbox, it is readily installed by copying the relevant directory on the CD to the local hard disk. It is also possible to run it straight from the CD.

1.2.7.2. *Datamine: automated identification and restriction of the cointegration space.* This program is also includes on the CD. It performs the algorithm in chapter 3 in a relatively user-friendly way in Matlab.

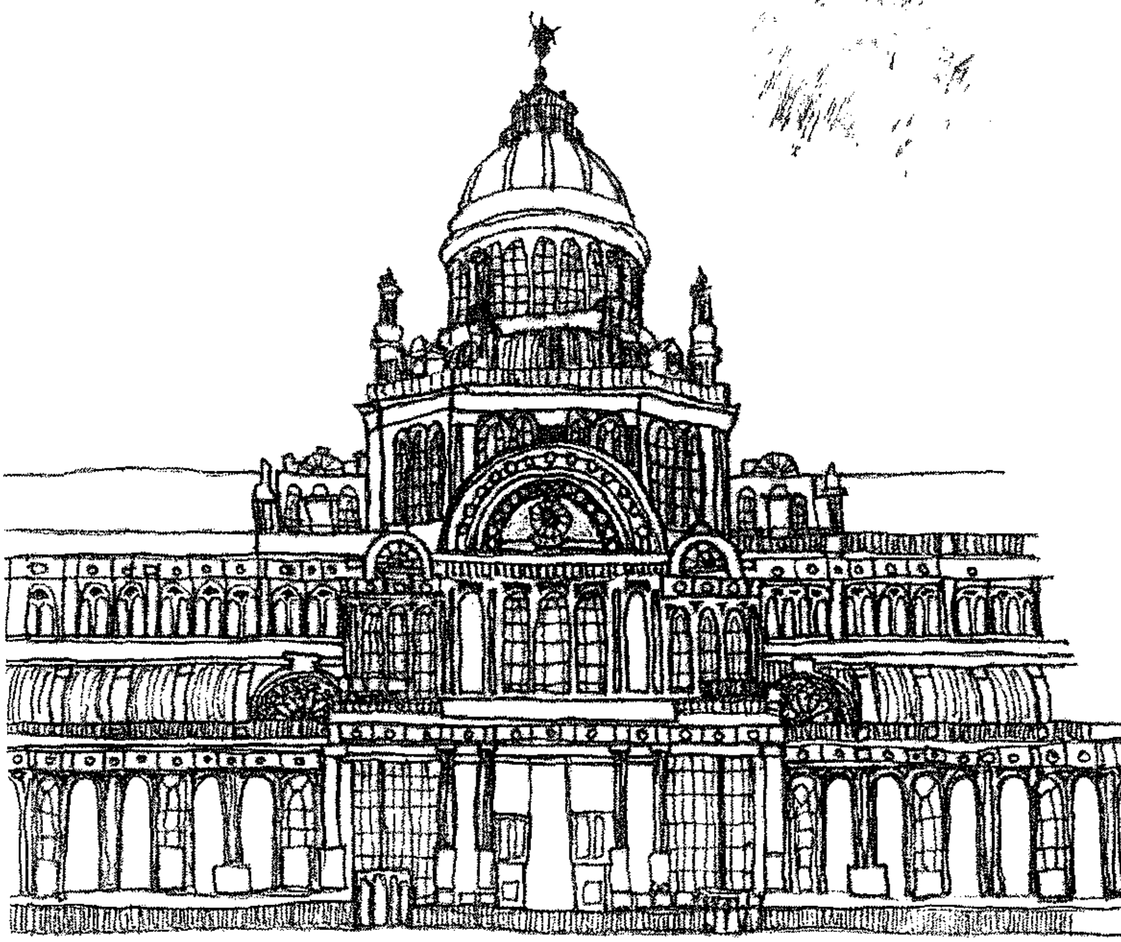
Chapters 6 and 7 contain research notes on how to replicate the results in those chapters and thus on how to use these two software packages.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

# Part 1

# Automated model selection

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609



A M S T E R D A M


DOI: 10.2870/52609

CHAPTER 2

# Automatic identification of simultaneous equations models

## 2.1. Introduction

Consider the simultaneous equations model:

$$\beta' z_t = A' y_t + B' x_t = u_t,\ t = 1, \ldots, T \qquad (2.1)$$
$$u_t \backsim iidN(0, \Omega),$$

where $y_t$ is a vector of length $r$ with endogenous variables, $x_t$ a vector of length $q$ with predetermined variables and $\beta = [A', B']'$ a $p \times r$ matrix of coefficients ($p = r + q$). Assume that $A$ and therefore $\beta$ is of full rank and that $x_t$ and $u_t$ are independent.

Likelihood inference on $(\beta, \Omega)$ is possible, but it is readily verified that a parameter point $(\beta^1, \Omega^1)$ is not uniquely identified (which means that there is at least one other parameter point $(\beta^2, \Omega^2)$, with whom it shares the same probability measure). For any non-singular matrix $C$, $(\beta^1 C, C\Omega^1 C')$ has an identical probability measure. To uniquely identify a space we need to put restrictions on the parameter space. In this article we shall consider only within-equation restrictions on $\beta$ without putting restrictions on $\Omega$. More precisely, we consider

$$\beta = [H_1\varphi_1, \ldots, H_r\varphi_r] \qquad (2.2)$$

where $H_i$ are $p \times s_i$ matrices of full column rank. Defining $R_i = (H_i)_\perp$, an equivalent expression of these restrictions is given by:

$$R_i'\beta_i = 0 \text{ for } i = 1, \ldots, r \qquad (2.3)$$

We repeat the well known fact that the likelihood of the model is invariant under premultiplication by a full rank matrix $S$.

The Wald condition states a necessary and sufficient condition for identification:

THEOREM 1. *The parameter value $(\beta, \Omega)$ is uniquely identified (up to a normalization of one of the elements in each vector of $\beta$) if and only if for any $i = 1, \ldots r$*

$$rank(R_i'\beta) = r - 1 \qquad (2.4)$$

There are two problems in practice with this theorem: it depends on the a priori unknown parameters $\beta$ and it does not give an indication as how to identify a model if (2.4) fails. The first problem was tackled by Johansen (1995a) , who proved the following theorem:

THEOREM 2. *If the only restrictions imposed on the parameters are (2.3) a set of necessary and sufficient conditions for the parameter value $\beta$ and hence $\Omega$ to be uniquely identified (up to a normalization of one of the elements in each vector $\beta$) is:*

$$rank\left(R_j'\left[H_{k_1}, ..., H_{k_n}\right]\right) \geq n \qquad (2.5)$$

*for* $n = 1, \ldots r - 1$,
*for all* $j \in \{1, \ldots, r\}$,
*and for every set* $\{k_1, \ldots, k_n\}$ *not containing* $j$,



Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

This theorem gives conditions which only depend on the restrictions, not on the parameters. If all the conditions (2.5) are satisfied, then there are $\sum_{i=1}^{r}(p - r - s_i + 1)$ degrees of freedom for testing the hypothesis (2.2).

However if one of the rank conditions (2.5) fails, serious problems arise not just in the interpretation of potential estimates, but in the maximization of the likelihood function and testing process itself. To my knowledge no analytical method exists to determine the number of restrictions imposed by (2.3) on the model.

This paper provides a simple algorithm to determine identifying restrictions, when (2.5) fails. There are four main applications of this algorithm:

(1) A device for counting the number of restrictions in a particular model (if the restrictions are not identifying).
(2) An instrument to be used for estimation algorithms, which require identification. For instance the algorithms of Johansen and Juselius (1994) and Johansen (1995a) require identification. Without identification, they seem to work more than 99% of the time: for automated model selection, this is however not sufficient. Other algorithms for estimation based on general optimization methods, do not require identification (Doornik, 1995).
(3) Only for identified models can (asymptotic) standard errors be given for all estimated parameters. The algorithm finds an identification scheme, which is not necessarily unique. One can thus employ the algorithm to find standard errors of the estimated parameters. If there are multiple identification schemes, we can scan them all (usually there are only a few) and use the standard errors of all schemes to decide where to put additional restrictions.
(4) In the cointegrated VAR model Davidson (1998a) provides an algorithm to find all possible restricted cointegration vectors (using Wald testing). However he only considers one restricted vector at the time, but not a combination of them. Using the results in this paper and likelihood ratio tests, Omtzigt (2001) tests not only one restricted vector at the time, but also all possible combinations of them. The switching algorithm never breaks down, once the model has been generically identified and the automated model selection procedure results in one preferred restricted model only (with possibly equivalent formulations).

For further discussions on the (restricted) simultaneous equations models, we refer to Koopmans et al. (1950), Fisher (1966), Hsiao (1983) and Sargan (1988) and references therein. For potential applications to the I(1) model we refer to Johansen (1995a) and for the I(2) model to Johansen (2000b).

The outline of the paper is as follows. Section 2.2 contains the main theoretical result. In section 2.3 the algorithm is presented and illustrated by means of an example. Section 2.4 provides an empirical illustration of its use in cointegrated VAR models and section 2.5 concludes. The appendices contain all proofs and a Matlab program implementing the algorithm.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

## 2.2. Results

In this paper we shall refer to (2.5) as rank conditions of order $n$. They can be given the following logical ordering (from order 1 to order $r$):

$$rank\left(R_j' H_{k_1}\right) \geqq 1, \quad j \neq k_1 \tag{2.6}$$

$$rank\left(R_j'\left[H_{k_1}, H_{k_2}\right]\right) \geqq 2, \quad j \neq k_1 \neq k_2 \tag{2.7}$$

$$\vdots$$

$$rank\left(R_j'\left[H_{k_1}, H_{k_2}, \ldots, H_{k_{m-1}}\right]\right) \geqq m-1, \quad j \neq k_1 \neq \ldots \neq k_{m-1} \tag{2.8}$$

$$\vdots$$

$$rank\left(R_j'\left[H_{k_1}, \ldots, H_{k_{r-1}}\right]\right) \geqq r-1, \quad j \neq k_1 \neq \ldots \neq k_{r-1} \tag{2.9}$$

There are $r$ rank conditions of order 1, $r(r-1)/2$ rank conditions of order two and $r$ rank conditions of order $r$. In case the rank conditions do not hold, many different ones may fail at the same time. Let $m$ be the lowest order for which at least one rank condition breaks down:

$$rank\left(R_j'\left[H_{k_1}, .., H_{k_m}\right]\right) = m-1, \quad j \neq k_1 \neq \ldots \neq k_m \tag{2.10}$$

The rank deficiency in (2.10) must be exactly one as all the lower rank conditions (2.6)-(2.8) hold and in particular

$$rank\left(R_j'\left[H_{k_1}, .., H_{k_{m-1}}\right]\right) = m-1, \quad j \neq k_1 \neq \ldots \neq k_{m-1}$$

Let the columns of $H_j$ be $h_{j1}, ..., h_{js_j}$ and let $H_{j,-i} = [h_{11}, \ldots h_{1i-1}, h_{1i+1}, \ldots h_{1s}]$, that is $H_j$ without column $h_{ji}$. Furthermore let $k_{ji} = h_{ij} - H_{j,-i}\left(H_{j,-i}' H_{j,-i}\right)^{-1} H_{j,-i}' h_{ji}$.

The following theorem shows that we can always 'repair' this rank condition by deleting one column from matrix $H_j$ and adjusting $R_j$ accordingly. Not any column can be deleted, but at least one of the columns repairs the rank condition.

THEOREM 3. *If (2.6)-(2.8) and (2.10) hold, then for at least one of the columns $h_{ji}$ of $H_j$,*

$$rank\left(\left[R_j, k_{ji}\right]'\left[H_{k_1}, .., H_{k_m}\right]\right) = m. \tag{2.11}$$

Without loss of generality, we shall assume that a condition involving $R_1$ is the first one for which the rank condition fails to hold and that $h_{1d}$ is the column in Theorem 3.

The next theorem shows that we can rotate the columns of any matrix $\beta$ which is restricted as in (2.2) to find a matrix $\beta^*$ which obeys all the previous restrictions implied by (2.2) and the new restriction, caused by shifting $h_{1d}'$ from $H_1$ to $R_1$.

THEOREM 4. *If (2.6)-(2.8) and (2.10) hold and $h_{1d}'$ satisfies (2.11) then for any $\beta = [H_1\varphi_1, ..., H_r\varphi_r]$ there exists almost surely $\beta^* = [H_{1,-d}\varphi_1^*, H_2\varphi_2, ..., H_r\varphi_r]$ such that $sp(\beta) = sp(\beta^*)$.*

The result has been split into two parts on purpose: theorem 3 only involves the restrictions, whereas theorem 4 shows that whatever the parameter value, the additional restriction can be satisfied. This means that we are only putting an extra identifying constraint on the model and do not put additional binding restrictions on it.

The idea of the proof is that if the rank condition of order $m$ fails (and all the lower ones hold), then we can find exactly one linear combination of $(\beta_{k_1}, .., \beta_{k_m})$, say $\gamma$ which lies in the space of $\beta_j$. Let $\beta_j = H_j\varphi_j$ and $\gamma = H_j\psi$. To distinguish $\beta_j$ from $\gamma$ we put one additional restriction on the $\beta_j$.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

## 2.3. Algorithm

Together these last two theorems give rise to an operational algorithm to identify the space, given by any set of restrictions. Each time the rank condition is not satisfied by $(H_1, ..., H_r)$, we are able to take away a column of one of the $H's$ without imposing further restrictions. We repeat the operation until we have identifying restrictions (the algorithm is guaranteed to end as the number of columns of the matrices $H$ is finite).

Formally we propose the following algorithm:

ALGORITHM 1.

(1) *Check the rank conditions (2.6)-(2.9), for identification, starting with the lowest one, (2.6).*
(2) *If all rank conditions are satisfied, go to 4.*
(3) *When the first rank condition is broken, as in (2.10), find a column $h_{ij}$ such that (2.11) is satisfied. Cancel this column from $H_i$ and then go to 1.*
(4) *The space is generically identified*

An implementation of this algorithm in Matlab is available in the appendix of the chapter. Note the loop structure in which all the rank conditions are checked, starting from the lowest one. If a rank condition does not hold, we see which of the columns $h_{ij}$ we can eliminate from $H_i$ to satisfy it. The checking of all the rank conditions then starts again.

**2.3.1. An example.** A detailed example of how the algorithm works in practice clarifies the exact functioning of the algorithm. Among other things it shows that if a rank condition of order $m$ is repaired at step $t$, then at time $t+1$ it may be necessary to repair a lower order rank condition. It is thus absolutely vital that all conditions are checked in each round. The example is also just simple enough to be done by hand, but a computer will just do it quite a bit faster.

Consider the following matrix $\beta$ with 5 rows and 3 columns, on which we impose within-equation restrictions (2.2) by means of the following matrices $H_i$: (Note that of each of the three matrices $H_f$ the columns are mutually orthogonal, such that $h_{fi} = k_{fi}$.):

$$H_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, H_2 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, H_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \tag{2.12}$$

$$H_1 = [h_{11}, h_{12}, h_{13}], H_2 = [h_{21}, h_{22}, h_{23}], H_3 = [h_{31}, h_{32}, h_{33}]$$

As bases of orthogonal complements to these matrices we choose:

$$R_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ -1 & 0 \end{bmatrix}, R_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, R_3 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ -1 & 0 \end{bmatrix}$$

$$R_1 = [r_{11}, r_{12}], R_2 = [r_{21}, r_{22}], R_3 = [r_{31}, r_{32}]$$

The algorithm now runs as follows:

FIRST ROUND

**Check the first-order rank conditions**

$$rank(R_1' H_2) = rank \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} = 1$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

$$rank(R_1'H_3) = rank\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = 1$$

$$rank(R_2'H_1) = rank\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} = 1$$

$$rank(R_2'H_3) = rank\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} = 1$$

$$rank(R_3'H_1) = rank\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = 1$$

$$rank(R_3'H_2) = rank\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = 1$$

**Check the second-order rank conditions**

As all first-order rank conditions are satisfied, we check the second-order rank conditions:

$$rank(R_1'[H_2, H_3]) = rank\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} = 1$$

This rank condition fails, which means that we must apply step 3 of the algorithm.

**Find a column of $H_1$ that satisfies (2.11)**

We add one of the columns of $H_1$ to $R_1$ and see whether this particular rank condition is repaired. Try $H_1^* = [h_{12}, h_{13}]$ and $R_1^* = [r_{11}, r_{12}, h_{11}]$. The rank condition becomes:

$$rank(R_1^{*\prime}[H_2, H_3]) = rank\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \end{pmatrix} = 2$$

The rank condition is now satisfied and we take $H_1 = H_1^*$ and $R_1^* = R_1$ (leaving the other matrices as they were before) and start the algorithm at point 1:

**SECOND ROUND**

**Check the first order rank conditions**

$$rank(R_1'H_2) = rank\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 1$$

$$rank(R_1'H_3) = rank\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 0 \end{pmatrix} = 2$$

$$rank(R_2'H_1) = rank\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0$$

This rank condition fails.

**Find a column of $H_2$ that satisfies (2.11)**

When we move the first column of $H_2$ to $R_2$ we obtain the following candidates $H_2^* = [h_{22}, h_{23}]$ and $R_1^* = [r_{21}, r_{22}, h_{21}]$. The rank condition then reads:

$$rank(R_2^{*\prime}H_1) = rank\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = 0$$

It is still not satisfied, so we try shifting the second column of $H_2$: $H_2^* = [h_{21}, h_{23}]$ and $R_2^* = [r_{21}, r_{22}, h_{22}]$. This results in the following rank condition:

$$rank(R_2^{*\prime}H_1) = rank\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix} = 1$$

The rank condition now holds and we take $H_2 = H_2^*$ and $R_2 = R_2^*$ to go back to step 1 of the algorithm:

**THIRD ROUND**

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

**Check the first-order rank conditions**

It is easily verified that of all the first-order rank conditions are satisfied, with the exception of

$$rank(R_3'H_2) = rank\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0$$

**Find a column of $H_3$ that satisfies (2.11)**

Shifting the first column of $H_3$ to $R_3$ would clearly not work, as that would imply $sp(H_2) = sp(H_3)$. (In this case even $H_2 = H_3$). We therefore shift the second column of $H_3 : H_3^* = [h_{31}, h_{33}]$ and $R_3^* = [r_{31}, r_{32}, h_{32}]$. The rank condition is now satisfied:

$$rank(R_3^{*\prime}H_2) = rank\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix} = 1$$

For the next round take $H_3 = H_3^*$ and $R_3 = R_3^*$.

**FOURTH ROUND**

**Check the first and second-order rank conditions**

All 6 first order and 3 second order rank conditions are satisfied, such that we conclude that the restrictions identify the model: The conditions of Theorem 3 in Johansen (1995a) now hold for this example

For completeness we shall also give the matrices $S$ from Theorem 4. If we have the matrices (2.12), then we can write the matrix $\beta$ as:

$$\beta = \begin{bmatrix} \varphi_{11} & 0 & \varphi_{31} \\ 0 & \varphi_{21} & \varphi_{32} \\ \varphi_{12} & \varphi_{22} & 0 \\ \varphi_{13} & \varphi_{23} & \varphi_{33} \\ \varphi_{11} & 0 & \varphi_{31} \end{bmatrix}$$

The combination $\varphi_{32}\beta_2 - \varphi_{21}\beta_3 \equiv \gamma \in sp\,(H_1)$. Post-multiplying $\beta$ by the full-rank matrix

$$S_1 = \begin{bmatrix} 1 & 0 & 0 \\ \frac{\varphi_{32}}{\varphi_{31}} & 1 & 0 \\ -\frac{\varphi_{11}}{\varphi_{31}} & 0 & 1 \end{bmatrix}$$

gives way to

$$\beta^* = \begin{bmatrix} 0 & 0 & \varphi_{31} \\ 0 & \varphi_{21} & \varphi_{32} \\ \varphi_{12}^* & \varphi_{22} & 0 \\ \varphi_{13}^* & \varphi_{32} & \varphi_{33} \\ 0 & 0 & \varphi_{31} \end{bmatrix} \tag{2.13}$$

which satisfies the restrictions after the first round of the algorithm. Note that this transformation is not defined if $\varphi_{22} = 0$ or $\varphi_{31} = 0$.

Taking away the stars in the last expression, we can post-multiply again by

$$S_2 = \begin{bmatrix} 1 & \frac{\varphi_{22}}{\varphi_{12}} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

to obtain a matrix, satisfying the restrictions at the end of the second round. This step inserts a zero in place of $\varphi_{22}$ in (2.13). In the last step, the matrix $S_3$ is given by:

$$S_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{\varphi_{32}}{\varphi_{21}} \\ 0 & 0 & 1 \end{bmatrix}.$$

Post-multiplication leads to the following general matrix:

$$\beta = \begin{bmatrix} 0 & 0 & \varphi_{31} \\ 0 & \varphi_{21} & 0 \\ \varphi_{12} & 0 & 0 \\ \varphi_{13} & \varphi_{32} & \varphi_{33} \\ 0 & 0 & \varphi_{31} \end{bmatrix} \tag{2.14}$$

which satisfies all the rank conditions and is therefore generically identified.

**2.3.2. Discussion.** Making a change in a broken rank condition can cause a previously satisfied rank condition to fail. In the example above, all rank conditions of first order are satisfied in the first round, but the change made causes first-order rank conditions to fail subsequently. This demonstrates that in every round we have to start checking the lowest order rank conditions.

In the second round, we note that not any column can be eliminated from $H$, but we can still choose between deleting the second and the third column. This implies that the restrictions imposed by the algorithm are in general not unique. We thus find but only one of many ways to identify this space. It may be hard to attach an economic meaning to a particular identification in any one application. In some way this is the only weak point of the algorithm: in automatic search algorithms and other applications, the researcher may look for an different identification scheme to make economic sense of it. This however can easily be achieved by making available all equivalent identification schemes.

## 2.4. An application

As a practical application we consider the $p$-dimensional cointegrated VAR-model with $k$ lags:

$$\Delta X_t = \alpha\beta' \begin{pmatrix} X_{t-1} \\ t \end{pmatrix} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Psi d_t + \mu + \varepsilon_t \tag{2.15}$$

where $\beta'$ are the cointegration vectors, $t$ is a time trend and $d_t$ are dummy variables.

We note that the likelihood function depends on $\Pi = \alpha\beta'$. This means that we can take $\alpha^* = \alpha\kappa$ and $\beta^{*\prime} = \kappa^{-1}\beta'$, where $\kappa$ is any invertible matrix. The likelihood is unchanged after this transformation as $\Pi^* = \Pi$. We thus have exactly the same identification problem for $\beta$ in this model as in the simultaneous equations model (2.1).

For tests of the kind (2.2), a number of computer packages have implemented the switching algorithm of Johansen and Juselius (1994) , henceforth JJ. CATS[1] by Hansen and Juselius (1994) performed better than PcFiml version 9.3[2] by Doornik and Hendry (1997) and my own implementation of the switching algorithm (which does not put identifying restrictions): The maximum in the likelihood function CATS found was the highest in all the examples considered below.

We compare the CATS implementation of the switching algorithm of JJ (which does not put identifying restrictions on the cointegration space) and the implementation of a Matlab program,

[1]We use version 1 of this CATS in RATS 4.3

[2]Version 10.1 of PcGive does not give the user the option to use beta switching. We thus used the latest version of the program that did.

| Hypothesis tested | m | y | p | i3 | i10 | t | CATS LR-test | CATS # iterations | New Algorithm LR-test | New Algorithm # iterations |
|---|---|---|---|---|---|---|---|---|---|---|
| $\mathcal{H}_1$ : | 1 | -1 | -1 | 0 | 0 | * | 11.30 | 36 | 11.28 | 49 |
| $\mathcal{H}_2 : \mathcal{H}_1 +$ | 0 | 0 | 0 | * | * | * | 12.63 | 93 | 12.61 | 80 |
| $\mathcal{H}_3 : \mathcal{H}_1 +$ | 0 | 0 | 0 | 1 | -1 | * | 15.88 | 200 | 13.44 | 53 |
| $\mathcal{H}_4 : \mathcal{H}_1 +$ | 0 | 0 | 0 | 1 | -1 | 0 | 16.62 | 2 | 16.10 | 158 |
| $\mathcal{H}_5 : \mathcal{H}_1 +$ | 0 | 0 | 0 | 1 | * | 0 | 15.20 | 195 | 15.15 | 87 |
| $\mathcal{H}_6 : \mathcal{H}_1 +$ | 0 | 0 | 0 | 0 | 1 | * | 17.28 | 200 | 17.13 | 200 |
| $\mathcal{H}_7 : \mathcal{H}_1 +$ | 0 | 0 | 0 | 1 | 0 | * | 16.37 | 197 | 16.32 | 152 |
| $\mathcal{H}_8 : \mathcal{H}_1 +$ | 1 | 0 | * | 0 | * | 0 | 11.40 | 89 | 11.62 | 200 |

Table 2.1: Comparison between optimization in CATS (without identification) and optimization with identification

which executes the switching algorithm after having imposed identifying restrictions by means of algorithm 1. We consider an Australian data set, first analyzed by JJ and also used by Doornik (1995) to illustrate his alternative numerical method. It consists of the log of nominal money (m), the log of real national income (y), the log of the GDP deflator (p), a three month interest rate (i3) and the 10 year government bond rate (i10). $d_t$ contains centered seasonal dummies and a dummy which takes value 0 until 1982, 2nd quarter and 1 afterwards.

JJ fit a VAR with 2 lags for the period 1976-1 until 1991-1 (61 effective observations). The trace and rank test point to a rank of at most one, but JJ choose three as their preferred rank.They test a a number of hypotheses, which we have tested in PcGive 9.3, PcFiml 10.1, CATS and our own program. All of them give identical answers. We then considered testing a number of hypothesis, where restrictions were put on only one or two of the cointegration vectors. By definition these restrictions are not identifying. In the table 2.1 we report the results of testing that the inverse velocity of money (m-y-p) is trend stationary on its own and in combination with all possible cointegration relations between the interest rates and an unrelated hypothesis $\mathcal{H}_8$. Both algorithms have the same convergence criterium and number of maximum iterations (200).

For hypothesis $\mathcal{H}_1 - \mathcal{H}_7$ the new algorithm in Matlab does remarkably better. It needs less iterations and finds a higher maximum in the likelihood, resulting in a lower Likelihood Ratio test statistic. Unlike CATS it also reports the degrees of freedom of the likelihood ratio test. Just to show that there is no mathematical guarantee, we also report $\mathcal{H}_8$, where CATS does better than the new method.[3] We have checked these LR-tests against all other methods implemented in PcGive 9.3, namely linear switching (Boswijk, 1995) and the Broyden-Fletcher-Goldfarb-Shanno method (Doornik, 1995) . Neither of them found better maxima. (and the first one did notably worse in cases $\mathcal{H}_6 - \mathcal{H}_7$). Both these methods have the advantage that they are able to cope with more general restrictions on both $\alpha$ and $\beta$.

A partial explanation for the result is the difficulty of the the optimization problem at hand. We impose three cointegration relationship, where the evidence of the second and third is weak, such that those relationships are hard to find in the data. Prices and money are often modelled as I(2) and this would probably be better in the current data set as well. And the hypotheses tested are all soundly rejected at the 5% level by any method. The likelihood in the region we are searching is extremely flat. Yet this is the ideal situation to put algorithms to the test. In easy situations all of them find the same maximum, which in all likelihood is the global one.

[3]If the maximum number of iterations is lifted, it does find a maximum after 725 iterations for an LR-test statistic of 11.35

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

## 2.5. Conclusions

We have presented a way of identifying an under-identified parameter space in simultaneous equations models and hence rendered estimation by the means of the switching methods of Johansen (1995a) possible. In over $10^6$ test executed so far in simulations Omtzigt (2002b), the method has not broken down once, such that it is ideal in automated model selection. It is reasonably fast, calculates the degree of freedoms for the likelihood ratio test automatically and analytically (no separate procedure is needed) and allows for calculating standard errors on all the estimated parameters. The procedure is very easy to implement and a Matlab version is attached to this chapter.

## 2.A. Proofs

The following lemma is needed for the proof of Theorem 3:

LEMMA 1. *If the rank conditions (2.6)-(2.8) hold*

$$rank\left(\left[H_{k_1},\ldots,H_{k_j}\right]\right) \geqq j, j=1,\ldots,m-1,m \tag{2.16}$$

PROOF. For $j=1,\ldots,m-1$ the result follows directly from (2.6)-(2.8): for instance (2.8) implies that

$$m-1 \leq rank\left(R_j'\left[H_{k_1},H_{k_2},\ldots,H_{k_{m-1}}\right]\right) \leq rank(H_1,\ldots,H_{m-1})$$

such that

$$m-1 \leq rank(H_1,\ldots,H_{m-1})$$

For $j=m$ let us assume that the lemma does not hold, i.e. that $rank(H_1,\ldots,H_m) \leq m-1$. We find

$$m-1 \leq rank(H_1,\ldots,H_{m-1}) \leq rank(H_1,\ldots,H_m) \leq m-1 \tag{2.17}$$

such that equality holds throughout and $rank(H_1,\ldots,H_m)=m-1$ This leads to the existence of $h_1,\ldots,h_{m-1} \in sp(H_1,\ldots,H_m)$ so that $H_i=(h_1,\ldots,h_{m-1})M_i$. From (2.17) we see that

$$m-1 = rank(H_1,\ldots,H_{m-1}) = rank(h_1,\ldots,h_{m-1})$$

such that $(H_1,\ldots,H_{m-1})_\perp = sp(h_1,\ldots,h_{m-1})_\perp$, and hence $(H_1,\ldots,H_{m-1})'_\perp H_m = 0$, which contradicts (2.8), since $H_m$ is a non-null matrix. □

PROOF OF THEOREM 3. By lemma 1 we know that

$$rank\left(\left[H_{k_1},..,H_{k_m}\right]\right) \geqq m \tag{2.18}$$

$$rank\left(\left[R_j,H_j\right]'\left[H_{k_1},..,H_{k_m}\right]\right) \geqq m$$

as $(R_j,H_j)$ is a matrix of full rank. As $H_j$ is of full column rank, $\left[k_{j1},\ldots,k_{js_j},R_j\right]$ is a square, full rank matrix, which together with (2.18) implies that

$$rank\left(\left[k_{j1},\ldots,k_{js_j},R_j\right]'\left(H_{k_1},..H_{k_m}\right)\right) \geqq m$$

This combined with (2.10) means that

$$rank\left(\left[R_j,k_{ji}\right]'\left[H_{k_1},..,H_{k_m}\right]\right) = m$$

for at least one column of $H_j$. □

We note that

$$rank\left(k_{1d}'\left[H_{k_1},..H_{k_m}\right]\right) = 1 \tag{2.19}$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

PROOF OF THEOREM 4. $rank(R_1' [\beta_{k_1}, ..., \beta_{k_m}]) = m - 1$, (2.10), implies that there exists an $m \times 1$ vector $a_\perp$, such that $R_1' [\beta_{k_1}, ..., \beta_{k_m}] a_\perp = 0$. Without loss of generality we assume that $k_1 = 2, \ldots, k_m = m + 1$.

Therefore $[\beta_2, ..., \beta_{m+1}] a_\perp \equiv \gamma \in sp(H_1)$
As $rank([R_1, h_{1d}]' [\beta_2, ..., \beta_{m+1}]) = m$, $h_{1d}'\gamma \neq 0$. This implies that we can take $\beta_1^* = \beta_1 - \gamma (h_{1d}'\gamma)^{-1} h_{1d}'\beta_1$. This transformation is of the kind $\beta^* = \beta S$, where

$$S = \begin{bmatrix} 1 & 0 & 0 \\ -a_\perp (h_{1d}'\gamma)^{-1} h_{1d}'\beta_1 & I_m & 0 \\ 0 & 0 & I_{r-m-1} \end{bmatrix}$$

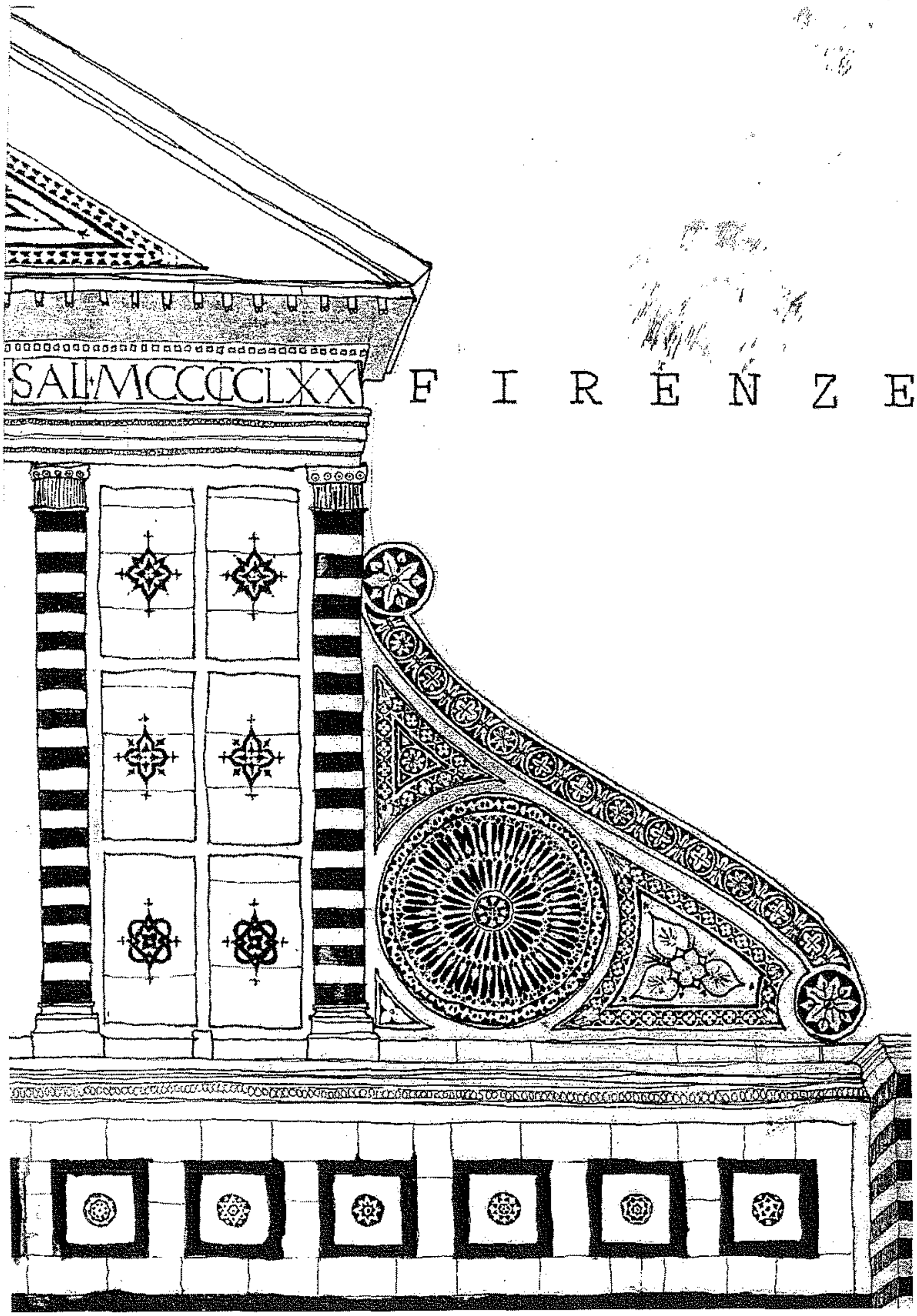
This matrix does not exist if $(h_{1d}'\gamma) = 0$, but this only happens on a set of Lebesgue measure zero. When it exists it is clearly of full rank, which means that $sp(\beta) = sp(\beta^*)$ □

## 2.B. Matlab program

```
function [Hblockout,Rblock] = identify(Hblock)


% For a given set of linear restrictions of the kind
% beta =[Hblclock{1}*phi1,...,Hblock{r}*phir]
% (without normalizations), this function provides an equivalent
% identifying set of restrictions Hblockout


r = size(Hblock,2);
p =size(Hblock{1},1);
% Get the orthogonal complements (see equation (2.3))
for f=1:r
 Rblock{f}= null(Hblock{f}');
end


% The main loop of the program
identification = 0;
% As long as there is no identification run the following loop
while identification == 0
 [Hblock,Rblock,identification] = mainloop(Hblock,Rblock,r);
end




%*********************************************
% Internal function:
%*********************************************
function [Hblock,Rblock,identification] = mainloop(Hblock,Rblock,r)
identification = 1;
% Set identification flag to one.  If one the rank conditions fails
% we repair it and set it to zero (no identification)
% Start with rank condition of order 1 (for which k=2)
 M = nchoosek(1:r,k);
 % one of the indices,j, on the left (R) others (in C) on the right (H's)
 for j=1:size(M,1)
  for m=1:k
   C = setdiff(M(j,:),M(j,m));
   right = zeros(size(Hblock{1},1),0);
```

```
   for m2=1:k-1
   right = [right,Hblock{C(m2)}];
  end
  % Check whether rank condition is satisfied.
  if rank(Rblock{M(j,m)}'*right, 0.00001)< k-1
   % if not, check which column of H can be shifted
   sizeH = size(Hblock{M(j,m)},2);
   H = Hblock{M(j,m)};
   for s2 = 1:sizeH
   H(:,1:s2-1);
   H(:,s2+1:sizeH);
   testblockH = [H(:,1:s2-1),H(:,s2+1:sizeH)];
   testblockR = null(testblockH');
   if rank(testblockR'*right, 0.00001) == k-1
    %this column can be shifted!
    Hblock{M(j,m)}=testblockH;
    Rblock{M(j,m)}=testblockR;
    identification = 0; % no identification
   %model has been changed, such that there is
   %no guarantee all rank conditions are satisfied
    break,end
   end
  end
 end
 end
end
```

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609




F I R E N Z E


DOI: 10.2870/52609

CHAPTER 3

# Automatic identification and restriction of the cointegration space

## 3.1. Introduction

The introduction of cointegration (Engle and Granger, 1987) has led to the development of a wide variety of methods to analyze cointegrated systems. The method of Johansen (1988, 1991) is frequently applied as part of a wider General-to-Specific Modelling strategy which is advocated by the LSE-school in econometrics.

The process of arriving at the specific model is often long and arduous, especially in a VAR with more than 3 variables. Many sequential decisions are taken in the modelling process: which dummy should be included? which accepted restrictions on the cointegration parameters should be tested jointly? Criticism against the LSE methodology often targets these procedures as leaving too many options for the individual researcher and turning the process into an art form instead of science. From the LSE practitioner point of view, the process is tedious and indeed difficult to replicate. Starting with a particular data set once analyzed, one does not necessarily take the same decisions again and it is quite hard to exactly replicate the analysis.

After the article by Lovell (1983), data mining and automated model selection were seen in a bad light. Recently Hoover and Perez (1999) and Hendry and Krolzig (2003) re-ran the original experiment, changed the decision rules and found that it was possible to recover the original DGP with a very high probability. The last article also contains an exhaustive justification of automated modelling and the LSE-framework.

Whereas the three cited articles in the last paragraph are concerned with selecting the right variables in a single equation regression, Brüggeman and Lütkepohl (2001) consider lag length selection in a vector autoregressive model. Davidson (1998a) automates finding the zero-restrictions[1] in the cointegration space. His aim is very close to the one pursued in this paper. Yet his method differs from ours in a number of key elements, which will be fully discussed in section 3.5.4.

In this chapter we will consider the cointegrated VAR-model and automate the search for some within-equation restrictions in the cointegration space.

In the next section we will outline the model, show how identification and restriction are intimately connected and describe the standard modelling strategy employed. In section 3.3 the algorithm is stated. A worked-out example, a Monte Carlo simulation and an application to UK money demand data then amply illustrate its use. Extensive comments on the uses of the procedure are given before the concluding remarks in the last section. Proofs follow in the only appendix to this chapter, section 3.A.

## 3.2. The cointegrated VAR model

The $p$-dimensional cointegrated VAR model is given by:

$$\Delta X_t = \alpha\beta' \begin{pmatrix} X_{t-1} \\ D_t \end{pmatrix} + \sum_{i=1}^{k-1} \Gamma_i \begin{pmatrix} \Delta X_{t-i} \\ \Delta D_{t-i} \end{pmatrix} + \Phi d_t + \varepsilon_t \tag{3.1}$$

$$\varepsilon_t \sim iidN(0, \Omega), \qquad t = 1, \dots, T$$

[1]That is he finds overidentifying restrictions in the cointegration space, but only exclusion-restrictions.



Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

where $\alpha$ is of dimension $p \times r$ and $\beta$ is of dimension $q \times r$ and $D_t$ a vector with deterministic variables, which are included in the cointegration space. The lagged differenced variables inside the cointegration space are explicitly included outside the cointegration space. Some authors and computer packages take this convention (CATS in RATS, Hansen and Juselius 1994), whereas others do not include $\Delta D_{t-i}$, the lagged differenced deterministic variables, outside the cointegration space (like PcFiml, Doornik and Hendry 1997). The $d_t$ contains other dummies like seasonal and blip[2] dummies.

Neither $\alpha$ nor $\beta$ is identified, but only their product $\Pi = \alpha\beta'$ is. The likelihood does not change if we take $\alpha^* = \alpha\kappa$ and $\beta^{*\prime} = \kappa^{-1}\beta'$ for an arbitrary invertible matrix $\kappa$. We assume that all the identifying restrictions are put on $\beta$ and that there are either no restrictions on $\alpha$ or only restrictions of the kind $\alpha = Ha$. The last kind of restrictions, which corresponds to weak exogeneity of $H_{\perp}'X_t$ for $\alpha$ and $\beta$, does not bring any identification onto the system.

For $\beta$ we consider only the following restrictions:

$$\beta = (H_1\varphi_1, \psi) \tag{3.2}$$

or

$$\beta = (H_1\varphi_1, \ldots, H_r\varphi_r) \tag{3.3}$$

where $H_i$ are $q \times s_i$ matrices, $s_i \leq q + 1 - r$.

**3.2.1. Identification and restriction.** Identification and restriction are so intimately related, that it is impossible to separate them out completely. Let's take the following cointegration space in a DGP as an example:

$$\beta'_{DGP} = \begin{bmatrix} 1 & 3.2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

If we decide to identify the space by putting an identity matrix in the top of $\beta$:

$$\beta'_{M1} = \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & c \end{bmatrix}$$

then by successive testing whether $a = 0$ and or $c = 0$ we are unable to recover the DGP, as the statistical model does not contain the DGP as a special case. If we had started with the following identification scheme however:

$$\beta'_{M2} = \begin{bmatrix} 1 & d & 0 \\ 0 & e & 1 \end{bmatrix}$$

we would have been able to recover the DGP. As it is impossible to know the DGP before hand, we shall have to treat the problem of identification and restriction contemporaneously in the algorithm.

The algorithm we propose shall mimic the strategy which Juselius (2002) explicitly employs in her papers. First she finds which restrictions of the kind (3.2) are accepted. She then combines the accepted restrictions to (over)-identify the whole parameter space (3.3) and reports the final combination which is supported by the data. Usually she will test a great many of these combinations, but for lack of space it is normally only possible to report one or two final models. We shall systematically search for such combinations.

**3.2.2. Modelling Strategy.** This paragraph gives a short overview as how a small cointegrated VAR is often modelled in an LSE-type fashion. This is no definite guide, but more an attempt to describe current practice. We refer once again to the work of Juselius as an example. The modelling process can graphically be represented as in figure 3.1.

[2] They take value 1 in one (or two) periods and are zero elsewhere.

 DOI: 10.2870/52609

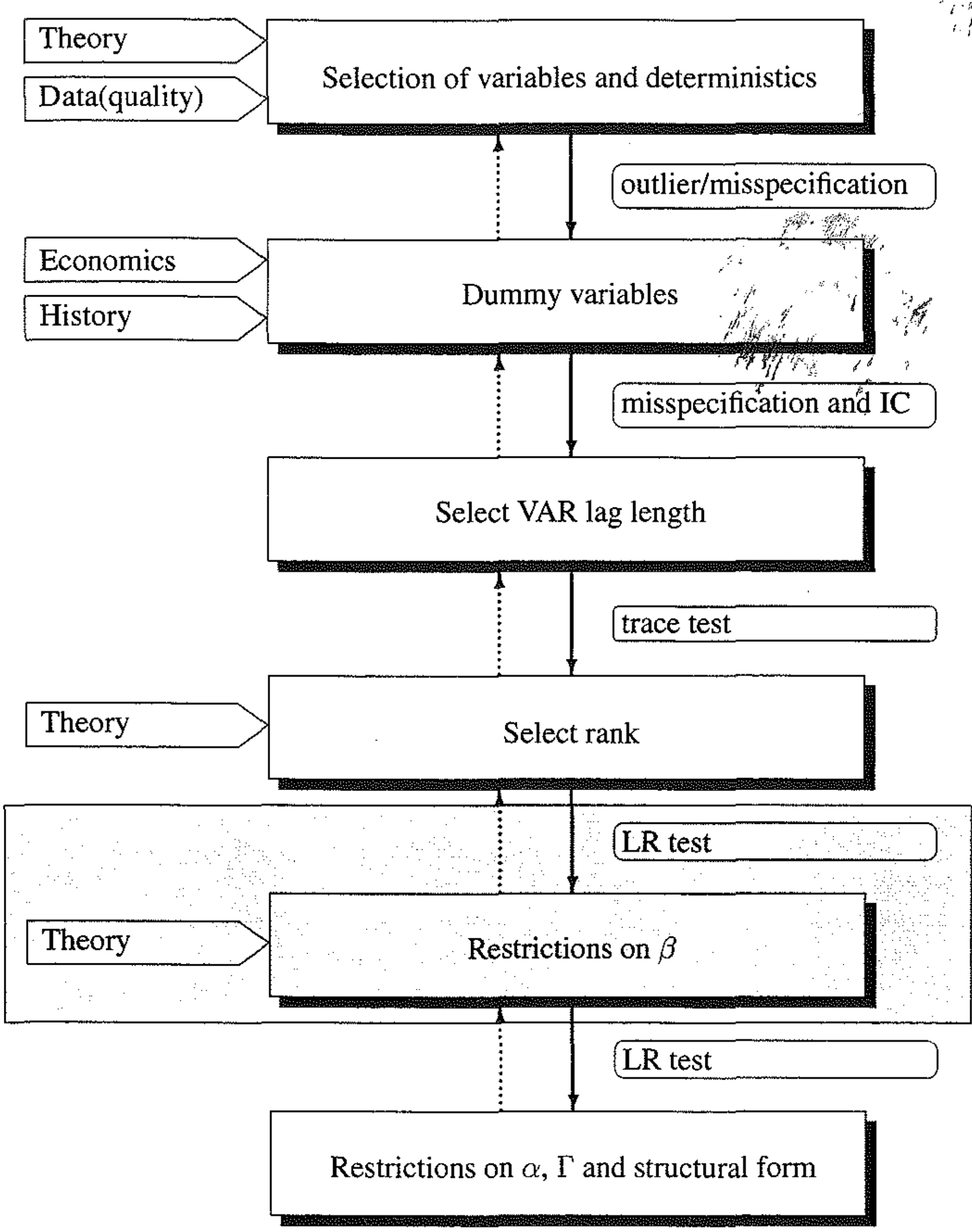



Figure 3.1: Modelling a cointegrated VAR

(1) The selection of variables is based on the theories to be tested or the part of the economy to be modelled. If money demand and monetary transmission mechanisms are the object of study, then logically money, an interest rate, income and prices/inflation should be part of the data vector to get a meaningful model: these variables are dictated by the standard

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

textbook theories on money. Money demand can for instance depend on inflation, total income and the relevant opportunity cost, which is measured by means of an interest rate. Often many measures of money (and income) are available. When abstract theories do not offer a convincing choice, data quality often does. Series may contain breaks in the form of changes of definition, changes of bank concerned etcetera. Long series of high quality are often hard to find.
An additional problem is possible I(2)-ness in the data. If an I(1) model is to be preferred, then some pretesting is warranted. This can take the form of unit root tests on series. The augmented Dickey Fuller test and the Phillips-Perron test are often used. Alternatively one can do a full I(2)-rank test. If at least one I(2)-trend is found, a test for a so-called nominal-to-real transformation (Kongsted, 1998, 2002) can be executed to transform the model into an I(1) model.
In choosing the deterministics of the model, preference is often given to so-called star-models, which allow for the same kind of trend in both the stationary and the non-stationary direction. This means either a constant in the cointegration space or a trend in the cointegration space $D_t$ and a constant (differenced trend) outside. This fits in naturally with our formulation of the problem. For these specifications, the rank test is asymptotically similar with respect to the actual trends in the data, such that rank and deterministics determination can be separated (Nielsen and Rahbek, 2000). Restrictions on the trend/constant can be tested in 5 and 6.

(2) The procedure we are using is based on a Gaussian likelihood function. If there are large outliers, then these have to be modelled explicitly. Tests on the residuals will reveal whether any misspecification is present. Outliers often have a meaning and several solutions to outlier problems can be employed. If the outliers are caused by the say the oil shocks, then the economic and historical knowledge can lead either to the inclusion of one or two dummies for these shocks or to the inclusion of the oil price as an exogenous variable in the model.
The dotted arrow between dummy variables and selection of variables on the one hand and deterministics on the other indicates the idea that it is possible to go back one stage in the modelling process and rethink which variables should be included. In fact at any stage in the modelling process the researcher can go back to previous stages.

(3) If the model is well-specified, the lag length of the process is set: in the model the lag length is equal on all variables (though this can be changed later). This can be done either on the basis of an information criterion like Aikaike's Information Criterion or by testing successively that the coefficients to the last lag present in the model are zero.

(4) The rank of the matrix $\Pi$ is selected using the trace test Johansen (1995b) . The maximum eigenvalue is not frequently used. If there are strong theoretical priors they can be included in the procedure, see Paruolo (2001) .

(5) Restrictions on the cointegration space are tested by means of likelihood ratio test. These restrictions are usually motivated by economic theory. If real money and real income are the first two variables in our data vector and we want to consider whether velocity of money is stationary, we test an hypothesis of the kind (3.2) with $H_1 = \begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix}^3$. Consequently accepted hypotheses are combined to restrict the whole cointegration space.

(6) Restrictions on the short run parameters $\alpha, \Gamma$ or on the so-called structural form, can be tested for known $\beta$, as the last are estimated in a superconsistent way and the former are not. If interest only lies in testing theories which involve just the cointegration parameters, then this step can be omitted.

[3] For convenience we have assumed 4 variables and at least a constant outside the cointegration space

 DOI: 10.2870/52609

The various parts of the modelling process do not take up equal amounts of time. In applied work especially number 5, the testing of restrictions often requires a large amount of time. This part will be object of study in the next paragraph and then be automated.

## 3.3. Identification and restriction of $\beta$

Current practice is to find sets of individual restrictions, based on economic theory. One vector is restricted and the other vectors are left to vary freely. We thus have a test of the kind (3.2). Hypotheses, which are accepted at a set significance level (often 5%) are then tested jointly. To my knowledge, few authors, if any, actually report which combinations they have tested and why and how they have selected the final model. Usually one reports only the final combination of restrictions accepted and the p-value associated with it.

The following algorithm mimics and automates this process. It is graphically represented in figure 3.2.

(1) Let the user specify $q \times 1$ restriction columns of the kind $f_i = \begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix}', i = 1, \ldots, F$. These can be interpreted as new variables, that might be stationary themselves or enter into a stationary relationship. Examples include the real rate of interest, velocity and interest rate spread.

(2) Letting $e_j = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}', j = 1, \ldots, q$ denote the unit vector with 1 in the $j$th position, we take all possible matrices $H_k, k = 1, \ldots, K$ whose columns are combinations of $f_i, i = 1, \ldots, F$ and $e_j, j = 1, \ldots, q$ such that:
   (a) The number of columns of $H_k$ is smaller than or equal to $q + 1 - r$ (this ensures that each of these matrices will put at least one over-identifying restriction).
   (b) $H_k$ is of full column rank
   (c) $sp(H_k) \neq sp(H_l)$ for $k \neq l$ (because they would represent the same restriction).
   (d) If the system contains both stochastic and deterministic variables, then that part of $H_k$, which premultiplies the stochastic variables contains at least one non-null element. We thereby avoid testing whether there exists a stationary combination among the deterministic variables only.

(3) For $k = 1, \ldots, K$ test whether $\beta = (H_k\varphi_1, \psi)$ is accepted or not. Reject if the $p$-value $p(k)$ is smaller than 1%. Define $C_1 = \{1, \ldots, i, j, \ldots c_1\}$ as the ordered set of accepted restrictions, that is $p(i) > p(j)$ if $i < j$. If $r = 1$, go to step 7

(4) For every combination $(\{i, j\}, i < j, i \in C_1, j \in C_1)$ test whether $\beta = (H_i\varphi_1, H_j\varphi_2, \psi)$ is accepted or not at the 1% level. Define $C_2 = \{\{i, j\}_l, l = 1, \ldots, c_2\}$ as the set of restrictions accepted. If $r = 2$, go to step 7

(5) For every combination of three restrictions $(\{i, j, k\}, \{i, j\} \in C_2, k \in C_1, i < j < k)$ test whether $\beta = (H_i\varphi_1, H_j\varphi_2, H_k\varphi_3, \psi)$ is accepted or not at the 1% level. Define $C_3 = \{\{i, j, k\}_l, l = 1, \ldots, c_3\}$ as the set of restrictions accepted.

(6) Repeat step 5 until $C_r$ or until one of the $C$'s is an empty set.

(7) Select the final model in the following way
   (a) Among the families of sets $B = \{C_1, C_2 \ldots C_r\}$ select all the models, which are accepted at the 5% level, which defines the set $B_2$.
   (b) Order all the models in $B_2$ according to the number of accepted restrictions, starting with the one with the highest number of accepted restrictions. In case of parity, rank the one with the highest $p$-value first.

### 3.3.1. Remarks on the algorithm.

A number of comments on the algorithm will clarify certain choices inside the procedure and replication of the experiment.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

| | User | | Program | | | | |
|---|---|---|---|---|---|---|---|
| | $f_1$ | $f_2$ | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_5$ |
| $m-p$ | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| $\Delta p$ | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| $i$ | 0 | $-1$ | 0 | 0 | 1 | 0 | 0 |
| $y$ | $-1$ | 0 | 0 | 0 | 0 | 1 | 0 |
| $trend$ | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

↓

Take all combinations of vectors $\{f_1, f_2, e_1, \ldots, e_5\}$ to obtain $H_1, \ldots, H_K$
- Max $p - r + 1$ vectors
- No overlap, such that if $sp(H_i) = sp(H_j)$, then only include one of them: here include only one of $[e_1, e_4], [f_1, e_1]$ and $[f_1, e_4]$
- Exclude possibility that deterministic term is cointegration relations: exclude $[e_5]$
- $H_i$ must have full column rank: exclude for instance $[e_1, e_4, f_1]$

↓

Test $\beta_i = [H_i\varphi, \psi]$ for $i = 1, \ldots, k$
Accept if p-value $\geq 0.01$
Create ordered (highest $p$-value first) set of accepted restrictions $C_1 = \{1, \ldots, c_1\}$

↓

Test $\beta_i = [H_i\varphi, H_j\varphi, \psi]$ for $i \in C_1, j \in C_1, i < j$ and only if $\{H_i, H_j\}$ generically identify a bi-dimensional space (use test Johansen 1995a)
Accept if p-value $\geq 0.01$
Create ordered (highest p-value first) set of accepted restrictions $C_2 = \{\{i,j\}_l\, l = 1, \ldots, c_2\}$

↓

Test $\beta_i = [H_i\varphi, H_j\varphi, H_k\varphi, \psi]$ for $\{i, j\} \in C_2, k \in C_1, k > \max(i, j)$ and only if $\{H_i, H_j, H_k\}$ generically identify a tridimensional space (use test Johansen 1995a)
Accept if p-value $\geq 0.01$
Create ordered (highest p-value first) set of accepted restrictions $C_3 = \{\{i,j\}_l\, l = 1, \ldots, c_3\}$

↓

.........

↓

Test $\beta_i = [H_i\varphi, H_{r-1}\varphi, \ldots, H_r\varphi]$ for $\{i, j, \ldots, r-1\} \in C_{r-1}, r \in C_1, r > \max(i, j, \ldots, r-1)$ and only if $\{H_i, H_j, \ldots, H_r\}$ generically identify an $r$-dimensional space
Accept if p-value $\geq 0.01$
Create ordered (highest p-value first) set of accepted restrictions $C_r = \{\{i, j, \ldots, r\}_l\, l = 1, \ldots, c_r\}$

↓

Combine family of sets $B_1 = \{C_1, \ldots, C_r\}$ select all models, which are accepted at 5% level: $B_5$
Rank models in $B_5$ according to the following criteria:
- highest number of overidentifying restrictions
- in case of parity highest p-value first

Figure 3.2: Algorithm for automatically restricting the cointegration space

- Every time we take a combination of restrictions (as in step 4, 5 and 6), we have to check whether the restrictions are generically identifying. This is done by checking the condition in Johansen (1995a). If the restrictions implied by say $H_1$ and $H_2$ are not generically identifying, then we can eliminate one column of $H_1$ or $H_2$ and still have

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

the same model, see Omtzigt (2002b) . As an example consider that in step 3, we have accepted both $H_1 = [e_1, e_2]$ and $H_2 = [e_2]$. Then testing $\beta = (H_1\varphi_1, H_2\varphi_2)$ is equivalent to testing $\beta = (H_3\varphi_1, H_2\varphi_2)$, where $H_3 = [e_1]$. In the algorithm above, this combination is only tested if both $H_1$ or $H_3$ are accepted in step 3.

- There is often more than one way to impose the same restrictions. If we define $f_1 = \begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix}$ and in the DGP the fist two variables are stationary and the other two are not (and not cointegrated), then we may accept three different combinations of $H$'s:$\{[f_1], [e_1]\}$,$\{[f_1], [e_2]\}$ and $\{[e_1], [e_2]\}$. In the final step one should check whether models are equivalent.
  Furthermore if one of the restrictions is rejected at an intermediate stage (say $[e_1]$), then the other restrictions could still be accepted (say both $[f_1]$ and $[e_2]$) such that it is still possible to recover at least one equivalent model. Thus the existence of equivalent models can increase the chance of accepting the DGP as the preferred model.

It is possible to prove the following theorem on the asymptotic size and power of the complete procedure

THEOREM 5. *The asymptotic size of the algorithm is smaller than or equal to* $0.03 + 0.02r$, *while the asymptotic power is* $1$.

The last theorem sets an upper limit to the size of the test procedure. It is very likely that the true size is considerably lower for a number of reasons:

(1) The results of the individual tests in the procedure are most likely positively correlated.
(2) There may be more than one set of $H$'s which identifies the model, as was seen in one of the previous comments: just one of them needs to be accepted.
(3) In 5, we took $(\{i, j, k\}, \{i, j\} \in C_2, k \in C_1, i < j < k)$ on purpose. If i,j and k are the three are the three restrictions of the DGP, which we want to recover, then we really want to arrive at the last step, where we test them jointly. To maximize this chance and minimize the number of calculations (taking all possible triples would take too much time, as would testing all couples and then adding any restriction $k$, not just $i < j < k$), we test the combination of $i$ and $j$, the two restrictions, which were most easily accepted by the data, first. Conditional on that combination being accepted, we add restriction $k$, the restriction with the lowest $p$-value of the three when tested individually.

   We can in fact prove the following theorem, which is not operational, as the number of restrictions is not known a priori, but shows that in special cases the asymptotic size of the algorithm is in fact 5%

THEOREM 6. *In the special case that only one relation contains over-identifying restrictions and in the special case that only two relations contain only 1 over-identifying restriction each, the asymptotic size of the algorithm equals* $0.05$

In fact it is possible to lower the intermediate rejection probability, such that the asymptotic size of the test procedure becomes 5%:

THEOREM 7. *If at the intermediate steps (3) and (5) of the algorithm the critical value is set at* $\kappa = \chi^2_{0.95}(qr - r^2)$, *the asymptotic size of the algorithm is 0.05, while the asymptotic power is* $1$.

In the following we shall not use theorem 7, because the critical value $\kappa$ can become so large in practice, that virtually none of the tested restrictions are rejected at the intermediate levels. Consequently the computational costs explodes and only a small benefit is obtained in size if is it used.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

**3.3.2. An example.** Let us consider a trivariate DGP with a trend inside the cointegration space. We have rightly determined the cointegration rank to be two and try to identify the parameter space. In the DGP we have that the first variable is trend-stationary, whereas the second one is stationary and the third one non-stationary, as in

$$\beta' = \begin{bmatrix} 1 & 0 & 0 & 0.023 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

(1) Based on economic theory, the following two vectors are indicated as possible combinations in a stationary relationship:

$$f_1' = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}$$
$$f_2' = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix}$$

(2) As possible matrices $H$ we take all matrices consisting of one column, choosing from the set $\{f_1, f_2, e_1, e_2, e_3, e_4\}$ and any combination of two columns with the following exceptions:
   (a) We do not test $e_4$ on its own, because then we would test whether a linear trend is stationary by itself.
   (b) We test only one out of the combinations $\{[e_1, e_2], [f_1, e_1], [f_1, e_2]\}$ as they are equivalent restrictions. We choose the first combination. Similarly we only take one of the following set $\{[e_2, e_3], [f_2, e_2], [f_2, e_3]\}$.
   In total we thus have 16 matrices $H$ which we test individually in the next step.

(3) We find that the following restrictions are accepted. The p-values are between brackets:

| | | | | | |
|---|---|---|---|---|---|
| $C_{1,1}$ | $H_1 := \begin{bmatrix} e_1 & e_2 \end{bmatrix}$ | (0.80) | $C_{1,4}$ | $H_4 := \begin{bmatrix} e_1 & f_2 \end{bmatrix}$ | (0.25) |
| $C_{1,2}$ | $H_2 := \begin{bmatrix} e_2 & e_3 \end{bmatrix}$ | (0.40) | $C_{1,5}$ | $H_5 := \begin{bmatrix} e_2 \end{bmatrix}$ | (0.06) |
| $C_{1,3}$ | $H_3 := \begin{bmatrix} e_2 & e_4 \end{bmatrix}$ | (0.35) | $C_{1,6}$ | $H_6 := \begin{bmatrix} f_1 & e_4 \end{bmatrix}$ | (0.04) |

We note that $H_4$ has erroneously been accepted, whereas the hypothesis $H_x := \begin{bmatrix} e_1 & e_4 \end{bmatrix}$ has been wrongly rejected. Its $p$-value was lower than 0.01.

(4) There are 15 potential combinations to be tested. However we do not test $\{H_1, H_5\}$ as this combination does not satisfy the conditions for generic identification in Johansen (1995a): it is easily seen that an equivalent identified model would be $\{[e_1], H_5\}$. The same comment applies to the combinations $\{H_2, H_5\}$ and $\{H_3, H_5\}$.

We find that the following combinations are accepted:

| | | | | | |
|---|---|---|---|---|---|
| $C_{2,1}$ | $\{H_1, H_3\}$ | (0.52) | $C_{2,5}$ | $\{H_2, H_4\}$ | (0.23) |
| $C_{2,2}$ | $\{H_1, H_6\}$ | (0.52) | $C_{2,6}$ | $\{H_5, H_6\}$ | (0.05) |
| $C_{2,3}$ | $\{H_3, H_6\}$ | (0.52) | $C_{2,7}$ | $\{H_2, H_4\}$ | (0.03) |
| $C_{2,4}$ | $\{H_2, H_6\}$ | (0.36) | $C_{2,8}$ | $\{H_1, H_4\}$ | (0.03) |

We note that the last two have been erroneously accepted, but that no error of type I was made in this round.

(7) We take $C_1, \ldots, C_5$ and $C_{2,1}, \ldots, C_{2,6}$ as they were accepted at 5% or more. First we note that $C_{2,1}, C_{2,2}$ and $C_{2,3}$ share the same $p$-value. It turns out that the models are equivalent, such that they are the same model. We give the first five models according to our selection criteria:

| $Model$ | d.o.f. | $p - value$ |
|---|---|---|
| $C_{2,6}$ | 3 | (0.05) |
| $C_{2,1}, C_{2,2}, C_{2,3}$ | 2 | (0.52) |
| $C_{2,4}$ | 2 | (0.36) |
| $C_{2,5}$ | 2 | (0.23) |
| $C_{1,5}$ | 2 | (0.06) |

Closer inspection shows that model $C_{2,6}$ is an equivalent model to the DGP: even though

 DOI: 10.2870/52609

$H_x$ was rejected at step 3, we find an equivalent representation. The next four accepted models all impose 2 out of the 3 restrictions we should have found. This is also a relative success as the restrictions we found actually hold. We just failed to get the last one in those cases.

## 3.4. Monte Carlo evidence

We used the following 5-variable DGP with $k$ lags and a trend in the cointegration relation to test the algorithm:

$$\prod_{i=1}^{k}(1-\phi_i L)\,X_{1t} = \alpha_{11}\left(X_{2t-1}+X_{3t-1}+0.007t\right)+\varepsilon_{1t}$$

$$\prod_{i=1}^{k}(1-\phi_i L)\,X_{3t} = \alpha_{32}\left(X_{4t-1}+X_{5t-1}\right)+\varepsilon_{3t}$$

$$\Delta X_{2t} = \varepsilon_{2t}$$

$$\Delta X_{4t} = \varepsilon_{4t}$$

$$\Delta X_{5t} = \varepsilon_{5t}$$

where

$$\alpha_{11} = \alpha_{32} = \prod_{i=1}^{k}(1-\phi_i)$$

$$\varepsilon_t \sim iidN\left(0, I_5 \times 10^{-4}\right)$$

There are two cointegration relations in the DGP and the roots of each stationary equations are equal to $\phi_1$ to $\phi_k$. The main reason for proposing this DGP is that when adding a lag, all the existent roots can be kept constant.

The cointegration space is equal for any number of lags:

$$\beta' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0.02 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

In the experiments that follow we vary three parts of the procedure:

(1) The series of roots $\{\phi_i\}_{i=1}^{k}$. In all but one case we model the right number of lags, but in one case we model only 2 lags, whereas the DGP contains three lags.

(2) The economic theory input: either we do not specify any $f_i$ vectors, in which case we should be able to find the 5 zeros in the DGP, which is equivalent to 3 restrictions or we specify only one, namely:

$$f_1' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{3.4}$$

In the last case recovering the DGP is equivalent to finding five over-identifying restrictions.

(3) The length of the time series: we take $T = 100$ as our bench mark case, but have one experiment with $T = 1000$ to check the asymptotic behaviour of the procedure.

The success of the procedure is measured in two different ways:

(1) Taking the first model, we look whether
   (a) the restrictions of the model exactly identify the DGP
   (b) the model found is a submodel of the DGP (that it all the restrictions of the model are present in the DGP), but the model misses one restriction
   (c) the model found is a submodel of the DGP, but it misses two restrictions

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

(d) the model found has all the over-identifying restrictions of the DGP plus one restriction too many.
(e) the model found has all the over-identifying restrictions plus two.
(f) a completely different model is selected
(g) no model at all is selected (as all of them are rejected)
(a) is an outright success, but especially (b) can also be classified as a success as the DGP is nearly recovered. (f) is complete failure.

(2) As in the example above, it is possible to give five selected model to the researcher and let her make the choice. We thus measure how often the model, which exactly identifies the DGP is among the those first five.

The degree of identification is conditional on the $f$-vectors specified. In our particular example three overidentifying restrictions can be found, if no $f$-vectors are specified, whereas 5 restrictions can be found if the $f-$vector (3.4) is given as input. In the first case recovering the DGP is equivalent to accepting $\{[e_1, e_2, e_3, e_6], [e_3, e_4, e_5]\}$, whereas in the second case it is $\{[f_1, e_6], [e_3, e_4, e_5]\}$.

In the last column of the table we report one minus the rejection probability of a straight test of these last models, when testing at the nominal 5% level. These numbers give an indication of the small sample performance of these tests and also provide an upper bound to how often the DGP can be among the first five models or indeed how often the DGP can be the first selected model.

The results of the Monte Carlo simulations are based on 500 replications. For each replication the first 100 observations were discarded. The results are reported in table 3.1.

In our benchmark case, number 1, the systems contains two residual roots of 0.6 and two roots of 0.2. We model the right number of lags, namely two and do not provide any input in the form of an $f$-vector. In 53.6% of the cases the proposed algorithm is able to recover the restrictions in the cointegration space, whereas in 7.8% it only misses one. If we consider the first five selected model, then the DGP is among them in fully 72.8% of the cases. This is a high success rate when one considers than the estimated size of testing just all the right restrictions is 19.2% (see the last column of the table) for a nominal 5% test. 80.8% is an upper limit to how often the selected model can be recovered (column a) or how often it can be among the first five.

The second and third DGP show that increasing the roots of the process leads to a complete breakdown of the algorithm. On the other hand DGP's with one lag (number 4 and 5) have a relatively good performance. In the case of number 4, in fully 98% of cases do we either recoup the original DGP or miss just one restriction.

It is fairly common to select a low lag length even if it is believed that the true lag length of the DGP is longer, possibly even infinite: to our knowledge no asymptotic results are known for the consistency of the estimated cointegration parameters in this case. (obviously the short run parameters are not estimated unbiased any more in that case). In experiment 6 and 7 we check the effect the effect of underselecting the true lag length and remarkably the algorithm does better by all measures, when the lag length is underestimated. This lends support to the view that a short model should be fitted to the data.

In experiment 8 we give some quasi-economic input in that we specify the vector $f_1' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$.

The results should be compared to those of experiment 1: even though the size distortion of this test is large (23.8%!), the true restrictions (5 in this case) or all but one of them are recovered in fully 80% of all cases. Furthermore the model with the true restrictions is among the first five selected in 74.6% of cases and thus almost reaches its upper bound.

Experiment 9 shows that the procedure works asymptotically. The fact that the true model is among the first five in only 93.8% of cases, compared to an upper bound of 94.2% is caused

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

| | DGP $\{\phi_i\}$ | Lags Mod | $T$ | $f$ ? | The first model a | b | c | d | e | f | g | DGP in 5? | 1-Size |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $\{0.6, 0.2\}$ | 2 | 100 | no | 53.6 | 7.8 | 0.6 | 11.2 | 2.0 | 24.6 | | 72.8 | 80.8 |
| 2 | $\{0.8, 0.2\}$ | 2 | 100 | no | 5.8 | 0.8 | 0.2 | 9.4 | 10.2 | 73.6 | | 19.6 | 72.2 |
| 3 | $\{0.8, 0.6\}$ | 2 | 100 | no | 0.4 | 0 | 0 | 1.4 | 6.6 | 91.6 | | 2.4 | 56.4 |
| 4 | $\{0.6\}$ | 1 | 100 | no | 91.0 | 7.0 | 1.2 | 0.4 | 0 | 0.4 | | 91.6 | 91.6 |
| 5 | $\{0.8\}$ | 1 | 100 | no | 44.8 | 4.2 | 0.2 | 13.4 | 3.4 | 34.0 | | 66.4 | 83.0 |
| 6 | $\{0.6, 0.2, 0.2\}$ | 3 | 100 | no | 27.4 | 6.4 | 0.6 | 9.0 | 4.6 | 51.8 | | 42.6 | 68.0 |
| 7 | $\{0.6, 0.2, 0.2\}$ | 2 | 100 | no | 37.8 | 5.2 | 0.2 | 13.4 | 5.2 | 38.2 | | 61.4 | 82.0 |
| 8 | $\{0.6, 0.2\}$ | 2 | 100 | yes | 65.0 | 15.0 | 1.4 | 3.2 | 0.6 | 14.8 | | 74.6 | 76.2 |
| 9 | $\{0.6, 0.2\}$ | 2 | 1000 | no | 93.8 | 5.4 | 0.8 | 0 | 0 | 0 | | 93.8 | 94.2 |

Table 3.1: Monte Carlo study of automated model selection

by the fact that one of the two vectors individually can be rejected at the 1% level, whereas the combination of the two restrictions can still be accepted at the 5% level. This happened in our example above, but in that case there were equivalent models. Unfortunately there are none in the Monte Carlo study.

**3.4.1. An empirical example.** We take the data on money demand in the UK as analyzed by Hendry and Doornik (1994) and many others. The data consist of log of real output (TFE), money (M1), inflation and an interest rate differential. Furthermore there are two dummies: one for output shocks (Dout) and one for oil shocks (Doil). For full details of the variables and the data the reader is referred to the original paper. The documented data and the programs to run the original analysis are available at the web site of David Hendry[4]. The cointegration space they find is the first one reported in table 3.2.

After formal testing the authors decide that the model should have 2 lags and 2 cointegration relations. They then decide that the oil dummy is outside the cointegration space and the output dummy inside. We shall put them both inside the space and thereby test their possible exclusion. Their reported test is a combined restriction of the short run matrix (four zero restrictions), a restriction that the second vector is completely specified and the following $H$-matrix on the first vector:

$$H_1' = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

We leave the second vector completely unrestricted throughout and in our first run of the algorithm we impose no $f$−vectors, that is we ran the algorithm without any theory input. The first model, the algorithm selects, is reported in table 3.2.

In the second run we use the theory input from Hendry and Doornik by specifying the following two vectors:

$$f_1' = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$f_2' = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

and feeding them into the algorithm. In the table we report only the fifth model. The first four models all have $p$-values of between 5% and 6%. They have different second cointegration vectors, which consists of changing combinations of inflation, output dummy and trend. In view of the outcome of the previous run of the algorithm (and the next run), the considerably higher

[4] http://www.nuff.ox.ac.uk/users/hendry/

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

| | H&D | | Ex1, M1 | | Ex 2, M5 | | Ex 3, M1 | |
|---|---|---|---|---|---|---|---|---|
| output | −1 | 1 | −1 | 1 | −1 | 1 | −1 | 1 |
| money | 1 | 0 | 0.89 | 0 | 1 | 0 | 1 | 0 |
| inflation | 6.91 | −3.4 | 7.23 | −1.98 | 6.90 | −1.86 | 7.12 | −1.90 |
| Rnet | 6.91 | 1.8 | 6.39 | 1.34 | 6.90 | 1.36 | 7.12 | 1.31 |
| trend/100 | 0 | −0.63 | 0 | −0.65 | 0 | −0.66 | 0 | −0.67 |
| Dout | 1.46 | −0.40 | 0 | −0.26 | 0 | −0.27 | 0 | −0.18 |
| Doil | $n/a$ | $n/a$ | 0 | 0 | 0 | 0 | 0 | 0 |
| p-value | 0.38[5] | | 0.65 | | 0.48 | | 0.14 | |

Table 3.2: Automated model selection in UK money demand

$p$-value and the economic interpretability of the vector as an demand equation, we report model 5.

In the third run we still have the two $f$-vectors, two cointegration vectors, but choose only one lag, as the previously reported simulation suggested this may help. We report the first model selected by the algorithm.

We thus found a model with one more over-identifying restriction, namely that the output dummy should be zero in first cointegration vector. The total analysis has only taken a few minutes on a Pentium-II computer.

## 3.5. Use of the algorithm

**3.5.1. Practical advantages.** The algorithm can be used to simply check whether the final result obtained by the traditional way of finding an identified cointegration space is good. There are however more promising uses: the procedure can be directly applied by the practitioner or referee to select the final model.

But the speed of the algorithm (in most cases the answer is given in a few minutes) allows for more possibilities. Sometimes the outcome of the lag length selection procedure or the rank-test is unclear and a choice has to be made. With the new procedure it is possible to find an identified model for both possibilities. Sensitivity check procedures of all kinds are now possible: if the inclusion of an extra dummy leads to a completely different model, then a problem of stability certainly exists in the data.

**3.5.2. Methodological advantages.** The presently used procedure to find identifying restrictions is long and cumbersome and many rules of the thumb have to be applied: not every combination of two vectors, which are accepted can possibly be tested in a reasonable time. Replication, even by the very same researcher who did the original study, is often difficult. With a standard, thorough procedure, like the here proposed algorithm, replication becomes possible.

The LSE-methodology itself is often attacked for being partially 'art' or worse, 'alchemy'. By clearly spelling out the modelling process, formalizing and automating it, the methodology itself is being strengthened. The number of key decisions in the identifying/restriction process is brought down to three:

(1) What economic theory should be tested in the form of $f$-vectors?
(2) Which of the top models should be chosen as the preferred model?
(3) Which sensitivity checks should be done?

All three decisions can be discussed and reported in a paper, whereas it is currently often impossible to report all tests on the cointegrated space done in the modelling process.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

Once more the dichotomy between on the one hand selecting the right variables, being explicit in each step which economic theory is being used and selecting the final model and on the other hand the testing procedures is stressed. The second part can be automated (and should be) to leave more time and space to do the first carefully.

**3.5.3. Improving the methods themselves.** The present algorithm is but a first proposal on how to automate a part of the modelling process. It can no doubt be improved, both in terms of computational speed and - far more importantly - in terms of internal decision rules. One possibility is to use sub-samples and sub-sets of the variables in the algorithm to find the building blocks for the total cointegration space.

The small Monte Carlo study has already pointed to two possible improvements in the methods: keeping in restrictions, which are accepted at the 1% level, but rejected at the 5% level and selecting fewer lags than are believed present in the DGP. The first rule can be applied in automated modelling only, but the second is certainly relevant for traditional modelling.

The very poor actual size performance of the applied tests is of great concern in the development of these methods. The only theorem underlying the algorithm relies on the asymptotic size. Omtzigt and Fachin (2002) clearly show that the currently available methods Bartlett corrections and bootstrapping, do not offer satisfactory solutions, so new ones will have to improve the algorithm. The size of this problem had been somewhat hidden in the literature due to the reliance on small DGPs, whose parameters can easily be controlled. The large number of parameters in the Monte Carlo DGP used here means that no effective exploration of the size in the whole parameter space could be executed. Yet this particular example shows that at least in some cases size distortion is an extremely serious problem.

Finally no evidence is yet available on how likely the modelling procedure is able to recover the original DGP. We show it only for part of the process (disregarding rank and lag selection for instance). Of course it is unlikely that the DGP falls within the class of models considered, but still measures on how well the modelling process performs, when the DGP is in the class of models considered, should be available.

**3.5.4. Comparison with Minimal.** Minimal by Davidson (1998a) follows a different approach: it tries to find all the smallest subsets of variables which are cointegrating. It does so by using the Wald test statistics for (3.2) as developed by Davidson (1998b) . His procedure is certainly faster, as the Wald test does not require restricted optimization. Yet the method proposed carries the following advantages:

(1) It allows for theory input in the form of $f$-vectors.
(2) The combinations of restrictions are tested in the algorithm presented, whereas in minimal the accepted cointegration vectors are just combined
(3) In minimal, the combination of accepted vectors can and does lead to conflicting evidence: it is perfectly possibly for minimal to accept $[e_1, e_2]$ and $[e_1, e_3]$ but reject $[e_2, e_3]$. Thus the space spanned by the cointegration vectors is not unambiguously defined.

The speed of the algorithm remains a cause for concern. With 2 cointegration vectors and 5 variables minimal or step 3 will consists of doing (a maximum of) $W = \sum_{i=1}^{3} \binom{5}{i} = 25$ tests. With 8 variables (either because variables are added or $f$-vectors defined), there are up to 246 tests. If many of them are accepted, the number of calculations explodes in the step 4 of the algorithm: with $r = 2$, there could be up to $W (W - 1) / 2$ combinations in that step. This will be a major concern in future development of the algorithm.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

## 3.6. Conclusions

We have presented an algorithm for the automatic identification and restriction of the parameter space of a cointegrated VAR. Its used was demonstrated both by means of a Monte Carlo study and an empirical example. One remarkable solution of the former is that underselection of the VAR lag length may lead to a higher probability of recovering the original cointegration space.

Throughout we have argued the methodological advantages of automation: separation of economic theory decision, which will have to be performed by the researcher and search procedures, which can be automated.

## 3.A. Proofs

PROOF OF THEOREM 5. If $r = 1$ we test all possible restrictions $\{H_i\}$. They fall in three categories:

(1) $\{H_i\}_t$ The true DGP (there could be more equivalent reprentations) This is accepted at the asymptotic 5% level. It has the maximum (true) number of overidentifying restrictions
(2) $\{H_i\}_s$ The restrictions hold true, but are less in number than in $\{H_i\}_t$. This means that if the model is accepted, it is classified below $\{H_i\}_t$ (if that last one is accepted)
(3) $\{H_i\}_f$ The restrictions do not hold true. The asymptotic power of the test is 1 (Johansen 1991):$L\left(\{H_i\}_f\right) \to \infty$ as $t \to \infty$ such that this test in always rejected asymptotically.
The result for $r = 1$ follows.

For $r > 1$ consider what happens to $\{H_1, \dots, H_r\}$ the DGP combination.
In step 3, we test all possible restrictions of the kind (3.2) and so each of the DGP restrictions is tested individually at the 1% level, that is accepted if $L(H_i) \leq c_{0.99}$ where $c_r$ is the 99%th percentile of the $\chi^2$ distribution. In step 4, we test whether $L(H_1, H_2) \leq c_{0.99}$ whereas in step 5 we test $L(H_1, H_{2,}, H_3) \leq c_{0.99}$ until $L(H_1, H_{2,}, \dots, H_{r-1}) \leq c_{0.99}$. In step 7, we then test whether $L(H_1, H_{2,}, \dots, H_r) \leq c_{0.95}$. To find the asymptotic size of this procedure, we use Bonferroni inequality:

$$P\left(\cap_{i=1}^{r}\left(L\left(H_i\right) \leq c_{0.99}\right) \cap \cap_{i=2}^{r-1}\left(L\left(H_1, \dots, H_i\right) \leq c_{0.99}\right) \cap L\left(H_1, H_{2,}, \dots, H_r\right) \leq c_{0.95}\right)$$
$$\geq 1 - \sum_{i=1}^{r} P\left(L\left(H_i\right) \geq c_{0.99}\right) - \sum_{i=2}^{r-1} P\left(L\left(H_1, \dots, H_i\right) \geq c_{0.99}\right) - P\left(L\left(H_1, H_{2,}, \dots, H_r\right) \geq c_{0.95}\right)$$
$$= 1 - 0.01r - 0.01(r-2) - 0.05$$
$$= 1 - 0.02r + 0.03$$

The comments of point 2 and 3 still hold true in this case. □

PROOF OF THEOREM 6. If only one relation contains overidentifying restrictions, the DGP can be written as

$$\beta = (H_x\phi, \psi)$$

and this is tested at round 3 and included in set $C_1$ if accepted. Then in step 7 it is tested again, such that we have:

$$P\left(\left(L\left(H_x\right) \leq c_{0.99}\right) \cap \left(L\left(H_x\right) \leq c_{0.95}\right)\right)$$
$$= P\left(L\left(H_x\right) \leq c_{0.95}\right)$$
$$= 0.95$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

If two relations contain exactly overidentifying relations, then we have the following DGP:

$$\beta = (H_x\phi_1, H_y\phi_2, \psi)$$

and we find

$$\begin{aligned}
&P\left(\left(L\left(H_x\right) \leq c_{0.99}\right) \cap \left(L\left(H_y\right) \leq c_{0.99}\right) \cap \left(L\left(H_x, H_y\right) \leq c_{0.99}\right) \cap \left(L\left(H_x, H_y\right) \leq c_{0.95}\right)\right) \\
&= P\left(\left(L\left(H_x\right) \leq 6.63\right) \cap \left(L\left(H_y\right) \leq 6.63\right) \cap \left(L\left(H_x, H_y\right) \leq 5.99\right)\right) \\
&= P\left(L\left(H_x, H_y\right) \leq 5.99\right) \\
&= 0.95
\end{aligned}$$

where we have used that $L(H_y) \leq L(H_x, H_y)$ and $L(H_x) \leq L(H_x, H_y)$. □

PROOF OF THEOREM 7. The proof proceeds as the proof of theorem 5 until

$$\begin{aligned}
&P\left(\cap_{i=1}^{r}\left(L\left(H_i\right) \leq \kappa\right) \cap \cap_{i=2}^{r-1}\left(L\left(H_1, \ldots, H_i\right) \leq \kappa\right) \cap L\left(H_1, H_{2,}, \ldots, H_r\right) \leq c_{0.95}\right) \\
&= P\left(L\left(H_1, H_{2,}, \ldots, H_r\right) \leq c_{0.95}\right) \\
&= 0.95
\end{aligned}$$

where in the second line we have used the fact that all hypotheses $\{H_i\}, i = 1, \ldots, r$ and $\{(H_1, \ldots, H_i)\}, i = 1, \ldots, r$ are subhypotheses of $\{(H_1, H_{2,}, \ldots, H_r)\}$ such that $L(H_i) \leq L(H_1, H_{2,}, \ldots, H_r), i = 1, \ldots, r$ and $L(H_1, \ldots, H_i) \leq L(H_1, H_{2,}, \ldots, H_r), i = 1, \ldots, r$. Furthermore the maximum number of overidentifying restrictions on the space is $qr - r^2$ (at which point each matrix $H_i$ has only one column), such that $\kappa \geq c_{0.95}$. □

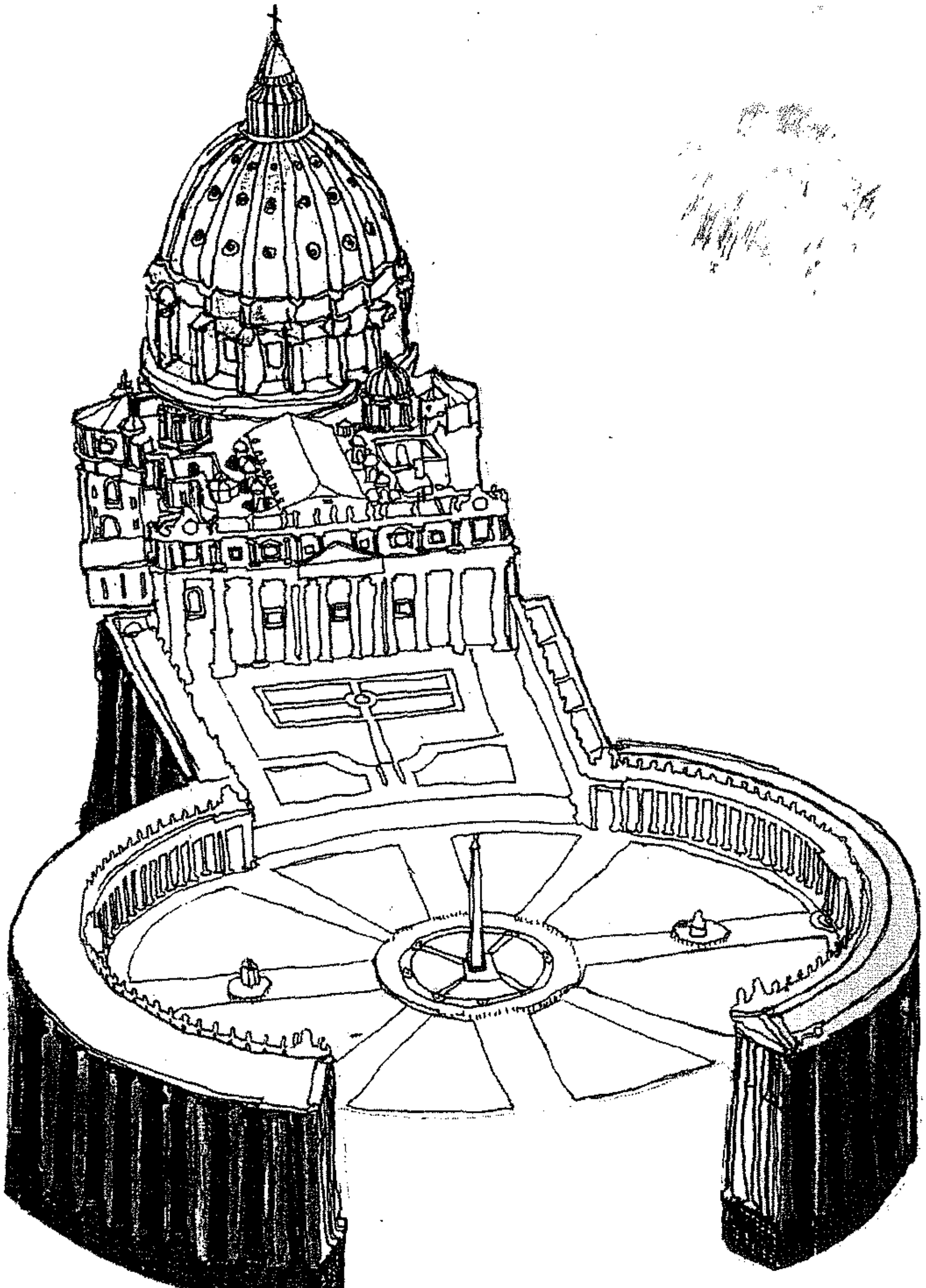
Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

# Part 2

# Small sample corrections


DOI: 10.2870/52609

R O M A



Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

CHAPTER 4

# Bootstrapping and Bartlett corrections in the cointegrated VAR model

## 4.1. Introduction

The small sample properties of tests on long-run coefficients in cointegrated systems are still a matter of concern to applied econometricians. Since the asymptotic procedures proposed by Johansen (1991) have been shown to suffer from severe size distortion (among others, see Gonzalo, 1994; Bewley et al., 1994; Li and Maddala, 1997) two natural and complementary solutions have been proposed: (i) applying Bartlett corrections to the test statistics, in the hope that the corrected statistic will follow a small sample distribution closer to the asymptotic one, and thus bring actual sizes closer to the nominal sizes (Johansen, 2000a); (ii), trying to estimate the actual small sample distribution by the bootstrap, a computer-intensive technique strictly linked with the Edgeworth expansion and indeed defined by Cribari-Neto and Cordeiro (1996) 'a simulation based alternative to Bartlett and Bartlett-type corrections' (Li and Maddala, 1996, 1997; Fachin, 2000; Gredenhoff and Jacobson, 2001).

For the time being, no definite solution has however appeared. Although the only aim of both the Bootstrap and the Bartlett correction is to get the actual size closer to the nominal size, the final aim of any testing procedure must be that of distinguishing between valid and invalid hypotheses: the proportion of Type II errors of corrected tests is therefore crucial. To the best of our knowledge no evidence on the power properties of Bartlett corrected tests in the cointegrated VAR model has appeared in the literature; the only available evidence on power for bootstrapped test statistics is in Fachin (2000) and shows that the type of bootstrap test examined may have a rather high Type II error. The aim of this paper is thus examining both the size and power properties of Bartlett-corrected and bootstrap tests. With respect to the latter, we also evaluate the feasible double bootstrap, recently proposed by Davidson and MacKinnon (2000). In either cases, a key result of the paper is that the Bartlett correction and the bootstrap tests should both be based on the *unrestricted* estimate of the cointegration vectors.

The chapter is organised as follows: in section 4.2 we shall briefly review the model, the structure of Bartlett-corrected and bootstrap tests, as well as a theoretical result, motivating us to base both procedures on unrestricted estimates. In section 4.3 we shall discuss the design of the Monte Carlo experiment and in section 4.4 present the results of the simulations.

Some conclusions, as well as tentative recommendations for applied work, are finally drawn in section 4.5.

## 4.2. Bartlett-corrected and bootstrap tests on cointegrating coefficients

**4.2.1. The model.** The cointegrated $p$-dimensional VAR model with $k$ lags in its autoregressive form is defined as:

$$\Delta X_t = \alpha\beta' \begin{pmatrix} X_{t-1} \\ D_t \end{pmatrix} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Psi d_t + \varepsilon_t \tag{4.1}$$

In this paper a linear trend is constrained to lie in the cointegration space and an unrestricted constant is included outside that space: $D_t = t$ and $d_t = 1$. We define $\gamma$ and $\rho$ by $\beta' = (\gamma', \rho')$,




DOI: 10.2870/52609

where $\gamma$ includes the coefficients linking the stochastic variables of the system and $\rho$ are the coefficients of the deterministic part.

Three assumptions are made to make sure this is a stable I(1) model:

**Assumption (Rank):** $\alpha$ and $\gamma$ are two full rank matrices of dimension $p \times r$, $p > r$;

**Assumption (No I(2)):** The matrix $\alpha'_\perp \left( I - \sum_{i=1}^{k-1} \Gamma_i \right) \gamma_\perp$ is of full rank;

**Assumption (No other roots):** The roots $z$ of the characteristic polynomial are either 1: $z = 1$ ($p - r$ roots are equal to unity) or larger than 1 in absolute value: $|z| > 1$.

The first two assumptions assure that the process is an I(1) process and not integrated of lower or higher order, while the third assumption excludes explosive behaviour and seasonal unit roots

The stationary, stochastic part of (4.1) can be written in a companion form[1]:

$$\begin{bmatrix} \gamma' X_t \\ \Delta X_t \\ \Delta X_{t-1} \\ \vdots \\ \Delta X_{t-k+2} \end{bmatrix} = \begin{bmatrix} I_r + \gamma'\alpha & \gamma'\Gamma_1 & \cdots & \gamma'\Gamma_{k-2} & \gamma'\Gamma_{k-1} \\ \alpha & \Gamma_1 & \cdots & \Gamma_{k-2} & \Gamma_{k-1} \\ 0 & I & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & I & 0 \end{bmatrix} \begin{bmatrix} \gamma' X_{t-1} \\ \Delta X_{t-1} \\ \Delta X_{t-2} \\ \vdots \\ \Delta X_{t-k+1} \end{bmatrix} + \begin{bmatrix} \gamma' \\ I \\ 0 \\ \vdots \\ 0 \end{bmatrix} \varepsilon_t$$

or

$$Y_t = PY_{t-1} + F\varepsilon_t \tag{4.2}$$

The Bartlett correction, which shall be discussed in the next section, depends crucially on the matrix $P$.

**4.2.2. The Bartlett correction.** The idea behind the Bartlett correction Bartlett (1937) is both simple and appealing. Suppose the aim is testing the following null hypothesis on the parameters $\Theta$, $\mathcal{H}_0 : \Theta_0 \subset \Theta$. In regular cases, the LR test statistic $s = -2\ln(LR(\Theta_0|\Theta))$ has an expected value of

$$E_{\hat{\theta}_0}\left[-2\ln(LR(\Theta_0|\Theta))\right] = E_{\hat{\theta}_0}\left[l_{\hat{\theta}} - l_{\hat{\theta}_0}\right] = h\left(1 + \frac{1}{T}g(\theta_0)\right) + O\left(\frac{1}{T^2}\right) \tag{4.3}$$

where $h$ denotes the number of restrictions tested. Then dividing the test statistic $s$ by $\left(1 + \frac{1}{T}g(\theta_0)\right)$ we may obtain the modified test statistic $s_B$ and expect the resulting distribution to be closer to a $\chi^2$ distribution. This division is called a Bartlett correction and $\frac{1}{T}g(\theta_0)$ will be referred to as the Bartlett factor.

We obviously do not know $\theta_0$, the true value of the parameters, $\theta$, and thus we substitute a consistent estimate of $\theta$, $\tilde{\theta}$, in expression (4.3) and thus get the Bartlett factor $g\left(\tilde{\theta}\right)$.

The arguments in the following pages will revolve around which consistent estimate should substitute $\theta_0$: $\hat{\theta}_0$ the maximum likelihood estimate under the null hypothesis or $\hat{\theta}$ the unconstrained maximum likelihood estimate. We shall argue that we need to substitute $\hat{\theta}$ and not $\hat{\theta}_0$ in the problem at hand. Even though the size correction works better with $\hat{\theta}_0$ which is more efficient under the null, we find the power of the Bartlett corrected test-statistic with $\hat{\theta}$ extremely poor. We demonstrate this both by means of a theory and simulations in section 4.3. To see the differences in practice between using these two estimates, we refer to page 49, where in figure 4.3 we have plotted power curves for both estimates. (the DGP-value is 1 and the curves are drawn for the 5% significance level).

[1]The deterministic part can be taken account of by adding an extra term in $d_t$ and $D_t$.


DOI: 10.2870/52609

The problems of the Bartlett section and their solutions, carry over to the bootstrap section as well.

Lawley (1956) and Barndorff-Nielsen and Hall (1988) proved that under certain regularity conditions (which exclude cointegrated VAR models and thus the problem at hand) for any real number $x$

$$p\left(s_B \leq x\right) = p\left(\chi^2(h) \leq x\right) + O\left(\frac{1}{T^2}\right) \tag{4.4}$$

So the whole $\chi^2$ distribution is better approximated after the correction.

Jensen and Wood (1997) showed that for the Dickey Fuller distribution (4.4) does not hold. This however does not mean that the size correction is not useful in practice. In fact Nielsen (1997) showed that a Bartlett correction in an AR(1) process with a unit root, does provide an improvement to the size of the test.

Under the assumption:

**Assumption (Deterministics):** there exist matrices $K$ and $M$ such that $d_t = Md_{t-1}$ and $\Delta D_t = Kd_t$ where all the eigenvalues of the matrix $M$ equal 1 in absolute value

Johansen (2000a) derived the Bartlett correction for three different kind of hypotheses on $\beta$ in (4.1), namely:

(1) $\beta = \beta^0$, a simple hypothesis on all the cointegration vectors;
(2) $\beta_1 = \beta_1^0$ where $\beta_1^0$ are the first $r_1$ relations ($1 \leq r_1 < r$) and the other cointegration relations are unrestricted;
(3) $\gamma = H\varphi$ where $H$ is a $(p \times r)$ matrix of full rank and $s < r$. This hypothesis implies the same restriction on all relations in $\gamma$.

Corrections for other kinds of hypotheses, like restrictions of the kind $\beta_1 = H_1\varphi_1$ do not yet exist.

We therefore limit ourselves to confronting the corrections 1 and 2 with the bootstrap in this paper and do not put any dummy variables in our DGP.

The correction term itself, for which we refer to the aforementioned article, depends crucially on the total number of parameters, the variance of $Y_t$ in (4.2) and a number of times on $\sum_{i=0}^{\infty} P^i$. We do not have the true value of the parameters, so we substitute estimates. Now under the null the matrix $P$ only contains eigenvalues strictly smaller than unity in absolute value as $Y_t$ in (4.2) is a stationary process. Yet under the alternative, the restricted estimate will contain at least one additional unit root, because one of the relations $\gamma' X_t$ is no longer stationary. Consequently the Bartlett correction is no longer defined as the sum $\sum_{i=0}^{\infty} P^i$ diverges. We prove this fact in the following theorem:

THEOREM 8. *Under the false null hypothesis $\beta = b = (b_1, b_2)$ where $b_1 \in sp\left(\beta^0\right), b_2 \notin sp\left(\beta^0\right)$ ($\beta^0$ is the true value of $\beta$), $\hat{\alpha}$ the restricted maximum likelihood estimate of $\alpha$, will have reduced rank $s < r$ in the limit. Consequently the matrix $\hat{P}$ contains additional unit roots in the limit*

PROOF. Partition the maximum likelihood estimate of $\alpha$, $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2)$ conformably with $b$. It is found by ordinary least squares:

$$\hat{\alpha} = S_{01}\left(b_1, b_2\right)\begin{pmatrix} b_1' S_{11} b_1 & b_1' S_{11} b_2 \\ b_2' S_{11} b_1 & b_2' S_{11} b_2 \end{pmatrix}^{-1}$$

where $S_{01}$ and $S_{11}$ are defined in standard fashion (see Johansen 1995b, page 90-91). From Chan and Wei (1988) we find that $S_{01}b_1, b_1'S_{11}b_1 \in O\left(1\right)$, $S_{01}b_2, b_2'S_{11}b_1 \in O\left(T\right)$ and $b_2'S_{11}b_2 \in O\left(T^2\right)$. Using standard inverse matrix formula, we find that $\hat{\alpha}_2 = \left(S_{01}b_2 - S_{01}b_1\left(b_1'S_{11}b_1\right)^{-1}b_1'S_{11}b_2\right) \times \left(\left(b_2'S_{11}b_2\right)^{-1} - \left(b_2'S_{11}b_1\right)\left(b_1'S_{11}b_1\right)^{-1}\left(b_1'S_{11}b_2\right)\right)^{-1} \in O\left(T^{-1}\right)$ which implies $\hat{\alpha}_2 \xrightarrow{P} 0$ □

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

This means that if we use the restricted estimate $\hat{\theta}_0$ and the null hypothesis is false, the Bartlett correction is not defined. We shall see that in practice the absolute value of the roots is underestimated, such that the additional root is estimated to be close to 1. This means that the estimated Bartlett factor can be calculated, but becomes extremely large and the null hypothesis is then easily accepted.

We thus seek an estimator which

- Whenever the null hypothesis is true is consistent.
- Whenever the null hypothesis is false, the matrix $\hat{P}$ should have stable roots, such that the Bartlett correction is defined. If possible these roots should in some sense be as stable as possible, for when they are very close to unity, the Bartlett factor explodes and a false hypothesis is accepted.

SOLUTION 1. *Use the unrestricted estimates $\hat{\beta}$ of $\beta$ in (4.1) and not the restricted estimates in the Bartlett correction factor. The Bartlett correction factor for $\beta = \beta^0$ and $\gamma = H\varphi$ only depends on $\hat{\beta}$, such that this defines the solution in these cases.*
*In case 2 ($\beta_1 = \beta_1^0$, only some of the cointegration relations are restricted) we need estimates for both the restricted and the unrestricted vectors. In this case $\beta_1^0$ and the associated restricted estimate $\beta_2(\beta_1^0)$ should not be used, as this will lead to instability of the $\hat{P}$ matrix when $\mathcal{H}_0$ is false. Instead, we find a matrix $b_1$ for which $sp(b_1) \subset sp(\hat{\beta})$ and as close to $\beta_1^0$ as possible. This means that we find a matrix $\xi$ such that*

$$\xi = \left(\hat{\beta}'\hat{\beta}\right)^{-1}\hat{\beta}'\beta_1^0 \tag{4.5}$$

*Then the estimators $b_1 = \hat{\beta}\xi$ and $b_2 = \hat{\beta}\xi_\perp$ are consistent when the null hypothesis is true and the companion matrix $\hat{P}$ is stable when it is false.*

**4.2.3. Bootstrap methods.** In principle, the great advantage of the bootstrap[2] is that it can offer immediate solutions to new problems. However, in practice its ability to deliver good alternatives when reliable small sample parametric procedures are lacking must be accurately tested before its use may be recommended. This is especially true for the problem we are trying to solve, as the asymptotics of the bootstrap applied to integrated data is still largely unexplored: Horowitz (2002) summarises his survey stating that 'at present (...) there are no theoretical results on the ability of the bootstrap to provide asymptotic refinements for tests or confidence intervals when the data are integrated or cointegrated'. Recent developments in this direction covering specific cases are Chang et al. (2001), Davidson (2002), Paparoditis and Politis (2001) and Inoue and Kilian (2002). At the opposite, a striking example of how blind implementations of the bootstrap can deliver entirely wrong results is given by Phillips (2001) for the case of spurious regression with integrated variables.

The general idea underlying bootstrap tests is to assess the value of the test statistic $s$ obtained from the empirical analysis on the basis of the distribution of a large number of statistics $s^*$ computed from suitably constructed pseudodata, with the null hypothesis of the former consistent with the data generating process (DGP) of the latter. To this end, $\mathcal{H}_0$ may be imposed when generating the pseudodata (as in some examples in Efron and Tibshirani, 1993), or, vice versa, the chosen DGP taken as the null hypothesis (as recommended by Hall, 1992). In both cases, $\mathcal{H}_0$ is true for the pseudodata, and thus, assuming for simplicity a one-sided test, the proportion of $s^*$ more extreme than $s$ in the relevant direction is a natural estimate of the $p$-value of the test.

[2]General introductions to the bootstrap are provided, *inter alia*, by Efron and Tibshirani (1993), Hall (1995) and Horowitz (2002), while a recent review especially addressed at time series applications is Berkowitz and Kilian (2000).


DOI: 10.2870/52609

With cointegrated VARs and a simple hypothesis on the long-run coefficients $\mathcal{H}_0 : \beta = \beta^0$, the two approaches entail respectively:

(*a*) estimating a VAR constrained under $H_0 : \beta = \beta^0$, generating the pseudodata on the basis of the estimated *constrained* estimates $\hat{\theta}_0$ and a set of random noises (we will discuss the choice of these below), and testing $\mathcal{H}_0 : \beta = \beta^0$ both on the original data and on the pseudodata;

(*b*) estimating an unconstrained VAR, generating the pseudodata on the basis of the estimated *unconstrained* estimatres $\hat{\theta}$ and a set of random noises, testing $\mathcal{H}_0 : \beta = \beta^0$ on the original data and $\mathcal{H}_0^* : \beta = \widehat{\beta}$ (where $\widehat{\beta}$ are the unconstrained estimates of $\beta$) on the pseudodata.

So far, approach (*a*) has been favoured with no exception in the applications of interest here. However, a point of crucial importance for testing in the maximum likelihood estimation of cointegrated VARs seems to have gone unnoticed: although both approaches are valid and asymptotically equivalent under $H_0$, this is not true any more when it is false. To see this, consider the case of a test $\mathcal{H}_0 : \beta = \beta^0$ in a model without lags and just one cointegration vector. If this vector is misspecified, then $\beta^{0\prime} X_{t-1}$ is clearly an $I(1)$ process, whereas $\Delta X_t$ is $I(0)$. The only congruent values for the loading factors $\alpha$ are therefore zero. Hence all the element of the matrix $\hat{\Pi} = \hat{\alpha}\beta^{0\prime}$ equal zero (asymptotically) and the rank of such a matrix is 0 not 1. If one were to use this matrix for the Bootstrap DGP, one would generate just random walks without any cointegration (this is essentially a different version of exactly the same issue already discussed in the previous subsection with respect to the computation of the Bartlett factor when $\mathcal{H}_0$ is false). Thus, we will consider bootstrap tests of type (*b*).

With respect to the noise, there are again essentially two alternatives: either generating it under some parametric hypothesis (typically, $MIIDN$) or by resampling from the set of residuals of a VAR. In the latter case the natural choice are the residuals of the unconstrained VAR, empirically $MIIDN$. Gredenhoff and Jacobson (2001) favoured the parametric option, while Fachin (2000) and Li and Maddala (1997) the non-parametric one[3]. Here we will consider both alternatives. Block-resampling methods, such as the "Continuous-Path Block Bootstrap" proposed by Paparoditis and Politis (2001), which may be potentially powerful in dealing with the stochastic trends present in the system, will be the subject of future research.

Defining $\Theta$ the entire parameter set of the VAR and assuming we are interested in the test $\mathcal{H}_0 : \beta = \beta^0$ we thus implement the following bootstrap procedure, which is graphically represented in figure 4.1

- *Bootstrap test*

(1) Estimate VAR on data $X$; for given cointegrating rank obtain unrestricted estimates $\hat{\theta}$, unrestricted residuals $\widehat{\varepsilon}$, restricted estimates $\hat{\theta}_0$, restricted residuals $\widehat{\varepsilon}_0$ and test statistic $s$ for the hypothesis $H_0 : \beta = \beta^0$;

(2) Construct pseudodata: $X^* = \phi(\hat{\theta}, \varepsilon^*)$, $\varepsilon^*$ drawn at random with replacement from $\widehat{\varepsilon}$ or $NID$.

(3) Estimate VAR on pseudodata $X^*$; obtain $\hat{\theta}^*, \widehat{\varepsilon}^*$, $\hat{\theta}_0^*, \widehat{\varepsilon}_0^*$ and test statistic $s^*$ for the hypothesis $H_0^*$: $\beta = \hat{\beta}$;

*Repeat (2)-(3) a large number of times*

(4) Compute bootstrap $p$-value: $p^* = prop(s^* > s)$.

[3]Note that there is a possible source of confusion here, as the terms 'parametric' and 'non-parametric' have been used in the bootstrap literature with different meanings. We define procedures based on resampling from estimated residuals as 'non parametric', and that involving drawings from a theoretical distribution as 'parametric'.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

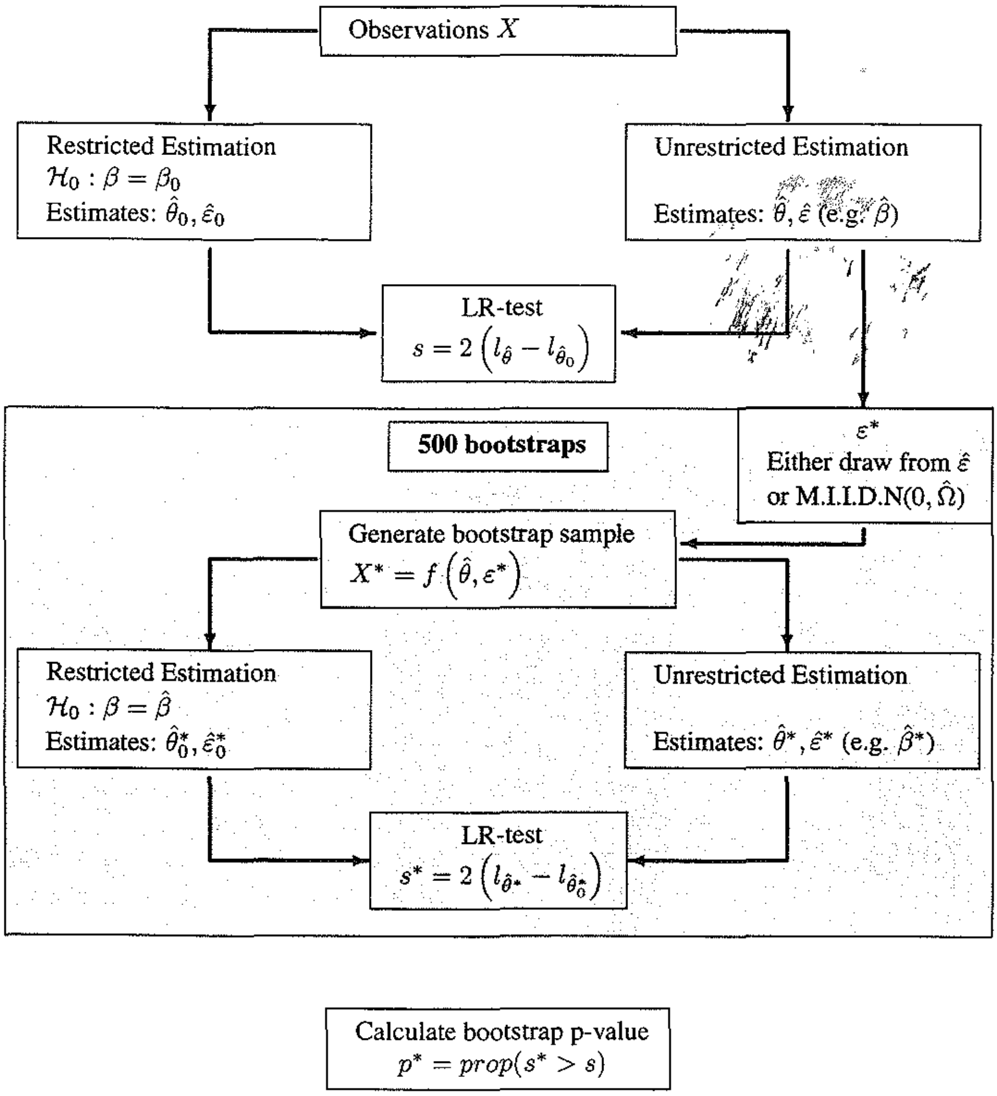



Figure 4.1: Bootstrap procedure for tests on the cointegration parameters

The test statistic is the likelihood ratio test (which is the only one allowing a Bartlett correction).

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

If we have a simple hypothesis on only part of the cointegration space, $\beta_1 = \beta_1^0$, we take the following null hypothesis in step 3:

$$\beta_1 = \hat{\beta}'(\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}'\beta_1^0 \tag{4.6}$$

which is easily seen to converge to $\beta_1^0$ if $H_0$ is true.

As mentioned in the introduction, Davidson and MacKinnon (2000) recently put forth a computationally cheap double bootstrap procedure which may deliver results superior to the standard bootstrap just outlined[4]. The idea behind the double bootstrap, proposed by Beran (1988) is that of correcting the possible bias in the bootstrap procedure implemented by a second application of the bootstrap. For instance, in the case of a test the aim of the second-level application of the bootstrap would be to estimate, and thus correct for, the bias $(p_i - i)$, where $p_i$ is the $p$-value of the $i$-level bootstrap test. Although the principle is certainly attractive, it is also very expensive, as it involves the construction of a bootstrap pseudo-population for each bootstrap redraw. It is thus practically impossible to evaluate by means of Monte Carlo experiments with the currently available computing power. On the contrary, in Davidson and MacKinnon's method there is only one second level bootstrap redraw for each first level one, so that the computing time is of the same order of magnitude of the standard bootstrap. Monte Carlo experiments are thus feasible. Going into the details of the method is clearly beyond the scope of this paper. However, the basic intuition is very simple: if the bootstrap estimate $p^* = prop(s^* > s)$ of true $p$-value of the test is distorted, we may get a better estimate by replacing $s$ with some $\widetilde{s}$ chosen so to counterbalance the distortion. Now, $s$ is by definition the $p^* - th$ quantile of the distribution of the $s^*$; hence, an obvious candidate for $\widetilde{s}$ is the same quantile of the distribution of a *second-level* bootstrap distribution. If $p^*$ is distorted downwards, such a quantile will tend to be larger than the true quantile $s$, and viceversa, thus delivering the desired effect.

The general structure of the fast double bootstrap test we shall implement is the following: (see figure 4.2 for a graphical representation)

- *Fast Double Bootstrap test*

(1) Estimate VAR on data $X$; for given cointegrating rank obtain estimates $\hat{\theta}, \widehat{\varepsilon}, \hat{\theta}_0, \widehat{\varepsilon}_0$ and test statistic $s$ for the hypothesis $H_0 : \beta = \beta^0$;
(2) Construct pseudodata: $X^* = \phi(\hat{\theta}, \varepsilon^*), \varepsilon^*$ drawn at random with replacement from $\widehat{\varepsilon}$ or $NID$;
(3) Estimate VAR on pseudodata $X^*$; obtain $\hat{\theta}^*, \widehat{\varepsilon}^*$, $\hat{\theta}_0^*, \widehat{\varepsilon}_0^*$ and test statistic $s^*$ for the hypothesis $H_0^*$: $\beta = \hat{\beta}$;
(4) Construct second-level pseudodata $X^{**} = \phi(\hat{\theta}^*, \varepsilon^{**}), \varepsilon^{**}$ drawn at random with replacement from $\widehat{\varepsilon}^*$ or $NID$;
(5) Estimate VAR on second-level pseudodata $X^{**}$; obtain $\hat{\theta}^{**}, \widehat{\varepsilon}^{**}$, $\hat{\theta}_0^{**}, \widehat{\varepsilon}_0^{**}$ and test statistic $s^{**}$ for the hypothesis $H_0^{**}$: $\beta = \hat{\beta}^*$;

Repeat (2)-(5) a large number of times

(6) Compute bootstrap $p$-value: $p^* = prop(s^* > s)$.
(7) Compute fast double bootstrap $p$-value type 1: $p_1^{**} = prop(s^* > Q_{p^*}^{**})$,where $Q_{p^*}^{**}$ is the $p^*$ quantile of the $s^{**}$'s.

A (costless) further step is advisable:

(8) Compute fast double bootstrap $p$-value type 2: $p_2^{**} = 2p^* - prop(s^{**} > s)$.

[4]Although Davidson and MacKinnon's analytical results are valid only for one-sided tests with asymptotic N(0,1) distributions, some simulation evidence suggests that the properties may extend to the asymptotic $\chi^2$ of interest here.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

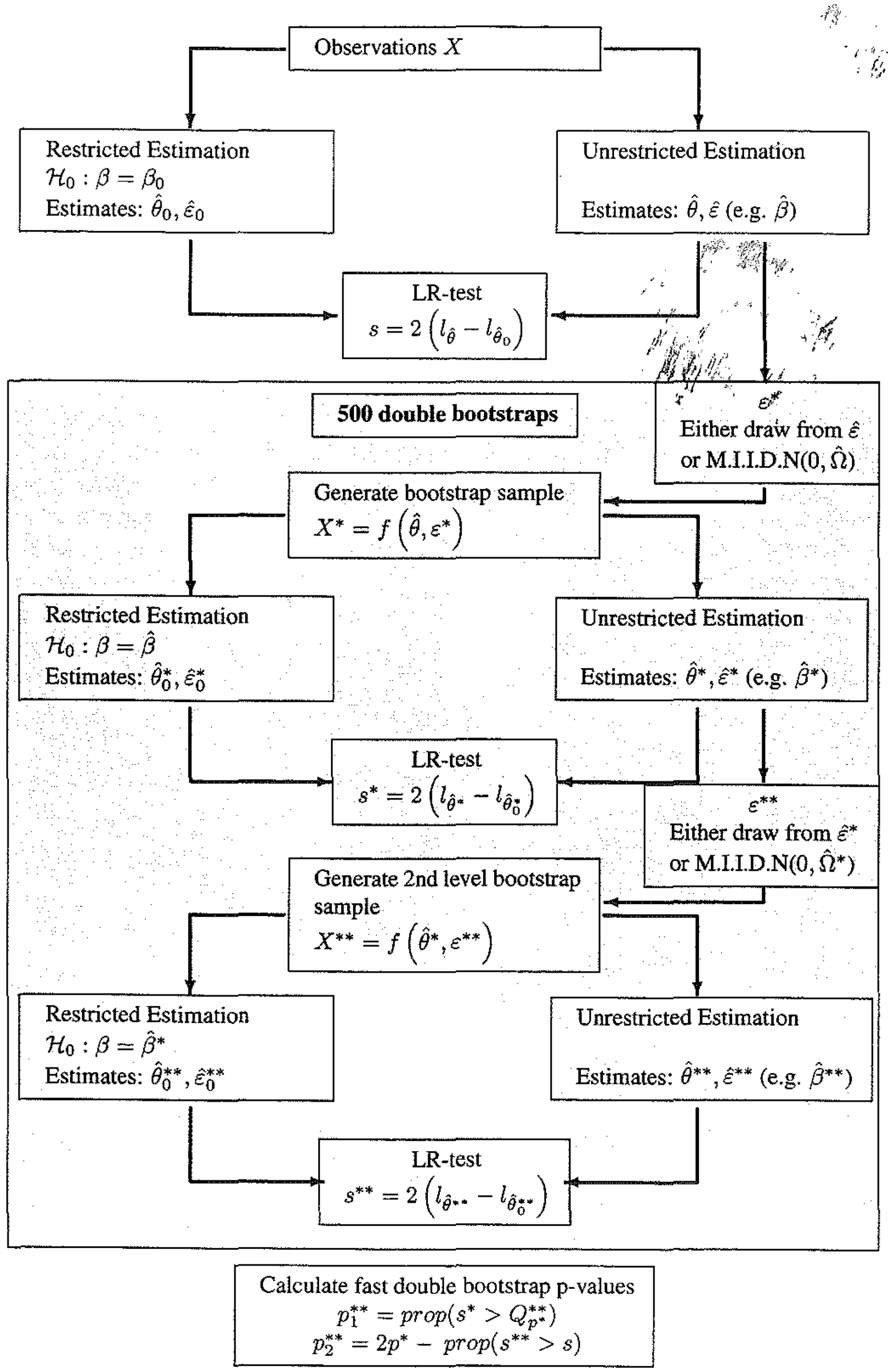



Figure 4.2: Fast Double Bootstrap procedure for tests on the cointegration parameters

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

Again, the intuition here is that if for instance $p^* > p$, we can expect $prop(s^{**} > s) > p^*$, so that $p_2^{**}$ will be closer to $p$ than $p^*$. However, $p_2^{**}$ may not be greater than $2p^*$ and it may be negative, two undesirable features that suggest limiting its use to a reliability check: if the difference between the two $p$-values is sizebale neither of them should be trusted.

### 4.3. Design of the Monte Carlo experiment

On the basis of the simulation results reported by Gredenhoff and Jacobson (2001) and Fachin (2000), the key characteristics of the DGP to be controlled in the experiments are the dimension of the system, i.e. number of variables and lags, and its long-run structure, i.e. number of the cointegrating relationships and the speed at which the system adjusts to them. Estimation of systems of higher dimension (both in terms of number of variables and lags) demand more from the data, and thus it is (ex-post) not surprising to see that both the asymptotic test and the bootstrap test proposed by Gredenhoff and Jacobson (2001) perform better in smaller systems. A crucial remark here is that the simple bivariate DGPs employed in virtually all simulation studies do suffer from loss of generality, a fact not suspected so far. The experimental design adopted here will thus generalize to a multivariate system the classical DGP used by a number of studies starting with Engle and Granger (1987), which allows an easy control of the speed of adjustment. We shall consider systems including $p = 5$ random variables and with $r = 1$ or 2 cointegrating relationships. Let $x_t = [x_{1t} \dots x_{5t}]'$ be the column vector of the realizations of the random variables of interest at time $t = 1, \dots, T$ , $u_t = [u_{1t} \dots u_{5t}]'$ the errors, $\epsilon_t = [\varepsilon_{1t} \dots \varepsilon_{5t}]'$ the noise, whose stochastic structure will be discussed in detail below, and $t$ a time trend. Our DGP is then given by

$$\begin{bmatrix} \beta_1' \\ \vdots \\ \beta_5' \end{bmatrix} \begin{bmatrix} x_t \\ t \end{bmatrix} = u_t \tag{4.7}$$

$$\Phi u_t = \epsilon_t \tag{4.8}$$

with
$\Phi = diag(\phi), \phi = \begin{bmatrix} \phi_1(L) & \phi_2(L) & \phi_3(L) & \phi_4(L) & \phi_5(L) \end{bmatrix}.$

Although the Bartlett corrections do depend on the parameters of the system, in order to keep the size of the experiment within manageable dimensions in the size simulations the cointegrating coefficients will be kept fixed across trials to either zero or 1, with the vectors resembling quite closely those used by Haug (1996), while in the power simulations we shall consider a few values in the range $[0.5, 1.5]$. Given that we are using a full-information method we do not need to worry about endogeneity; we shall thus consider a very simple structure, with one stochastic trend ($X_p$) transmitted to the first $r$ variables of the system, while the remaining $p-r-1$ follow independent random walks. The details in the two cases are as follows:

(*a*) $r = 1$
$\beta_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & \beta_{15} & 0.01 \end{bmatrix}$ is the cointegration vector.
All the other relations are non-stationary:
$\beta_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}; \beta_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix};$
$\beta_4 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}; \beta_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix};$
$\phi_1(L) = (1, \varphi_1 L, \dots, \varphi_k L^k);$
$\phi_2(L) = \phi_3(L) = \phi_4(L) = \phi_5(L) = (1, -L).$


DOI: 10.2870/52609

(*b*) $r = 2$
$\beta_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0.01 \end{bmatrix}$ becomes a cointegration vector.
All the other $\beta's$ are as in case $(a)$.
$\phi_1(L) = \phi_2(L) = (1, \varphi_1 L, \ldots, \varphi_k L^k)$;
$\phi_3(L) = \phi_4(L) = \phi_5(L) = (1, -L)$.

Some simple considerations will allow great simplification of the design as far as the $\varepsilon$'s are concerned. First of all, in previous work on the related topic of stationary VARs Fachin and Bravetti (1996) found that the shape of the distribution of the shocks does not appear to have a significant impact on the performances of asymptotic procedures. Further, the expectation that with a full-information method, their covariance structure should not matter either has been confirmed in the case of a simple bivariate DGP by Fachin (2000). We shall thus assume $\varepsilon = [\varepsilon_1 \ldots \varepsilon_p] \sim MIIDN(0, I)$.

Finally, the number of both Monte Carlo replications and bootstrap redrawings has been fixed to 500: on the basis of previous work and some pilot experiments we concluded that the gain in precision deliver by higher numbers of either was not worth the higher computing costs and longer calendar time required. At 0.05 the Monte Carlo standard error will thus be about 0.010.

In table 4.1 we give an overview of the parameter values in the various experiments. In the benchmark case we have 1 cointegration vector and test that the cointegration parameters is known. Furthermore the model has (case (a)), 2 lags, and fairly slow adjustment ($\varphi_1 = \varphi_2 = -0.35$) We have 100 observations ($T = 100$), and for the bootstrap algorithm, we resample with replacement from the estimated errors $\hat{\varepsilon}$ (and in the second level bootstrap from $\hat{\varepsilon}^*$ ).

Furthermore $\beta_{15} = 1$ and we thus test $\mathcal{H}_0 : \beta_1^0 = \beta_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0.01 \end{bmatrix}$. For the benchmark case and each of the other cases we also execute a power experiment in which we set $\beta_{15} = 0.5$ and test $\beta_1^0 = \beta_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0.01 \end{bmatrix}$.

The complexity of the DGP is such that we are unable to execute a full factorial design over all variations, we consider relevant. We thus provide

Each time we only deviate in one respect from our benchmark DGP, which is the first experiment. In the second we test case (b), that is two cointegration vectors and test that either one or both vectors are known. Next we increase the sample size to find out whether the corrections are working with 400 observations: we do not regularly find such large samples in time series analysis, but find that even with that many observations, the asymptotic tests do not work well. In the fourth experiment we increase the lag length of the VAR to four. The sum of the adjustment coefficients is kept constant at 0.7. Subsequently we test the effect of an increase in the speed of adjustment to equilibrium. This increases the signal to noise ratio and the performance of the asymptotic test (remember that the Bartlett correction depends on $\sum_{i=0}^{\infty} P^i$). In the last experiment we try the parametric bootstrap.

Finally we compute a power curve for the benchmark case and $\beta_{15}$ in $(0.5, 1.50)$, with $\beta_{15}^0 = 1$ as usual, both for the bootstrap and Bartlett correction we propose, that is those based on the estimates under the alternative and one based on estimates under the null.

## 4.4. Results

Although the results of the simulations amount to a considerable mass, their essence is quite simple, and summarised in Table 4.2; in the following tables a few details are highlighted, with baseline results repeated in different tables in order to facilitate comparisons. In all the cases reported and discussed in this section the nominal significance level of the tests is always 5%, with results for different values available on request.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

First of all, in the sample sizes we typically encounter in applied econometric work (100 observations) the asymptotic tests deliver disastrous performance as far as type I errors are concerned: at times they exceed 50% at the nominal 5% level. All the alternative procedures (Bartlett correction, simple and fast double bootstrap) are able to reduce substantially the size distortion in all our experiments, but are unable to eliminate it: in the case of rank=1 the minimum rejection rate, delivered by the fast double bootstrap type 1, is 26%, while in the case of rank=2 and test on one vector the rejection rate of all bootstrap procedures and the Bartlett corrected test is 15%. The two types of fast double bootstrap $p$-values are always very close, confirming that the procedure is reliable in our context. The power loss from using the procedures with lower size distortion is acceptable, with the rejection rates always over 70%. This finding will be confirmed by the power curve reported in table 4.7. To understand the point of basing both the bootstrap and Bartlett correction on the unrestricted estimates, a glance at the power curves in table 4.8 suffices: whereas the size performance is indeed slightly better than in table 4.7, the level of type II errors is unacceptably high, such that the power curves are almost flat. In the last row we report the number of cases where the highest estimated root is explosive: when the discrepancy between DGP and model becomes large, this percentage rises rapidly and corroborates theorem 1 of this paper[5].

Another key point from Table 4.2, is that the size performance of all test procedures of $\mathcal{H}_0 : \beta_1 = \beta_1^0$ in a model with 2 vectors is markedly better than the hypothesis $\mathcal{H}_0 : \beta = \beta^0$ in a model with one cointegrating vector.

For $T = 400$ all corrected tests achieve correct size and 100% power while the Type I error of the asymptotic test is still higher than the nominal size (cf. Table 4.3).

Increasing the length of the VAR also has large adverse effects on the test (cf. Table 4.4): thus, contrary to somehow common wisdom and in accord with Abadir et al. (1999), parsimony in the estimation of the VAR seems to be a rather important virtue.

How sensitive are the performances of the tests to the speed of adjustment to equilibrium? Unsurprisingly, the answer is, a lot. Cutting $\phi$ (the sum of the coefficients of the autoregressive polynomial describing the dynamics of the errors in the cointegrating relationships) from 0.7 to

[5] $tr(\sum_{i=0}^{\infty} P^i)$ does not converge if $P$ contains an explosive root. However computationally we use the standard formula $(\sum_{i=0}^{\infty} \lambda^i) = \frac{1}{1-\lambda}$ to calculate the Bartlett correction both in the convergent case (when it is valid) and the non-convergent case.

There is nothing, which prevents the Bartlett correction factor from being smaller than -1: this is a known problem in the literature. We assign a p-value of 1 to these cases. In all the published and unpublished simulations we did, this only happened in those of table 4.8.

| | Benchmark | Variation | Table | Page |
|---|---|---|---|---|
| Cointegration rank | $r = 1$ | $r = 2$ | 4.2 | 50 |
| Number of vectors tested | $s = 1$ | $s = 1, 2$ | 4.2 | 50 |
| Sample size | $T = 100$ | $T = 400$ | 4.3 | 50 |
| VAR lag length | $k = 2$ | $k = 4$ | 4.4 | 50 |
| | $\varphi_1 = \varphi_2 = -\frac{0.7}{2}$ | $\varphi_1 = -\frac{0.7}{2}, \varphi_2 = -\frac{0.7}{3}, \varphi_3 = \varphi_4 = -\frac{0.7}{16}$ | | |
| Speed of adjustment | $\varphi_1 = \varphi_2 = -\frac{0.7}{2}$ | $\varphi_1 = \varphi_2 = -\frac{0.4}{2}$ | 4.5 | 51 |
| Resampling of errors | parametric | resample from $\widehat{\varepsilon}$ and $\widehat{\varepsilon}^*$ | 4.6 | 51 |
| Power curve | based on $\hat{\theta}$ | | 4.7 | 52 |
| Power curve | | based on $\hat{\theta}_0$ | 4.8 | 52 |

Table 4.1: Design of Monte Carlo experiment for small sample corrections

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

0.4 causes generally a more than proportional fall of the Type I error (for instance, that of the fast double bootstrap type 1 falls from 26% to 10%, see table 4.5).

Given the good results delivered by the bootstrapped tests, it is of some interest to check if using resampled or parametrically generated errors makes any difference. The results reported in Table 4.6 suggest that it does not, and thus the parametric bootstrap (easier to implement) may be adopted in practice. However, some caution is needed here, as in our experiments the same parametric hypothesis (normality) is used both in the generation of the Monte Carlo and bootstrap errors. Further research with different error processes for the Monte Carlo and bootstrap DGPs (for instance a leptokurtic error distribution in the DPG and resampling for a normal distribution) is needed.

Finally, a noteworthy finding is that the power curves of all the variants of bootstrap tests are rather steep (table 4.7 and figure 3(b)). Although these results are specific to a single signal/noise ratio, they do suggest that the risk of unacceptable power losses from using some type of bootstrap test rather than the asymptotic or Bartlett corrected tests is likely to be remote.

## 4.5. Conclusions

We have compared different variants of bootstrap and Bartlett-corrected tests in a DGP which is relatively unfavourable, but reproduces some features of real life empirical applications: a relatively large system (5 variables and 2 or 4 lags), and rather slow adjustment to long-run equilibrium. With such a complex DGP the caveats common to all simulation studies are even more important than usual. Our design depends on over 120 parameters, the vast majority of which had to be kept fixed across all experiments, and thus we must be extremely cautious in reaching any conclusion.

Further, the type of tests examined assumes full knowledge of the tested cointegrating vectors, a rare event in practice: however, they are the only tests for which the Bartlett correction is available. Indeed, the Bartlett correction has not been derived yet for many cases of strong empirical interest (e.g., hypotheses of the kind $\beta_i = H_i\varphi_i$ and in general models with impulse dummies) and hence the bootstrap may in fact be the only alternative to the asymptotic $p$-values. With all these caveats, our recommendations are the following:

*(i)* Asymptotic tests should be used in no circumstance;

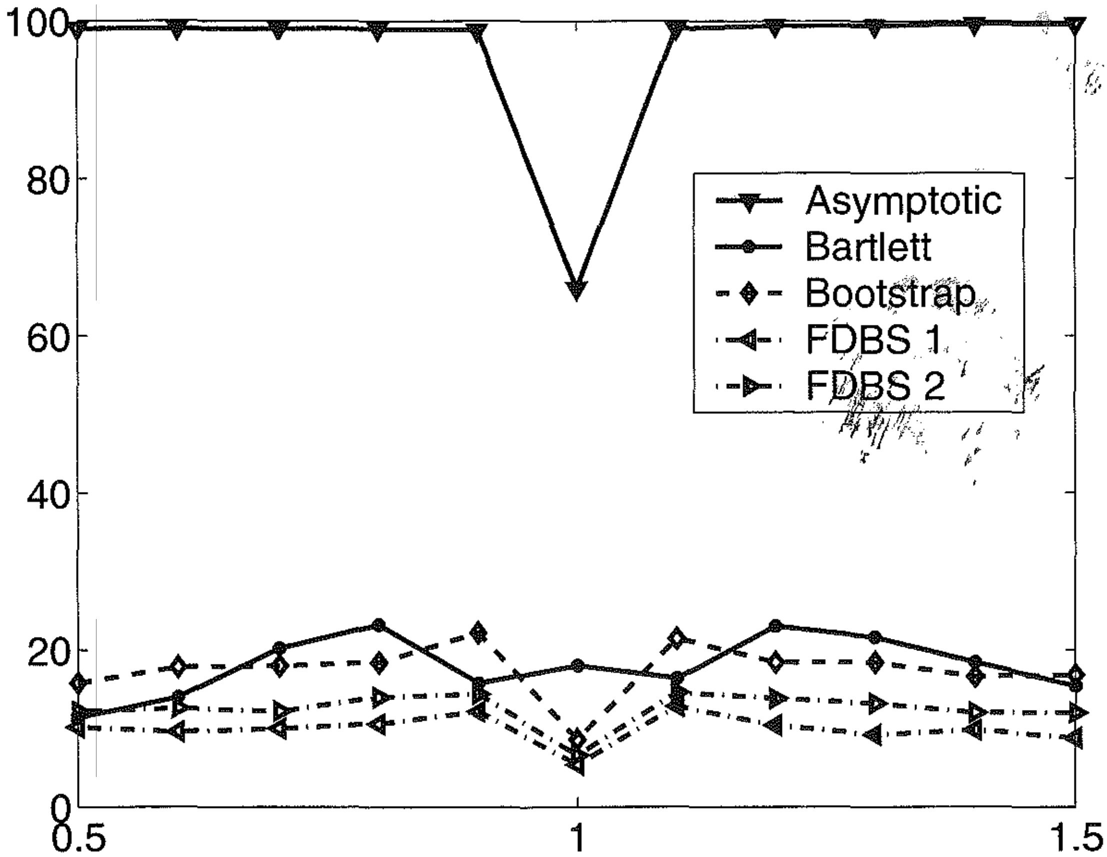
*(ii)* Bartlett-corrected tests may be used provided considerable caution is exercised, as their Type I error is often much larger than the nominal size;

*(iii)* Bootstrap tests, with a somehow lower size distortion than the Bartlett corrected tests accompanied by limited power losses, may also be used; the fast double bootstrap of Davidson and MacKinnon (2000) delivers the best performance, and thus it appears to be a powerful tool for applied work, especially in the many cases when the Bartlett correction is not available.

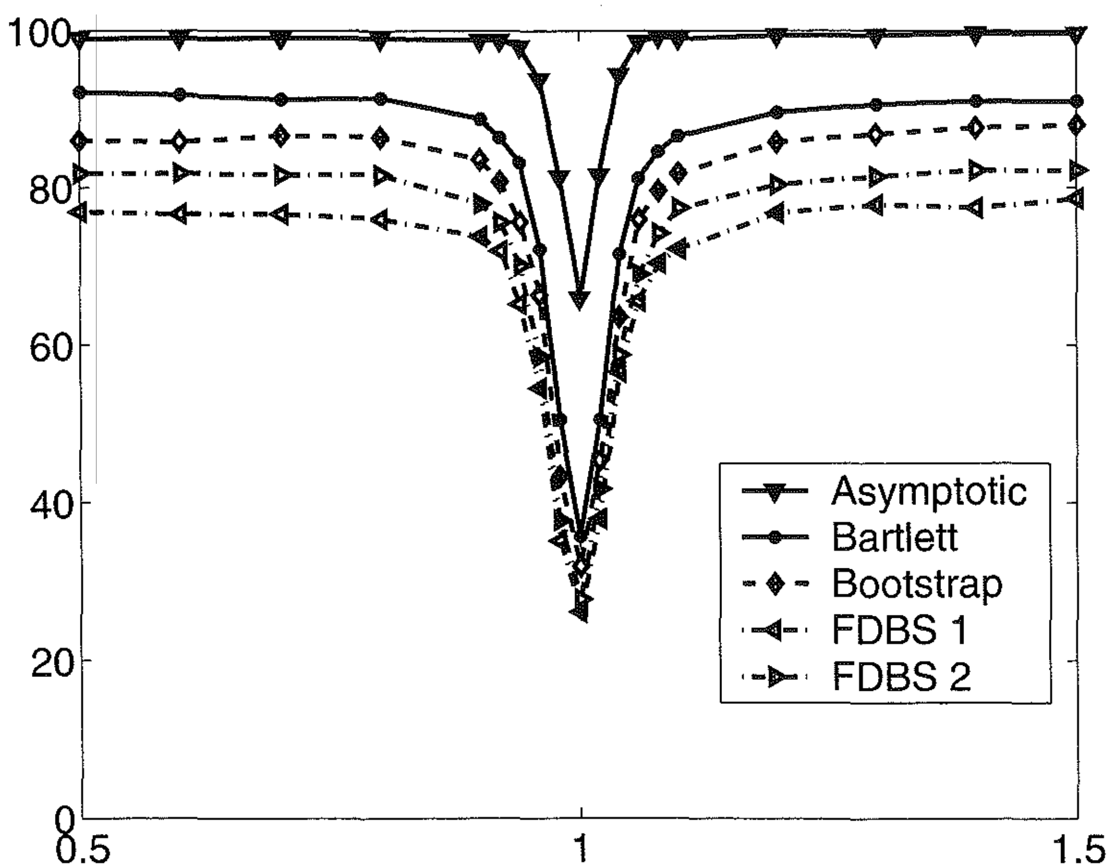
We stress that both the Bartlett correction and the bootstrap should always be based on the unrestricted estimate of $\beta$.

Among the many points that remain open, two are especially important: $(a)$ the development of equivalent hypothesis, like (4.6) for $\mathcal{H}_0 : \beta_1 = \beta_1^0$ for more general restrictions on $\beta$, like $\beta_i = H_i\varphi_i$ with an accurate Monte Carlo study of their properties and $(b)$ theoretical results on the asymptotics of the (fast double) bootstrap in cointegrated systems.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609




(a) power curves based on restricted estimates $\hat{\theta}_0$




(b) power curves based on unrestricted estimates $\hat{\theta}$

Figure 4.3: Power curves for test on cointegration coefficients

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

| *1 to 2 cointegration vectors, test on 1 to 2 vectors* $\phi = 0.7, T = 100, k = 2$ | | | | | | |
|---|---|---|---|---|---|---|
| *rank, tested vectors* | 1,1 | | 2,1 | | 2,2 | |
| *Test* | *size* | *power* | *size* | *power* | *size* | *power* |
| Asymptotic | 66.0 | 99.0 | 39.2 | 97.6 | 68.6 | 98.2 |
| Bartlett | 35.8 | 92.2 | 15.8 | 79.2 | 33.2 | 82.2 |
| Bootstrap | 32.0 | 86.0 | 15.2 | 77.2 | 28.2 | 74.6 |
| $FDB_1$ | 26.2 | 76.0 | 13.4 | 68.0 | 20.0 | 62.2 |
| $FDB_2$ | 27.8 | 81.8 | 14.2 | 71.4 | 23.6 | 68.2 |

*nominal significance level: 5%;* $FDB_i$: Fast Double Bootstrap type $i$
*power simulations:*
*case (1,1)* $H_0 : \beta_1^0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \end{bmatrix}$, DGP: $\beta_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0.5 \end{bmatrix}$
*case (2,1)*: as case (1,1) with DGP: $\beta_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \end{bmatrix}$
*case (2,2)*: as case (2,1) with $H_0 : \beta_2^0 = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \end{bmatrix}$

Table 4.2: Benchmark case small sample correction for tests on cointegration vectors

| *1 cointegrating vector, test on 1 vector* $\phi = 0.7, T = 100$ and 400, $k = 2$ | | | | |
|---|---|---|---|---|
| *T* | 100 | | 400 | |
| *Test* | *size* | *power* | *size* | *power* |
| Asymptotic | 66.0 | 99.0 | 11.0 | 100.0 |
| Bartlett | 35.8 | 92.2 | 5.6 | 100.0 |
| Bootstrap | 32.0 | 86.0 | 6.2 | 100.0 |
| $FDB_1$ | 26.2 | 76.0 | 5.6 | 100.0 |
| $FDB_2$ | 27.8 | 81.8 | 5.8 | 100.0 |

*nominal significance level: 5%*
*power simulations: see Table 4.2*

Table 4.3: Sample size and small sample corrections

| *1 cointegrating vector, test on 1 vector* $\phi = 0.7, T = 100, k = 2$ and 4 | | | | |
|---|---|---|---|---|
| *lags* | 2 | | 4 | |
| *Test* | *size* | *power* | *size* | *power* |
| Asymptotic | 66.0 | 99.0 | 85.4 | 99.4 |
| Bartlett | 35.8 | 92.2 | 53.2 | 91.0 |
| Bootstrap | 32.0 | 86.0 | 39.2 | 82.0 |
| $FDB_1$ | 26.2 | 76.0 | 32.4 | 68.8 |
| $FDB_2$ | 27.8 | 81.8 | 35.6 | 74.4 |

*nominal significance level: 5%*
*power simulations: see Table 4.2*

Table 4.4: Lag length and small sample corrections

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

| *1 cointegrating vector, test on 1 vector* | | | | |
|---|---|---|---|---|
| $\phi = 0.7$ and $0.4$, $T = 100$, $k = 2$ | | | | |
| $\phi$ | 0.7 | | 0.4 | |
| *Test* | *size* | *power* | *size* | *power* |
| Asymptotic | 66.0 | 99.0 | 33.0 | 99.6 |
| Bartlett | 35.8 | 92.2 | 17.0 | 97.6 |
| Bootstrap | 32.0 | 86.0 | 14.2 | 96.8 |
| $FDB_1$ | 26.2 | 76.0 | 10.8 | 94.4 |
| $FDB_2$ | 27.8 | 81.8 | 11.8 | 95.6 |
| *nominal significance level: 5%* | | | | |
| *power simulations: see Table 4.2* | | | | |

Table 4.5: Speed of adjustment and small sample corrections

| 1 cointegrating vector, test on 1 vector | | | | |
|---|---|---|---|---|
| $\phi = 0.7, T = 100, k = 2$ | | | | |
| Type of bootstrap | Non-Parametric | | Parametric | |
| Test | size | power | size | power |
| Asymptotic | 66.0 | 99.0 | 66.0 | 99.0 |
| Bartlett | 35.8 | 92.2 | 35.8 | 92.2 |
| Bootstrap | 32.0 | 86.0 | 32.0 | 86.4 |
| $FDB_1$ | 26.2 | 76.0 | 25.0 | 76.2 |
| $FDB_2$ | 27.8 | 81.8 | 27.0 | 80.2 |
| nominal significance level: 5% | | | | |
| power simulations: see Table 4.2 | | | | |

Table 4.6: Non-parametric bootstrap and small sample corrections

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

| $\phi = 0.7, T = 100, k = 2$ | | | | | |
|---|---|---|---|---|---|
| *Test* / $\beta_{15}$ | *Asymptotic* | *Bartlett* | *Bootstrap* | $FDB_1$ | $FDB_2$ |
| 0.5 | 99.0 | 92.2 | 86.0 | 76.9 | 81.8 |
| 0.6 | 99.0 | 91.8 | 85.8 | 76.6 | 81.8 |
| 0.7 | 99.0 | 91.2 | 86.6 | 76.6 | 81.6 |
| 0.8 | 99.0 | 91.4 | 86.4 | 76.0 | 81.6 |
| 0.9 | 98.8 | 88.8 | 83.6 | 73.8 | 78.0 |
| 0.92 | 98.8 | 86.4 | 80.8 | 72.0 | 75.6 |
| 0.94 | 98.0 | 83.2 | 75.6 | 65.2 | 70.2 |
| 0.96 | 93.8 | 72.2 | 66.4 | 54.6 | 58.6 |
| 0.98 | 81.4 | 50.6 | 43.4 | 35.2 | 37.8 |
| *1.0* | *66.0* | *35.8* | *32.0* | *26.2* | *27.8* |
| 1.02 | 81.6 | 50.6 | 45.4 | 38.0 | 41.8 |
| 1.04 | 94.4 | 71.6 | 63.6 | 56.6 | 58.8 |
| 1.06 | 98.6 | 81.2 | 76.0 | 65.6 | 69.0 |
| 1.08 | 99.2 | 84.6 | 79.6 | 70.4 | 74.2 |
| 1.1 | 99.0 | 86.6 | 81.8 | 72.2 | 77.4 |
| 1.2 | 99.4 | 89.6 | 85.8 | 76.8 | 80.4 |
| 1.3 | 99.4 | 90.6 | 86.8 | 77.8 | 81.4 |
| 1.4 | 99.6 | 91.0 | 87.6 | 77.4 | 82.2 |
| 1.5 | 99.6 | 91.0 | 88.0 | 78.6 | 82.2 |
| *nominal significance level: 5%* | | | | | |

Table 4.7: Power curve based on unrestricted estimates

| $\phi = 0.7, T = 100, k = 2$ | | | | | | |
|---|---|---|---|---|---|---|
| *Test* / $\beta_{15}$ | *Asymptotic* | *Bartlett* | *Bootstrap* | $FDB_1$ | $FDB_2$ | Explosive Roots *(% of simulations)* |
| 0.5 | 99.0 | 11.4 | 15.8 | 10.2 | 12.2 | 34.8 |
| 0.6 | 99.0 | 14.0 | 17.8 | 9.6 | 12.6 | 30.4 |
| 0.7 | 99.0 | 20.2 | 18.0 | 10.0 | 12.2 | 23.4 |
| 0.8 | 99.0 | 23.2 | 18.4 | 10.6 | 14.0 | 12.6 |
| 0.9 | 98.8 | 15.8 | 22.2 | 12.2 | 14.4 | 0.2 |
| *1.0* | *66.0* | *18.0* | *8.6* | *5.4* | *6.6* | *0* |
| 1.1 | 99.0 | 16.4 | 21.4 | 12.8 | 14.6 | 0.6 |
| 1.2 | 99.4 | 23.0 | 18.4 | 10.4 | 13.8 | 9.2 |
| 1.3 | 99.4 | 21.6 | 18.4 | 9.2 | 13.2 | 20.4 |
| 1.4 | 99.6 | 18.4 | 16.6 | 9.8 | 12.0 | 28.8 |
| 1.5 | 99.6 | 15.4 | 16.8 | 8.8 | 12.0 | 32.2 |
| *nominal significance level: 5%* | | | | | | |

Table 4.8: Power curve based on restricted estimates

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

K Ø B E N H A V N


DOI: 10.2870/52609

CHAPTER 5

# A Bartlett correction in stationary autoregressive models

## 5.1. Introduction

Vector Autoregressive Models (VAR) are widely applied both in macroeconomics and econometrics. Estimation of these models is often done by means of maximum likelihood methods. For almost every test statistics only asymptotic results are available regarding the distribution of the statistic under the null hypothesis. In small samples, the size distortion can be particularly large if large models (in terms of number of variables and lags) are used for relatively short spans of data series. A Bartlett correction (Bartlett, 1937) to a likelihood ratio test is one method to correct for the size distortion.

In this paper we consider the following multivariate model:

$$Y_t = AX_t + \eta_{2t}$$

where

$$X_t = Q(L)\eta_{t-1} = Q_0\eta_{t-1} + Q_1\eta_{t-2} + Q_2\eta_{t-3} + \dots$$
$$\eta_t = \left[\begin{array}{cc} \eta_{1t}' & \eta_{2t}' \end{array}\right]' \sim MIIDN(0, \Omega)$$

under the assumption that $Q(L)$ is an exponentially decreasing polynomial and we derive the Bartlett correction for a simple hypothesis on $A$ $\mathcal{H} : A = A_0$ both when $var(\eta_t)$ is known (theorem 10) and when it is unknown (theorem 9).

After a short introduction into Bartlett corrections and the two main theorems, we consider three specific applications. In section 5.4 we consider likelihood ratio tests for the absence of autocorrelation in a VAR model and in section 5.5 we consider a more general hypothesis on the autoregressive parameters of the VAR. Section 5.6 contains the Bartlett correction for two different tests of no long-run feedback in the cointegrated VAR model. These last three sections all contain Monte Carlo studies of the derived results.

Conclusions are drawn in section 5.7. The longest section, the proof of the two main theorems and two other theorems, is given in the only appendix of this chapter, section 5.A.

## 5.2. Bartlett corrections

Let $l_T(\theta), \theta = (\theta_1, \theta_2)$ denote the log likelihood function of $T$ observations. Then the log likelihood ratio ($W_T$) test statistic for the null hypothesis $\mathcal{H}_0 : \theta_1 = \theta_1^0$ equals

$$-2\ln LR\left[\theta_1 = \theta_1^0 | \theta\right] = W_T = -2\left(\max_{\theta_2} l_T\left(\theta_1^0, \theta_2\right) - \max_{\theta_1,\theta_2} l_T\left(\theta_1, \theta_2\right)\right)$$

Under a number of regularity conditions, this test statistic converges in distribution. In many cases this is the $\chi^2$-distribution, but it can also also be a different distribution; The rank test in cointegration analysis (Johansen, 1988, 1991) for instance, converges to an expression involving stochastic integrals.

In small samples, the asymptotic distribution does not necessarily provide a good approximation to the actual one. The idea of the Bartlett correction (Bartlett, 1937) is to expand the



Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

expectation of the LR-statistic:

$$E[W_T] = f\left(1 + \frac{B(\theta)}{T} + O\left(T^{-2}\right)\right)$$

where $f = \lim_{T\to\infty} E_\theta[W_T]$ and then to define the Bartlett adjusted likelihood ratio statistic $W_T^{BC}$ as:

$$W_T^{BC} = W_T / \left(1 + B(\theta)/T\right)$$

The term $B(\theta)/T$ shall be referred to as the Bartlett Factor (BF). It generally depends on the parameters of the model. When substituting values, it will sometimes make a difference whether we take the true values from the data generating process, the restricted estimates (that is the maximum likelihood estimates under the null hypothesis), or the unrestricted estimates.

Lawley (1956) proves that for stationary series and under a number of stochastic order conditions that the Bartlett Correction (BC) not only corrects the first moment up to $O(T^{-2})$, but also all higher moments. Barndorff-Nielsen and Hall (1988) prove the same result elegantly and demonstrate that is holds when $B\left(\hat{\theta}\right)$ replaces $B(\theta)$, where $\hat{\theta}$ is a $\sqrt{n}$-consistent estimator of $\theta$. Often small sample corrections are referred to as Bartlett correction only if the result of Lawley holds. We shall however also refer to any division of the likelihood ratio test statistic by its expectation as a Bartlett correction.

Nielsen (1997) and Johansen (2000a, 2002a,b) show that a Bartlett correction can be useful in models with unit roots. Jensen and Wood (1997) show by means of calculation of the first two moments that the result of Lawley does not hold for the Dickey-Fuller distribution. More precisely they show that $E[W_T] = f\left(1 + \frac{b_1}{T}\right) + O\left(T^{-2}\right)$ and $E[W_T^2] = 2f\left(1 + \frac{2b_2}{T}\right) + O\left(T^{-2}\right)$, but that $b_1 \neq b_2$.

General overviews of Bartlett and related corrections can be found in Jensen (1993) and Cribari-Neto and Cordeiro (1996) .

A large number of Bartlett correction concern univariate models, but Attfield (1995, 1998) derives a number of Bartlett corrections for simultaneous systems with fixed exogenous regressors. In this paper we consider multivariate models with lagged endogenous regressors.

### 5.3. The model and main results

Let us consider the following statistical model $\mathcal{K}_1$:

$$Y_t = AX_t + \eta_{2t} \tag{5.1}$$

where

$$\begin{aligned} X_t &= Q(L)\eta_{t-1} = Q_0\eta_{t-1} + Q_1\eta_{t-2} + Q_2\eta_{t-3} + \dots \\ \eta_t &= \left[\begin{array}{cc} \eta_{1t}' & \eta_{2t}' \end{array}\right]' \sim MIIDN(0, \Omega) \\ A &\in \mathcal{R}^{q\times n}, \Omega \in \mathcal{S}_{p\times p} \end{aligned}$$

and the null hypothesis

$$\mathcal{H}_0 : A = A_0$$

$\mathcal{R}$ is the space of real numbers $\mathcal{S}_{p\times p}$ the space of positive definite matrices of dimension $p\times p$. The process $\eta_t$ is of dimension $p$ and $\eta_{2t}$ is of dimension $q(\leq p)$. The independent variable $X_t$ $(1\times n)$ is a moving average process. The innovations $\eta_{2t}$ of the dependent variable $Y_t$ are a subset of the innovations $\eta_t$, which constitute the moving average process $X_t$. This model allows for the possibility that $X_t$ contains not only past values of $Y_t$, but also past value of exogenous variables, but not present values of exogenous variables. The model does not contain any deterministic terms.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

Define

$$C_i = Q_i \Omega^{\frac{1}{2}}, \qquad i = 0, 1, 2, \ldots \tag{5.2}$$

such that $X_t = C(L)\varepsilon_{t-1}$ and $\varepsilon_t = \Omega^{-\frac{1}{2}}\eta_t$ is distributed $MIIDN(0, I_p)$

Define the jth autocovariance matrix of $X_t$ as $\Gamma_j = E\left[X_t X_{t-j}'\right] = \sum_{\alpha=0}^{\infty} Q_{\alpha+j}\Omega Q_\alpha' = \sum_{\alpha=0}^{\infty} C_{\alpha+j}C_\alpha'$ and its variance $\Phi = \Gamma_0$.

In the examples, it will be clarified how seemingly more general situations, like multiple lags, are in fact special cases of the following theorem, which concerns a simple hypothesis on the parameter $A$ :

THEOREM 9. *For the statistical model $\mathcal{K}_1$, the expected value of the likelihood ratio test of the null hypothesis that $\mathcal{H}_0 : A = A_0$ equals:*

$$E\left[W_T\right] \stackrel{1}{=} nq + \frac{1}{T}\left(\beth(n,q) + \daleth(n,q,\{C_i\})\right) \tag{5.3}$$

*where:*

$$\beth = \frac{1}{2}\left(-4q + qn + q^2 n + qn^2\right) \tag{5.4}$$

$$\daleth = \sum_{\beta,\kappa=0}^{\infty} tr\left\{\left[C_\kappa'\Phi^{-1}\Gamma_{\kappa+1}'\Phi^{-1}\Gamma_{\beta+1}\Phi^{-1}C_\beta\right]_{22}\right\} \tag{t1}$$

$$+ 2\sum_{\beta,\kappa=0}^{\infty} tr\left\{\left[C_\kappa'\Phi^{-1}\Gamma_{\kappa+1}'\Phi^{-1}C_\beta\right]_{22}\right\} tr\left\{\Gamma_{\beta+1}'\Phi^{-1}\right\} \tag{t2}$$

$$+ \sum_{\beta,\kappa=0}^{\infty} tr\left\{\left[C_\kappa'\Phi^{-1}C_\beta\right]_{22}\right\} tr\left\{\Gamma_{\kappa+1}'\Phi^{-1}\right\} tr\left\{\Gamma_{\beta+1}'\Phi^{-1}\right\} \tag{t3}$$

$$+ \sum_{\beta,\kappa=0}^{\infty} tr\left\{\left[C_\beta'\Phi^{-1}\Gamma_{\kappa+1}'\Phi^{-1}\Gamma_{\beta+1}\Phi^{-1}C_\kappa\right]_{22}\right\} \tag{t4}$$

$$+ 2\sum_{\beta,\kappa=0}^{\infty} tr\left\{\left[C_\beta'\Phi^{-1}\Gamma_{\kappa+1}'\Phi^{-1}\Gamma_{\beta+1}'\Phi^{-1}C_\kappa\right]_{22}\right\} \tag{t5}$$

$$+ \sum_{\beta,\kappa=0}^{\infty} tr\left\{\left[C_\beta'\Phi^{-1}C_\kappa\right]_{22}\right\} tr\left\{\Gamma_{\beta+1}\Phi^{-1}\Gamma_{\kappa+1}\Phi^{-1}\right\} \tag{t6}$$

$$- 2\sum_{\beta,\kappa=0}^{\infty} tr\left\{\left[C_\kappa'\Phi^{-1}C_\beta\right]_{22}\right\} tr\left\{\Gamma_{\kappa+\beta+2}'\Phi^{-1}\right\} \tag{t7}$$

$$- 2\sum_{\beta,\kappa=0}^{\infty} tr\left\{\left[C_\kappa'\Phi^{-1}\Gamma_{\kappa+\beta+2}'\Phi^{-1}C_\beta\right]_{22}\right\} \tag{t8}$$

$$- 2\sum_{\beta,\kappa=0}^{\infty} tr\left\{\left[C_\kappa'\Phi^{-1}\Gamma_{\beta+1}\Phi^{-1}C_{\beta+\kappa+1}\right]_{22}\right\} \tag{t9}$$

$$- 2\sum_{\beta,\kappa=0}^{\infty} tr\left\{\left[C_\kappa'\Phi^{-1}\Gamma_{\beta+1}'\Phi^{-1}C_{\beta+\kappa+1}\right]_{22}\right\} \tag{t10}$$

PROOF. See the appendix □

With $[M]_{22}$ we indicate the lower right hand block of dimension $q \times q$ in the matrix $M$, which itself is of dimension $p \times p$. Thus $tr\left\{[M]_{22}\right\}$ is the sum of the last $q$ elements on the main diagonal of the matrix $M$.

The expression $\daleth(n, q, C(L))$ looks complicated, but it should be borne it mind that it needs to be programmed only once and is programmed and computed relatively quickly. Furthermore it simplifies considerably in most cases. The version in the theorem has been written down with an eye on programming: it contains only two loops. The loops in the theorem go to infinity, but in all the examples and corollaries contained in this paper, the expression for $\daleth$ simplifies, such that only finite loops remain. The following expression for $\daleth(n, q, C(L))$ is useful in the corollaries and examples that will follow (we just substitute $\sum_{\alpha=0}^{\infty} C_{\alpha+\eta}C_\eta'$ for $\Gamma_\alpha$):

 DOI: 10.2870/52609

$$
\begin{aligned}
&(\text{t1'}) && \daleth = \sum\nolimits_{\beta,\eta,\kappa,\zeta=0}^{\infty} tr\left\{\left[C'_\kappa\Phi^{-1}C_\zeta C'_{\kappa+\zeta+1}\Phi^{-1}C_{\beta+\eta+1}C'_\eta\Phi^{-1}C_\beta\right]_{22}\right\}\\
&(\text{t2'}) && +2\sum\nolimits_{\alpha,\eta,\kappa,\zeta=0}^{\infty} tr\left\{\left[C'_\kappa\Phi^{-1}C_\zeta C'_{\kappa+\zeta+1}\Phi^{-1}C_\alpha\right]_{22}\right\} tr\left\{C'_{\alpha+\eta+1}\Phi^{-1}C_\eta\right\}\\
&(\text{t3'}) && +\sum\nolimits_{\alpha,\eta,\lambda,\zeta=0}^{\infty} tr\left\{\left[C'_\lambda\Phi^{-1}C_\alpha\right]_{22}\right\} tr\left\{C'_\zeta\Phi^{-1}C_{\lambda+\zeta+1}\right\} tr\left\{C'_{\alpha+\eta+1}\Phi^{-1}C_\eta\right\}\\
&(\text{t4'}) && +\sum\nolimits_{\beta,\eta,\kappa,\zeta=0}^{\infty} tr\left\{\left[C'_\beta\Phi^{-1}C_\eta C'_{\kappa+\eta+1}\Phi^{-1}C_{\beta+\zeta+1}C'_\zeta\Phi^{-1}C_\kappa\right]_{22}\right\}\\
&(\text{t5'}) && +2\sum\nolimits_{\beta,\eta,\lambda,\zeta=0}^{\infty} tr\left\{\left[C'_\beta\Phi^{-1}C_\eta C'_{\lambda+\eta+1}\Phi^{-1}C_\zeta C'_{\beta+\zeta+1}\Phi^{-1}C_\lambda\right]_{22}\right\}\\
&(\text{t6'}) && +\sum\nolimits_{\alpha,\eta,\lambda,\zeta=0}^{\infty} tr\left\{\left[C'_\alpha\Phi^{-1}C_\lambda\right]_{22}\right\} tr\left\{C'_\zeta\Phi^{-1}C_{\lambda+\eta+1}C'_\eta\Phi^{-1}C_{\alpha+\zeta+1}\right\}\\
&(\text{t7'}) && -2\sum\nolimits_{\zeta,\eta,\kappa=0}^{\infty} tr\left\{\left[C'_\kappa\Phi^{-1}C_\zeta\right]_{22}\right\} tr\left\{C'_{\kappa+\zeta+\eta+2}\Phi^{-1}C_\eta\right\}\\
&(\text{t8'}) && -2\sum\nolimits_{\zeta,\eta,\lambda=0}^{\infty} tr\left\{\left[C'_\lambda\Phi^{-1}C_\eta C'_{\lambda+\zeta+\eta+2}\Phi^{-1}C_\zeta\right]_{22}\right\}\\
&(\text{t9'}) && -2\sum\nolimits_{\kappa,\eta,\alpha=0}^{\infty} tr\left\{\left[C'_\kappa\Phi^{-1}C_{\alpha+\eta+1}C'_\eta\Phi^{-1}C_{\alpha+\kappa+1}\right]_{22}\right\}\\
&(\text{t10'}) && -2\sum\nolimits_{\kappa,\zeta,\alpha=0}^{\infty} tr\left\{\left[C'_{\alpha+\kappa+1}\Phi^{-1}C_{\zeta+\alpha+1}C'_\zeta\Phi^{-1}C_\kappa\right]_{22}\right\}
\end{aligned}
$$

In most applications the variance of $\eta_t$ is unknown. There is however little difference in deriving the main result for known and unknown variance. In section 5.5 we shall encounter one instance of a result in the literature which deals with known variance. We thus include the version of the main theorem with known variance in this paper to make results comparable. Consider the following statistical model $\mathcal{K}_2$:

$$Y_t = AX_t + \varepsilon_{2t} \tag{5.5}$$

where

$$
\begin{aligned}
X_t &= C(L)\varepsilon_{t-1} = C_0\varepsilon_{t-1} + C_1\varepsilon_{t-2} + C_2\varepsilon_{t-3} + \ldots\\
\varepsilon_t &= \left[\begin{array}{cc}\varepsilon'_{1t} & \varepsilon'_{2t}\end{array}\right]' \sim MIIDN(0, I_p)\\
A &\in \mathcal{R}^{q\times n}
\end{aligned}
$$

and the null hypothesis

$$\mathcal{H}_0 : A = A_0$$

THEOREM 10. *For the statistical model $\mathcal{K}_2$, the expected value of the likelihood ratio test of the null hypothesis that $\mathcal{H}_0 : A = A_0$ equals:*

$$E\left[W_T\right] \stackrel{1}{=} nq + \frac{1}{T}\left(\beth_2\left(n,q\right) + \daleth\left(n,q,\{C_i\}\right)\right) \tag{5.6}$$

*where:*

$$\beth_2 = -2\sum_{\zeta=0}^{\infty} tr\left\{\left[C'_\zeta\Phi^{-1}C_\zeta\right]_{22}\right\}$$

*and $\daleth\left(n,q,\{C_i\}\right)$ is given in theorem 9.*

PROOF. See the appendix, section 5.A.11. □

All corollaries that follow will be of theorem 9. The only exceptions is corollary 5, which follows from theorem 10.

The following three sections carry examples of increasing complexity, of the main result. Each section contains at least one simulation study to see how useful the correction is in practice.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

## 5.4. Autocorrelation

**5.4.1. First order autocorrelation.** A first illustration of the theory is the test that a certain $p$-dimensional process $u_t$ is white noise versus the alternative that it contains first order autocorrelation. The model is

$$\begin{aligned} u_t &= B_1 u_{t-1} + \eta_t \\ \eta_t &\sim MIIDN(0,\Omega) \end{aligned} \tag{5.7}$$

and the hypothesis: $\mathcal{H}_0 : B_1 = 0$. Note that Maximum Likelihood and Ordinary Least Squares coincide in this case (see also lemma 2 in the appendix) and that $q = n = p$. The last equality implies that $tr\{M_{22}\} = tr\{M\}$. Under the null hypothesis (5.7) collapses to $u_{t-1} = \eta_{t-1}$, such that we find that $Q_0 = I_p$ and $C_i = \Gamma_i = 0$ for all $i \geq 1$. This implies $C_0 = \Omega^{\frac{1}{2}}$ and $\Phi_0 = \Omega$. Now each of the terms t1-t10 in $\urcorner$ has at least one term, whose summation starts at $t = 1$, for instance $\Gamma'_{\kappa+1}$ or $C_{\beta+\kappa+1}$. Therefore each term in all 10 summations is zero and thus $\urcorner = 0$. So we obtain:

COROLLARY 1. *The likelihood ratio test that $B_1 = 0$ in model (5.7) has the following expected value*

$$E[W_T] \stackrel{1}{=} p^2 + \frac{1}{2T}\left(p^2 + 2p^3 - 4p\right)$$

The Bartlett correction here does not depend on the parameters of the model, that is $B(\theta) = B$. The correction only depends on $p$, the dimension of the system. In this simple example we therefore do not encounter any problem as to which estimate for the parameters we should take.

By means of a Monte Carlo Study we investigate how well the Bartlett correction performs. As parameters of choice we take $\Omega = I_n, n \in \{1,2,\ldots,8\}$ and $T \in \{25,50,100\}$. The results are reported by means of QQ-plots for half of the experiments, that is for $n \in \{1,3,5,7\}, T \in \{25,50,100\}$ whereas all the results are reported in table 5.1 and are based on $10^6$ Monte Carlo replications each.

For each experiment we report $E[W_T]$, $E\left[W_T^{BC}\right]$, the Bartlett Factor and the empirical rejection probabilities at the nominal 10%, 5% and 1% level of both the asymptotic and Bartlett adjusted test statistic. We note that the Bartlett corrections brings the rejection probability close to the nominal one, except for the area $T \in \{25\}, n \in \{5,6,7,8\}$ where at the 5% nominal rejection probability the empirical rejection probability is still above 8% after the correction. Yet it does come down from values as high as 81% to at most 25%.

The QQ-plots show that $W_T$ is a straight line, which makes it ideally suited for the Bartlett correction. A Bartlett correction, which does not depend on the estimated parameters, rotates the QQ-plot around the origin. If it is negative (as it is for $p = 1$) it rotates the line anti-clockwise and if it is positive it rotates it clockwise. Success is measured in how well the rotated line coincides with the 45-degree line. In the QQ-plots in figure 5.1, we see that with the possible exceptions of subfigures 5.1(j),5.1(g) and 5.1(h) the rotated line is virtually indistinguishable from the 45-degree line.

**5.4.2. Fourth order autocorrelation.** A second illustration is a test that fourth order autocorrelation is absent

$$\begin{aligned} u_t &= B_4 u_{t-4} + \eta_t \\ \eta_t &\sim MIIDN(0,\Omega) \end{aligned} \tag{5.8}$$

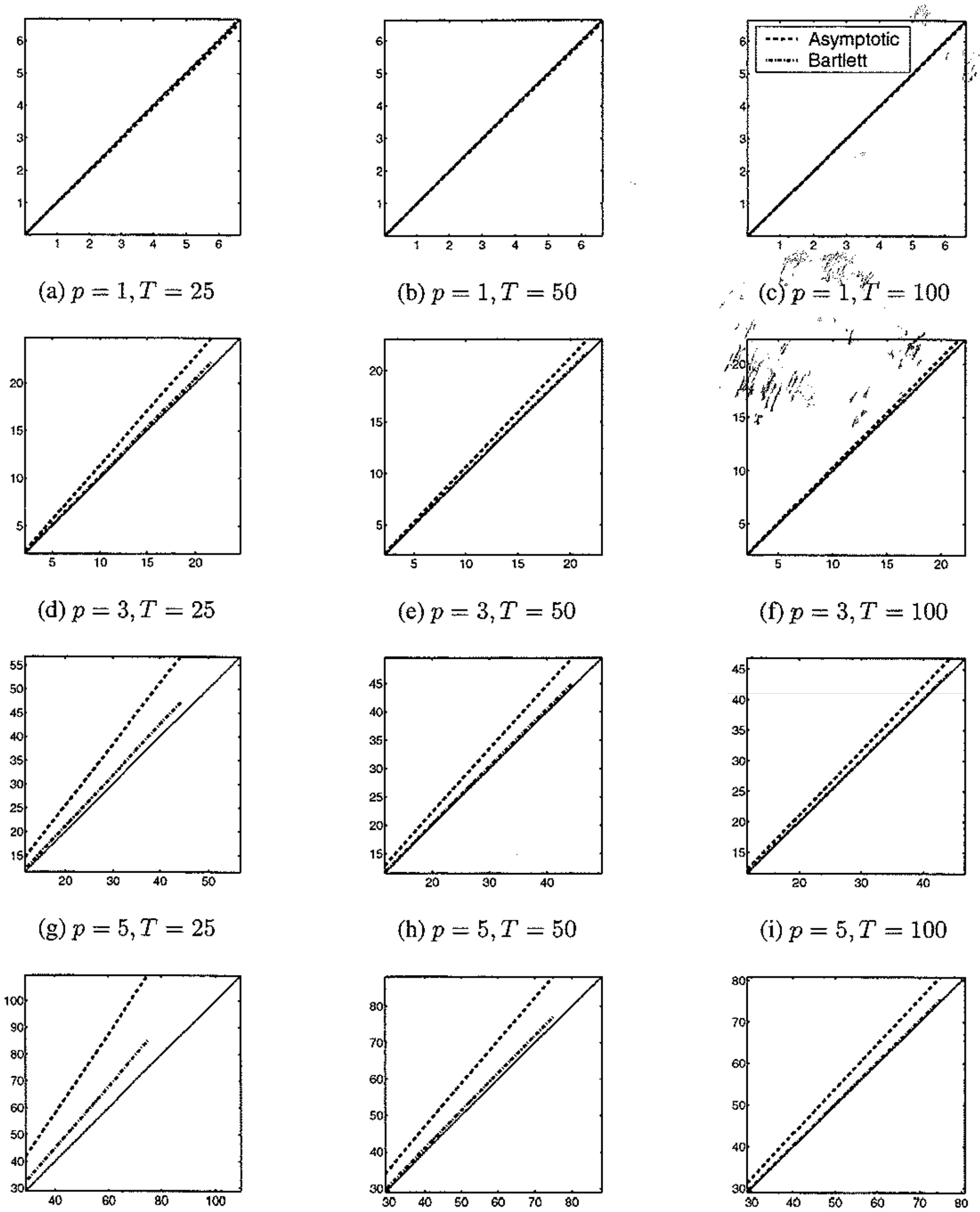
Now our null hypothesis is $\mathcal{H}_0 : B_4 = 0$. We find that $Q_3 = I$ and $Q_i = 0$ for $i \in \{0,1,2,4,5,\ldots\}$. As a consequence $\Phi = \Omega$. Let us now define $F_{\alpha\beta} = (C'_\alpha \Phi^{-1} C_\beta) = \left(\Omega^{\frac{1}{2}} Q'_\alpha \Phi^{-1} Q_\beta \Omega^{\frac{1}{2}}\right)$. It is immediately clear that in this example $F_{\alpha,\beta} = I_p$ iff $\alpha = \beta = 3$ and $F_{\alpha,\beta} = 0$ otherwise. Now

 DOI: 10.2870/52609

| $T$ | 25 | | 50 | | 100 | |
|---|---|---|---|---|---|---|
| | $W_T$ | $W_T^{BC}$ | $W_T$ | $W_T^{BC}$ | $W_T$ | $W_T^{BC}$ |
| $p=1$ | | | | | | |
| $E[LR]$ | 0.9836 | 1.0036 | 0.9906 | 1.0006 | 0.9938 | 0.9988 |
| $BF$ | −0.0200 | | −0.0100 | | −0.0050 | |
| 10% | 9.70 | 10.04 | 9.85 | 10.02 | 9.92 | 10.00 |
| 5% | 4.78 | 5.01 | 4.90 | 5.02 | 4.94 | 4.99 |
| 1% | 0.95 | 1.02 | 0.96 | 1.02 | 0.99 | 1.00 |
| $p=2$ | | | | | | |
| $E[LR]$ | 4.2749 | 4.0329 | 4.1322 | 4.0118 | 4.0652 | 4.0051 |
| $BF$ | 0.0600 | | 0.0300 | | 0.0150 | |
| 10% | 12.20 | 10.26 | 11.03 | 10.09 | 10.48 | 10.01 |
| 5% | 6.44 | 5.18 | 5.65 | 5.06 | 5.31 | 5.01 |
| 1% | 1.44 | 1.05 | 1.19 | 1.01 | 1.08 | 1.00 |
| $p=3$ | | | | | | |
| $E[LR]$ | 10.2294 | 9.1881 | 9.5609 | 9.0482 | 9.2612 | 9.0060 |
| $BF$ | 0.1133 | | 0.0567 | | 0.0283 | |
| 10% | 16.64 | 10.98 | 12.90 | 10.28 | 11.30 | 10.05 |
| 5% | 9.45 | 5.65 | 6.86 | 5.16 | 5.83 | 5.02 |
| 1% | 2.50 | 1.20 | 1.37 | 1.05 | 1.24 | 1.00 |
| $p=4$ | | | | | | |
| $E[LR]$ | 19.2543 | 16.5986 | 17.4245 | 16.1338 | 16.6750 | 16.0336 |
| $BF$ | 0.1600 | | 0.0800 | | 0.0400 | |
| 10% | 24.10 | 12.28 | 15.65 | 10.53 | 12.55 | 10.16 |
| 5% | 14.87 | 6.47 | 8.67 | 5.31 | 6.60 | 5.09 |
| 1% | 4.69 | 1.44 | 2.16 | 1.09 | 1.48 | 1.02 |
| $p=5$ | | | | | | |
| $E[LR]$ | 31.8993 | 26.4944 | 27.9155 | 25.3317 | 26.3618 | 25.0826 |
| $BF$ | 0.2040 | | 0.1020 | | 0.0510 | |
| 10% | 35.88 | 14.66 | 19.61 | 10.94 | 14.12 | 10.23 |
| 5% | 24.40 | 8.02 | 11.40 | 5.59 | 7.63 | 5.14 |
| 1% | 9.44 | 1.96 | 3.15 | 1.17 | 1.81 | 1.06 |
| $p=6$ | | | | | | |
| $E[LR]$ | 48.9253 | 39.2449 | 41.2386 | 36.7109 | 38.4106 | 36.1796 |
| $BF$ | 0.2467 | | 0.1223 | | 0.0617 | |
| 10% | 52.68 | 19.46 | 25.41 | 11.73 | 16.29 | 10.44 |
| 5% | 39.88 | 10.96 | 15.67 | 6.07 | 9.06 | 5.26 |
| 1% | 19.46 | 3.00 | 4.88 | 1.30 | 2.27 | 1.07 |
| $p=7$ | | | | | | |
| $E[LR]$ | 71.1650 | 55.2278 | 57.5895 | 50.3279 | 52.8800 | 49.3218 |
| $BF$ | 0.2886 | | 0.1443 | | 0.0721 | |
| 10% | 71.97 | 25.92 | 33.05 | 12.78 | 19.00 | 10.66 |
| 5% | 60.51 | 16.14 | 21.74 | 6.72 | 10.92 | 5.41 |
| 1% | 37.35 | 5.16 | 7.69 | 1.50 | 2.92 | 1.11 |
| $p=8$ | | | | | | |
| $E[LR]$ | 100.0981 | 75.2617 | 77.2226 | 66.2855 | 69.8832 | 64.5572 |
| $BF$ | 0.3300 | | 0.1650 | | 0.0825 | |
| 10% | 88.69 | 37.26 | 42.92 | 14.31 | 22.48 | 10.97 |
| 5% | 81.58 | 25.49 | 30.18 | 7.72 | 13.38 | 5.59 |
| 1% | 62.77 | 9.95 | 12.42 | 1.83 | 3.86 | 1.17 |

Table 5.1: Bartlett corrections for the test of absence of first order autocorrelation

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609




(a) $p = 1, T = 25$ (b) $p = 1, T = 50$ (c) $p = 1, T = 100$

(d) $p = 3, T = 25$ (e) $p = 3, T = 50$ (f) $p = 3, T = 100$

(g) $p = 5, T = 25$ (h) $p = 5, T = 50$ (i) $p = 5, T = 100$

(j) $p = 7, T = 25$ (k) $p = 7, T = 50$ (l) $p = 7, T = 100$

Figure 5.1: QQ-plots of LR-tests (asymptotic and Bartlett corrected) for residual autocorrelation

rewrite $t_1' = \sum_{\beta,\eta,\kappa,\zeta=0}^{\infty} tr\left\{F_{\kappa,\zeta}F_{\kappa+\zeta+1,\beta+\eta+1}F_{\eta,\beta}\right\}$. For any of the terms in this summation to be different from zero, we need $\kappa = \zeta = \kappa + \zeta + 1 = 3$, such that we conclude that $t_1' = 0$. In similar fashion we see that all other nine terms $t2' - t10'$ equal zero as well, such $\urcorner = 0$ and we obtain the same expression as in the last parapraph:

COROLLARY 2. *The likelihood ratio test that $B_4 = 0$ in model (5.8) has the following expected value*

$$E\left[W_T\right] \stackrel{1}{=} p^2 + \frac{1}{2T}\left(p^2 + 2p^3 - 4p\right)$$


DOI: 10.2870/52609

which once again does not depend on the parameters of the model.

**5.4.3. First to kth order autocorrelation.** Third we test whether there is no first up to $k$th order autocorrelation:

(5.9)
$$\begin{aligned} u_t &= B_1 u_{t-1} + \ldots + B_k u_{t-k} + \eta_t \\ \eta_t &\sim MIIDN(0, \Omega) \end{aligned}$$

The null hypothesis is thus $\mathcal{H}_0 : B_1 = \ldots = B_k = 0$. We see that the regressors in the model $u_{t-1}$ are all independently identically distributed with mean 0 and variance-covariance matrix $\Omega$. The polynomial matrices $Q$ are of dimension $pk \times p$ and read:

$$\begin{aligned} Q_0 &= [I_p : 0 : 0 : \cdots : 0 : 0]' \\ Q_1 &= [0 : I_p : 0 : \cdots : 0 : 0]' \\ &\vdots \\ Q_{k-1} &= [0 : 0 : 0 : \cdots : 0 : I_p]' \\ Q_j &= [0 : 0 : 0 : \cdots : 0 : 0]' \text{ for } j \geq k \end{aligned}$$

This implies $\Phi = (I_k \otimes \Omega)$. Realizing that $Q_i' Q_j = I_p$ iff $i = j, i \leq k-1$ and 0 otherwise, we check each of the ten terms in turn to find out which ones are non-zero. As in the last paragraph we define $F_{\alpha,\beta} = (C_\alpha' \Phi^{-1} C_\beta) = \left(\Omega^{\frac{1}{2}} Q_\alpha' \Phi^{-1} Q_\beta \Omega^{\frac{1}{2}}\right)$ and see that $F_{\alpha,\beta} = I_p$ if $Q_i' Q_j = I_p$ and equals zero when $Q_i' Q_j = 0$.

For the first term $t_1' = \sum_{\beta,\eta,\kappa,\zeta=0}^{\infty} tr \{F_{\kappa,\zeta} F_{\kappa+\zeta+1,\beta+\eta+1} F_{\eta,\beta}\}$ we see that each time $\kappa = \zeta, \kappa+\zeta+1 = \beta+\eta+1 \leq k-1$ and $\beta = \eta$ simultaneously, this term equals $p$. In all other cases it equals zero. Thus we look for how many combination there are for which $\kappa = \zeta \geq 0$ and $\kappa + \zeta + 1 \leq k-1$ hold true. There are $\left\lfloor \frac{k}{2} \right\rfloor$ such that this term equals $p \left\lfloor \frac{k}{2} \right\rfloor$.

The second term is $t_2' = \sum_{\alpha,\eta,\kappa,\zeta=0}^{\infty} tr \{F_{\kappa,\zeta} F_{\kappa+\zeta+1,\alpha}\} tr \{F_{\alpha+\eta+1,\eta}\}$. By definition $\alpha + \eta + 1 \neq \eta$, leading to the conclusion that this term is zero. Similarly we see that $t_3'$, $t_5'$, $t_6'$, $t_7'$, $t_8'$ and $t_{10}'$ are zero.

For any of the terms in the summation $t_4' = \sum_{\beta,\eta,\kappa,\zeta=0}^{\infty} tr \{F_{\beta,\eta} F_{\kappa+\eta+1,\beta+\zeta+1} F_{\zeta,\beta}\}$ to be different from zero we need $\kappa = \zeta$ and $\eta = \beta$ and $\kappa + \eta + 1 = \beta + \zeta + 1 \leq k-1$. There are $\sum_{i=1}^{k-1} i$ such combination, giving a contribution of $\frac{1}{2} pk(k-1)$. Similarly $t_{10}' = -\sum_{\kappa,\zeta,\alpha=0}^{\infty} tr \left\{\left(C_{\alpha+\kappa+1}' C_{\zeta+\alpha+1}\right)\left(C_\zeta' C_\kappa\right)\right\}$ equals $-p$ iff $\kappa = \zeta$ and $\alpha+\kappa+1 = \zeta+\alpha+1 \leq k-1$ which is possible in $\frac{1}{2} k(k-1)$ ways.

For the problem at hand we see that $q = p$ and $n = pk$. Substituting all these terms in the expression in theorem 9 we obtain the following result:

COROLLARY 3. *The likelihood ratio test that $\mathcal{H}_0 : B_1 = \ldots = B_k = 0$ in model (5.9) has the following expected value*

$$E[W_T] \doteq kp^2 + \frac{1}{2T}\left(p^2 k + p^3 k^2 + p^3 k - 4p\right) + \frac{1}{T}\left(p \left\lfloor \frac{k}{2} \right\rfloor - \frac{1}{2} pk(k-1)\right)$$

Once more we notice that the Bartlett factor does not depend on any of the parameters.

## 5.5. Multivariate AR(1) process

Let us once more consider the $p$-dimensional AR(1) model and denote it by $\mathcal{L}_1$:

(5.10)
$$\begin{aligned} X_t &= BX_{t-1} + \eta_t \\ \eta_t &\sim MIIDN(0, \Omega) \end{aligned}$$


DOI: 10.2870/52609

The parameters of this model are $\theta = (B, \Omega) \in (\mathcal{R}^{p\times p}, \mathcal{S}_{p\times p})$ and we test the hypothesis $\mathcal{H}_0$ : $B = \rho_0 I$, where $|\rho_0| < 1$. Under $\mathcal{H}_0$ the dependent variable $X_{t-1}$ has the following moving average representation:

$$X_{t-1} = \sum_{i=0}^{\infty} \rho_0^i I \eta_{t-1-i}$$

from which we see directly that $Q_i = I\rho_0^i$ for $i \geq 0$. Then $C_i = \Omega^{\frac{1}{2}} (I\rho_0^i)$ and $\Phi = \sum_{i=0}^{\infty} (I\rho_0^{2i}) \Omega = \left(I\frac{1}{1-\rho_0^2}\right) \Omega$. Now take the first term

$$\begin{aligned}
t_1' &= \sum_{\beta,\eta,\kappa,\zeta=0}^{\infty} tr \left\{C_\kappa' \Phi^{-1} C_\zeta C_{\kappa+\zeta+1}' \Phi^{-1} C_{\beta+\eta+1} C_\eta' \Phi^{-1} C_\beta\right\} \\
&= \sum_{\beta,\eta,\kappa,\zeta=0}^{\infty} tr \left\{\rho_0^k \left(1-\rho_0^2\right) \rho_0^\zeta \rho_0^{\kappa+\zeta+1} \left(1-\rho_0^2\right) \rho_0^{\beta+\eta+1} \rho_0^\eta \left(1-\rho_0^2\right) \rho_0^\beta I_p\right\} \\
&= \sum_{\beta,\eta,\kappa,\zeta=0}^{\infty} tr \left\{\rho_0^2 \left(1-\rho_0^2\right)^3 \rho_0^{2\zeta} \rho_0^{2\kappa} \rho_0^{2\beta} \rho_0^{2\eta} I_p\right\} \\
&= tr \left\{\rho_0^2 \left(1-\rho_0^2\right)^{-1} I_p\right\} \\
&= \frac{\rho_0^2}{(1-\rho_0^2)} p
\end{aligned}$$

The third term is derived in the following way:

$$\begin{aligned}
t_3' &= \sum_{\alpha,\eta,\lambda,\zeta=0}^{\infty} tr \left\{C_\lambda' \Phi^{-1} C_\alpha\right\} tr \left\{C_\zeta' \Phi^{-1} C_{\lambda+\zeta+1}\right\} tr \left\{C_{\alpha+\eta+1}' \Phi^{-1} C_\eta\right\} \\
&= \sum_{\alpha,\eta,\lambda,\zeta=0}^{\infty} tr \left\{\rho_0^\lambda \left(1-\rho_0^2\right) \rho_0^\alpha I\right\} tr \left\{\rho_0^\zeta \left(1-\rho_0^2\right) \rho_0^{\lambda+\zeta+1}\right\} tr \left\{\rho_0^{\alpha+\eta+1} \left(1-\rho_0^2\right) \rho_0^\eta\right\} \\
&= \sum_{\alpha,\eta,\lambda,\zeta=0}^{\infty} p^3 \left(\rho_0^2 \left(1-\rho_0^2\right)^3 \rho_0^{2\alpha} \rho_0^{2\eta} \rho_0^{2\lambda} \rho_0^{2\zeta}\right) \\
&= \frac{\rho_0^2}{(1-\rho_0^2)} p^3
\end{aligned}$$

The other 8 terms are derived in an entirely analogous manner. In fact each of them gives a contribution equal to $\left(\frac{\rho_0^2}{(1-\rho_0^2)}\right) p^s$ where $s$ is the number of different traces in the expression. We obtain the following result:

COROLLARY 4. *The likelihood ratio test that $\mathcal{H}_0 : B = \rho_0 I$ in model $\mathcal{L}_1$ (5.10) has the following expected value*

$$E[W_T] \stackrel{1}{=} p^2 + \frac{1}{2T} \left(p^2 + 2p^3 - 4p\right) + \frac{1}{T} \left(p^3 + p^2 - 2p\right) \left(\frac{\rho_0^2}{(1-\rho_0^2)}\right) \tag{5.11}$$

The expected value of the likelihood depends on the parameters $\theta_1$ ($B$ in this case) but not on the parameters $\theta_2$ ($\Omega$). This means that when using this correction, no estimated parameters have to be substituted in the Bartlett correction.

Now consider model $\mathcal{L}_2$:

$$\begin{aligned}
X_t &= BX_{t-1} + \varepsilon_t \\
\varepsilon_t &\sim MIIDN(0, I_p)
\end{aligned} \tag{5.12}$$

with the parameters $\theta = B \in \mathcal{R}^{p\times p}$.

Taniguchi (1988, 1991) derives Bartlett corrections for univariate ARMA-processes and in the special case of an AR(1) process with known variance, finds that the expected value of the likelihood ratio equals $1 - \frac{2}{T}$. We thus also state the corollary for model $\mathcal{L}_2$ which is based on theorem 10:

 DOI: 10.2870/52609

COROLLARY 5. *The likelihood ratio test that $\mathcal{H}_0 : B = \rho_0 I$ in model $\mathcal{L}_2$ (5.12) has the following expected value*

$$E[W_T] \doteq p^2 - \frac{2p}{T} + \frac{1}{T}\left(p^3 + p^2 - 2p\right)\left(\frac{\rho_0^2}{(1-\rho_0^2)}\right) \tag{5.13}$$

and conclude that the result of Taniguchi is a special case of (5.13) with $p = 1$.

Both expectations, (5.11) and (5.13) have a pole for $|\rho_0| = 1$. Even though the Bartlett correction is only valid when $|\rho_0| < 1$, it is of interest how close to the pole the Bartlett correction is still of practical use. We thus perform a Monte Carlo study for both corollary 4 and 5.

The DGP is

$$\begin{aligned} X_t &= (\rho I_p)\, X_{t-1} + \varepsilon_t \\ \varepsilon_t &\sim MIIDN(0, I_p) \end{aligned} \tag{5.14}$$

and the parameters of choice are $T = \{100\}$, $\rho = \{-0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9\}$, $p = \{1, 5\}$ and we test the hypothesis $\mathcal{H}_0 : B = \rho_0 I$ both when $\Omega$ is unknown and when it is known. The results are reported in table 5.2 and are based on $10^5$ replications. The Bartlett factor for the case of a one-dimensional process does not depend on any of the parameters and is thus constant over the choice of $\rho$. For the 5-dimensional VAR, we see that when $|\rho|$ approaches unity, the uncorrected test becomes severely oversized. The Bartlett correction does however somewhat overcorrect, which is what we expected with the pole in the expression. Overall the Bartlett corrected test is closer to the nominal size of the test than the uncorrected one in 69 out of 84 cases.

### 5.6. No level feedback in the cointegrated VAR

Let us consider the cointegrated VAR model in the Equilibrium Correction form:

$$\begin{aligned} \Delta X_t &= \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \eta_t \\ \eta_t &\sim MIIDN(0, \Omega) \end{aligned} \tag{5.15}$$

with the following assumptions:

(1) Every root $z$ of the characteristic polynomial of $X_t$ satisfies $z = 1$ or $|z| > 1$.
(2) $\Pi := -A(1) = \alpha\beta'$, where $\alpha$ and $\beta$ are $p \times r$ matrices of full rank $r < p$.
(3) $\alpha'_\perp \Gamma \beta_\perp$ has full rank $p - r$, where $\Gamma := I - \sum_{i=1}^{k-1} \Gamma_i$.

We consider maximum likelihood estimation as proposed by Johansen (1988).

Divide the variable-vector $X_t$ in two, $X_{1t}$ of dimension $p-s$ and $X_{2t}$ of dimension $s$ $(\leq p - r)$ and the parameters $\alpha$ and $\Gamma_i$ conformably, that is $\alpha = [\alpha'_1, \alpha'_2]'$. We then obtain the following system of equations:

$$\Delta X_{1t} = \alpha_1 \beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_{1i} \Delta X_{t-i} + \eta_{1t} \tag{5.16}$$

$$\Delta X_{2t} = \alpha_2 \beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_{2i} \Delta X_{t-i} + \eta_{2t} \tag{5.17}$$

$$\eta_t = \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix} \sim MIIDN(0, \Omega), \quad \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$$

Conditioning on $\Delta X_{2t}$ in equation (5.16) we obtain the following system.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

| $T = 100$ | $\Omega$ unknown (corollary 4) | | | | $\Omega$ known (corollary 5) | | | |
|---|---|---|---|---|---|---|---|---|
| | $p = 1$ | | $p = 5$ | | $p = 1$ | | $p = 5$ | |
| | $W_T$ | $W_T^{BC}$ | $W_T$ | $W_T^{BC}$ | $W_T$ | $W_T^{BC}$ | $W_T$ | $W_T^{BC}$ |
| $\rho = -0.9$ | | | | | | | | |
| $E[LR]$ | 0.997 | 1.002 | 31.31 | 24.28 | 0.983 | 1.003 | 29.33 | 23.75 |
| $BF$ | −0.005 | | 0.290 | | −0.020 | | 0.235 | |
| 10% | 9.95 | 10.03 | 32.85 | 6.96 | 9.61 | 9.96 | 24.23 | 5.72 |
| 5% | 4.95 | 5.00 | 21.04 | 3.06 | 4.76 | 4.99 | 14.21 | 2.36 |
| 1% | 1.00 | 1.02 | 7.00 | 0.43 | 0.98 | 1.05 | 3.88 | 0.33 |
| $\rho = -0.6$ | | | | | | | | |
| $E[LR]$ | 0.996 | 1.001 | 27.10 | 25.04 | 0.981 | 1.001 | 25.57 | 24.88 |
| $BF$ | −0.005 | | 0.083 | | −0.020 | | 0.028 | |
| 10% | 9.95 | 10.03 | 16.61 | 10.15 | 9.65 | 10.01 | 10.62 | 9.64 |
| 5% | 5.00 | 5.06 | 9.32 | 5.16 | 4.80 | 5.02 | 5.95 | 4.76 |
| 1% | 1.00 | 1.01 | 2.34 | 1.03 | 0.94 | 1.01 | 1.22 | 0.87 |
| $\rho = -0.3$ | | | | | | | | |
| $E[LR]$ | 0.992 | 0.997 | 26.47 | 25.05 | 0.977 | 0.997 | 25.00 | 24.96 |
| $BF$ | −0.005 | | 0.057 | | −0.020 | | 0.002 | |
| 10% | 9.86 | 9.94 | 14.51 | 10.15 | 9.62 | 9.97 | 10.04 | 9.3 |
| 5% | 4.93 | 5.00 | 7.96 | 5.16 | 4.77 | 4.99 | 4.95 | 4.88 |
| 1% | 0.97 | 0.99 | 1.88 | 1.03 | 0.93 | 0.99 | 0.99 | 0.97 |
| $\rho = 0$ | | | | | | | | |
| $E[LR]$ | 0.990 | 0.995 | 26.33 | 25.05 | 0.975 | 0.995 | 24.88 | 24.98 |
| $BF$ | −0.005 | | 0.051 | | −0.020 | | −0.004 | |
| 10% | 9.90 | 10.00 | 14.19 | 10.29 | 9.64 | 9.99 | 9.64 | 9.93 |
| 5% | 4.87 | 4.92 | 7.58 | 5.12 | 4.73 | 4.97 | 4.76 | 4.92 |
| 1% | 0.96 | 0.97 | 1.79 | 1.04 | 0.90 | 0.98 | 0.93 | 0.99 |
| $\rho = 0.3$ | | | | | | | | |
| $E[LR]$ | 0.990 | 0.995 | 26.48 | 25.06 | 0.976 | 0.996 | 25.01 | 24.97 |
| $BF$ | −0.005 | | 0.057 | | −0.020 | | 0.002 | |
| 10% | 9.91 | 9.98 | 14.54 | 10.15 | 9.66 | 10.03 | 9.95 | 9.85 |
| 5% | 4.90 | 4.95 | 7.83 | 5.11 | 4.72 | 4.94 | 4.99 | 4.92 |
| 1% | 0.94 | 0.96 | 1.88 | 1.03 | 0.88 | 0.94 | 0.96 | 0.94 |
| $\rho = 0.6$ | | | | | | | | |
| $E[LR]$ | 0.993 | 0.998 | 27.13 | 25.06 | 0.978 | 0.998 | 25.60 | 24.91 |
| $BF$ | −0.005 | | 0.083 | | −0.020 | | 0.028 | |
| 10% | 10.04 | 10.13 | 16.71 | 10.08 | 9.78 | 10.09 | 11.64 | 9.58 |
| 5% | 5.00 | 5.05 | 9.26 | 4.97 | 4.83 | 5.06 | 5.91 | 4.65 |
| 1% | 0.98 | 1.00 | 2.34 | 1.02 | 0.92 | 0.99 | 1.21 | 0.91 |
| $\rho = 0.9$ | | | | | | | | |
| $E[LR]$ | 0.999 | 1.004 | 31.34 | 24.30 | 0.984 | 1.004 | 29.37 | 23.79 |
| $BF$ | −0.005 | | 0.290 | | −0.020 | | 0.235 | |
| 10% | 9.93 | 10.02 | 33.01 | 6.94 | 9.67 | 9.99 | 24.35 | 5.74 |
| 5% | 5.00 | 5.06 | 21.09 | 3.06 | 4.83 | 5.04 | 14.40 | 2.47 |
| 1% | 1.05 | 1.07 | 6.99 | 0.44 | 0.99 | 1.06 | 3.98 | 0.34 |

Table 5.2: Bartlett corrections of tests on the autoregressive parameters in the multivariate AR(1) model with unknown and known variance

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

$$\Delta X_{1t} = \omega \Delta X_{2t} + (\alpha_1 - \omega\alpha_2)\beta' X_{t-1} + \sum_{i=1}^{k-1} (\Gamma_{1i} - \omega\Gamma_{2i}) \Delta X_{t-i} + \tilde{\eta}_{1t} \tag{5.18}$$

$$\Delta X_{2t} = \alpha_2 \beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_{2i} \Delta X_{t-i} + \eta_{2t} \tag{5.19}$$

$$\tilde{\eta}_t = \begin{bmatrix} \tilde{\eta}_{1t} \\ \eta_{2t} \end{bmatrix} \sim MIIDN(0, \tilde{\Omega}), \quad \tilde{\Omega} = \begin{bmatrix} \Omega_{11} - \omega\Omega_{21} & 0 \\ 0 & \Omega_{22} \end{bmatrix}$$

where we have defined $\omega = \Omega_{12}\Omega_{22}^{-1}$. Furthermore define $\Psi_1^* = (\Gamma_{11} - \omega\Gamma_{21}, \ldots, \Gamma_{1k-1} - \omega\Gamma_{2k-1}$ and $\Psi_2 = (\Gamma_{21}, \ldots, \Gamma_{2k-1})$. The parameters in the conditional equation (5.18) are $\theta_{con} = (\alpha_1 - \omega\alpha_2, \beta, \Psi_1^*, \omega, \Omega_{11} - \omega\Omega_{21})$ and those in the marginal model (5.19) read $\theta_{mar} = (\alpha_2, \beta, \Psi_2,$ $\theta_{con}$ and $\theta_{mar}$ do not vary in a product space, such that for inference the whole system (5.15) needs to be analyzed.

The following concept will offer a way to analyze partial systems:

DEFINITION 1. *There is No Level Feedback (NLF) from the cointegration relations $\beta' X_{t-1}$ to $\Delta X_{2t}$ , when $\Delta X_{2t}$ does not react to a disequilibrium in the cointegration relations $\beta' X_{t-1}$ that is when $\alpha_2 = 0$.*

This means that the differences $\Delta X_{2t}$ do not react directly to a disequilibrium in the cointegration relation. Of course they may still react to past changes in the differences as under NLF $\Psi_2$ does not necessarily equal zero.

If NLF holds, then the parameters in the marginal equation become $\theta_{mar}^2 = (\Psi_2, \Omega_{22})$. Johansen(1996, theorem 8.1) proves that if $\alpha_2 = 0$, that is NLF from $\beta' X_{t-1}$ to $\Delta X_{2t}$, then the maximum likelihood estimates of $\beta$ (and $\alpha_1$) are obtained from the conditional equation (5.18) only, as $\theta_{mar}^2$ and $\theta_{con}$ do vary in a product space.

There are two moments, one can test for NLF: before and after determination of the cointegration space. Even though both tests have the same asymptotic distribution under the null, namely $\chi^2_{s(p-r)}$ they do not have the same small sample properties.

The first test is the one proposed by Harbo et al. (1998) as an ex-post misspecification test after analyzing a conditional system. The second one is a test on the adjustment parameters $\alpha$ before inference on $\beta$ is made. If the test does not reject conditional inference can be made afterwards. First we shall outline each of these tests in turn and their Bartlett correction. A Monte Carlo simulation study will illustrate the use of the Bartlett correction in each case and show remarkable differences between the two tests.

**5.6.1. Testing NLF after determination of the cointegration space.** Harbo et al. (1998) propose to use economic arguments to determine which $s\,(\leq r)$ variables $\Delta X_{2t}$ do not react to disequilibria in the cointegration relations. Having assumed NLF from $\beta' X_{t-1}$ to $\Delta X_{2t}$ they suggest estimating the rank from the conditional model (5.18), as this is maximum likelihood estimator if NLF holds. They then go on and and restrict the cointegration space, still using only the conditional model.

After this they propose to do a misspecification test to check whether the initial assumption of NLF was correct. Defining $Z_t = \beta' X_t$ this is done by testing $\mathcal{H}_0 : \alpha_2 = 0$ in

$$\Delta X_{2t} = \alpha_2 Z_{t-1} + \sum_{i=1}^{k-1} \Gamma_{2i} \Delta X_{t-i} + \eta_{2t} \tag{5.20}$$

by means of a likelihood ratio test. The parameter space in this model is $\theta_{mar}^3 = (\alpha_2, \Psi_2, \Omega_{22})$. The null hypothesis only concerns $\alpha_2$ and not $\Psi_2$ such that we cannot apply theorem 9 directly.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

We can however write the expectation of the desired test as the difference between two tests, that are each special cases of theorem 9.

Define the following three models, which successively restrict the parameter space in the marginal model (5.20):

(1) $\mathcal{M}_1$ : unrestricted parameters $\alpha_2$, $\Psi_2$ and $\Omega_{22}$.
(2) $\mathcal{M}_2 : \alpha_2 = 0$, but $\Psi_2$ and $\Omega_{22}$ unrestricted.
(3) $\mathcal{M}_3 : \alpha_2 = 0$, $\Psi_2 = \Psi_{20}$ and $\Omega_{22}$ unrestricted.

Let $\left(\tilde{\alpha}_2, \tilde{\Psi}_2\right)$ be the maximum likelihood estimators of $\mathcal{M}_1$ and $\hat{\Psi}_2$ those of $\mathcal{M}_2$. Then the test that $\alpha_2 = 0$ in $\mathcal{M}_1$, that is $\mathcal{M}_2$ in $\mathcal{M}_1$ can be written as:

$$LR\left(\mathcal{M}_2|\mathcal{M}_1\right) = \frac{L\left(\alpha_2 = 0, \hat{\Psi}\right)}{L\left(\tilde{\alpha}_2, \tilde{\Psi}\right)} = \frac{L\left(\alpha_2 = 0, \hat{\Psi}\right)}{L\left(\alpha_2 = 0, \Psi = \Psi_0\right)} \times \frac{L\left(\alpha_2 = 0, \Psi = \Psi_0\right)}{L\left(\tilde{\alpha}_2, \hat{\Psi}\right)}$$

This means that the log-likelihood ratio test can be written as the difference between two log-likelihood ratio tests:

$$-2\ln LR\left(\mathcal{M}_2|\mathcal{M}_1\right) = -2\ln LR\left(\mathcal{M}_3|\mathcal{M}_1\right) + 2\ln LR\left(\mathcal{M}_3|\mathcal{M}_2\right)$$

such that to get the Bartlett correction, we just have to take the difference between the two expectations. To see how these tests are both special cases of theorem 9, rewrite the stationary part of the cointegrated VAR model (5.15) in the following error correction form:

(5.21)

$$\begin{bmatrix} \Delta X_t \\ \Delta X_{t-1} \\ \vdots \\ \Delta X_{t-k+3} \\ \Delta X_{t-k+2} \\ Z_t \end{bmatrix} = \begin{bmatrix} \Gamma_1 & \Gamma_2 & \dots & \dots & \Gamma_{k-1} & \alpha \\ I_p & 0 & \dots & \dots & \dots & 0 \\ 0 & \ddots & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ 0 & \dots & 0 & I_p & 0 & 0 \\ \beta'\Gamma_1 & \dots & \dots & \dots & \beta'\Gamma_{k-1} & \beta'\alpha + I_r \end{bmatrix} \begin{bmatrix} \Delta X_{t-1} \\ \Delta X_{t-2} \\ \vdots \\ \Delta X_{t-k+2} \\ \Delta X_{t-k+1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} I_p \\ 0 \\ \vdots \\ \vdots \\ 0 \\ \beta' \end{bmatrix} \eta_t$$

(5.22) $$Y_t = DY_{t-1} + E\eta_t$$

(5.23) $$\eta_t \sim MIIDN(0, \Omega)$$

The regressors in $\mathcal{M}_1$ are $Y_{t-1}$. These can be written in terms of the $MIIDN(0, \Omega)$ process $\eta_t$ as $Y_{t-1} = \sum_{i=0}^{\infty} G_i \eta_{t-1-i}$ where

(5.24) $$G_i = D^i E \qquad \text{for } i = 0, 1, \dots$$

(5.25) $$H_i = D^i E \Omega^{\frac{1}{2}} = D^i F \qquad \text{for } i = 0, 1, \dots$$

In the last line we defined $\{H_i\}$ by postmultiplying $\{G_i\}$by $\Omega^{\frac{1}{2}}$, just as we postmultiplied $\{Q_i\}$ to obtain $\{C_i\}$ and then expressed the theorems in terms of $\{C_i\}$. Next define the matrix $S$ which selects the first differences and the lagged first differences, but not the cointegration relationships from $Y_{t-1}$ as:

(5.26) $$S = \left[I_{p(k-1)}, 0_{p(k-1)\times r}\right]'$$

such that $S'Y_{t-1}$ are the regressors in $\mathcal{M}_2$ and we obtain the following expressions for its polynomial

$$N_i = S'D^iE \qquad \text{for } i = 0, 1, \dots \tag{5.27}$$

$$O_i = S'D^iE\Omega^{\frac{1}{2}} = S'D^iF \qquad \text{for } i = 0, 1, \dots \tag{5.28}$$

For future reference we also define the variance of the process $Y$ as $\Sigma_{yy}$:

$$\Sigma_{yy} = var\,(Y_t) \tag{5.29}$$

In $\mathcal{M}_1$ the dimension of the coefficient matrix is $s \times ((k-1)\,p + r)$, whereas in $\mathcal{M}_2$ it is $s \times (k-1)\,p$. The null hypothesis is $\mathcal{H}_0 : \alpha_2 = 0$. Consequently the Bartlett factor can be used and the expectation of the likelihood ratio is given in the following corollary:

COROLLARY 6. *The likelihood ratio for* $\mathcal{H}_0 : \alpha_2 = 0$ *in* (5.19) *has the following expected value:*

$$E\left[-2\ln LR\left(\mathcal{M}_2|\mathcal{M}_1\right)\right] \doteq sr + \frac{1}{2T}\left(sr + s^2r + sr^2 + 2rsp(k-1)\right)$$
$$+\frac{1}{T}\urcorner\left(\left(k-1\right)p + r, s, \{H_i\}\right) - \frac{1}{T}\urcorner\left(\left(k-1\right)p, s, \{O_i\}\right)$$

*where* $H_i$ *and* $O_i$ *are defined in* (5.25) *and* (5.28) *respectively and* $\urcorner$ *is defined in theorem 9.*

The two expression for $\urcorner$ in corollary 6 contain infinite loops, but due to their structure $\{H_i\}$ and $\{O_i\}$ in equations (5.25) and (5.28) can be simplified, such that the expressions can be computed exactly.

Let

$$v_1, \dots, v_n \tag{5.30}$$

be the (possibly complex) eigenvalues of $D$ and $w_1, \dots, w_n$ the corresponding eigenvectors. Then define:

$$V = \begin{bmatrix} w_1 & \dots & w_n \end{bmatrix} \tag{5.31}$$

$$\Lambda = \begin{bmatrix} v_1 & & \\ & \ddots & \\ & & v_n \end{bmatrix} \tag{5.32}$$

$$\Lambda^{ro} = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} \tag{5.33}$$

$$\Lambda^{co} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \tag{5.34}$$

$$\imath_{n\times 1} = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}' \tag{5.35}$$

A number of terms, which are expressed in terms of $v_i$,$V$,$\Lambda$,$\Lambda^{ro}$and $\Lambda^{co}$ are given in table 5.3. They are used in the following two theorems.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

| | | |
|---|---|---|
| $\Psi = \sum_{i=0}^{\infty} D^i F F' D^{i\prime}$ | $A_3 = V' S' \Phi^{-1} S V \Lambda$ | $A_7 = (I - \Lambda^2)^{-1}$ |
| $P = S' (S \Psi S')^{-1} S \Psi$ | $A_4 = V^{-1} P' V$ | $A_8 = (ll' - \Lambda^{co} \Lambda^{ro})$ |
| $A_1 = V^{-1} F I_{22} F' V^{-1\prime}$ | $A_5 = V' S' \Phi^{-1} S V$ | $A_{9i} = (I_n - v_i \Lambda)^{-1}$ |
| $A_2 = V' P V^{-1\prime} \Lambda$ | $A_6 = V' \Phi^{-1} V$ | $A_{9j} = (I_n - v_j \Lambda)^{-1}$ |

Table 5.3: Definition of a number of terms for theorems 11 and 12

THEOREM 11. *If $C_i = S' D^i F$ for $i \geq 0$ then the expression for* $\daleth$ *in theorem 9 simplifies to:*

$$\begin{aligned}
\daleth &= tr\left\{A_1 \left(A_2 \oslash A_8\right) A_3 \left(A_4 \oslash A_8\right)\right\} \\
&+ 2\sum_{i=1}^{n} \left(A_2\right)_{ii} tr\left\{A_1 \left(A_2 \oslash A_8\right) A_5 A_{9i}\right\} \\
&+ \sum_{i,j=1}^{n} \left(A_2\right)_{ii} \left(A_2\right)_{jj} tr\left\{A_1 A_{9i} A_5 A_{9j}\right\} \\
&+ \sum_{i,j,k,m=1}^{n} \frac{\left(A_1\right)_{ij} \left(A_2\right)_{jk} \left(A_3\right)_{km} \left(A_4\right)_{mi}}{(1 - v_j v_m)(1 - v_i v_k)} \\
&+ 2\sum_{i,j,k,m=1}^{n} \frac{\left(A_1\right)_{ij} \left(A_2\right)_{jk} \left(A_2\right)_{km} \left(A_5\right)_{mi}}{(1 - v_j v_m)(1 - v_i v_k)} \\
&+ \sum_{i,j=1}^{n} \left(A_2\right)_{ji} \left(A_2\right)_{ij} tr\left\{A_1 A_{9i} A_5 A_{9j}\right\} \\
&- 2\sum_{i=1}^{n} \left(A_4\right)_{ii} v_i^2 tr\left\{A_1 A_{9i} A_5 A_{9i}\right\} \\
&- 2tr\left\{A_1 \left(A_4' \oslash A_8\right) \Lambda^2 \left(A_5 \oslash A_8\right)\right\} \\
&- 2tr\left\{\left(A_1 \oslash A_8\right) A_3 \left(A_4 \oslash A_8\right) \Lambda\right\} \\
&- 2tr\left\{\left(A_1 \oslash A_8\right) \Lambda \left(A_3 \oslash A_8\right) A_4\right\}
\end{aligned}$$

*where relevant definitions are given in equations (5.30)-(5.34) and in table 5.3. $\oslash$ denotes Hadamard division. For three matrices $A, B$ and $C$ of equal dimension $C = A \oslash B$ is the matrix with entries $c_{ij} = a_{ij}/b_{ij}$.*

PROOF. see section 5.A.12 □

THEOREM 12. *If $C_i = D^i F$ for $i \geq 0$ then the expression for* $\daleth$ *in theorem 9 simplifies to:*

$$\begin{aligned}
\daleth &= tr\left\{A_1 \Lambda A_7 A_6 \Lambda A_7\right\} \\
&+ 2\sum_{i=1}^{n} v_i tr\left\{A_1 \Lambda A_7 A_6 A_{9i}\right\} \\
&+ \sum_{i,j=1}^{n} v_i v_j tr\left\{A_1 A_{9i} A_6 A_{9j}\right\} \\
&- tr\left\{\left(\left(\Lambda A_1 \Lambda\right) \oslash A_8\right) \left(A_6 \oslash A_8\right)\right\} \\
&- 2tr\left\{A_1 A_7 \Lambda^2 \left(A_6 \oslash A_8\right)\right\} \\
&- \sum_{i=1}^{n} v_i^2 tr\left\{A_1 A_{9i} A_6 A_{9i}\right\}
\end{aligned}$$

*where relevant definitions are given in equations (5.30)-(5.34) and and in table 5.3.*

PROOF. see section 5.A.13. □

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

Both expressions are quickly programmed and as they contain only finite loops[1], the first order expansion of the expectation of the likelihood ratio test statistic can be calculated exactly.

**5.6.2. Testing NLF before determination of the cointegration space.** Under the assumption of NLF from $\beta' X_{t-1}$ to $\Delta X_{2t}$ the parameters of the conditional model (5.18) $\theta_{con}$ and those in the marginal model (5.19) $\theta^2_{mar}$ vary in a product space, such that $\Delta X_{2t}$ is weakly exogenous for $\beta$. The aim of the test for NLF is thus to be able to do inference on $\beta$ in the conditional model only.

We can find the Bartlett correction for that test, but once again we need to take differences between likelihood ratio tests to be able to find a first order approximation to the expectation of the test of interest. Define the following models:

(1) $\mathcal{N}_{-1}$ : matrix $\Pi$ is of full rank $p$.
(2) $\mathcal{N}_0$ : unrestricted parameters in the cointegrated VAR, equation (5.15)
(3) $\mathcal{N}_1 : \beta = \beta_0 \phi$
(4) $\mathcal{N}_{1a} : \alpha = \alpha_0 \psi$
(5) $\mathcal{N}_2 : \beta = \beta_0, \alpha = \alpha_0$

where $\phi$ and $\psi$ are $(r \times r)$ matrices of full rank.

The difference between $\mathcal{N}_2$ and $\mathcal{M}_2$ is that in $\mathcal{M}_2$ $s$ $(\leq p - r)$ rows equal zero and the others are estimated freely. In $\mathcal{N}_2$ the whole column space of $\alpha$ is fixed. This implies that $LR(\mathcal{N}_2|\mathcal{N}_1)$ is a special case of $LR(\mathcal{M}_2|\mathcal{M}_1)$.

Our interest focuses on $LR(\mathcal{N}_{1a}|\mathcal{N}_0)$ which can be written as:

$$LR(\mathcal{N}_{1a}|\mathcal{N}_0) = \frac{L(\mathcal{N}_{1a})}{L(\mathcal{N}_2)} \times \frac{L(\mathcal{N}_2)}{L(\mathcal{N}_1)} \times \frac{L(\mathcal{N}_1)}{L(\mathcal{N}_0)} \tag{5.36}$$

such that we find:

$$-2\ln LR(\mathcal{N}_{1a}|\mathcal{N}_0) = +2\ln LR(\mathcal{N}_2|\mathcal{N}_{1a}) - 2\ln LR(\mathcal{N}_2|\mathcal{N}_1) - 2\ln LR(\mathcal{N}_1|\mathcal{N}_0)$$

In this section we have already derived the first order approximation to the expectation of $-2\ln LR(\mathcal{N}_2|\mathcal{N}_1$ whereas Johansen (2000a) derives that of $-2\ln LR(\mathcal{N}_1|\mathcal{N}_0)$ and Johansen (2002a) contains the one for $-2\ln LR(\mathcal{N}_2|\mathcal{N}_{1a})$. We can simply add up the three expectations of these terms to find the Bartlett correction of the test for $-2\ln LR(\mathcal{N}_{1a}|\mathcal{N}_0)$.

All three tests concern the whole system of equations, namely (5.15), but $-2\ln LR(\mathcal{N}_2|\mathcal{N}_1)$ is done in the marginal equation only, as we saw in the last paragraph. Adding up the three expressions we obtain:

COROLLARY 7. *For unknown cointegration parameter $\beta$ the likelihood ratio for $\mathcal{H}_0 : \alpha_2 = 0$ in* (5.19) *has the following expected value:*

$$E[-2\ln LR(\mathcal{N}_{1a}|\mathcal{N}_0)] \doteq r(p-r) + \frac{1}{2T}\left(r^2 + 2r^3 + 2r^2 p(k-1)\right)$$
$$+ \frac{1}{T}\urcorner((k-1)p + r, r, \{H_i\}) - \frac{1}{T}\urcorner((k-1)p, r, \{O_i\})$$
$$+ \frac{1}{T} r^2 tr\left\{\left(\alpha'\Omega^{-1}\alpha\right)^{-1} S'_\perp \Sigma_{yy}^{-1} S_\perp\right\}$$

*where $H_i$, $O_i$, $S$ and $\Sigma_{yy}$ are defined in* $(5.25) - (5.29)$ *and $\urcorner$ is defined in theorem 9.*

For completeness we state that Johansen (2002b) derives the Bartlett correction for the rank test, that is for $LR(\mathcal{N}_0|\mathcal{N}_{-1})$ and graphically represent this information in figure 5.2.

[1]Note that $\Psi = V\left(\left(V^{-1}FF'V^{-1\prime}\right) \oslash \left(ll' - \Lambda^{co}\Lambda^{ro}\right)\right)V'$ such that only finite loops remain for the expression in table 5.3.


DOI: 10.2870/52609

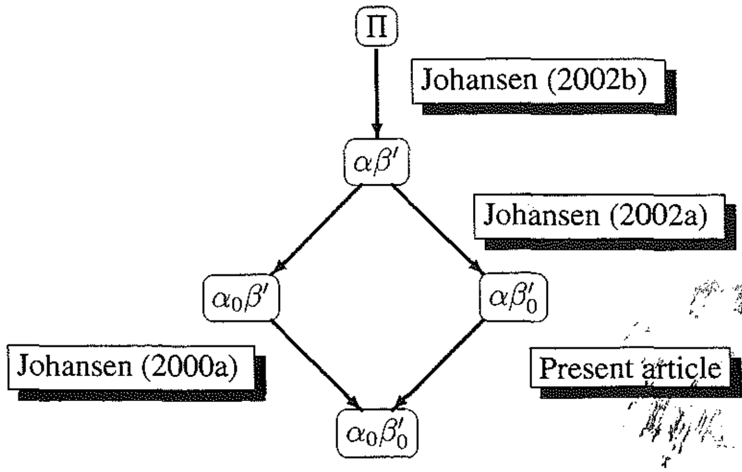



Figure 5.2: Overview of Bartlett corrections in the cointegrated VAR

Equation (5.36) shows that we are able to Bartlett correct the one test in the diagram, for which the Bartlett correction has not been derived explicitly. We do stress that whereas the Bartlett corrections in Johansen (2000a, 2002a,b) allow for certain deterministic terms, the one in this paper does not and is therefore somewhat less general.

**5.6.3. A Monte Carlo study of the test for NLF.** We perform a Monte Carlo study of the two tests for no long-run feedback and use the following 5-dimensional Data Generating Process:

$$\begin{aligned} \phi_1(L)\, X_{1t} &= \varepsilon_{1t} \\ \phi_2(L)\, X_{2t} &= \varepsilon_{2t} \\ g(L)\, X_{it} &= \varepsilon_{it} \qquad \text{for } i = 3,4,5 \\ \varepsilon_t &\sim MIIDN(0, I_n) \end{aligned}$$

where

$$\begin{array}{lll} \phi_1(L) = \prod_{i=1}^{k}(1-\phi_{1i}L) & \varphi_1 = \left[\begin{array}{ccc}\phi_{11} & \ldots & \phi_{1k}\end{array}\right] & max(|\phi_{1i}|) < 1 \\ \phi_2(L) = \prod_{i=1}^{k}(1-\phi_{2i}L) & \varphi_2 = \left[\begin{array}{ccc}\phi_{21} & \ldots & \phi_{2k}\end{array}\right] & max(|\phi_{2i}|) < 1 \\ g(L) = \prod_{i=1}^{k}(1-g_iL) & \gamma = \left[\begin{array}{ccc}g_1 & \ldots & g_k\end{array}\right] & max(|g_i|) = 1 \end{array}$$

The first two variables are stationary, whereas the last three each contain exactly one unit root. As the calculation of the Bartlett correction is computer-intensive (in a simulation framework) and in order to keep the size of this experiment under control, we have opted for a benchmark case and then varied one or two aspects of the benchmark DGP.

When we rewrite the model in the equilibrium correction form (5.15), then $\alpha$ and $\beta$ take the following values:

$$\alpha' = \left[\begin{array}{ccccc}\alpha_{11} & 0 & 0 & 0 & 0 \\ 0 & \alpha_{22} & 0 & 0 & 0\end{array}\right], \quad \beta' = \left[\begin{array}{ccccc}1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0\end{array}\right]$$

$$\alpha_{11} = \sum_{i=1}^{k} \phi_{1i} - 1$$

$$\alpha_{22} = \sum_{i=1}^{k} \phi_{2i} - 1$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

| | | $\beta$ known (corollary 6) | | | | $\beta$ unknown (corollary 7) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $W_T$ | $W_T^{BC}$ | | | $W_T$ | $W_T^{BC}$ | | |
| | | | $\theta$ | $\hat{\theta}_r$ | $\hat{\theta}_u$ | | $\theta$ | $\hat{\theta}_r$ | $\hat{\theta}_u$ |
| Experiment 1 | $E[LR]$ | 6.71 | 6.07 | 6.08 | 6.10 | 11.49 | 7.68 | 8.65 | 9.21 |
| $\varphi_1 = [0.8, 0.6]$ | $BF$ | 0.106 | | | | 0.496 | | | |
| $\varphi_2 = [0.8, 0.6]$ | 10% | 14.8 | 10.4 | 10.4 | 10.5 | 51.0 | 19.8 | 27.9 | 32.6 |
| $\gamma = [1, \epsilon]$ | 5% | 8.0 | 5.2 | 5.2 | 5.2 | 37.6 | 10.5 | 16.4 | 21.0 |
| $T = 100$ | 1% | 1.8 | 0.7 | 0.7 | 0.8 | 16.5 | 1.8 | 4.2 | 6.7 |
| Experiment 2 | $E[LR]$ | 6.28 | 5.96 | 5.97 | 5.97 | 8.53 | 6.83 | 7.08 | 7.28 |
| $\varphi_1 = [0.8, 0.6]$ | $BF$ | 0.053 | | | | 0.248 | | | |
| $\varphi_2 = [0.8, 0.6]$ | 10% | 11.7 | 9.8 | 9.8 | 9.8 | 27.0 | 14.7 | 16.35 | 17.9 |
| $\gamma = [1, \epsilon]$ | 5% | 6.3 | 5.3 | 5.3 | 5.3 | 17.2 | 8.6 | 9.5 | 10.8 |
| $\mathbf{T = 200}$ | 1% | 2.0 | 1.4 | 1.4 | 1.5 | 7.2 | 2.4 | 2.8 | 3.3 |
| Experiment 3 | $E[LR]$ | 6.09 | 5.93 | 5.93 | 5.93 | 6.91 | 6.15 | 6.21 | 6.27 |
| $\varphi_1 = [0.8, 0.6]$ | $BF$ | 0.026 | | | | 0.124 | | | |
| $\varphi_2 = [0.8, 0.6]$ | 10% | 11.0 | 10.0 | 10.0 | 10.0 | 16.2 | 11.0 | 11.4 | 11.8 |
| $\gamma = [1, \epsilon]$ | 5% | 5.4 | 4.8 | 4.8 | 4.8 | 8.7 | 5.6 | 5.6 | 6.1 |
| $\mathbf{T = 400}$ | 1% | 1.0 | 0.8 | 0.8 | 0.8 | 2.6 | 1.3 | 1.4 | 1.6 |
| Experiment 4 | $E[LR]$ | 7.46 | 6.11 | 6.25 | 6.31 | 12.59 | 7.82 | 9.08 | 9.72 |
| $\varphi_1 = [0.8, 0.6]$ | $BF$ | 0.221 | | | | 0.611 | | | |
| $\varphi_2 = [0.8, 0.6]$ | 10% | 21.2 | 11.6 | 11.6 | 12.2 | 59.0 | 20.1 | 31.6 | 36.6 |
| $\gamma = [1, \mathbf{0.6}]$ | 5% | 12.0 | 5.8 | 5.8 | 6.3 | 45.3 | 10.0 | 18.4 | 24.4 |
| $T = 100$ | 1% | 3.1 | 0.9 | 0.9 | 1.0 | 21.6 | 2.3 | 5.5 | 7.7 |
| Experiment 5 | $E[LR]$ | 6.74 | 6.06 | 6.07 | 6.09 | 13.03 | 7.19 | 9.54 | 10.29 |
| $\varphi_1 = [0.8, \mathbf{0.8}]$ | $BF$ | 0.112 | | | | 0.812 | | | |
| $\varphi_2 = [0.8, \mathbf{0.8}]$ | 10% | 14.8 | 10.3 | 10.3 | 10.4 | 61.7 | 15.0 | 35.6 | 42.1 |
| $\gamma = [1, \epsilon]$ | 5% | 8.2 | 5.4 | 5.4 | 5.6 | 47.3 | 6.7 | 23.1 | 28.2 |
| $T = 100$ | 1% | 1.8 | 0.9 | 0.9 | 0.9 | 24.8 | 0.9 | 5.9 | 10.2 |
| Experiment 6 | $E[LR]$ | 6.48 | 6.08 | 6.06 | 6.06 | 9.78 | 7.28 | 7.80 | 8.13 |
| $\varphi_1 = [0.8, \mathbf{-0.6}]$ | $BF$ | 0.066 | | | | 0.343 | | | |
| $\varphi_2 = [0.8, \mathbf{-0.6}]$ | 10% | 13.1 | 10.2 | 10.2 | 10.2 | 37.9 | 17.0 | 21.7 | 24.0 |
| $\gamma = [1, \epsilon]$ | 5% | 6.8 | 5.3 | 5.3 | 5.3 | 25.2 | 9.3 | 12.2 | 14.9 |
| $T = 100$ | 1% | 1.8 | 1.3 | 1.3 | 1.3 | 9.6 | 2.0 | 2.9 | 4.3 |
| Experiment 7 | $E[LR]$ | 6.45 | 6.02 | 6.01 | 6.01 | 7.22 | 6.28 | 6.34 | 6.39 |
| $\varphi_1 = [\mathbf{0.6}, \mathbf{-0.6}]$ | $BF$ | 0.072 | | | | 0.150 | | | |
| $\varphi_2 = [\mathbf{0.6i}, \mathbf{-0.6i}]$ | 10% | 13.4 | 10.7 | 10.7 | 10.7 | 18.8 | 12.4 | 12.9 | 13.6 |
| $\gamma = [1, \epsilon]$ | 5% | 7.4 | 5.4 | 5.4 | 5.3 | 11.2 | 6.0 | 6.2 | 6.4 |
| $T = 100$ | 1% | 1.2 | 0.8 | 0.8 | 0.8 | 2.5 | 1.2 | 1.3 | 1.4 |
| Experiment 8 | $E[LR]$ | 7.92 | 6.34 | 6.41 | 6.45 | 14.37 | 8.53 | 9.87 | 10.65 |
| $\varphi_1 = [0.8, 0.6, \mathbf{0.2}, \mathbf{0.2}]$ | $BF$ | 0.250 | | | | 0.684 | | | |
| $\varphi_2 = [0.8, 0.6, \mathbf{0.2}, \mathbf{0.2}]$ | 10% | 23.5 | 12.5 | 12.5 | 12.7 | 67.6 | 27.2 | 38.1 | 43.8 |
| $\gamma = [1, \epsilon, \epsilon, \epsilon]$ | 5% | 14.7 | 6.7 | 6.7 | 7.0 | 54.7 | 16.2 | 25.6 | 31.2 |
| $T = 100$ | 1% | 4.7 | 1.2 | 1.2 | 1.2 | 31.8 | 4.1 | 8.45 | 12.6 |

Table 5.4: Bartlett corrections for two tests of no level feedback in the cointegrated VAR. The variations with respect to Experiment 1 are given in bold face. $\epsilon = 10^{-4}$ (If $\epsilon$ were equal to zero, $\Phi$ would be of reduced rank in the DGP)

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

We vary the following aspects of the DGP: $T$ (the number of observations), $k$ (the number of lags) and $\varphi_1, \varphi_2$ and $\gamma$. For each experiment we report two tests (in their uncorrected and corrected versions): the test that the last three rows of the adjustment parameters $\alpha$ are zero for known cointegration space $\beta$ and for unknown $\beta$. Under $\mathcal{H}_0$ both tests asymptotically have a $\chi^2$-distribution with six degrees of freedom and the Bartlett correction for the first test is given in corollary 6.Corrollary 7 provides the expression for the second test.

Each of these Bartlett corrections depends on the parameters of the model. We calculate the Bartlett correction based on

(1) The true (DGP) value of the parameters, $\theta$.
(2) The maximum likelihood estimates of the parameters under $\mathcal{H}_0, \hat{\theta}_r$.
(3) The maximum likelihood estimates of the parameters under the alternative, $\hat{\theta}_u$.

Omtzigt and Fachin (2002) argue that for the test of corollary 7, one needs to use $\hat{\theta}_u$ as $lim_{T\to\infty} tr\left\{\left(\alpha'\Omega^{-1}\alpha\right)^{-1} S_\perp'\Sigma_{yy}^{-1}S_\perp\right\}$ is not defined under the alternative. Their point does not apply to the test in corollary 6.

The simulation is based on 2000 replications and for each test we report the expected value of the likelihood ratio test, as well as the expected value of the Bartlett corrected test based on $\theta, \hat{\theta}_r$ and $\hat{\theta}_u$.We also give the Bartlett factor based on $\theta$. As before we report the empirical rejection probabilities at the nominal 10%, 5% and 1% level.

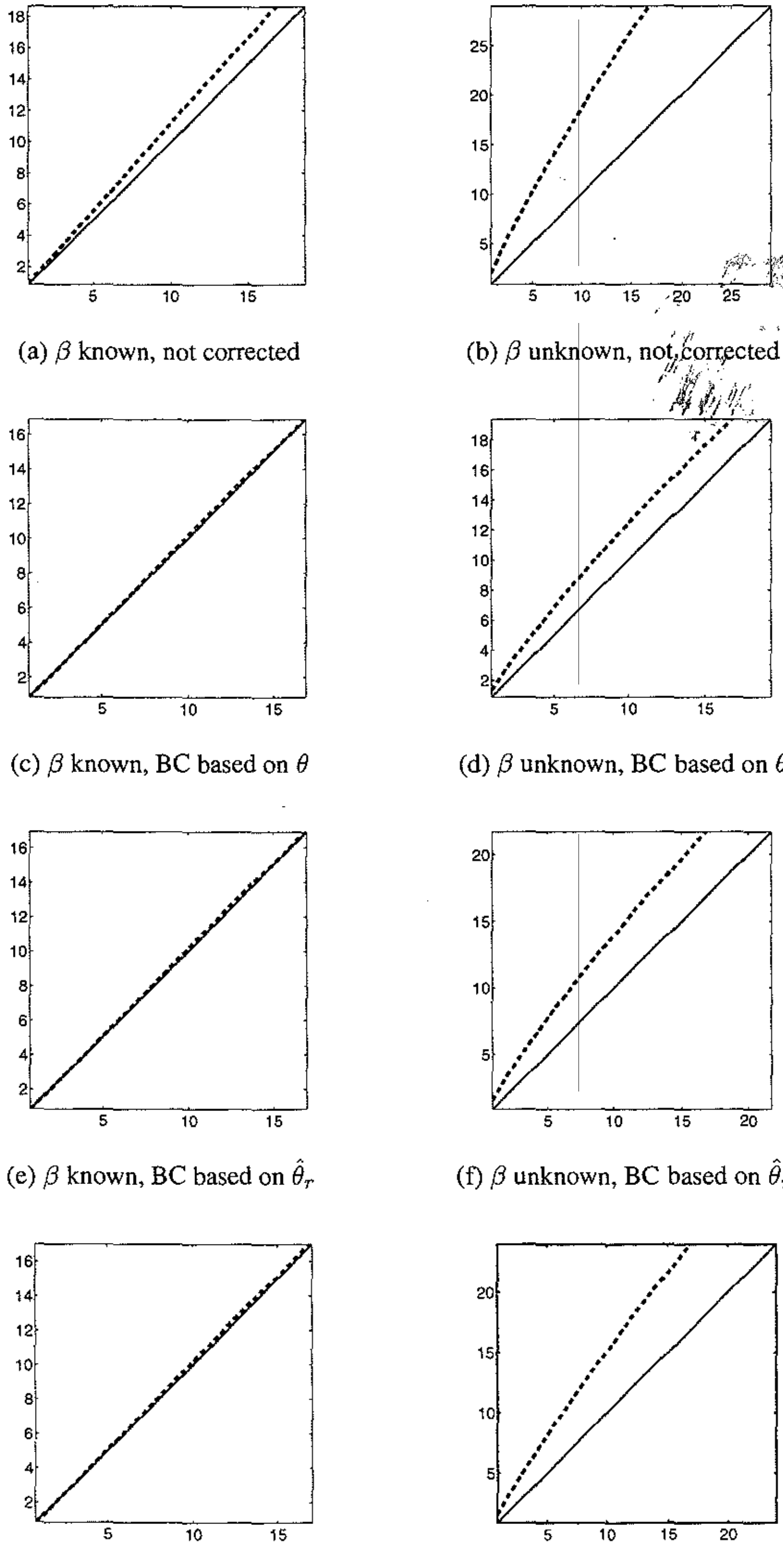
In the benchmark model (experiment 1), both stationary variables, $X_{1t}$ and $X_{2t}$ have relatively large residual roots at 0.6 and 0.8. The other three series are pure random walks[2] and we have 100 observations. The first block-row of table 5.4 shows that the Bartlett correction in the test for known $\beta$ performs well: at the 5% nominal level, it corrects from 8.0% to 5.2% for all three Bartlett corrected tests. For unknown $\beta$ the results are different. The original size distortion is considerably larger, as the empirical size of the asymptotic test at the nominal 5% level is 37.6%. The Bartlett correction based on the true value brings this down to 10.5%, but those based on the restricted and unrestricted estimates only bring it down to 16.4% and 21.0% respectively. Even for T=200 (experiment 2) the corrected test remains size distorted. Four hundred observations (experiment 3) are needed for the corrected test to reach a rejection probability close to 5%. In experiments 4 and 5, the smallest residual roots in the non-stationary and stationary variables respectively are raised. The Bartlett correction for the test based on known $\beta$ continues to perform well, but the one based on unknown $\beta$ does even worse than in the benchmark case. If the roots are more scattered on the real line (experiment 6) or inside the unit circle (experiment 7), the performance of the Bartlett corrected test with unknown $\beta$ is more acceptable. The size corrections perform worse with a longer lag length (experiment 8), which is in line with the findings in Omtzigt and Fachin (2002) .

Overall the Bartlett correction performs better when $\beta$ is known than when it is unknown, though this may be specific to the Monte Carlo design chosen and the larger size distortion of the non-corrected test.

In figure 5.3 we give the QQ-plots of the uncorrected and corrected test in experiment 1, based on 20000 replications. We observe that the plots on the left hand side, which correspond to corollary 6 are straight and that all three corrected test virtually coincide with the 45 degree line, showing the effectiveness of the Bartlett correction. The plots on the right hand side correspond to the case where $\beta$ is unknown and in none of the four plots does the empirical QQ-plot coincide with the 45 degree line. However all four plots are almost straight lines. (In the bottom two rows, the Bartlett correction depends on the estimated parameters, such that the Bartlett correction does

[2] There are one or three very small extra small roots in the polynomial, which are $\epsilon = 10^{-4}$. They serve no other purpose than to ensure invertibility of the matrix $\Phi$.


DOI: 10.2870/52609



(a) $\beta$ known, not corrected

(b) $\beta$ unknown, not corrected

(c) $\beta$ known, BC based on $\theta$

(d) $\beta$ unknown, BC based on $\theta$

(e) $\beta$ known, BC based on $\hat{\theta}_r$

(f) $\beta$ unknown, BC based on $\hat{\theta}_r$

(g) $\beta$ known, BC based on $\hat{\theta}_u$

(h) $\beta$ unknown, BC based on $\hat{\theta}_u$

Figure 5.3: QQ-plots of LR-tests (asymptotic and Bartlett corrected) for 'no level feedback' in the cointegrated VAR model

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

not just rotate the QQ-plot. Potentially it can also change the curvature). The relatively straight line and the fact that the correction functions with 400 observations are consistent with the view that a higher order expansion of the expectation of the likelihood ratio test is needed in this case. Nielsen (1997) and Johansen (2002b) provide examples of Bartlett corrections in which higher order terms are needed to make the Bartlett correction function.

## 5.7. Conclusions

We have derived the Bartlett correction for a simple hypothesis on the regression parameters in a multivariate stationary autoregressive process. Three applications illustrate the use of the correction: the test for absence of autocorrelation of any order, a simple hypothesis on the autoregressive parameters and two tests for no long-run feedback in the cointegrated VAR model. In the first of these last two tests, the cointegration space is known, in the second it is not. In all sections explicit expressions for the Bartlett correction are given.

The Bartlett correction performs well in all simulation studies, except in the one of the last test, that is a test for weak exogeneity in the cointegrated VAR with an unknown cointegration space. In that particular case a second order expansion might improve the Bartlett correction.

## 5.A. Derivation of the main results

In this appendix we prove theorem 9 and 10. In the first subsection we derive a number of useful lemma's, which will be applied over and again in the theorems. Then theorem 9 is derived. Theorem 10 is derived in subsection 5.A.11: the short proof is in some way a special case of theorem 9. Theorems 12 and 11 are derived in subsections 5.A.13 and 5.A.12 respectively.

**5.A.1. Lemma's.** To prove the two theorems and their corollaries, we shall state a few useful lemma's. The first one states that in all the estimation problems we consider in this paper, the Ordinary Least Squares (OLS) estimator and Maximum Likelihood (ML) estimators coincide:

LEMMA 2. *If A varies unrestricted in a product space, that is* $A \in R^{n\times(q+r)}$ *in the model:*

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} X_t + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix} \tag{5.37}$$
$$\varepsilon_{1t} \sim MIIDN(0,\Omega)$$

*then the maximum likelihood estimator of* $A$, $\tilde{A}$ *and the OLS-estimator of* $A, \hat{A}$ *coincide. Furthermore* $\hat{A}_2 = A_2$

PROOF. In the first sub model $Y_{1t} = A_1X_t + \varepsilon_{1t}$, $\hat{A}_1 = (X'X)^{-1}X'Y_1$ whereas in the full model (5.37) $\hat{A} = (X'X)^{-1}X'Y$ which implies $\hat{A}_1 = (I,0)(X'X)^{-1}X'Y = (X'X)^{-1}X'Y_1$. Therefore the OLS estimators in the two small submodels coincide with the OLS estimator of the large model (5.37)
The variance-covariance matrix of $\begin{bmatrix} \varepsilon_{1t} & 0 \end{bmatrix}'$ is trivially block-diagonal with $\Omega$ and 0 as diagonal elements. Therefore maximization of the likelihood function of (5.37) is the same as the separate maximization of the likelihood functions of the two submodels.
In the second sub model $Y_{2t} = A_2X_t$,$\hat{A}_2 = (X'X)^{-1}X'Y_2 = A_2$ as we are estimating an identity. This estimator trivially equals the maximum likelihood estimator. The ML-estimator of the first submodel equals the OLS-estimator as $A_1 \in R^{n\times q}$ □

Next we state two standard result on the products of the errors in the multivariate normal distribution:

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

LEMMA 3. *Let $\varepsilon_i = [\varepsilon_{1i}', \varepsilon_{2i}']', i = 1, ..., T$ be $(n \times 1)$ vectors, distributed i.i.d.$N(0, I_n)$ and let $\varepsilon_{2i}$ be of dimension $q \leq n$. Further let $M$ be an $(n \times n)$ matrix.*
*Then:*

$$E[\varepsilon_i' M \varepsilon_j] = \begin{cases} tr\{M\} & if\, i = j \\ 0 & otherwise \end{cases}$$

$$E[D] = E[\varepsilon_{2i}\varepsilon_j' M \varepsilon_k \varepsilon_{2l}'] = \begin{cases} M_{22} + M_{22}' + I_q tr\{M\} & if\ i = j = k = l \\ I_q tr\{M\} & if\ i = l \neq j = k \\ M_{22}' & if\ i = k \neq j = l \\ M_{22} & if\ i = j \neq k = l \\ 0 & otherwise, \end{cases}$$

PROOF. First let $\varepsilon_i = \begin{bmatrix} \epsilon_i^1 & \epsilon_i^2 & \cdots & \epsilon_i^n \end{bmatrix}', \varepsilon_{2i} = \begin{bmatrix} \epsilon_i^{n-q+1} & \cdots & \epsilon_i^n \end{bmatrix}'$ and denote the element in row $a$ and column $b$ of matrix $M$ as $m^{ab}$. Then let $L_{q \times n} = [0, I_q]$. Throughout we use the fact that the first and third moment of this normal distribution are zero and that $E\left[\epsilon_i^a \epsilon_j^b\right] = 1$ iff $a = b$ and $i = j$.

- $E[\varepsilon_i' M \varepsilon_i] = E\left[\sum_{a=1}^n \sum_{b=1}^n \epsilon_i^a m^{ab} \epsilon_i^b\right] = E\left[\sum_{a=1}^n \epsilon_i^a m^{aa} \epsilon_i^a\right] = tr\{M\}$
- If $i = j \neq k = l$, then $E\left[\varepsilon_{2i}\varepsilon_j' M \varepsilon_k \varepsilon_{2l}'\right] = E[\varepsilon_{2i}\varepsilon_i'] M E[\varepsilon_k \varepsilon_{2k}'] = LML' = M_{22}$
- If $i = l \neq j = k$, then $E\left[\varepsilon_{2i}\varepsilon_j' M \varepsilon_k \varepsilon_{2l}'\right] = E[\varepsilon_{2i}\varepsilon_{2i}'] tr\{M\} = I_q tr\{M\}$
- If $i = k \neq j = l$, then $E\left[\varepsilon_{2i}\varepsilon_j' M \varepsilon_k \varepsilon_{2l}'\right] = E[\varepsilon_{2i}\varepsilon_i'] M' E\left[\varepsilon_j \varepsilon_{2j}'\right] = LM'L' = M_{22}'$
- If $i = j = k = l$, Consider $D^* = L'DL$. Then only the entries in the lower right hand part of the matrix are non zero.Let $\delta$ be the Kronecker delta, such that $\delta_{\alpha\beta} = 1$ iff $\alpha = \beta$ and zero otherwise to find: $\left[d^{*ab}\right] = (1 - \delta_{ab}) E\left[(\epsilon_i^a)^2 \left(m^{ab} + m^{ba}\right) \left(\epsilon_i^b\right)^2\right] +$
  $\delta_{ab} E\left[(\epsilon_i^a)^2 \left(\epsilon_i^b\right)^2 \sum_{\substack{b=1 \\ b \neq a}}^n m^{bb} + (\epsilon_i^a)^4 m^{aa}\right]$
  $= (1 - \delta_{ab}) \left(m^{ab} + m^{ba}\right) + \delta_{ab} \sum_{\substack{b=1 \\ b \neq a}}^n m^{bb} + \delta_{ab} 3 m^{aa}$ for $a, b \geq n - q + 1$
  otherwise $E\left[d^{*ab}\right] = 0$
  We thus find that $E[D^*] = L'ML + L'M'L + L'L \times tr\{M\}$.
  $E[D] = M + M' + I_q \times tr\{M\}$.

- If we have a $\varepsilon-$vector, whose index does not coincide with the index of another $\varepsilon-$vector, then by independence the expectation of the whole expression becomes zero.

□

LEMMA 4. *Let $\varepsilon_i = [\varepsilon_{1i}, \varepsilon_{2i}], i = 1, ..., T$ be $(n \times 1)$ vectors, distributed i.i.d.$N(0, I_n)$ and let $\varepsilon_{2i}$ be of dimension $q \leq n$. Further let $S$ be an $(q \times q)$ matrix and $L_{q \times n} = [0, I_q]$. Define $S^* = L'SL$. Then:*

$$E[D] = E[\varepsilon_i \varepsilon_{2j}' S \varepsilon_{2k} \varepsilon_l'] = \begin{cases} S^* + S^{*\prime} + I_n tr\{S\} & if\ i = j = k = l \\ I_n tr\{S\} & if\ i = l \neq j = k \\ S^{*\prime} & if\ i = k \neq j = l \\ S^* & if\ i = j \neq k = l \\ 0 & otherwise \end{cases}$$

PROOF. First let $\varepsilon_i' = \begin{bmatrix} \epsilon_i^1 & \epsilon_i^2 & \cdots & \epsilon_i^n \end{bmatrix}, \varepsilon_{2i}' = \begin{bmatrix} \epsilon_i^{n-q+1} & \cdots & \epsilon_i^n \end{bmatrix}$ and denote the element in row $a$ and column $b$ of matrix $S$ as $s^{ab}$.

- If $i = j \neq k = l$, then $E\left[\varepsilon_i \varepsilon_{2j}' S \varepsilon_{2k} \varepsilon_l'\right] = E[\varepsilon_i \varepsilon_{2i}'] M E[\varepsilon_{2k} \varepsilon_k'] = L'SL = S^*$
- If $i = l \neq j = k$,then $E\left[\varepsilon_{2i}\varepsilon_j' M \varepsilon_k \varepsilon_{2l}'\right] = E[\varepsilon_{2i}\varepsilon_{2i}'] tr\{S\} = I_n tr\{S\}$
- If $i = k \neq j = l$,then $E\left[\varepsilon_{2i}\varepsilon_j' M \varepsilon_k \varepsilon_{2l}'\right] = E[\varepsilon_{2i}\varepsilon_i'] M' E\left[\varepsilon_j \varepsilon_{2j}'\right] = LS'L' = S^{*\prime}$

- For $i = j = k = l$, et $\delta_2 = 1$ iff $\alpha, \beta \geq n - 1 + 1$ and 0 otherwise then $E\left[\varepsilon_i \varepsilon_{2j}' S \varepsilon_{2k} \varepsilon_2'\right]$
$= \delta_{ab}\delta_2 E\left[\left(\epsilon_i^a\right)^2\left(\epsilon_i^b\right)^2 \sum_{\substack{b=1 \\ b\neq a}}^q s^{bb} + \left(\epsilon_i^a\right)^4 s^{aa}\right] + (1-\delta_{ab})\,\delta_2\left[\left(\epsilon_i^a\right)^2\left(s^{ab}+s^{ba}\right)\left(\epsilon_i^b\right)^2\right]$
$+\delta_{ab}(1-\delta_2)\,E\left[\left(\epsilon_i^a\right)^2\left(\epsilon_i^b\right)^2 \sum_{b=1}^q s^{bb}\right] + (1-\delta_{ab})(1-\delta_2)\,0$
such that we find $E\left[\varepsilon_{2i}\varepsilon_i' S \varepsilon_i \varepsilon_{2l}'\right] = S^* + S^{*\prime} + I_n tr\{S\}$
- If we have a $\varepsilon-$vector, whose index does not coincide with the index of another $\varepsilon-$vector, then by independence the expectation of the whole expression becomes zero.

□

LEMMA 5. *Let $\varepsilon_i, i = 1, ..., T$ be $(n \times 1)$ vectors, distributed i.i.d. $N(0, I_n)$ and $M$ and $N$ $(n \times n)$ matrices.*
*Then:*

$$E\left[\varepsilon_i' M \varepsilon_j \varepsilon_k' N \varepsilon_l\right] = \begin{cases} tr\{MN\} + tr\{MN'\} + tr\{M\}\,tr\{N\} & \text{if } i=j=k=l \\ tr\{M\}\,tr\{N\} & \text{if } i=j\neq k=l \\ tr\{MN'\} & \text{if } i=k\neq j=l \\ tr\{MN\} & \text{if } i=l\neq j=k \\ 0 & \text{otherwise} \end{cases} \tag{5.38}$$

PROOF. We proceed as in the last two lemma's and refer to them for notation:

- If $i = j \neq k = l$, then $E\left[\varepsilon_i' M \varepsilon_j \varepsilon_k' N \varepsilon_l\right] = E\left[\varepsilon_i' M \varepsilon_j\right] E\left[\varepsilon_k' N \varepsilon_l\right] = tr\{M\}\,tr\{N\}$
- If $i = l \neq j = k$, then $E\left[\varepsilon_i' M \varepsilon_j \varepsilon_k' N \varepsilon_l\right] = tr\{MN\}$
- If $i = k \neq j = l$, then $E\left[\varepsilon_i' M \varepsilon_j \varepsilon_k' N \varepsilon_l\right] = E\left[\varepsilon_i' M \varepsilon_j \varepsilon_l' N' \varepsilon_k\right] = tr\{MN'\}$
- If $i = j = k = l$, then $E\left[\varepsilon_i' M \varepsilon_j \varepsilon_k' N \varepsilon_l\right] = E\left[\varepsilon_i' M \varepsilon_i \varepsilon_i' N \varepsilon_i\right]$
$= E\left[\sum_{a,b,c,d=1}^n \epsilon_i^a m^{ab} \epsilon_i^b \epsilon_i^c n^{cd} \epsilon_i^d\right]$
$= E\left[\sum_{a=1}^n \left(\epsilon_i^a\right)^4 m^{aa} n^{aa}\right] + E\left[\sum_{\substack{a,c=1 \\ a\neq c}}^n \left(\epsilon_i^a\right)^2\left(\epsilon_i^c\right)^2 m^{aa} n^{cc}\right]$
$+E\left[\sum_{\substack{a,b=1 \\ a\neq b}}^n \left(\epsilon_i^a\right)^2\left(\epsilon_i^b\right)^2\left(m^{ab}n^{ab} + m^{ab}n^{ba}\right)\right]$
$= tr\{MN\} + tr\{MN'\} + tr\{M\}\,tr\{N\}$
- If we have a $\varepsilon-$vector, whose index does not coincide with the index of another $\varepsilon-$vector, then by independence the expectation of the whole expression becomes zero. Throughout we have used the fact that the first and third moments of the normal distribution is zero.

□

**5.A.2. Proof of Theorem 9.** We first consider the model of theorem 9, which concerns a simple hypothesis $\mathcal{H}_0 : A = A_0$

$$Y_t = AX_t + \eta_{2t} \tag{5.39}$$

where

$$X_t = Q(L)\eta_{t-1}$$
$$\eta_t = \left[\begin{array}{cc} \eta_{1t}' & \eta_{2t}' \end{array}\right]' \sim MIIDN(0, \Omega)$$

where $\eta_t$ is of dimension $n$, whereas $\eta_{2t}$ is of dimension $q$. Furthermore under $\mathcal{H}_0$, $H = Y - XA$, where with capitals we denote the stacked vectors. For instance $Y = [y_1, \ldots, y_T]'$,$U = [\varepsilon_{21}, \ldots, \varepsilon_{2T}]'$, $H = [\eta_{21}, \ldots, \eta_{2T}]'$. Also $\varepsilon_{2t} = \Omega_{22}^{-\frac{1}{2}}\eta_{2t}$ and $\varepsilon_t = \Omega^{-\frac{1}{2}}\eta_t$.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

It is well-known that the ordinary least squares estimator and the maximum likelihood estimator coincide in this model, such that the maximum likelihood estimator can be written as: $\hat{A} = A + (X'X)^{-1}(X'H)$. We substitute this in the likelihood ratio test for $\mathcal{H}_0 : A = A_0$ and expand it, keeping only first order terms:

$$
\begin{aligned}
-2\ln LR\,(A = A_0) &= -T\log\left|(Y - X\hat{A}')'(Y - X\hat{A}')\right|\left|(H'H)\right|^{-1} \\
&= -T\log\left|I_q - \left(\Omega_{22}^{\frac{1}{2}}U'U\Omega_{22}^{\frac{1}{2}}\right)^{-1}\left(\Omega_{22}^{\frac{1}{2}}U'X\right)(X'X)^{-1}\left(X'U\Omega_{22}^{\frac{1}{2}}\right)\right| \\
&= -T\log\left|I_q - \Omega_{22}^{-\frac{1}{2}}(U'U)^{-1}(U'X)(X'X)^{-1}(X'U)\Omega_{22}^{\frac{1}{2}}\right| \\
&= -T\log\left|\Omega_{22}^{-\frac{1}{2}}\right|\left|I_q - \Omega_{22}^{-\frac{1}{2}}(U'U)^{-1}(U'X)(X'X)^{-1}(X'U)\,\Omega_{22}^{\frac{1}{2}}\right|\left|\Omega_{22}^{\frac{1}{2}}\right| \\
&= -T\log\left|I_q - (U'U)^{-1}(U'X)(X'X)^{-1}(X'U)\right| \\
&= -T\left|I_q - K\right| \\
&\stackrel{1}{=} tr(K) + \frac{1}{2T}tr(K^2)
\end{aligned}
$$

where we have defined $K \equiv T\,(U'U)^{-1}(U'X)(X'X)^{-1}(X'U)$.
The probability limits of the two matrices, whose inverses enter $K$, are:

$$
\left(\frac{1}{T}U'U\right) \xrightarrow{P} I_q
$$

$$
\left(\frac{1}{T}X'X\right) \xrightarrow{P} \Phi = \sum_{\eta=0}^{\infty} C_\eta C_\eta' = Var(X_t)
$$

and their first order expansions are:

$$
\begin{aligned}
\left(\frac{1}{T}U'U\right)^{-1} &= \left(I_q - \left(I_q - \frac{1}{T}U'U\right)\right)^{-1} \\
&\stackrel{1}{=} I_q + \left(I_q - \frac{1}{T}U'U\right) + \left(I_q - \frac{1}{T}U'U\right)^2 \qquad (5.40)
\end{aligned}
$$

$$
\begin{aligned}
\left(\frac{1}{T}X'X\right)^{-1} &= \left(\Phi - \left(\Phi - \frac{1}{T}X'X\right)\right)^{-1} \\
&\stackrel{1}{=} \Phi^{-1} + \Phi^{-1}\left(\Phi - \frac{1}{T}X'X\right)\Phi^{-1} + \Phi^{-1}\left(\Phi - \frac{1}{T}X'X\right)\Phi^{-1}\left(\Phi - \frac{1}{T}X'X\right)\Phi^{-1} \qquad (5.41)
\end{aligned}
$$


DOI: 10.2870/52609

Using (5.40) and (5.41) we can write the first order expansion of the expected value of $K$ as:

$$
\begin{aligned}
E\left[tr(K)\right] &\doteq tr\left\{E\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\} \\
&+ tr\left\{E\left(I_q-\frac{1}{T}U'U\right)\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\} \\
&+ tr\left\{E\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\Phi-\frac{1}{T}X'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\} \\
&+ tr\left\{E\left(I_q-\frac{1}{T}U'U\right)\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\Phi-\frac{1}{T}X'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\} \\
&+ tr\left\{E\left(I_q-\frac{1}{T}U'U\right)^2\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\} \\
&+ tr\left\{E\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\Phi-\frac{1}{T}X'X\right)\Phi^{-1}\left(\Phi-\frac{1}{T}X'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\}
\end{aligned}
$$

The names of these terms shall be $D_1$ to $D_6$. Together with $\frac{1}{2T}E\left[tr(K^2)\right]$ these terms form the expansion of the expectation of the likelihood ratio test. Their expectations are worked out one by one in the following pages.

### 5.A.3. Derivation of $D_1$.

$$
\begin{aligned}
& tr\left\{E\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\} \\
&= tr\left\{E\left[\frac{1}{T}\sum_{t,s=1}^{T}\sum_{\zeta,\eta=0}^{\infty}\varepsilon_{2t}\varepsilon_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right]\right\}
\end{aligned}
$$

There is only one way in which this terms gives a non-zero expectation: $t=s,\eta=\zeta$. We then get:

$$
\begin{aligned}
& tr\left\{\frac{1}{T}E\left[\sum_{s=1}^{T}\varepsilon'_{2s}\varepsilon_{2s}\right]E\left[\sum_{\eta=0}^{\infty}\varepsilon'_{s-1-\eta}C'_\zeta\Phi^{-1}C_\zeta\varepsilon_{s-1-\eta}\right]\right\} \\
&= q\times tr\left\{\sum_{\eta=0}^{\infty}C'_\zeta\Phi^{-1}C_\zeta\right\} \\
&= q\times tr\left\{I_n\right\} \\
&= qn
\end{aligned}
$$

$$\boxed{D_1 = qn}$$

### 5.A.4. Derivation of $D_2$.

$$
\begin{aligned}
& tr\left\{E\left(I_q-\frac{1}{T}U'U\right)\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\} \\
&= -tr\left\{\frac{1}{T^2}E\sum_{t,s,r=1}^{T}\sum_{\zeta,\eta=0}^{\infty}\left(\varepsilon_{2r}\varepsilon'_{2r}-I_q\right)\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}
\end{aligned}
$$

There are two ways in which this combination gives has an expectation of at least $O(\frac{1}{T})$: Either $t=s=r$ and $\eta=\zeta$ or $t=s$ and $s-1-\eta=t-1-\zeta=r$.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

5.A.4.1. *The first combination.* $t = s = r$ and $\eta = \zeta$
I find

$$- tr\left\{\frac{1}{T^2}E\left[\sum_{s=1}^{T}\varepsilon_{2s}\varepsilon_{2s}'\varepsilon_{2s}\varepsilon_{2s}' - \sum_{s=1}^{T}\varepsilon_s\varepsilon_s'\right]E\left[\sum_{\eta=0}^{\infty}\varepsilon_{s-1-\eta}'C_\eta'\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\right]\right\}$$
$$= -tr\left\{\frac{1}{T^2}\left(T\left(q+1\right)\times I_q\right)\times tr\left\{I_n\right\}\right\}$$
$$= -\frac{q^2n+nq}{T}$$

where we applied lemma 3 in the second line.

$$\boxed{D_{21} = -\frac{q^2n+nq}{T}}$$

5.A.4.2. *The second combination.* $t = s$ and $s - 1 - \eta = t - 1 - \zeta = r$.
Here the reasoning goes as follows:

$$- tr\left\{\frac{1}{T^2}E\sum_{t,s,r=1}^{T}\sum_{\zeta,\eta=0}^{\infty}\left(\varepsilon_{2r}\varepsilon_{2r}' - I_q\right)\varepsilon_{2t}\varepsilon_{t-1-\zeta}'C_\zeta'\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon_{2s}'\right\}$$
$$= -tr\left\{\frac{1}{T^2}E\left[\sum_{\eta=0}^{\infty}\varepsilon_{2s-1-\eta}\varepsilon_{s-1-\eta}'C_\eta'\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon_{2s-1-\eta}'\right]E\left[\sum_{s=1}^{T}\varepsilon_{2s}\varepsilon_{2s}'\right]\right\}$$
$$+ tr\left\{\frac{1}{T^2}E\left[\sum_{\eta=0}^{\infty}\varepsilon_{s-1-\eta}'C_\eta'\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\right]E\left[\sum_{s=1}^{T}\varepsilon_{2s}\varepsilon_{2s}'\right]\right\}$$
$$= -tr\left\{\frac{1}{T^2}\left(\left(n+2\right)\times I_q\right)\times\left(T\times I_q\right)\right\} + \frac{nq}{T}$$
$$= -\frac{2q}{T}$$

where we have applied lemma 4 in the third passage. We thus conclude that

$$\boxed{D_{22} = -\frac{2q}{T}}$$

**5.A.5. Derivation of $D_3$.**

$$tr\left\{E\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\Phi - \frac{1}{T}X'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\}$$
$$= -tr\left\{\frac{1}{T^2}E\sum_{t,s,v=1}^{T}\sum_{\zeta,\eta,\kappa,\lambda=0}^{\infty}\varepsilon_{2t}\varepsilon_{t-1-\zeta}'C_\zeta'\Phi^{-1}C_\kappa\left(\varepsilon_{v-1-\kappa}\varepsilon_{v-1-\lambda}' - \delta_{\kappa\lambda}I_n\right)C_\lambda'\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon_{2s}'\right\}$$

There are five ways in which this expression gives an expectation of at least $O(\frac{1}{T})$ :

(1) $s = t = v - 1 - \lambda = v - 1 - \kappa$ and $t - 1 - \zeta = s - 1 - \eta$ and $\lambda = \kappa$
(2) $s = t$ and $t - 1 - \zeta = s - 1 - \eta = v - 1 - \lambda = v - 1 - \kappa$
(3) $v - 1 - \kappa = s - 1 - \eta \neq v - 1 - \lambda = t - 1 - \zeta$ and $s = t$ (also change $\kappa$ and $\lambda$ to get two combinations in total)

(4) $s = v-1-\kappa$ and $t = s-1-\eta$ and $v-1-\lambda = t-1-\zeta$ (also change both $\kappa$ and $\lambda$ and $\zeta$ and $\eta$ for four combinations)

(5) $t = v-1-\kappa$ and $s = v-1-\lambda$ and $t-1-\zeta = s-1-\eta$ (also change $\kappa$ and $\lambda$, deriving two expressions)

These five combinations, some of them consisting of subcombinations, shall now be dealt with one by one:

5.A.5.1. *Derivation of* $D_{31}$. $s = t = v-1-\lambda = v-1-\kappa$ and $\zeta = \eta$ and $\lambda = \kappa$

$$
\begin{aligned}
&-tr\left\{\frac{1}{T^2}E\sum_{t,s,v=1}^{T}\sum_{\zeta,\eta,\kappa,\lambda=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\left(\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}-\delta_{\kappa\lambda}I_n\right)C'_\lambda\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}\\
&=-tr\left\{\frac{1}{T^2}E\left[\sum_{s=1}^{T}\varepsilon_s\varepsilon'_{2s}\varepsilon_{2s}\varepsilon'_s-\sum_{s=1\kappa\lambda}^{T}I_n\varepsilon'_s\varepsilon_s\right]E\left[\sum_{\kappa=0}^{\infty}\sum_{\eta=0}^{\infty}C'_\kappa\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{s-1-\eta}C'_\eta\Phi^{-1}C_\kappa\right]\right\}\\
&=-tr\left\{\frac{1}{T^2}\left(2T\times I_q^*\right)\times\sum_{\kappa=0}^{\infty}C'_\kappa\Phi^{-1}C_\kappa\right\}\\
&=-\frac{2}{T}tr\left\{\left[\sum\nolimits_{\kappa=0}^{\infty}C'_\kappa\Phi^{-1}C_\kappa\right]_{22}\right\}
\end{aligned}
$$

where we defined the $(n\times n)$ matrix $I_q^*$ as:

$$
I_q^* = \begin{bmatrix} 0 & 0 \\ 0 & I_q \end{bmatrix}
$$

$$
\boxed{D_{31} = -\tfrac{2}{T}tr\left\{\left[\sum\nolimits_{\kappa=0}^{\infty}C'_\kappa\Phi^{-1}C_\kappa\right]_{22}\right\}}
$$

5.A.5.2. *Derivation of* $D_{32}$. $s = t$ and $t-1-\zeta = s-1-\eta = v-1-\lambda = v-1-\kappa$

$$
\begin{aligned}
&-tr\left\{\frac{1}{T^2}E\sum_{t,s,v=1}^{T}\sum_{\zeta,\eta,\kappa,\lambda=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\left(\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}-\delta_{\kappa\lambda}I_n\right)C'_\lambda\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}\\
&=-\frac{1}{T^2}E\left[\sum_{s=1}^{T}\varepsilon'_{2s}\varepsilon_{2s}\right]E\left[\sum_{\kappa=0}^{\infty}\sum_{\eta=0}^{\infty}\varepsilon'_uC'_\eta\Phi^{-1}C_\kappa\left(\varepsilon_u\varepsilon'_u-I_n\right)C'_\kappa\Phi^{-1}C_\eta\varepsilon_u\right]\\
&=-\frac{q}{T}E\left[\sum_{\kappa=0}^{\infty}\sum_{\eta=0}^{\infty}\varepsilon'_uC'_\eta\Phi^{-1}C_\kappa\varepsilon_u\varepsilon'_uC'_\kappa\Phi^{-1}C_\eta\varepsilon_u\right]+\frac{q}{T}E\left[\sum_{\kappa=0}^{\infty}\sum_{\eta=0}^{\infty}\varepsilon'_uC'_\eta\Phi^{-1}C_\kappa C'_\kappa\Phi^{-1}C_\eta\varepsilon_u\right]\\
&=-\frac{q}{T}\sum_{\kappa=0}^{\infty}\sum_{\eta=0}^{\infty}tr\left\{\left(C'_\eta\Phi^{-1}C_\kappa\right)^2\right\}-\frac{q}{T}\sum_{\kappa=0}^{\infty}\sum_{\eta=0}^{\infty}tr\left\{\left(C'_\eta\Phi^{-1}C_\kappa C'_\kappa\Phi^{-1}C_\eta\right)\right\}\\
&\frac{q}{T}\sum_{\kappa=0}^{\infty}\sum_{\eta=0}^{\infty}tr^2\left\{\left(C'_\kappa\Phi^{-1}C_\eta\right)\right\}+\frac{q}{T}\sum_{\kappa=0}^{\infty}\sum_{\eta=0}^{\infty}tr\left\{\left(C'_\eta\Phi^{-1}C_\kappa C'_\kappa\Phi^{-1}C_\eta\right)\right\}
\end{aligned}
$$

where we have applied lemma 5 in the third passage, such that we conclude that the total contribution of D32 is equal to:

$$
\boxed{D_{32} = -\tfrac{q}{T}\sum\nolimits_{\kappa=0}^{\infty}\sum\nolimits_{\eta=0}^{\infty}tr\left\{\left(C'_\eta\Phi^{-1}C_\kappa\right)^2\right\}-\tfrac{q}{T}\sum\nolimits_{\kappa=0}^{\infty}\sum\nolimits_{\eta=0}^{\infty}tr^2\left\{\left(C'_\kappa\Phi^{-1}C_\eta\right)\right\}}
$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

5.A.5.3. *Derivation of* $D_{33}$.

First combination. This is the way to derive the first combination of D33 $v-1-\lambda=s-1-\eta\neq v-1-\kappa=t-1-\zeta$ and $s=t\,,\kappa\neq\lambda$ which means that:

$$\boxed{\zeta+\lambda=\eta+\kappa\,,\kappa\neq\lambda}$$

$$-\,tr\left\{\frac{1}{T^2}E\sum_{t,s,v=1}^{T}\sum_{\zeta,\eta,\kappa,\lambda=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\left(\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}-\delta_{\kappa\lambda}I_n\right)C'_\lambda\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}$$

$$=-\frac{1}{T}\sum_{s}^{T}\sum_{\substack{\kappa+\eta=\lambda+\zeta\\ \kappa\neq\lambda}}tr\left\{E\left[\varepsilon_{2s}\varepsilon'_wC'_\zeta\Phi^{-1}C_\kappa\varepsilon_w\varepsilon'_uC'_\lambda\Phi^{-1}C_\eta\varepsilon_u\varepsilon'_{2s}\right]\right\}$$

$$=-\frac{q}{T}\sum_{\substack{\kappa+\eta=\lambda+\zeta\\ \kappa\neq\lambda}}tr\left\{C'_\zeta\Phi^{-1}C_\kappa\right\}tr\left\{C'_\lambda\Phi^{-1}C_\eta\right\}$$

So the total contribution of the first part of the D33 term is:

$$-\frac{q}{T}\sum_{\substack{\kappa+\eta=\lambda+\zeta\\ \kappa\neq\lambda}}tr\left\{C'_\zeta\Phi^{-1}C_\kappa\right\}tr\left\{C'_\lambda\Phi^{-1}C_\eta\right\}$$

Second combination. $v-1-\kappa=s-1-\eta\neq v-1-\lambda=t-1-\zeta$ and $s=t\,,\kappa\neq\lambda$ Stated alternatively:

$$\boxed{\zeta+\kappa=\eta+\lambda\,,\kappa\neq\lambda}$$

$$-\,tr\left\{\frac{1}{T^2}E\left[\sum_{t,s,v=1}^{T}\sum_{\zeta,\eta,\kappa,\lambda=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\left(\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}-\delta_{\kappa\lambda}I_n\right)C'_\lambda\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right]\right\}$$

(5.42)

$$=-\frac{q}{T}tr\left\{E\left[\sum_{\substack{\lambda+\eta=\kappa+\zeta\\ \kappa\neq\lambda}}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\right]\right\}$$

(5.43)

$$=-\frac{q}{T}\sum_{\substack{\lambda+\eta=\kappa+\zeta\\ \kappa\neq\lambda}}tr\left\{C'_\zeta\Phi^{-1}C_\kappa C'_\eta\Phi^{-1}C_\lambda\right\}$$

Total

The total summation of $D_{33}$ is equal to:

$$\boxed{\begin{aligned}D_{33}=&-\tfrac{q}{T}\textstyle\sum_{\substack{\kappa+\eta=\lambda+\zeta\\ \kappa\neq\lambda}}tr\left\{C'_\zeta\Phi^{-1}C_\kappa\right\}tr\left\{C'_\lambda\Phi^{-1}C_\eta\right\}\\&-\tfrac{q}{T}\textstyle\sum_{\substack{\lambda+\eta=\kappa+\zeta\\ \kappa\neq\lambda}}tr\left\{C'_\zeta\Phi^{-1}C_\kappa C'_\eta\Phi^{-1}C_\lambda\right\}\end{aligned}}$$

5.A.5.4. *Derivation of* $D_{34}$.


DOI: 10.2870/52609

First combination. $s = v - 1 - \kappa$ and $t = s - 1 - \eta$ and $v - 1 - \lambda = t - 1 - \zeta$
Note that $\kappa \neq \lambda$ and

$$\boxed{\lambda = \kappa + \zeta + \eta + 2}$$

$$- tr \left\{ \frac{1}{T^2} E \left[ \sum_{t,s,v=1}^{T} \sum_{\zeta,\eta,\kappa,\lambda=0}^{\infty} \varepsilon_{2t} \varepsilon'_{t-1-\zeta} C'_\zeta \Phi^{-1} C_\kappa \left( \varepsilon_{v-1-\kappa} \varepsilon'_{v-1-\lambda} - \delta_{\kappa\lambda} I_n \right) C'_\lambda \Phi^{-1} C_\eta \varepsilon_{s-1-\eta} \varepsilon'_{2s} \right] \right\}$$

$$= -tr \left\{ \frac{1}{T^2} E \left[ \sum_{s=1}^{T} \sum_{\zeta,\eta,\kappa=0}^{\infty} \varepsilon_{2t} \varepsilon'_{t-1-\zeta} C'_\zeta \Phi^{-1} C_\kappa \varepsilon_{v-1-\kappa} \varepsilon'_{v-1-\lambda} C'_\lambda \Phi^{-1} C_\eta \varepsilon_{s-1-\eta} \varepsilon'_{2s} \right] \right\}$$

$$= -tr \left\{ \frac{1}{T^2} E \left[ \sum_{s=1}^{T} \sum_{\zeta,\eta,\kappa=0}^{\infty} \varepsilon_{2s} \varepsilon'_{v-1-\kappa} C'_\kappa \Phi^{-1} C_\zeta \varepsilon_{t-1-\zeta} \varepsilon'_{v-1-\lambda} C'_{\kappa+\zeta+\eta+2} \Phi^{-1} C_\eta \varepsilon_{s-1-\eta} \varepsilon'_{2t} \right] \right\}$$

$$= -\frac{1}{T} tr \left\{ \left[ \sum_{\zeta,\eta,\kappa=0}^{\infty} C'_\kappa \Phi^{-1} C_\zeta C'_{\kappa+\zeta+\eta+2} \Phi^{-1} C_\eta \right]_{22} \right\}$$

The second combination. $s = v - 1 - \lambda$ and $t = s - 1 - \eta$ and $v - 1 - \kappa = t - 1 - \zeta$
Note that $\kappa \neq \lambda$ and

$$\boxed{\kappa = \lambda + \zeta + \eta + 2}$$

$$- tr \left\{ \frac{1}{T^2} E \left[ \sum_{t,s,v=1}^{T} \sum_{\zeta,\eta,\kappa,\lambda=0}^{\infty} \varepsilon_{2t} \varepsilon'_{t-1-\zeta} C'_\zeta \Phi^{-1} C_\kappa \left( \varepsilon_{v-1-\kappa} \varepsilon'_{v-1-\lambda} - \delta_{\kappa\lambda} I_n \right) C'_\lambda \Phi^{-1} C_\eta \varepsilon_{s-1-\eta} \varepsilon'_{2s} \right] \right\}$$

$$= -tr \left\{ \frac{1}{T^2} \sum_{s=1}^{T} \sum_{\zeta,\eta,\lambda=0}^{\infty} E \left[ \varepsilon'_{t-1-\zeta} C'_\zeta \Phi^{-1} C_{\lambda+\zeta+\eta+2} \varepsilon_{v-1-\kappa} \right] E \left[ \varepsilon_{2t} \varepsilon'_{s-1-\eta} \right] C'_\eta \Phi^{-1} C_\lambda E \left[ \varepsilon_{v-1-\lambda} \varepsilon'_{2s} \right] \right\}$$

$$= -\frac{1}{T} \sum_{\zeta,\eta,\lambda=0}^{\infty} tr \left\{ C'_\zeta \Phi^{-1} C_{\lambda+\zeta+\eta+2} \right\} tr \left\{ \left[ C'_\eta \Phi^{-1} C_\lambda \right]_{22} \right\}$$

The third combination. $t = v - 1 - \kappa$ and $s = t - 1 - \zeta$ and $v - 1 - \lambda = s - 1 - \eta$
Note that $\kappa \neq \lambda$ and

$$\boxed{\lambda = \kappa + \zeta + \eta + 2}$$

$$- tr \left\{ \frac{1}{T^2} E \left[ \sum_{t,s,v=1}^{T} \sum_{\zeta,\eta,\kappa,\lambda=0}^{\infty} \varepsilon_{2t} \varepsilon'_{t-1-\zeta} C'_\zeta \Phi^{-1} C_\kappa \left( \varepsilon_{v-1-\kappa} \varepsilon'_{v-1-\lambda} - \delta_{\kappa\lambda} I_n \right) C'_\lambda \Phi^{-1} C_\eta \varepsilon_{s-1-\eta} \varepsilon'_{2s} \right] \right\}$$

$$= -tr \left\{ \frac{1}{T^2} \sum_{t=1}^{T} \sum_{\zeta,\eta,\kappa=0}^{\infty} E \left[ \varepsilon_{2t} \varepsilon'_{v-1-\kappa} \right] C'_\kappa \Phi^{-1} C_\zeta E \left[ \varepsilon_{t-1-\zeta} \varepsilon'_{2s} \right] E \left[ \varepsilon'_{v-1-\lambda} C'_{\kappa+\zeta+\eta+2} \Phi^{-1} C_\eta \varepsilon_{s-1-\eta} \right] \right\}$$

$$= -\frac{1}{T} \sum_{\zeta,\eta,\kappa=0}^{\infty} tr \left\{ \left[ C'_\kappa \Phi^{-1} C_\zeta \right]_{22} \right\} tr \left\{ C'_{\kappa+\zeta+\eta+2} \Phi^{-1} C_\eta \right\}$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

The fourth combination. $t = v - 1 - \lambda$ and $s = t - 1 - \zeta$ and $v - 1 - \kappa = s - 1 - \eta$
Note that $\kappa \neq \lambda$ and $\kappa = \lambda + \zeta + \eta + 2$

$$- tr \left\{ \frac{1}{T^2} E \left[ \sum_{t,s,v=1}^{T} \sum_{\zeta,\eta,\kappa,\lambda=0}^{\infty} \varepsilon_{2t} \varepsilon'_{t-1-\zeta} C'_\zeta \Phi^{-1} C_\kappa \left( \varepsilon_{v-1-\kappa} \varepsilon'_{v-1-\lambda} - \delta_{\kappa\lambda} I_n \right) C'_\lambda \Phi^{-1} C_\eta \varepsilon_{s-1-\eta} \varepsilon'_{2s} \right] \right\}$$
$$= -tr \left\{ \frac{1}{T^2} \sum_{t=1}^{T} \sum_{\zeta,\eta,\lambda=0}^{\infty} E \left[ \varepsilon_{2t} \varepsilon'_{v-1-\lambda} \right] C'_\lambda \Phi^{-1} C_\eta E \left[ \varepsilon_{s-1-\eta} \varepsilon'_{v-1-\kappa} \right] C'_{\lambda+\zeta+\eta+2} \Phi^{-1} C_\zeta E \left[ \varepsilon_{t-1-\zeta} \varepsilon'_{2s} \right] \right\}$$
$$= -\frac{1}{T} \sum_{\zeta,\eta,\lambda=0}^{\infty} tr \left\{ \left[ C'_\lambda \Phi^{-1} C_\eta C'_{\lambda+\zeta+\eta+2} \Phi^{-1} C_\zeta \right]_{22} \right\}$$

Total. The total contribution of D34 is therefore:

$$\boxed{\begin{aligned} D_{34} = & -\tfrac{2}{T} \textstyle\sum_{\zeta,\eta,\kappa=0}^{\infty} tr \left\{ [C'_\kappa \Phi^{-1} C_\zeta]_{22} \right\} tr \left\{ C'_{\kappa+\zeta+\eta+2} \Phi^{-1} C_\eta \right\} \\ & -\tfrac{2}{T} \textstyle\sum_{\zeta,\eta,\lambda=0}^{\infty} tr \left\{ \left[ C'_\lambda \Phi^{-1} C_\eta C'_{\lambda+\zeta+\eta+2} \Phi^{-1} C_\zeta \right]_{22} \right\} \end{aligned}}$$

5.A.5.5. *Derivation of* $D_{35}$.
First combination. $t = v - 1 - \kappa$ and $s = v - 1 - \lambda$ and $t - 1 - \zeta = s - 1 - \eta$ and $\kappa \neq \lambda$

$$\boxed{\kappa + \zeta = \eta + \lambda}$$

$$- tr \left\{ \frac{1}{T^2} E \left[ \sum_{t,s,v=1}^{T} \sum_{\zeta,\eta,\kappa,\lambda=0}^{\infty} \varepsilon_{2t} \varepsilon'_{t-1-\zeta} C'_\zeta \Phi^{-1} C_\kappa \left( \varepsilon_{v-1-\kappa} \varepsilon'_{v-1-\lambda} - \delta_{\kappa\lambda} I_n \right) C'_\lambda \Phi^{-1} C_\eta \varepsilon_{s-1-\eta} \varepsilon'_{2s} \right] \right\}$$
$$= -tr \left\{ \frac{1}{T^2} \sum_{t=1}^{T} \sum_{\substack{\lambda+\eta=\kappa+\zeta \\ \kappa\neq\lambda}} E \left[ \varepsilon_{2t} \varepsilon'_{v-1-\kappa} \right] C'_\kappa \Phi^{-1} C_\zeta E \left[ \varepsilon_{t-1-\zeta} \varepsilon'_{s-1-\eta} \right] C'_\eta \Phi^{-1} C_\lambda E \left[ \varepsilon_{v-1-\lambda} \varepsilon'_{2s} \right] \right\}$$
$$- \frac{2}{T} \sum_{\kappa=0}^{\infty} \sum_{\lambda=\kappa+1}^{\infty} \sum_{\eta=0}^{\infty} tr \left\{ \left[ C'_\kappa \Phi^{-1} C_{\lambda+\eta-\kappa} C'_\eta \Phi^{-1} C_\lambda \right]_{22} \right\}$$
$$= -\frac{2}{T} \sum_{\kappa,\eta,\alpha=0}^{\infty} tr \left\{ \left[ C'_\kappa \Phi^{-1} C_{\alpha+\eta-1} C'_\eta \Phi^{-1} C_{\alpha+\kappa+1} \right]_{22} \right\}$$

Second combination. $t = v - 1 - \lambda$ and $s = v - 1 - \kappa$ and $t - 1 - \zeta = s - 1 - \eta$ and $\kappa \neq \lambda$

$$\boxed{\kappa + \eta = \zeta + \lambda}$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

$$- tr\left\{\frac{1}{T^2}E\left[\sum_{t,s,v=1}^{T}\sum_{\zeta,\eta,\kappa,\lambda=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\left(\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}-\delta_{\kappa\lambda}I_n\right)C'_\lambda\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right]\right\}$$

$$= -tr\left\{\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\substack{\kappa+\eta=\zeta+\lambda\\ \kappa\neq\lambda}}E\left[\varepsilon_{2t}\varepsilon'_{v-1-\lambda}\right]C'_\lambda\Phi^{-1}C_\eta E\left[\varepsilon_{s-1-\eta}\varepsilon'_{t-1-\zeta}\right]C'_\zeta\Phi^{-1}C_\kappa E\left[\varepsilon_{v-1-\kappa}\varepsilon'_{2s}\right]\right\}$$

$$= -\frac{2}{T}\sum_{\kappa=0}^{\infty}\sum_{\lambda=\kappa+1}^{\infty}\sum_{\zeta=0}^{\infty}tr\left\{\left[C'_\lambda\Phi^{-1}C_{\zeta+\lambda-\kappa}C'_\zeta\Phi^{-1}C_\kappa\right]_{22}\right\}$$

$$= -\frac{2}{T}\sum_{\kappa,\alpha,\zeta=0}^{\infty}tr\left\{\left[C'_{\alpha+\kappa+1}\Phi^{-1}C_{\zeta+\alpha+1}C'_\zeta\Phi^{-1}C_\kappa\right]_{22}\right\}$$

The total contribution of D35 is therefore:

$$D_{35} = -\tfrac{2}{T}\textstyle\sum_{\kappa,\eta,\alpha=0}^{\infty} tr\left\{\left[C'_\kappa\Phi^{-1}C_{\alpha+\eta-1}C'_\eta\Phi^{-1}C_{\alpha+\kappa+1}\right]_{22}\right\} -\tfrac{2}{T}\textstyle\sum_{\kappa,\alpha,\zeta=0}^{\infty} tr\left\{\left[C'_{\alpha+\kappa+1}\Phi^{-1}C_{\zeta+\alpha+1}C'_\zeta\Phi^{-1}C_\kappa\right]_{22}\right\}$$

### 5.A.6. Derivation of $D_4$.

$$tr\left\{E\left(I_1-\frac{1}{T}U'U\right)\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\Phi-\frac{1}{T}X'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\}$$

$$= tr\left\{\frac{1}{T^3}E\sum_{t,s,r,\nu=1}^{T}\sum_{\zeta,\eta,\kappa,\lambda=0}^{\infty}\left(\varepsilon_{2r}\varepsilon'_{2r}-I_q\right)\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\left(\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}-\delta_{\kappa\lambda}I_n\right)C'_\lambda\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}$$

I only find one combination in this case:
$t-1-\zeta = s-1-\eta$ and $v-1-\lambda = v-1-\kappa = r$ which implies that $\lambda=\kappa$.

Take the four separate terms one by one, starting with the case in which we take both identity matrices:

$$tr\left\{\frac{1}{T^3}E\sum_{t,\nu=1}^{T}\sum_{\zeta,\lambda=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\lambda C'_\lambda\Phi^{-1}C_\zeta\varepsilon_{s-1-\zeta}\varepsilon'_{2t}\right\}$$

$$= \frac{1}{T}qn$$

Then

$$- tr\left\{\frac{1}{T^3}E\sum_{t,\nu=1}^{T}\sum_{\zeta,\lambda=0}^{\infty}\varepsilon_{2r}\varepsilon'_{2r}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\lambda C'_\lambda\Phi^{-1}C_\zeta\varepsilon_{s-1-\zeta}\varepsilon'_{2t}\right\}$$

$$= -\frac{1}{T}qn$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

$$
\begin{aligned}
&- tr\left\{\frac{1}{T^3}E\sum_{t,r=1}^{T}\sum_{\zeta,\lambda=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\lambda\varepsilon_r\varepsilon'_rC'_\lambda\Phi^{-1}C_\zeta\varepsilon_{s-1-\zeta}\varepsilon'_{2t}\right\}\\
&= -\frac{1}{T}qn
\end{aligned}
$$

and the most complicated one:

$$
\begin{aligned}
& tr\left\{\frac{1}{T^3}E\sum_{t,r=1}^{T}\sum_{\zeta,\lambda=0}^{\infty}\varepsilon_{2r}\varepsilon'_{2r}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_{v-1-\lambda}\varepsilon_r\varepsilon'_rC'_{v-1-\lambda}\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2t}\right\}\\
&= tr\left\{\frac{1}{T^3}E\sum_{t,r=1}^{T}\sum_{\zeta,\lambda=0}^{\infty}\left[\varepsilon_{2t}\varepsilon'_{2t}\right]\varepsilon_{2r}\varepsilon'_rC'_{v-1-\lambda}\Phi^{-1}C_\zeta\varepsilon_{t-1-\zeta}\varepsilon'_{s-1-\zeta}C'_\eta\Phi^{-1}C'_{v-1-\lambda}\varepsilon_r\varepsilon'_{2r}\right\}\\
&= \frac{2}{T}tr\left\{\sum_{\zeta=0}^{\infty}\left[C'_\zeta\Phi^{-1}C_\zeta\right]_{22}\right\}+\frac{1}{T}qn
\end{aligned}
$$

The total expression then becomes equal to:

$$
\boxed{D_4 = \tfrac{2}{T}tr\left\{\textstyle\sum_{\zeta=0}^{\infty}\left[C'_\zeta\Phi^{-1}C_\zeta\right]_{22}\right\}}
$$

### 5.A.7. Derivation of $D_5$.

$$
\begin{aligned}
& tr\left\{E\left(I_q-\frac{1}{T}U'U\right)^2\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\}\\
&= tr\left\{\frac{1}{T^3}E\sum_{t,s,y,r=1}^{T}\sum_{\zeta,\eta=0}^{\infty}\left(\varepsilon_{2y}\varepsilon'_{2y}-I_q\right)\left(\varepsilon_{2r}\varepsilon'_{2r}-I_q\right)\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C'_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}
\end{aligned}
$$

Here I find only one combination:
$y=r, s-1-\eta=t-1-\zeta, t=s$

$$
\begin{aligned}
& tr\left\{\frac{1}{T^3}E\sum_{t,r=1}^{T}\sum_{\zeta=0}^{\infty}\left(\varepsilon_{2r}\varepsilon'_{2r}-I_q\right)\left(\varepsilon_{2r}\varepsilon'_{2r}-I_q\right)\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C'\varepsilon_{t-1-\zeta}\varepsilon'_{2t}\right\}\\
&= \frac{1}{T}tr\left\{I_n\right\}tr\left\{E\left[\left(\varepsilon_{2r}\varepsilon'_{2r}-I_q\right)\left(\varepsilon_{2r}\varepsilon'_{2r}-I_q\right)\right]\right\}\\
&= \frac{1}{T}\left(nq+nq^2\right)
\end{aligned}
$$

Therefore:

$$
\boxed{D_5 = \tfrac{1}{T}\left(nq+nq^2\right)}
$$


DOI: 10.2870/52609

**5.A.8. Derivation of $D_6$.**

$$tr\left\{E\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\Phi-\frac{1}{T}X'X\right)\Phi^{-1}\left(\Phi-\frac{1}{T}X'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\}$$

$$=\frac{1}{T^3}Etr\left\{\sum_{t,s,v,w=1}^{T}\sum_{\zeta,\eta,\kappa,\lambda,\alpha,\beta=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_{v-1-\kappa}\left(\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}-\delta_{\kappa\lambda}I_n\right)C'_{v-1-\lambda}\Phi^{-1}\right.$$
$$\left.\times C_\alpha\left(\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}-\delta_{\alpha\beta}I_n\right)C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_s\right\}$$

The combinations, which give non-zero expectations of order $\frac{1}{T}$can be logically subdivided in three groups:

(1) The $\left(\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}-\delta_{\kappa\lambda}I_n\right)$ and $\left(\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}-\delta_{\alpha\beta}I_n\right)$ all coincide. At the same time $s=t$ and $t-1-\zeta=s-1-\eta$.
(2) One of $\varepsilon_{v-1-\kappa}$ and $\varepsilon'_{v-1-\lambda}$ coincides with one of $\varepsilon_{w-1-\alpha}$ /$\varepsilon'_{w-1-\beta}$. The two remaining ones then also coincide. Obviously we have two different combinations and $s=t$ and $t-1-\zeta=s-1-\eta$.
(3) $\varepsilon'_s$ coincides with one of $\varepsilon_{v-1-\kappa}$ and $\varepsilon'_{v-1-\lambda}$. $\varepsilon_{s-1-\eta}$ then coincides with the other. Similarly $\varepsilon_t$ and $\varepsilon'_{t-1-\zeta}$ each coincide with one of $\varepsilon_{w-1-\alpha}$ /$\varepsilon'_{w-1-\beta}$. Note that there are eight of such combinations, which are listed one by one below in the derivation of C63

Each of these possibilities shall now be dealt with in turn.

5.A.8.1. *Derivation of $D_{61}$.*

$$\frac{1}{T^3}Etr\left\{\sum_{t,s,v,w=1}^{T}\sum_{\zeta,\eta,\kappa,\lambda,\alpha,\beta=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\left(\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}-\delta_{\kappa\lambda}I_n\right)C'_\lambda\Phi^{-1}\right.$$
$$\left.\times C_\alpha\left(\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}-\delta_{\alpha\beta}I_n\right)C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}$$

For the first combination, we have $t-1-\zeta=s-1-\eta, s=t, v-1-\kappa=v-1-\lambda=w-1-\alpha=w-1-\beta=y$
which can be rephrased as:

$$\boxed{s=t,\zeta=\eta,\alpha=\beta,\kappa=\lambda}$$

$w$ can also vary.
For simplicity we shall just use the $\sum$ for now.

$$\sum tr\left\{\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\left(\varepsilon_y\varepsilon'_y-I_n\right)C'_\kappa\Phi^{-1}\right.$$
$$\left.\times C_\alpha\left(\varepsilon_y\varepsilon'_y-I_n\right)C'_\alpha\Phi^{-1}C_\zeta\varepsilon_{t-1-\zeta}\varepsilon'_{2t}\right\}$$
$$=\frac{q}{T}tr\left\{\sum_{\alpha,\kappa}C'_\alpha\Phi^{-1}C_\kappa\left(\varepsilon_y\varepsilon'_y-I_n\right)C'_\kappa\Phi^{-1}C_\alpha\left(\varepsilon_y\varepsilon'_y-I_n\right)\right\}$$
$$=\frac{q}{T}\sum_{\alpha,\kappa}tr^2\left\{C'_\alpha\Phi^{-1}C_\kappa\right\}+\frac{q}{T}tr\sum_{\alpha,\kappa}\left\{\left(C'_\alpha\Phi^{-1}C_\kappa\right)^2\right\}$$

So we have that the total is equal to:

$$\boxed{D_{61}=\frac{q}{T}\sum_{\alpha,\kappa}tr^2\left\{C'_\alpha\Phi^{-1}C_\kappa\right\}+\frac{q}{T}tr\sum_{\alpha,\kappa}\left\{\left(C'_\alpha\Phi^{-1}C_\kappa\right)^2\right\}}$$

 DOI: 10.2870/52609

5.A.8.2. *Derivation of* $D_{62}$.

$$\frac{1}{T^3}Etr\left\{\sum_{t,s,v,w=1}^{T}\sum_{\zeta,\eta,\kappa,\lambda,\alpha,\beta=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_{\zeta}\Phi^{-1}C_{\kappa}\left(\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}-\delta_{\kappa\lambda}I_n\right)C'_{\lambda}\Phi^{-1}\right.$$
$$\left.\times C_{\alpha}\left(\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}-\delta_{\alpha\beta}I_n\right)C'_{\beta}\Phi^{-1}C_{\eta}\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}$$

There are two combinations:

(1) $t=s,\eta=\zeta,$
$v-1-\kappa=w-1-\alpha$
$v-1-\lambda=w-1-\beta$
$\kappa\neq\lambda$

(2) $t=s,\eta=\zeta,$
$v-1-\lambda=w-1-\alpha$
$v-1-\kappa=w-1-\beta$
$\kappa\neq\lambda$

First combination. This combination implies that:

$$\boxed{\beta+\kappa=\alpha+\lambda,\kappa\neq\lambda}$$

$$\frac{1}{T^3}Etr\left\{\sum_{t,v,w=1}^{T}\sum_{\zeta=0}^{\infty}\sum_{\substack{\beta+\kappa=\alpha+\lambda\\ \kappa\neq\lambda}}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_{\zeta}\Phi^{-1}C_{\kappa}\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}C'_{\lambda}\Phi^{-1}\right.$$
$$\left.\times C_{\alpha}\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}C'_{\alpha+\lambda-\kappa}\Phi^{-1}C_{\zeta}\varepsilon_{t-1-\zeta}\varepsilon'_{2t}\right\}$$
$$=\frac{q}{T}Etr\left\{\sum_{\substack{\beta+\kappa=\alpha+\lambda\\ \kappa\neq\lambda}}\Phi^{-1}C_{\kappa}\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}C'_{\lambda}\Phi^{-1}C_{\alpha}\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}C'_{\beta}\right\}$$
$$=\frac{q}{T}tr\left\{\sum_{\substack{\beta+\kappa=\alpha+\lambda\\ \kappa\neq\lambda}}C'_{\beta}\Phi^{-1}C_{\kappa}C'_{\alpha}\Phi^{-1}C_{\lambda}\right\}$$

Note that this expression is exactly the opposite of expression D33. So we conclude that the expectation of this combination is equal to:

$$\frac{q}{T}tr\left\{\sum_{\substack{\beta+\kappa=\alpha+\lambda\\ \kappa\neq\lambda}}C'_{\beta}\Phi^{-1}C_{\kappa}C'_{\alpha}\Phi^{-1}C_{\lambda}\right\}$$

Second combination. Combining the conditions, we obtain:

$$\boxed{\beta+\lambda=\alpha+\kappa,\kappa\neq\lambda}$$


DOI: 10.2870/52609

$$tr\left\{E\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\Phi-\frac{1}{T}X'X\right)\Phi^{-1}\left(\Phi-\frac{1}{T}X'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\}$$

$$=\frac{1}{T^3}Etr\left\{\sum_{t,v,w=1}^{T}\sum_{\zeta,\kappa,\lambda,\alpha,\beta=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}\right.$$

$$\left.\times C_\alpha\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}C'_{\alpha-\lambda+\kappa}\Phi^{-1}C_\zeta\varepsilon_{t-1-\zeta}\varepsilon'_{2t}\right\}$$

$$=\frac{q}{T}\sum_{\substack{\beta+\lambda=\alpha+\kappa\\ \kappa\neq\lambda}}tr\left\{C'_\beta\Phi^{-1}C_\kappa\right\}tr\left\{C'_\lambda\Phi^{-1}C_\alpha\right\}$$

Total. The total expectation of this term is therefore equal to:

$$\boxed{\begin{aligned}D_{62}=&\tfrac{q}{T}tr\textstyle\sum_{\substack{\kappa,\lambda,\alpha=0\\ \kappa\neq\lambda}}^{\infty}\left\{C'_{\alpha+\lambda-\kappa}\Phi^{-1}C_\kappa C'_\alpha\Phi^{-1}C_\lambda\right\}\\ &+\tfrac{q}{T}\textstyle\sum_{\substack{\beta+\lambda=\alpha+\kappa\\ \kappa\neq\lambda}}tr\left\{C'_\beta\Phi^{-1}C_\kappa\right\}tr\left\{C'_\lambda\Phi^{-1}C_\alpha\right\}\end{aligned}}$$

5.A.8.3. *Derivation of $D_{63}$.*

$$tr\left\{E\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\Phi-\frac{1}{T}X'X\right)\Phi^{-1}\left(\Phi-\frac{1}{T}X'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\}$$

$$=\frac{1}{T^3}Etr\left\{\sum_{t,s,v,w=1}^{T}\sum_{\zeta,\eta,\kappa,\lambda,\alpha,\beta=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\left(\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}-\delta_{\kappa\lambda}I_n\right)C'_\lambda\Phi^{-1}\right.$$

$$\left.\times C_\alpha\left(\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}-\delta_{\alpha\beta}I_n\right)C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}$$

There are eight possible constellations, which give rise to first order terms:

| | | | | |
|---|---|---|---|---|
| 1. | $s=w-1-\beta$ $\quad s-1-\eta=w-1-\alpha$ | $t=v-1-\kappa$ | $t-1-\zeta=v-1-\lambda$ |
| 2. | $s=w-1-\alpha$ $\quad s-1-\eta=w-1-\beta$ | $t=v-1-\kappa$ | $t-1-\zeta=v-1-\lambda$ |
| 3 | $s=w-1-\beta$ $\quad s-1-\eta=w-1-\alpha$ | $t=v-1-\lambda$ | $t-1-\zeta=v-1-\kappa$ |
| 4 | $s=w-1-\alpha$ $\quad s-1-\eta=w-1-\beta$ | $t=v-1-\lambda$ | $t-1-\zeta=v-1-\kappa$ |
| 5 | $s=v-1-\kappa$ $\quad s-1-\eta=v-1-\lambda$ | $t=w-1-\beta$ | $t-1-\zeta=w-1-\alpha$ |
| 6 | $s=v-1-\lambda$ $\quad s-1-\eta=v-1-\kappa$ | $t=w-1-\beta$ | $t-1-\zeta=w-1-\alpha$ |
| 7 | $s=v-1-\kappa$ $\quad s-1-\eta=v-1-\lambda$ | $t=w-1-\alpha$ | $t-1-\zeta=w-1-\beta$ |
| 8 | $s=v-1-\lambda$ $\quad s-1-\eta=v-1-\kappa$ | $t=w-1-\alpha$ | $t-1-\zeta=w-1-\beta$ |

In all of these eight constellations we have that $\kappa\neq\lambda,\alpha\neq\beta$. We shall now take them one by one:

First combination. This combination implies that:

$$\boxed{\alpha=\beta+\eta+1\text{ and }\lambda=\kappa+\zeta+1}$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

$$\frac{1}{T^3}Etr\left\{\sum \varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}\right.$$
$$\left.\times C_\alpha\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}$$
$$=\frac{1}{T^3}Etr\sum\left\{\varepsilon_{2t}\varepsilon'_{v-1-\kappa}C'_\kappa\Phi^{-1}C_\zeta\varepsilon_{t-1-\zeta}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}C_\alpha\varepsilon_{w-1-\alpha}\right.$$
$$\left.\times\varepsilon'_{s-1-\eta}C'_\eta\Phi^{-1}C_\beta\varepsilon_{w-1-\beta}\varepsilon'_{2s}\right\}$$
$$=\frac{1}{T}tr\left\{\sum_{\beta,\eta,\kappa,\zeta}\left[C'_\kappa\Phi^{-1}C_\zeta C'_{\kappa+\zeta+1}\Phi^{-1}C_{\beta+\eta+1}C'_\eta\Phi^{-1}C_\beta\right]_{22}\right\}$$

Second combination.

$$\boxed{\beta=\alpha+\eta+1 \text{ and } \lambda=\kappa+\zeta+1}$$

$$\frac{1}{T^3}Etr\left\{\sum \varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}\right.$$
$$\left.\times C_\alpha\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}$$
$$=\frac{1}{T^3}E\sum tr\left\{\varepsilon_{2t}\varepsilon'_{v-1-\kappa}C'_\kappa\Phi^{-1}C_\zeta\varepsilon_{t-1-\zeta}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}C_\alpha\varepsilon_{w-1-\alpha}\varepsilon'_{2s}\right\}\times$$
$$tr\left\{\varepsilon'_{w-1-\beta}C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\right\}$$
$$=\frac{1}{T}\sum_{\alpha,\eta,\kappa,\zeta=0}^{\infty}tr\left\{\left[C'_\kappa\Phi^{-1}C_\zeta C'_{\kappa+\zeta+1}\Phi^{-1}C_\alpha\right]_{22}\right\}tr\left\{C'_{\alpha+\eta+1}\Phi^{-1}C_\eta\right\}$$

Third combination.

$$\boxed{\alpha=\beta+\eta+1 \text{ and } \kappa=\lambda+\zeta+1}$$

$$\frac{1}{T^3}Etr\left\{\sum \varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}\right.$$
$$\left.\times C_\alpha\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}$$
$$=\frac{1}{T^3}E\sum tr\left\{\varepsilon_{2t}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}C_\alpha\varepsilon_{w-1-\alpha}\varepsilon'_{s-1-\eta}C'_\eta\Phi^{-1}C_\beta\varepsilon'_{w-1-\beta}\varepsilon'_{2s}\right\}\times$$
$$tr\left\{\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\right\}$$
$$=\frac{1}{T}\sum_{\beta,\eta,\lambda,\zeta=0}^{\infty}tr\left\{\left[C'_\lambda\Phi^{-1}C_{\beta+\eta+1}C'_\eta\Phi^{-1}C_\beta\right]_{22}\right\}tr\left\{C'_\zeta\Phi^{-1}C_{\lambda+\zeta+1}\right\}$$

which incidentally is equal to the second combination

Fourth combination.

$$\boxed{\beta=\alpha+\eta+1 \text{ and } \kappa=\lambda+\zeta+1}$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

$$\frac{1}{T^3}Etr\left\{\sum \varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}\right.$$
$$\left.\times C_\alpha\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}$$
$$=\frac{1}{T^3}E\sum tr\left\{\varepsilon_{2t}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}C_\alpha\varepsilon_{w-1-\alpha}\varepsilon'_{2s}\right\}$$
$$\times\, tr\left\{\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\right\} tr\left\{\varepsilon'_{w-1-\beta}C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\right\}$$
$$=\frac{1}{T}\sum_{\alpha,\eta,\lambda,\zeta=0}^{\infty} tr\left\{\left[C'_\lambda\Phi^{-1}C_\alpha\right]_{22}\right\} tr\left\{C'_\zeta\Phi^{-1}C_{\lambda+\zeta+1}\right\} tr\left\{C'_{\alpha+\eta+1}\Phi^{-1}C_\eta\right\}$$

Fifth combination.

$$\boxed{\lambda=\kappa+\eta+1 \text{ and } \alpha=\beta+\zeta+1}$$

$$\frac{1}{T^3}Etr\left\{\sum \varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}\right.$$
$$\left.\times C_\alpha\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}$$
$$=\frac{1}{T^3}Etr\left\{\sum \varepsilon_{2t}\varepsilon'_{w-1-\beta}C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}C_\alpha\varepsilon_{w-1-\alpha}\right.$$
$$\left.\times\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\varepsilon'_{2s}\right\}$$
$$=\sum_{\beta,\eta,\kappa,\zeta=0}^{\infty} tr\left\{\left[C'_\beta\Phi^{-1}C_\eta C'_{\kappa+\eta+1}\Phi^{-1}C_{\beta+\zeta+1}C'_\zeta\Phi^{-1}C_\kappa\right]_{22}\right\}$$

Sixth combination.

$$\boxed{\kappa=\lambda+\eta+1 \text{ and } \alpha=\beta+\zeta+1}$$

$$\frac{1}{T^3}Etr\left\{\sum \varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}\right.$$
$$\left.\times C_\alpha\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}$$
$$=\frac{1}{T^3}E\sum tr\left\{\varepsilon_{2t}\varepsilon'_{w-1-\beta}C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{v-1-\kappa}C'_\kappa\Phi^{-1}C_\zeta\varepsilon_{t-1-\zeta}\varepsilon'_{w-1-\alpha}C'_\alpha\Phi^{-1}C_\lambda\varepsilon_{v-1-\lambda}\varepsilon'_{2s}\right\}$$
$$tr\left\{\varepsilon'_{w-1-\beta}C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\right\}$$
$$=\sum_{\beta,\eta,\lambda,\zeta=0}^{\infty} tr\left\{\left[C'_\beta\Phi^{-1}C_\eta C'_{\lambda+\eta+1}\Phi^{-1}C_\zeta C'_{\beta+\zeta+1}\Phi^{-1}C_\lambda\right]_{22}\right\} tr\left\{C'_\beta\Phi^{-1}C_\eta\right\}$$

Seventh combination.

$$\boxed{\lambda=\kappa+\eta+1 \text{ and } \beta=\alpha+\zeta+1}$$

$$\frac{1}{T^3}Etr\left\{\sum \varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}\right.$$
$$\left.\times C_\alpha\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}$$
$$=\frac{1}{T^3}E\sum tr\left\{\varepsilon_{2t}\varepsilon'_{w-1-\alpha}C'_\alpha\Phi^{-1}C_\lambda\varepsilon_{v-1-\lambda}\varepsilon'_{s-1-\eta}C'_\eta\Phi^{-1}C_\beta\varepsilon_{w-1-\beta}\right.$$
$$\left.\times\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\varepsilon'_{2s}\right\}$$
$$=\frac{1}{T}\sum_{\alpha,\eta,\kappa,\zeta=0}^{\infty} tr\left\{\left[C'_\alpha\Phi^{-1}C_{\kappa+\eta+1}C'_\eta\Phi^{-1}C_{\alpha+\zeta+1}C'_\zeta\Phi^{-1}C_\kappa\right]_{22}\right\}$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

Eighth combination.

$$\boxed{\kappa = \lambda + \eta + 1 \text{ and } \beta = \alpha + \zeta + 1}$$

$$\frac{1}{T^3} Etr\left\{\sum \varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\varepsilon'_{v-1-\lambda}C'_\lambda\Phi^{-1}\right.$$
$$\left.\times C_\alpha\varepsilon_{w-1-\alpha}\varepsilon'_{w-1-\beta}C'_\beta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\right\}$$
$$= \frac{1}{T^3}E\sum tr\left\{\varepsilon_{2t}\varepsilon'_{w-1-\alpha}C'_\alpha\Phi^{-1}C_\lambda\varepsilon_{v-1-\lambda}\varepsilon'_{2s}\right\}$$
$$tr\left\{\varepsilon_{w-1-\beta}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\kappa\varepsilon_{v-1-\kappa}\varepsilon'_{s-1-\eta}C'_\eta\Phi^{-1}C_\beta\right\}$$
$$= \frac{1}{T}\sum_{\alpha,\eta,\lambda,\zeta=0}^{\infty} tr\left\{\left[C'_\alpha\Phi^{-1}C_\lambda\right]_{22}\right\} tr\left\{C'_\zeta\Phi^{-1}C_{\lambda+\eta+1}C'_\eta\Phi^{-1}C_{\alpha+\zeta+1}\right\}$$

which is seen to equal the sixth combination

Total. The total of the $D_{63}$ term then becomes:

$$\boxed{\begin{aligned}
D_{63} &= \tfrac{1}{T}tr\left\{\textstyle\sum_{\beta,\eta,\kappa,\zeta}\left[C'_\kappa\Phi^{-1}C_\zeta C'_{\kappa+\zeta+1}\Phi^{-1}C_{\beta+\eta+1}C'_\eta\Phi^{-1}C_\beta\right]_{22}\right\}\\
&+\tfrac{2}{T}\textstyle\sum_{\alpha,\eta,\kappa,\zeta=0}^{\infty} tr\left\{\left[C'_\kappa\Phi^{-1}C_\zeta C'_{\kappa+\zeta+1}\Phi^{-1}C_\alpha\right]_{22}\right\} tr\left\{C'_{\alpha+\eta+1}\Phi^{-1}C_\eta\right\}\\
&+\tfrac{1}{T}\textstyle\sum_{\alpha,\eta,\lambda,\zeta=0}^{\infty} tr\left\{[C'_\lambda\Phi^{-1}C_\alpha]_{22}\right\} tr\left\{C'_\zeta\Phi^{-1}C_{\lambda+\zeta+1}\right\} tr\left\{C'_{\alpha+\eta+1}\Phi^{-1}C_\eta\right\}\\
&+\tfrac{1}{T}\textstyle\sum_{\beta,\eta,\kappa,\zeta=0}^{\infty} tr\left\{\left[C'_\beta\Phi^{-1}C_\eta C'_{\kappa+\eta+1}\Phi^{-1}C_{\beta+\zeta+1}C'_\zeta\Phi^{-1}C_\kappa\right]_{22}\right\}\\
&+\tfrac{2}{T}\textstyle\sum_{\beta,\eta,\lambda,\zeta=0}^{\infty} tr\left\{\left[C'_\beta\Phi^{-1}C_\eta C'_{\lambda+\eta+1}\Phi^{-1}C_\zeta C'_{\beta+\zeta+1}\Phi^{-1}C_\lambda\right]_{22}\right\}\\
&+\tfrac{1}{T}\textstyle\sum_{\alpha,\eta,\lambda,\zeta=0}^{\infty} tr\left\{[C'_\alpha\Phi^{-1}C_\lambda]_{22}\right\} tr\left\{C'_\zeta\Phi^{-1}C_{\lambda+\eta+1}C'_\eta\Phi^{-1}C_{\alpha+\zeta+1}\right\}
\end{aligned}}$$

**5.A.9. The second term:** $E\left[\frac{1}{2T}tr(K^2)\right]$**.** This term is already of the order $\frac{1}{T}$, so we just have to take the nullth order expansion of $F = E\left[tr(K^2)\right]$

$$tr\left\{E\left[K^2\right]\right\}$$
$$= tr\left\{E\left[\left(\frac{1}{T}U'U\right)^{-1}\left(\frac{1}{\sqrt{T}}U'X\right)\left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\left(\frac{1}{T}U'U\right)^{-1}\right.\right.$$
$$\left.\left.\times\left(\frac{1}{\sqrt{T}}U'X\right)\left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right]\right\}$$
$$\overset{0}{=} tr\left\{E\left[\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right]\right\}$$
$$= tr\left\{E\left[\frac{1}{T}\sum_{t,s,m,n=1}^{T}\sum_{\zeta,\eta,\gamma,\chi=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_\zeta\Phi^{-1}C_\eta\varepsilon_{s-1-\eta}\varepsilon'_{2s}\varepsilon_{2m}\varepsilon'_{m-1-\gamma}C'_\gamma\Phi^{-1}C_\chi\varepsilon_{n-1-\chi}\varepsilon'_{2n}\right]\right\}$$

There are three possible combinations:

(1) $t = s, \zeta = \eta, m = n, \gamma = \chi$
(2) $t = m, \zeta = \gamma, s = n, \eta = \chi$
(3) $t = n, \zeta = \chi, s = m, \gamma = \eta$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

5.A.9.1. *First combination.*

$$tr\left\{E\left[\frac{1}{T^2}\sum_{t,s,m,n=1}^{T}\sum_{\zeta,\eta,\gamma,\chi=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_{\zeta}\Phi^{-1}C_{\eta}\varepsilon_{s-1-\eta}\varepsilon'_{2s}\varepsilon_{2m}\varepsilon'_{m-1-\gamma}C'_{\gamma}\Phi^{-1}C_{\chi}\varepsilon_{n-1-\chi}\varepsilon'_{2n}\right]\right\}$$
$$=\frac{1}{T^2}E\sum_{t,m=1}^{T}\sum_{\zeta,\gamma=0}^{\infty}tr\left\{[\varepsilon_{2t}\varepsilon'_{2s}\varepsilon_{2m}\varepsilon'_{2n}]\right\}tr\left\{\varepsilon'_{t-1-\zeta}C'_{\zeta}\Phi^{-1}C_{\zeta}\varepsilon_{s-1-\eta}\right\}tr\left\{\varepsilon'_{m-1-\gamma}C'_{\gamma}\Phi^{-1}C_{\gamma}\varepsilon_{n-1-\chi}\right\}$$
$$=qn^2$$

5.A.9.2. *Second combination.*

$$tr\left\{E\left[\frac{1}{T^2}\sum_{t,s,m,n=1}^{T}\sum_{\zeta,\eta,\gamma,\chi=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_{\zeta}\Phi^{-1}C_{\eta}\varepsilon_{s-1-\eta}\varepsilon'_{2s}\varepsilon_{2m}\varepsilon'_{m-1-\gamma}C'_{\gamma}\Phi^{-1}C_{\chi}\varepsilon_{n-1-\chi}\varepsilon'_{2n}\right]\right\}$$
$$=\frac{1}{T^2}E\sum_{t,s=1}^{T}\sum_{\zeta,\eta=0}^{\infty}tr\left\{\varepsilon_{2t}\varepsilon'_{2m}\varepsilon_{2s}\varepsilon'_{2n}\right\}tr\left\{\varepsilon_{n-1-\gamma}\varepsilon'_{t-1-\zeta}C'_{\zeta}\Phi^{-1}C_{\eta}\varepsilon_{s-1-\eta}\varepsilon'_{m-1-\chi}C'_{\eta}\Phi^{-1}C_{\zeta}\right\}$$
$$=qn$$

5.A.9.3. *Third combination.*

$$tr\left\{E\left[\frac{1}{T^2}\sum_{t,s,m,n=1}^{T}\sum_{\zeta,\eta,\gamma,\chi=0}^{\infty}\varepsilon_{2t}\varepsilon'_{t-1-\zeta}C'_{\zeta}\Phi^{-1}C_{\eta}\varepsilon_{s-1-\eta}\varepsilon'_{2s}\varepsilon_{2m}\varepsilon'_{m-1-\gamma}C'_{\gamma}\Phi^{-1}C_{\chi}\varepsilon_{n-1-\chi}\varepsilon'_{2n}\right]\right\}$$
$$=\frac{1}{T^2}E\sum_{t,s=1}^{T}\sum_{\zeta,\eta=0}^{\infty}tr\left\{\varepsilon'_{2n}\varepsilon_{2t}\varepsilon'_{2s}\varepsilon_{2m}\right\}tr\left\{C_{\zeta}\varepsilon_{n-1-\chi}\varepsilon'_{t-1-\zeta}C'_{\zeta}\Phi^{-1}C_{\eta}\varepsilon_{s-1-\eta}\varepsilon'_{m-1-\gamma}C'_{\eta}\Phi^{-1}\right\}$$
$$=q^2n$$

So the total contribution of this term is:

$$\boxed{\frac{1}{2T}F=\frac{qn}{2T}+\frac{q^2n+qn^2}{2T}}$$

**5.A.10. Total.** Adding up all the terms, we find:

$$
\begin{aligned}
(5.44)\quad E[LR] = {} & qn + \frac{1}{2T}\left(-4q + qn + q^2 n + qn^2\right) \\
& + \frac{1}{T} tr\left\{\sum_{\beta,\eta,\kappa,\zeta=0}^{\infty}\left[C'_\kappa \Phi^{-1} C_\zeta C'_{\kappa+\zeta+1} \Phi^{-1} C_{\beta+\eta+1} C'_\eta \Phi^{-1} C_\beta\right]_{22}\right\} \\
& + \frac{2}{T}\sum_{\alpha,\eta,\kappa,\zeta=0}^{\infty} tr\left\{\left[C'_\kappa \Phi^{-1} C_\zeta C'_{\kappa+\zeta+1} \Phi^{-1} C_\alpha\right]_{22}\right\} tr\left\{C'_{\alpha+\eta+1} \Phi^{-1} C_\eta\right\} \\
& + \frac{1}{T}\sum_{\alpha,\eta,\lambda,\zeta=0}^{\infty} tr\left\{\left[C'_\lambda \Phi^{-1} C_\alpha\right]_{22}\right\} tr\left\{C'_\zeta \Phi^{-1} C_{\lambda+\zeta+1}\right\} tr\left\{C'_{\alpha+\eta+1} \Phi^{-1} C_\eta\right\} \\
& + \frac{1}{T}\sum_{\beta,\eta,\kappa,\zeta=0}^{\infty} tr\left\{\left[C'_\beta \Phi^{-1} C_\eta C'_{\kappa+\eta+1} \Phi^{-1} C_{\beta+\zeta+1} C'_\zeta \Phi^{-1} C_\kappa\right]_{22}\right\} \\
& + \frac{2}{T}\sum_{\beta,\eta,\lambda,\zeta=0}^{\infty} tr\left\{\left[C'_\beta \Phi^{-1} C_\eta C'_{\lambda+\eta+1} \Phi^{-1} C_\zeta C'_{\beta+\zeta+1} \Phi^{-1} C_\lambda\right]_{22}\right\} \\
& + \frac{1}{T}\sum_{\alpha,\eta,\lambda,\zeta=0}^{\infty} tr\left\{\left[C'_\alpha \Phi^{-1} C_\lambda\right]_{22}\right\} tr\left\{C'_\zeta \Phi^{-1} C_{\lambda+\eta+1} C'_\eta \Phi^{-1} C_{\alpha+\zeta+1}\right\} \\
& - \frac{2}{T}\sum_{\varsigma,\eta,\kappa=0}^{\infty} tr\left\{\left[C'_\kappa \Phi^{-1} C_\varsigma\right]_{22}\right\} tr\left\{C'_{\kappa+\zeta+\eta+2} \Phi^{-1} C_\eta\right\} \\
& - \frac{2}{T}\sum_{\varsigma,\eta,\lambda=0}^{\infty} tr\left\{\left[C'_\lambda \Phi^{-1} C_\eta C'_{\lambda+\zeta+\eta+2} \Phi^{-1} C_\varsigma\right]_{22}\right\} \\
& - \frac{2}{T}\sum_{\kappa,\eta,\alpha=0}^{\infty} tr\left\{\left[C'_\kappa \Phi^{-1} C_{\alpha+\eta-1} C'_\eta \Phi^{-1} C_{\alpha+\kappa+1}\right]_{22}\right\} \\
& - \frac{2}{T}\sum_{\kappa,\zeta,\alpha=0}^{\infty} tr\left\{\left[C'_{\alpha+\kappa+1} \Phi^{-1} C_{\zeta+\alpha+1} C'_\zeta \Phi^{-1} C_\kappa\right]_{22}\right\}
\end{aligned}
$$

substituting $\Gamma_j = E\left[X_t X'_{t-j}\right] = \sum_{\alpha=0}^{\infty} C_{\alpha+j} C'_\alpha$ where possible gives the expression in theorem 9 which is hereby proven.

**5.A.11. Proof of Theorem 10.** For theorem 10 we note that the log-likelihood equals:

$$l_T = -\frac{1}{2}Tq\log 2\pi - \frac{T}{2}\log|\Omega_{22}| - \frac{1}{2}tr\left\{\Omega_{22}(Y - XA')'(Y - XA')\right\}$$

Thus for a known variance-covariance matrix $\Omega_{22} = I$, the likelihood ratio statistic equals

$$
\begin{aligned}
-2\ln LR(A = A_0) &= tr\left\{(Y - XA'_0)'(Y - XA'_0)\right\} - tr\left\{(Y - X\hat{A}')'(Y - X\hat{A}')\right\} \\
&= tr\{U'U\} - tr\left\{(Y - XA'_0 + X\left(A_0 - \hat{A}\right))'(Y - XA'_0 + X\left(A_0 - \hat{A}\right))\right\} \\
&= tr\{U'U\} - tr\left\{(Y - X(X'X)^{-1}X'U)'(Y - X(X'X)^{-1}X'U)\right\} \\
&= tr\left\{(U'X)(X'X)^{-1}(X'U)\right\}
\end{aligned}
$$

where we have used that $\hat{A} = (X'X)^{-1}(X'Y) = A_0 + (X'X)^{-1}(X'U)$ and defined $U = Y - XA_0$. We thus obtain:

$$-2\ln LR(A = A_0) = tr\left\{(U'X)(X'X)^{-1}(X'U)\right\}$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

A first order expansion of this expression (using equation (5.40)) delivers

$$E\left[tr\left\{(U'X)(X'X)^{-1}(X'U)\right\}\right]$$
$$\stackrel{1}{=} tr\left\{E\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\}$$
$$+ tr\left\{E\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\Phi-\frac{1}{T}X'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\}$$
$$+ tr\left\{E\left(\frac{1}{\sqrt{T}}U'X\right)\Phi^{-1}\left(\Phi-\frac{1}{T}X'X\right)\Phi^{-1}\left(\Phi-\frac{1}{T}X'X\right)\Phi^{-1}\left(\frac{1}{\sqrt{T}}X'U\right)\right\}$$

or stated differently

$$E\left[tr\left\{(U'X)(X'X)^{-1}(X'U)\right\}\right] \stackrel{1}{=} D_1 + D_3 + D_6$$

Adding up the expressions for all these terms, which were calculated in the last paragraph, deliver the result in theorem 10.

**5.A.12. Proof of theorem 11.** We take the terms of theorem 9 one by one, substute $C_\beta = SD^\beta F$ and $D = V\Lambda V^{-1}$ and then simplify. In this proof all ten terms turn out to be different.

$$\begin{aligned}
t1' &= \sum\nolimits_{\beta,\eta,\kappa,\zeta=0}^{\infty} tr\left\{C'_\kappa\Phi^{-1}C_\zeta C'_{\kappa+\zeta+1}\Phi^{-1}C_{\beta+\eta+1}C'_\eta\Phi^{-1}C_\beta I_{22}\right\}\\
&= \sum\nolimits_{\beta,\kappa=0}^{\infty} tr\left\{F'D^{\kappa\prime}S'\left(S\Psi S'\right)^{-1}S\Psi D^{\kappa+1\prime}S'\left(S\Psi S'\right)^{-1}SD^{\beta+1}\Psi S'\left(S\Psi S'\right)^{-1}SD^\beta FI_{22}\right\}\\
&= \sum\nolimits_{\beta,\kappa=0}^{\infty} tr\left\{FI_{22}F'D^{\kappa\prime}S'\left(S\Psi S'\right)^{-1}S\Psi D^{\kappa+1\prime}S'\left(S\Psi S'\right)^{-1}SD^{\beta+1}\Psi S'\left(S\Psi S'\right)^{-1}SD^\beta\right\}\\
&= \sum\nolimits_{\beta,\kappa=0}^{\infty} tr\left\{\left(V^{-1}FI_{22}F'V^{-1\prime}\right)\Lambda^\kappa\left(V'PV^{-1\prime}\Lambda\right)\Lambda^\kappa\left(V'S'\Phi^{-1}SV\Lambda\right)\Lambda^\beta\left(V^{-1}P'V\right)\Lambda^\beta\right\}\\
&= tr\left\{A_1\left(A_2\oslash A_8\right)A_3\left(A_4\oslash A_8\right)\right\}
\end{aligned}$$

$$\begin{aligned}
t2' &= \sum\nolimits_{\alpha,\eta,\kappa,\zeta=0}^{\infty} tr\left\{C'_\kappa\Phi^{-1}C_\zeta C'_{\kappa+\zeta+1}\Phi^{-1}C_\alpha I_{22}\right\}tr\left\{C'_{\alpha+\eta+1}\Phi^{-1}C_\eta\right\}\\
&= \sum\nolimits_{\alpha,\kappa=0}^{\infty} tr\left\{F'D^{\kappa\prime}S'\Phi^{-1}S\Psi D^{\kappa+1\prime}S'\Phi^{-1}SD^\alpha FI_{22}\right\}tr\left\{S'\Phi^{-1}S\Psi D^{\alpha+1\prime}\right\}\\
&= \sum\nolimits_{\alpha,\kappa=0}^{\infty} tr\left\{\left(V'PV^{-1\prime}\Lambda\right)\Lambda^\alpha\right\}tr\left\{\left(V^{-1}FI_{22}F'V^{-1\prime}\right)\Lambda^\kappa\left(V'PV^{-1\prime}\Lambda\right)\Lambda^\kappa\left(V'S'\Phi^{-1}SV\right)\Lambda^\alpha\right\}\\
&= \sum\nolimits_{\alpha,\kappa=0}^{\infty} tr\left\{\left(V'PV^{-1\prime}\Lambda\right)\Lambda^\alpha\otimes\left(V^{-1}FI_{22}F'V^{-1\prime}\right)\Lambda^\kappa\left(V'PV^{-1\prime}\Lambda\right)\Lambda^\kappa\left(V'S'\Phi^{-1}SV\right)\Lambda^\alpha\right\}\\
&= tr\left\{\left(A_2\otimes A_1\right)\left(I\otimes\left(A_2\oslash\left(ll'-\Lambda^{co}\Lambda^{ro}\right)\right)A_5\right)\left(I_{n^2}-\Lambda\otimes\Lambda\right)^{-1}\right\}\\
&= \sum\nolimits_{i=1}^{n}\left(A_2\right)_{ii}tr\left\{A_1\left(A_2\oslash A_8\right)A_5A_{9i}\right\}
\end{aligned}$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

$$
\begin{aligned}
t3' &= \sum_{\alpha,\eta,\lambda,\zeta=0}^{\infty} tr\left\{C_\lambda'\Phi^{-1}C_\alpha I_{22}\right\} tr\left\{C_\zeta'\Phi^{-1}C_{\lambda+\zeta+1}\right\} tr\left\{C_{\alpha+\eta+1}'\Phi^{-1}C_\eta\right\} \\
&= \sum_{\alpha,\lambda=0}^{\infty} tr\left\{P'D^{\lambda+1}\right\} tr\left\{PD^{\alpha+1\prime}\right\} tr\left\{FI_{22}F'D^{\lambda\prime}S'\Phi^{-1}SD^{\alpha}\right\} \\
&= \sum_{\alpha,\lambda=0}^{\infty} tr\left\{\left(\Lambda V^{-1}P'V\right)\Lambda^\lambda\right\} tr\left\{\left(V'PV^{-1\prime}\Lambda\right)\Lambda^\alpha\right\} tr\left\{\left(V^{-1}FI_{22}F'V^{-1\prime}\right)\Lambda^\lambda\left(V'S'\Phi^{-1}SV\right)\Lambda^\alpha\right\} \\
&= \sum_{\alpha,\lambda=0}^{\infty} tr\left\{A_2'\Lambda^\lambda\right\} tr\left\{A_2\Lambda^\alpha\right\} tr\left\{A_1\Lambda^\lambda A_5\Lambda^\alpha\right\} \\
&= tr\left\{\left(A_2'\otimes A_2\otimes A_1\right)\left(I-\Lambda\otimes I\otimes\Lambda\right)^{-1}\left(I\otimes I\otimes A_5\right)\left(I-I\otimes\Lambda\otimes\Lambda\right)^{-1}\right\} \\
&= \sum_{i,j=1}^{n} \left(A_2\right)_{ii}\left(A_2\right)_{jj} tr\left\{A_1A_{9i}A_5A_{9j}\right\}
\end{aligned}
$$

$$
\begin{aligned}
t4' &= \sum_{\beta,\eta,\kappa,\zeta=0}^{\infty} tr\left\{C_\beta'\Phi^{-1}C_\eta C_{\kappa+\eta+1}'\Phi^{-1}C_{\beta+\zeta+1}C_\zeta'\Phi^{-1}C_\kappa I_{22}\right\} \\
&= \sum_{\beta,\eta,\kappa,\zeta=0}^{\infty} tr\left\{F'D^{\beta\prime}PD^{\kappa+1\prime}S'\Phi^{-1}SD^{\beta+1}P'D^{\kappa}FI_{22}\right\} \\
&= \sum_{\beta,\kappa=0}^{\infty} tr\left\{\left(V^{-1}FI_{22}F'V^{-1\prime}\right)\Lambda^\beta\left(V'PV^{-1\prime}\Lambda\right)\Lambda^\kappa\left(V'S'\Phi^{-1}SV\Lambda\right)\Lambda^\beta\left(V^{-1}P'V\right)\Lambda^\kappa\right\} \\
&= \sum_{\beta,\kappa=0}^{\infty} tr\left\{A_1\Lambda^\beta A_2\Lambda^\kappa A_3\Lambda^\beta A_4\Lambda^\kappa\right\} \\
&= \sum_{i,j,k,m=1}^{n} \frac{\left(A_1\right)_{ij}\left(A_2\right)_{jk}\left(A_3\right)_{km}\left(A_4\right)_{mi}}{(1-v_jv_m)(1-v_iv_k)}
\end{aligned}
$$

$$
\begin{aligned}
t5' &= \sum_{\beta,\eta,\lambda,\zeta=0}^{\infty} tr\left\{C_\beta'\Phi^{-1}C_\eta C_{\lambda+\eta+1}'\Phi^{-1}C_\zeta C_{\beta+\zeta+1}'\Phi^{-1}C_\lambda I_{22}\right\} \\
&= \sum_{\beta,\lambda=0}^{\infty} tr\left\{F'D^{\beta\prime}PD^{\lambda+1\prime}PD^{\beta+1\prime}S'\Phi^{-1}SD^{\lambda}FI_{22}\right\} \\
&= \sum_{\beta,\lambda=0}^{\infty} tr\left\{\left(V^{-1}FI_{22}F'V^{-1\prime}\right)\Lambda^\kappa\left(V'PV^{-1\prime}\Lambda\right)\Lambda^\lambda\left(V'PV^{-1\prime}\Lambda\right)\Lambda^\kappa\left(V'S'\Phi^{-1}SV\right)\Lambda^\lambda\right\} \\
&= \sum_{\beta,\lambda=0}^{\infty} tr\left\{A_1\Lambda^\kappa A_2\Lambda^\lambda A_2\Lambda^\kappa A_5\Lambda^\lambda\right\} \\
&= \sum_{i,j,k,m=1}^{n} \frac{\left(A_1\right)_{ij}\left(A_2\right)_{jk}\left(A_2\right)_{km}\left(A_5\right)_{mi}}{(1-v_jv_m)(1-v_iv_k)}
\end{aligned}
$$

$$
\begin{aligned}
t6' &= \sum_{\alpha,\eta,\lambda,\zeta=0}^{\infty} tr\left\{C_\alpha'\Phi^{-1}C_\lambda I_{22}\right\} tr\left\{C_\zeta'\Phi^{-1}C_{\lambda+\eta+1}C_\eta'\Phi^{-1}C_{\alpha+\zeta+1}\right\} \\
&= \sum_{\alpha,\lambda=0}^{\infty} tr\left\{D^{\alpha+1}P'D^{\lambda+1}P'\right\} tr\left\{F'D^{\alpha\prime}S'\Phi^{-1}SD^{\lambda}FI_{22}\right\} \\
&= \sum_{\alpha,\lambda=0}^{\infty} tr\left\{\left(\Lambda V^{-1}P'V\right)\Lambda^\alpha\left(\Lambda V^{-1}P'V\right)\Lambda^\lambda\right\} tr\left\{\left(V^{-1}FI_{22}F'V^{-1\prime}\right)\Lambda^\alpha\left(V'S'\Phi^{-1}SV\right)\Lambda^\lambda\right\} \\
&= \sum_{\alpha,\lambda=0}^{\infty} tr\left\{A_2'\Lambda^\alpha A_2'\Lambda^\lambda\right\} tr\left\{A_1\Lambda^\alpha A_5\Lambda^\lambda\right\} \\
&= tr\left\{\left(A_2'\otimes A_1\right)\left(I-\Lambda\otimes\Lambda\right)^{-1}\left(A_2'\otimes A_5\right)\left(I-\Lambda\otimes\Lambda\right)^{-1}\right\} \\
&= \sum_{i,j=1}^{n} \left(A_2\right)_{ji}\left(A_2\right)_{ij} tr\left\{A_1A_{9i}A_5A_{9j}\right\}
\end{aligned}
$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

$$
\begin{aligned}
t7' &= \sum\nolimits_{\zeta,\eta,\kappa=0}^{\infty} tr\left\{C'_{\kappa}\Phi^{-1}C_{\zeta}I_{22}\right\} tr\left\{C'_{\kappa+\zeta+\eta+2}\Phi^{-1}C_{\eta}\right\} \\
&= \sum\nolimits_{\zeta,\kappa=0}^{\infty} tr\left\{D^{\kappa+\zeta+2}P\right\} tr\left\{F'D^{\kappa\prime}S'\Phi^{-1}SD^{\zeta}FI_{22}\right\} \\
&= \sum\nolimits_{\zeta,\kappa=0}^{\infty} tr\left\{\left(V^{-1}PV\right)\Lambda^{\kappa}\Lambda^{2}\Lambda^{\zeta}\right\} tr\left\{\left(V^{-1}FI_{22}F'V^{-1\prime}\right)\Lambda^{\kappa}\left(V'S'\Phi^{-1}SV\right)\Lambda^{\zeta}\right\} \\
&= \sum\nolimits_{\zeta,\kappa=0}^{\infty} tr\left\{A_4\Lambda^{\kappa}\Lambda^{2}\Lambda^{\zeta}\right\} tr\left\{A_1\Lambda^{\kappa}A_5\Lambda^{\zeta}\right\} \\
&= \sum\nolimits_{\zeta,\kappa=0}^{\infty} tr\left\{(A_4\otimes A_1)(I-\Lambda\otimes\Lambda)^{-1}\left(\Lambda^{2}\otimes A_5\right)(I-\Lambda\otimes\Lambda)^{-1}\right\} \\
&= \sum\nolimits_{i=1}^{n} (A_4)_{ii}\, v_i^2 tr\left\{A_1A_{9i}A_5A_{9i}\right\}
\end{aligned}
$$

$$
\begin{aligned}
t8' &= \sum\nolimits_{\zeta,\eta,\lambda=0}^{\infty} tr\left\{C'_{\lambda}\Phi^{-1}C_{\eta}C'_{\lambda+\zeta+\eta+2}\Phi^{-1}C_{\zeta}I_{22}\right\} \\
&= \sum\nolimits_{\zeta,\lambda=0}^{\infty} tr\left\{FI_{22}F'D^{\lambda\prime}PD^{\lambda+\zeta+2\prime}S'\Phi^{-1}SD^{\zeta}\right\} \\
&= \sum\nolimits_{\zeta,\lambda=0}^{\infty} tr\left\{V^{-1}FI_{22}F'V^{-1\prime}\Lambda^{\kappa}V'PV^{-1\prime}\Lambda^{\kappa}\Lambda^{2}\Lambda^{\zeta\prime}V'S'\Phi^{-1}SV\Lambda^{\zeta}\right\} \\
&= \sum\nolimits_{\zeta,\lambda=0}^{\infty} tr\left\{\left(V^{-1}FI_{22}F'V^{-1\prime}\right)\Lambda^{\kappa}\left(V'PV^{-1\prime}\right)\Lambda^{\kappa}\left(\Lambda^{2}\right)\Lambda^{\zeta\prime}\left(V'S'\Phi^{-1}SV\right)\Lambda^{\zeta}\right\} \\
&= \sum\nolimits_{\zeta,\lambda=0}^{\infty} tr\left\{A_1\Lambda^{\kappa}A'_4\Lambda^{\kappa}\left(\Lambda^{2}\right)\Lambda^{\zeta\prime}A_5\Lambda^{\zeta}\right\} \\
&= tr\left\{A_1\left(A'_4\oslash\left(ll'-\Lambda^{co}\Lambda^{ro}\right)\right)\left(\Lambda^{2}\right)\left(A_5\oslash\left(ll'-\Lambda^{co}\Lambda^{ro}\right)\right)\right\} \\
&= tr\left\{A_1\left(A'_4\oslash A_8\right)\left(\Lambda^{2}\right)\left(A_5\oslash A_8\right)\right\}
\end{aligned}
$$

$$
\begin{aligned}
t9' &= \sum\nolimits_{\kappa,\eta,\alpha=0}^{\infty} tr\left\{C'_{\kappa}\Phi^{-1}C_{\alpha+\eta+1}C'_{\eta}\Phi^{-1}C_{\alpha+\kappa+1}I_{22}\right\} \\
&= \sum\nolimits_{\kappa,\alpha=0}^{\infty} tr\left\{FI_{22}F'D^{\kappa\prime}S\Phi^{-1}SD^{\alpha+1}P'D^{\alpha+\kappa+1}\right\} \\
&= \sum\nolimits_{\kappa,\alpha=0}^{\infty} tr\left\{\Lambda^{\kappa}\left(V^{-1}FI_{22}F'V^{-1\prime}\right)\Lambda^{\kappa}\left(V'S\Phi^{-1}SV\Lambda\right)\Lambda^{\alpha}\left(V^{-1}P'V\right)\Lambda^{\alpha}\Lambda\right\} \\
&= \sum\nolimits_{\kappa,\alpha=0}^{\infty} tr\left\{\Lambda^{\kappa}A_1\Lambda^{\kappa}A_3\Lambda^{\alpha}A_4\Lambda^{\alpha}\Lambda\right\} \\
&= tr\left\{(A_1\oslash A_8)A_3(A_4\oslash A_8)\Lambda\right\}
\end{aligned}
$$

$$
\begin{aligned}
t10' &= \sum\nolimits_{\kappa,\zeta,\alpha=0}^{\infty} tr\left\{C'_{\alpha+\kappa+1}\Phi^{-1}C_{\zeta+\alpha+1}C'_{\zeta}\Phi^{-1}C_{\kappa}I_{22}\right\} \\
&= \sum\nolimits_{\kappa,\alpha=0}^{\infty} tr\left\{FI_{22}F'D^{\alpha+\kappa+1\prime}S\Phi^{-1}SD^{\alpha+1}P'D^{\kappa}\right\} \\
&= \sum\nolimits_{\kappa,\alpha=0}^{\infty} tr\left\{\Lambda^{\kappa}\left(V^{-1}FI_{22}F'V^{-1\prime}\right)\Lambda^{\kappa}\Lambda\Lambda^{\alpha}\left(V'S\Phi^{-1}SV\Lambda\right)\Lambda^{\alpha}\left(V^{-1}P'V\right)\right\} \\
&= \sum\nolimits_{\kappa,\alpha=0}^{\infty} tr\left\{\Lambda^{\kappa}A_1\Lambda^{\kappa}\Lambda\Lambda^{\alpha}A_3\Lambda^{\alpha}A_4\right\} \\
&= tr\left\{(A_1\oslash A_8)\Lambda(A_3\oslash A_8)A_4\right\}
\end{aligned}
$$

Adding the ten terms up, we obtain the expression in theorem 11:

 DOI: 10.2870/52609

$$
\begin{aligned}
\urcorner = {} & tr\left\{A_1\left(A_2 \oslash A_8\right) A_3\left(A_4 \oslash A_8\right)\right\} \\
& +2 \sum\nolimits_{i=1}^n\left(A_2\right)_{ii} tr\left\{A_1\left(A_2 \oslash A_8\right) A_5 A_{9i}\right\} \\
& +\sum\nolimits_{i, j=1}^n\left(A_2\right)_{ii}\left(A_2\right)_{jj} tr\left\{A_1 A_{9i} A_5 A_{9j}\right\} \\
& +\sum\nolimits_{i, j, k, m=1}^n \frac{\left(A_1\right)_{ij}\left(A_2\right)_{jk}\left(A_3\right)_{km}\left(A_4\right)_{mi}}{\left(1-v_j v_m\right)\left(1-v_i v_k\right)} \\
& +2 \sum\nolimits_{i, j, k, m=1}^n \frac{\left(A_1\right)_{ij}\left(A_2\right)_{jk}\left(A_2\right)_{km}\left(A_5\right)_{mi}}{\left(1-v_j v_m\right)\left(1-v_i v_k\right)} \\
& +\sum\nolimits_{i, j=1}^n\left(A_2\right)_{ji}\left(A_2\right)_{ij} tr\left\{A_1 A_{9i} A_5 A_{9j}\right\} \\
& -2 \sum\nolimits_{i=1}^n\left(A_4\right)_{ii} v_i^2 tr\left\{A_1 A_{9i} A_5 A_{9i}\right\} \\
& -2 tr\left\{A_1\left(A_4^{\prime} \oslash A_8\right)\left(\Lambda^2\right)\left(A_5 \oslash A_8\right)\right\} \\
& -2 tr\left\{\left(A_1 \oslash A_8\right) A_3\left(A_4 \oslash A_8\right) \Lambda\right\} \\
& -2 tr\left\{\left(A_1 \oslash A_8\right) \Lambda\left(A_3 \oslash A_8\right) A_4\right\}
\end{aligned}
$$

**5.A.13. Proof of theorem 12.** Theorem 12 is a special case of 11 with $S = I$. Inserting this in the expressions in table 5.3 we see that $P = I$ and furthermore that $A_2 = \Lambda$, $A_3 = A_6\Lambda$, $A_4 = I$, $A_5 = A_6$ and for any diagonal matrix $G$, $G \oslash A_8 = GA_7$. We substitute this in the ten terms of $\urcorner$ in the last expression:

$$
\begin{aligned}
t1' &= tr\left\{A_1\left(A_2 \oslash A_8\right) A_3\left(A_4 \oslash A_8\right)\right\} \\
&= tr\left\{A_1 \Lambda A_7 A_6 \Lambda A_7\right\}
\end{aligned}
$$

$$
\begin{aligned}
t2' &= \sum\nolimits_{i=1}^n\left(A_2\right)_{ii} tr\left\{A_1\left(A_2 \oslash A_8\right) A_5 A_{9i}\right\} \\
&= \sum\nolimits_{i=1}^n v_i tr\left\{A_1 \Lambda A_7 A_6 A_{9i}\right\}
\end{aligned}
$$

$$
\begin{aligned}
t3' &= \sum\nolimits_{i,j=1}^n\left(A_2\right)_{ii}\left(A_2\right)_{jj} tr\left\{A_1 A_{9i} A_5 A_{9j}\right\} \\
&= \sum\nolimits_{i,j=1}^n v_i v_j tr\left\{A_1 A_{9i} A_6 A_{9j}\right\}
\end{aligned}
$$

$$
\begin{aligned}
t4' &= \sum\nolimits_{\beta,\kappa=0}^{\infty} tr\left\{A_1 \Lambda^\beta A_2 \Lambda^\kappa A_3 \Lambda^\beta A_4 \Lambda^\kappa\right\} \\
&= \sum\nolimits_{\beta,\kappa=0}^{\infty} tr\left\{A_1 \Lambda^{\beta+\kappa+1} A_6 \Lambda^{\beta+\kappa+1}\right\} \\
&= tr\left\{\left(\left(\Lambda A_1 \Lambda\right) \oslash A_8\right)\left(A_6 \oslash A_8\right)\right\}
\end{aligned}
$$

$$
\begin{aligned}
t5' &= \sum\nolimits_{\beta,\lambda=0}^{\infty} tr\left\{A_1 \Lambda^\kappa A_2 \Lambda^\lambda A_2 \Lambda^\kappa A_5 \Lambda^\lambda\right\} \\
&= \sum\nolimits_{\beta,\lambda=0}^{\infty} tr\left\{A_1 \Lambda^{2\kappa+\lambda+2} A_6 \Lambda^\lambda\right\} \\
&= tr\left\{A_1 A_7 \Lambda^2\left(A_6 \oslash A_8\right)\right\}
\end{aligned}
$$

$$
\begin{aligned}
t6' &= \sum\nolimits_{i,j=1}^{n} \left(A_2\right)_{ji} \left(A_2\right)_{ij} tr\left\{A_1 A_{9i} A_5 A_{9j}\right\} \\
&= \sum\nolimits_{i=1}^{n} v_i^2 tr\left\{A_1 A_{9i} A_6 A_{9i}\right\}
\end{aligned}
$$

$$
\begin{aligned}
t7' &= \sum\nolimits_{i=1}^{n} \left(A_4\right)_{ii} v_i^2 tr\left\{A_1 A_{9i} A_5 A_{9i}\right\} \\
&= \sum\nolimits_{i=1}^{n} v_i^2 tr\left\{A_1 A_{9i} A_6 A_{9i}\right\}
\end{aligned}
$$

$$
\begin{aligned}
t8' &= tr\left\{A_1 \left(A_4' \oslash A_8\right)\left(\Lambda^2\right)\left(A_5 \oslash A_8\right)\right\} \\
&= tr\left\{A_1 A_7 \Lambda^2 \left(A_6 \oslash A_8\right)\right\}
\end{aligned}
$$

$$
\begin{aligned}
t9' &= \sum\nolimits_{\kappa,\alpha=0}^{\infty} tr\left\{\Lambda^{\kappa} A_1 \Lambda^{\kappa} A_3 \Lambda^{\alpha} A_4 \Lambda^{\alpha} \Lambda\right\} \\
&= \sum\nolimits_{\kappa,\alpha=0}^{\infty} tr\left\{A_1 \Lambda^{\kappa} A_6 \Lambda^{\kappa+2\alpha+2}\right\} \\
&= tr\left\{\Lambda^{\kappa+2\alpha+2} A_6 \Lambda^{\kappa} A_1\right\} \\
&= tr\left\{A_1 A_7 \Lambda^2 \left(A_6 \oslash A_8\right)\right\}
\end{aligned}
$$

$$
\begin{aligned}
t10' &= tr\left\{\left(A_1 \oslash A_8\right) \Lambda \left(A_3 \oslash A_8\right) A_4\right\} \\
&= tr\left\{\left(\Lambda A_1 \Lambda \oslash A_8\right)\left(A_6 \oslash A_8\right)\right\}
\end{aligned}
$$

Noting that in this case $t_5' = t_8' = t_9', t_4' = t_{10}'$ and $t_6' = t_7'$ and adding up we find the result in theorem 12:

$$
\begin{aligned}
\daleth &= tr\left\{A_1 \Lambda A_7 A_6 \Lambda A_7\right\} \\
&+ 2\sum\nolimits_{i=1}^{n} v_i tr\left\{A_1 \Lambda A_7 A_6 A_{9i}\right\} \\
&+ \sum\nolimits_{i,j=1}^{n} v_i v_j tr\left\{A_1 A_{9i} A_6 A_{9j}\right\} \\
&- tr\left\{\left(\left(\Lambda A_1 \Lambda\right) \oslash A_8\right)\left(A_6 \oslash A_8\right)\right\} \\
&- 2tr\left\{A_1 A_7 \Lambda^2 \left(A_6 \oslash A_8\right)\right\} \\
&- \sum\nolimits_{i=1}^{n} v_i^2 tr\left\{A_1 A_{9i} A_6 A_{9i}\right\}
\end{aligned}
$$

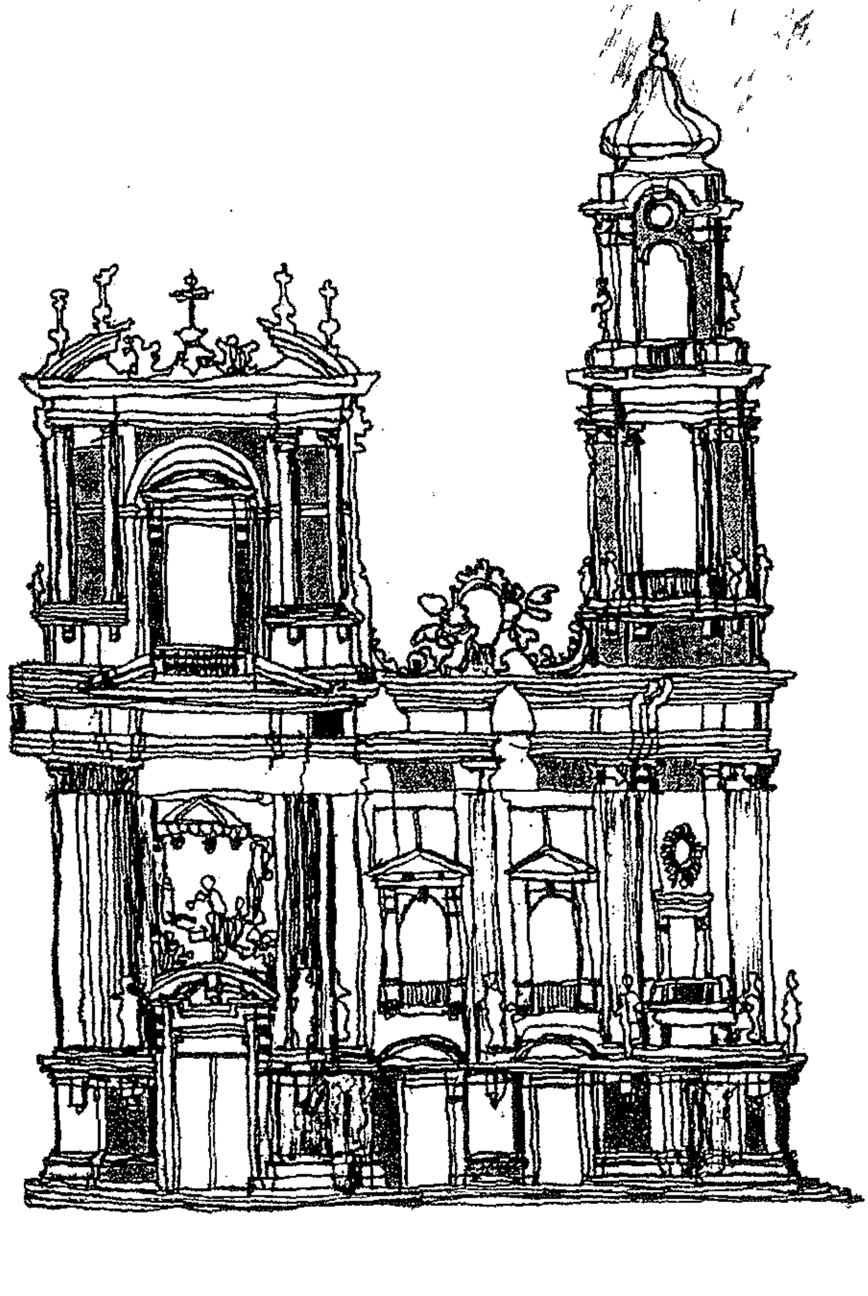
Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

# Part 3

# Impact Factors

 DOI: 10.2870/52609

M I L A N O




DOI: 10.2870/52609

CHAPTER 6

# Impact factors

## 6.1. Introduction

Forecasting is one of the major enterprizes in time-series econometrics, see Clements and Hendry (2002) and references therein. In this chapter we consider model-based long-run forecasts and their sensitivity with respect to information variables. We define a sensitivity indicator, called impact factors, IF. It is shown how this indicator allows to formulate questions on policy effectiveness and on the forecast uncertainty due to data revisions.

Sensitivity indicators have long been advocated in econometrics; see Banerjee and Magnus (1999, 2000) for recent references. By definition, they describe the sensitivity of a given procedure with respect e.g. to some possible source of model mis-specification. In the present case we apply this concept to mis-measurement of the information variables that are used in long-run forecasts.

Variations in the information variables can be caused by data revisions. Data revisions may alter the long-run forecasts of key macroeconomic indicators. Given that many economic decisions are based on forecasts made using preliminary data, it would be of interest to measure forecast uncertainty due to this source of data errors. Improving the quality of preliminary figures for variables to which forecasts are most sensitive would greatly improve the quality of the associated economic decisions. Conversely if data revisions on some variables do not have any impact on long run forecast, then there would be no need to obtain more timely or precise data.

Variations in the information variables may also be associated with the effects of policy interventions. In this perspective, it is of interest to find how long-run forecasts of key indicators are affected by possible policy actions. Absence of sensitivity would indicate long-run ineffectiveness of the policy measure. Although policy analysis and data revisions are the main economic areas of applications of this concept, the notion of IF can be defined and discussed in general for any dynamic system and forecast function.

The IF is not calculated on actual forecasts, but it is defined as a function of the model parameters and possibly of sample data. It measures long-run properties of the system; it is hence suggested as a tool of model interpretation, rather than of forecast performance. Quite obviously, the concept of IF does not account for the possible occurrence of model breaks between the past and the future.

The concept of IF is related to many standard econometric notions, like dynamic multipliers and impulse responses. Like a dynamic multiplier, the IF measures the sensitivity of a function. However, a dynamic multiplier is defined only between some endogenous variable $y$ and some exogenous variable $x$; impact factors, instead, are well defined for any dynamic systems, including VARs.

Moreover long-run multipliers are usually defined in terms of the static relation implied by a dynamic model for $y$ and $x$, see e.g. Hendry (1995, page 339), Gourerioux and Monfort (1995, pages 34-35), whereas the IF measures the accumulated effects on forecasts of perturbations in past information.

Impact factors turn out to be the limit of cumulated impulse responses (IR) in case of VARs. The definition of (economically meaningful) shocks is the subject of a vast debate in the literature,



Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

to which the present paper chapter does not contribute. We note here instead that, while IFs are defined in term of input variables, they are part of all limit cumulated IR of current use. Hence the analysis of the IF can be coupled with many possible definitions of structural shocks to obtain long-run sensitivity measures of forecast with respect to any particular shock definition. Moreover, the explicit expressions of the IF we derive in this chapter may be used to impose long-run restrictions on cumulated IR.

While the definition of IF is based on stationary processes, the concept is motivated and applied to non-stationary integrated systems. We consider I(1) and I(2) processes and compute IFs for these processes. For I(1) systems, the present chapter builds on ideas introduced in Bedini and Mosconi (2000). They defined the concept of 'long-run adjustment coefficients' with respect to the disequilibrium associated with an error correction term.

We here offer different insights on the I(1) case and extend the concept to I(2) systems. For the I(1) case we show how the long-run adjustment coefficients is related to the forecast function, and more in general to the concept of IF. This concept is linked to the choice of state vector and the timing of variables, and we discuss the relation among different choices.

Explicit expressions of the IF for the I(1) and I(2) case are given. These formulae do not involve infinite summations, and reveal the prominent role of the moving average impact matrices both in the I(1) and I(2) cases. Inference on these matrices has been considered in Paruolo (1997a,b, 2002a). These matrices and other parameters enter the expressions of the IF; this observation motivates the present extensions.

The explicit expressions of the IF allow to simplify and reduce the amount of computations needed to evaluate long-run effects. More importantly, the explicit expressions reveal the different contribution of various VAR parameters to the long-run effect. Several parameters are indeed shown to have no effect in the long run. Finally one can ascertain if there are any zero long run effects by analyzing the rank of specific blocks of the explicit form of the IF.

This chapter also analyzes the influence of timing in variable definitions on the long run effects. It is found that some IF are invariant to timing, while others are not.

Inference on the IF is presented in a compact and unified fashion for stationary, I(1) and I(2) systems. Wald tests on the IF are presented. Standard arguments imply that Wald tests on any smooth function of the IF are easily derived from the ones in this chapter through the delta method. We show how this analysis can be coupled with any definition of simultaneity structure to define sensitivity measures with respect to 'structural' perturbations.

The rest of the chapter is organized as follows. Section 6.2 reports relevant definitions and section 6.3 reports basic properties. Section 6.4 discusses impact factors in I(1) and I(2) processes. Section 6.5 discusses applications of this concept to policy analysis and data revisions. Section 6.6 discusses the estimation of IF, while section 6.7 reports as illustration an empirical analysis of a price system for Australia. Finally section 6.8 reports conclusions. All proofs are placed in 3 final appendices.

In the following $a := b$ and $b =: a$ indicate that $a$ is defined by $b$; $(a : b)$ indicates the matrix obtained by horizontally concatenating $a$ and $b$. $diag(A)$ is a matrix with elements on the main diagonal equal to the ones of $A$. For any full column rank matrices $H$, $A$, $B$, $sp(H)$ is the linear span of the columns of $H$, $\bar{H}$ indicates $H(H'H)^{-1}$ and $H_\perp$ indicates a basis of $sp(H)^\perp$, the orthogonal complement of $sp(H)$. $\|\cdot\|$ indicates a matrix norm and its associated vector norm. Moreover $P_H := H\bar{H}'$, $H_{AB} := \bar{A}'H\bar{B}$, $H_{AB.C} := H_{AB} - H_{AC}H_{CC}^{-1}H_{CB}$ while $H_A := H(A'H)^{-1}$. Finally $(\cdot)_{ij}$ indicates the $ij$-th element of the argument matrix, $vec$ is the column stacking operator, $\otimes$ is the Kronecker product (i.e. $A \otimes B$ is the matrix with generic block $a_{ij}B$, where $A := [a_{ij}]$) and $\overset{w}{\rightarrow}$ indicates weak convergence.




DOI: 10.2870/52609

## 6.2. Definition

Let $\{X_t\}_{t=-\infty}^{\infty}$ be a stationary $p$-variate time series, which contains the relevant information for the forecasting exercise. Let $Y_t$ be a $n \times 1$ vector of variables of interest, which are to be forecasted. Let $Y_{t+i|t}$ be the optimal forecast of $Y_{t+i}$ based on available information up and including time $t$, indicated by $X_{-\infty}^{t} := (X_t, X_{t-1}, \ldots)$, deemed to be the relevant information set.

The forecast $Y_{t+i|t}$ is a function, $g_i^\circ(\cdot)$ say, of $X_{-\infty}^t$, $Y_{t+i|t} = g_i^\circ(X_{-\infty}^t)$. Under quadratic loss, for instance, one has $Y_{t+i|t} = E(Y_{t+i}|X_{-\infty}^t)$, the conditional expectation.[1] We wish to summarize the sensitivity of the forecast function with respect to its inputs. Let $\widetilde{X}_t$ be a vector containing the relevant part of the information set retained in the forecast function, i.e. $Y_{t+i|t} = g_i^\circ(X_{-\infty}^t) = g_i(\widetilde{X}_t)$ for some function $g_i(\cdot)$. $\widetilde{X}_t$ is thus a 'sufficient statistic' for the information contained in $X_{-\infty}^t$; we call $\widetilde{X}_t$ the FS statistic ('Forecast Sufficient'), and indicate its dimension with $q$.

Let $\widetilde{v} := \widetilde{X}_t^c - \widetilde{X}_t$ be a perturbation in the FS statistic which induces a change $e_i(\widetilde{v}, \widetilde{X}_t) := g_i(\widetilde{X}_t^c) - g_i(\widetilde{X}_t)$ in the forecast function at forecast horizons $i = 1, \ldots, \ell$. We consider the cumulated changes $\sum_{i=1}^{\ell} e_i(\widetilde{v}, \widetilde{X}_t)$ up to some finite horizon $\ell$. If the sum converges for $\ell \to \infty$ we define the total effect, TE, of the perturbation as

$$\mathrm{TE}(\widetilde{v}, \widetilde{X}_t) := \sum_{i=1}^{\infty} e_i(\widetilde{v}, \widetilde{X}_t).$$

The quantity TE depends on $\widetilde{v}$ and possibly on $\widetilde{X}_t$; we wish to find a sensitivity measure of TE with respect to (small) changes $\widetilde{v}$, for fixed $\widetilde{X}_t$. This reflects the fact that the actual forecast takes place for given $\widetilde{X}_t$ and the sensitivity is measured locally, i.e. around a specific value for $\widetilde{X}_t$.

As a function of the perturbation $\widetilde{v}$, TE may be approximated by Taylor expansion around $\widetilde{v} = 0$ for fixed $\widetilde{X}_t$. This gives

$$\mathrm{TE}(\widetilde{v}, x) = \mathrm{TE}(0, x) + F(x)\widetilde{v} + R(\widetilde{v}, x) \tag{6.1}$$

where $R$ is a remainder term, which is of order $\|\widetilde{v}\|^2$ if TE is continuously differentiable up to order 2. Next note that, by definition, $\mathrm{TE}(0, x) = 0$ because $e_i(0, \widetilde{X}_t) = 0$. Hence

$$\mathrm{TE}(\widetilde{v}, x) = F(x)\widetilde{v} + R(\widetilde{v}, x).$$

We call

$$F := F(\widetilde{X}_t) = \left.\frac{\partial \mathrm{TE}(\widetilde{v}, \widetilde{X}_t)}{\partial \widetilde{v}'}\right|_{\widetilde{v}=0}$$

the Impact Factor, IF. It represents the coefficient of the linear approximation of $\mathrm{TE}(\widetilde{v}, x)$ as a function of the perturbation $\widetilde{v}$ close to $\widetilde{v} = 0$. Under the usual regularity conditions, differentiation and summation within TE may be interchanged; in this case $F(\widetilde{X}_t) = \sum_{i=1}^{\infty} \partial e_i(\widetilde{v}, \widetilde{X}_t)/\partial \widetilde{v}'$.

$F$ is a $p \times q$ matrix, where each entry gives a particular IF. Specifically $F_{ij}$ gives the IF of a perturbation in $\widetilde{X}_{jt}$, the $j$-th entry of $\widetilde{X}_t$, onto the long-run forecast of $Y_{it}$, the $i$-th element in $Y_t$. When $y_t$ and $x_t$ are subvectors of $Y_t$ and $\widetilde{X}_t$ respectively, we use the notation $F_{y,x} := F_{y_t, x_t}$ to indicate the corresponding submatrix of the IF matrix $F$; see also the following subsection 6.3.2.

Note that when $F := F(\widetilde{X}_t)$ does not depend on $\widetilde{X}_t$ it represents a global sensitivity measure. This is the case in linear systems, see subsection 6.3.3.

[1] Conditional expectations are defined up to a set of measure zero. In the following we will treat equalities concerning conditional expectations as a.s. equalities.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

## 6.3. Basic properties

In this section we derive basic properties of the IF defined in the previous section. As expressed in the definition, IFs are defined for stationary variables. Some of these properties are general, like the ones described in subsections 6.3.1, 6.3.2; others are specific for linear systems, which are discussed in subsections 6.3.3, 6.3.4, 6.3.5. The explicit derivation of IF in non-stationary systems of order 1 and 2 is considered in section 6.4. The reader mainly interested in the non-stationary applications may thus skip this section and go directly to section 6.4.

**6.3.1. Linear transformations.** Under quite unrestrictive assumptions on the forecast function, the IF matrix $F$ obeys a simple transformation rule under linear transformations of $Y_t$ and/or $\widetilde{X}_t$. Let $Y_t^* := N_Y Y_t$ , $\widetilde{X}_t^* := N_X \widetilde{X}_t$ be linear transformations of the original variables, where the $N$ matrices are square and non-singular. Let $F^*$ be the IF for the starred variables; we here show that

$$F^* = N_Y F N_X^{-1} \tag{6.2}$$

when the forecast function is equivariant with respect to linear combinations of the forecasts, i.e. when the prediction of $Y_{t+i}^*$ is equal to $N_Y g_i(\widetilde{X}_t)$ where $g_i(\widetilde{X}_t)$ is the forecast of $Y_{t+i}$. Conditional expectations e.g. possess this equivariant property. It is simple to see that the total effect of a change $\widetilde{v}$ to the input variables $\widetilde{X}_t$ on the forecast of $Y_{t+i}^*$ at all horizons is hence given by $\mathrm{TE}^* = N_Y \mathrm{TE}$, where $\mathrm{TE}(\widetilde{v}, \widetilde{X}_t)$ is the total effect on the forecast of $Y_t$.

The perturbations of the input variables are simply related by $\widetilde{v}^* := \widetilde{X}_t^{*c} - \widetilde{X}_t^* = N_X(\widetilde{X}_t^c - \widetilde{X}_t) = N_X \widetilde{v}$. Since $N_X$ is nonsingular, $\widetilde{v} = N_X^{-1}\widetilde{v}^*$, and $\widetilde{v} = 0$ iff $\widetilde{v}^* = 0$. Hence one can express the input $(\widetilde{v}, \widetilde{X}_t) = (N_X^{-1}\widetilde{v}^*, N_X^{-1}\widetilde{X}_t^*)$ in terms of $(\widetilde{v}^*, \widetilde{X}_t^*)$, giving

$$\mathrm{TE}^*(\widetilde{v}^*, \widetilde{X}_t^*) = N_Y \mathrm{TE}(N_X^{-1}\widetilde{v}^*, N_X^{-1}\widetilde{X}_t^*).$$

Thus, applying the definition of IF and the chain rule of differentiation,

$$\begin{aligned} F^* := & \left. \frac{\partial \mathrm{TE}^*(\widetilde{v}^*, \widetilde{X}_t^*)}{\partial \widetilde{v}^{*\prime}} \right|_{\widetilde{v}^*=0} = N_Y \left. \frac{\partial \mathrm{TE}(\widetilde{v}, \widetilde{X}_t)}{\partial \widetilde{v}'} \frac{\partial \widetilde{v}}{\partial \widetilde{v}^{*\prime}} \right|_{\widetilde{v}^*=0} = \\ = & \; N_Y \left. \frac{\partial \mathrm{TE}(\widetilde{v}, \widetilde{X}_t)}{\partial \widetilde{v}'} \right|_{\widetilde{v}=0} \cdot \frac{\partial N_X^{-1}\widetilde{v}^*}{\partial \widetilde{v}^{*\prime}} = N_Y F N_X^{-1}. \end{aligned}$$

Hence from the IF matrix $F$ one can derive all the IF implied by linear combinations of inputs and outputs applying the transformation (6.2).

**6.3.2. Subsets of variables.** Interest may be centered on some linear combinations $y_t = b'Y_t$ of $Y_t$ and/or on some linear combination $x_t := a'\widetilde{X}_t$ of $\widetilde{X}_t$. We can apply (6.2) to show that

$$F_{y_t, x_t} = b' F_{Y_t, \widetilde{X}_t} \bar{a}.$$

Consider the transformations $\widetilde{X}_t^* = N_X \widetilde{X}_t$, $Y_t^* = N_Y Y_t$ and choose $N_X := (a : a_\perp)'$, $N_Y := (b : b_\perp)'$. This gives $y_t$ as the first subvector of $Y_t^*$ and $x_t$ as the first subvector of $\widetilde{X}_t^*$; the other variables have been chosen so that $N_X$ and $N_Y$ are of full rank. Applying (6.2), the IF $F_{y_t,x_t}$ is the leading block in $F^* = N_Y F N_X^{-1}$, where $N_X^{-1} = (\bar{a} : \bar{a}_\perp)$ by the choice of $N_X$. This shows that $F_{y_t,x_t} = b' F_{Y_t,\widetilde{X}_t} \bar{a}$.

We next specialize the notion of IF to the case of a linear forecast function.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

**6.3.3. Linear forecast function.** When the forecast function is linear

$$g_i(\widetilde{X}_t) = a_i + B_i\widetilde{X}_t, \qquad (6.3)$$

it is simple to note that $e_i(\widetilde{v}, \widetilde{X}_t) := g_i(\widetilde{X}_t^c) - g_i(\widetilde{X}_t) = B_i(\widetilde{X}_t^c - \widetilde{X}_t) = B_i\widetilde{v}$, which depends on $(\widetilde{X}_t, \widetilde{v})$ only through $\widetilde{v}$.[2] Hence in this case, if $\{B_i\}$ is summable, one finds TE$(\widetilde{v}, \widetilde{X}_t)$ =TE$(\widetilde{v})$ = $(\sum_{i=1}^{\infty} B_i)\widetilde{v}$, and $F = \sum_{i=1}^{\infty} B_i$. Observe that the remainder term $R$ is zero because TE is a linear function of the perturbation $\widetilde{v}$ only. Here IF is a global sensitivity measure, since it is constant for all possible values of $\widetilde{X}_t$.

**6.3.4. Stationary VARs.** Let $X_t$ be generated by a VAR $A(L)X_t = \mu^* D_t^* + \epsilon_t$, with deterministic component $\mu^* D_t^*$, and i.i.d. $N(0, \Omega)$ errors $\epsilon_t$. Here and in the following we take $D_t^* := (t : 1 : d_t')'$, where $d_t := (d_{1,t} : ..d_{u-1,t})'$ is a vector of de-meaned seasonal dummies of the form $d_{i,t} = 1(t \bmod u = i) - 1/u$, $1(\cdot)$ is the indicator function and $u$ is the number of seasons.[3]

The associated state space representation is $\widetilde{X}_t = A\widetilde{X}_{t-1} + u_t$ with state vector $\widetilde{X}_t := (X_t' : X_{t-1}' : ... : X_{t-k+1}')'$, companion matrix

$$A := \begin{pmatrix} A_1 & A_2 & \dots & A_k \\ I & & & \\ & \ddots & & \\ & & I & 0 \end{pmatrix}.$$

and $u_t := J(\mu^* D_t^* + \epsilon_t)$, $J := (I_p : 0_{p \times p(k-1)})'$, $X_t = J'\widetilde{X}_t$.

Let the variables to be forecasted $Y_t$ coincide with $X_t$; in this case the forecast function is $Y_{t+i|t} = E(Y_{t+i}|X_{-\infty}^t) = J'A^i\widetilde{X}_t + \sum_{j=0}^{i-1} J'A^j J\mu^* D_{t+i-j}^*$. Note that $\widetilde{X}_t$ is the FS statistic, and that $Y_{t+i|t} = g_i(\widetilde{X}_t)$ is a linear function of it, as in (6.3), with $a_i := \sum_{j=0}^{i-1} J'A^j J\mu^* D_{t+i-j}^*$ and $B_i = J'A^i$. Hence $e_i = B_i\widetilde{v}$.

Assume also that the VAR process $X_t$ is stationary, which implies that all eigenvalues of $A$ are less than 1 in modulus. Then

$$\text{TE} = \sum_{i=1}^{\infty} B_i\widetilde{v} = J'\left(\sum_{i=1}^{\infty} A^i\right)\widetilde{v} = J'((I - A)^{-1} - I)\widetilde{v}$$

where the series is convergent because of the stationarity assumption. In this case the IF is equal to $F := J'((I - A)^{-1} - I)$, a simple function of the companion matrix.

If the variables to be forecasted are all the ones contained in the state vector, $Y_t = \widetilde{X}_t$, then the previous calculations reveal that TE $= ((I - A)^{-1} - I)\widetilde{v}$ and the IF is

$$F = (I - A)^{-1} - I. \qquad (6.4)$$

In the present case of stationary VARs the possibility to consider all of the state vector as $Y_t$ is not very interesting, because $Y_t$ contains the same variables $X_t$ at different lags. This possibility is instead of interest for non-stationary systems of order 1 and 2, considered in section 6.4 below.

We here observe that some selection of subvectors of $Y_t$ and $\widetilde{X}_t$ may include terms in TE, and hence in $F$, which are not associated with proper forecasting. To illustrate the point consider a VAR(2) process, $Y_t = \widetilde{X}_t = (X_t' : X_{t-1}')'$ and the selection $y_t := J_\perp' Y_t := (0 : I)Y_t = X_{t-1}$, $x_t := J'\widetilde{X}_t = (I : 0)\widetilde{X}_t = X_t$. Consider 1 step ahead forecasts $y_{t+1} := X_t$ given $x_t := X_t$. Obviously $E(X_t|X_t) = X_t$, and there is no forecast to be made, and no forecast error.

[2] Here the coefficients $a_i$ and $B_i$ may depend also on $t$; this is not included explicitly in the notation for simplicity.

[3] As noted by one of the referees, the vector $d_t$ may be assumed to contain other dummies $d_{it}$ which are bounded when cumulated once, $\left|\sum_{i=1}^t d_{it}\right| < c$. This would not change the asymptotics in the stationary, I(1) and I(2) cases.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

Hence changes in $x_t$ are identically transmitted to $e_1(v) = v$ which is then summed along with other $e_i(v)$ into TE. The IF thus contains one term not associated with actual forecasting. In order to discount the presence of this term, one could simply subtract the identity matrix after calculating the IF.

This phenomenon is also present in I(1) and I(2) systems; in the following we will call this the *lag-lead effect*. It can be eliminated by re-defining $Y_t$ and $\widetilde{X}_t$ through linear transformations as in section 6.3.1 or, more simply, by subtracting the extra terms not associated with forecasting after the calculation of the IF. This subtraction does not influence estimation and inference on the IF.

**6.3.5. Impulse responses.** Pesaran and Shin (1998) and Koop et al. (1996) defined the scaled generalized impulse responses (GIR) for stationary VARs as

$$J'A^iJ\Omega(diag(\Omega))^{-1/2} =: J'A^iJH^*$$

where $A$ is the companion matrix and $H^* := \Omega(diag(\Omega))^{-1/2}$. This definition of impulse response does not depend on orthogonalization of shocks. Taking the whole state vector as dependent variable, the GIR can be defined as $\psi^g(i) := A^iJH^*$. The cumulated GIR is

$$\sum_{i=1}^{\infty}\psi^g(i) = \sum_{i=1}^{\infty}A^iJH^* = ((I-A)^{-1}-I)JH^*$$

which is proportional to the IF matrix $F$ in eq. (6.4). A similar derivation applies to the cumulated impulse responses, which converge to an expression similar to $((I-A)^{-1}-I)JH^*$ with a different definition of the matrix $H^*$, which is usually a square root of $\Omega$.

In the case of structural IR, a model of the form $B\epsilon_t = C\eta_t$ is postulated, where $\eta_t$, with covariance equal to the identity, are called the "structural shocks". The matrices $B$ and $C$ are identified through restrictions, which ensure that $B$ is non-singular, see Amisano and Giannini (1997). In this case $H^*$ is taken as $B^{-1}C$, and the cumulated IRs with respect to $\eta_t$ are of the form $FJH^*$, where $F$ is the IF.

Despite being defined as a sensitivity measure of forecasts w.r.t. the FS variables, the IF thus present a central role also in the analysis of the cumulated effects of shocks in IR analysis, for any definition of shocks, whether structural, reduced form, generalized or standard. In section 6.6.3 we show how to extend the econometric tools to IR analysis.

We also note here that explicit expressions for $F$ derived in section 6.4 and the relation with the cumulated IR allow to impose long-run constraint on the IR by imposing them on the IF. See also Phillips (1998) on impulse responses in I(1) VARs.

**6.3.6. Linearity and superposition.** When the forecast function $g$ is linear, the principle of superposition applies, see Kailath (1980); this property is reviewed in this subsection. If one considers various perturbations $\widetilde{v}_0$, ..., $\widetilde{v}_j$, their cumulated effect is equal to $F\sum_{i=0}^{j}\widetilde{v}_i$. This equals the effect $F\widetilde{v}$ of a single perturbation $\widetilde{v}$ defined as the sum of the individual perturbations, $\widetilde{v} := \sum_{i=0}^{j}\widetilde{v}_i$, with the same IF $F$.

Consider next this equivalence specifically for VARs. Let perturbation $\widetilde{v}_i$ involve only the variables $X_{t+i}$ at lead $i = 0, 1, ...j$. In other words, at time $t+i$ only $X_{t+i}$ is perturbed. Then the superposition principle states that the IF of all these perturbations equals the IF of $\sum_{i=0}^{j}\widetilde{v}_i$ . In this sense, therefore, impact factors are insensitive to the timing of the perturbations. Obviously this does not need to be the case for non-linear forecast functions.

## 6.4. Cointegrated systems

In this section we apply the definition of IF to VARs integrated of order one and two, I(1) and I(2), and possibly cointegrated, CI. We refer to Johansen (1995b) for notation and definitions of

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

I(1) and I(2) VAR systems. In the rest of the chapter we assume that the forecast function is the conditional expectation and that $X_t$ is generated by a (possibly non stationary) VAR.

**6.4.1. Cointegrated I(1) VAR.** Consider the following equilibrium correction (EC) form of the VAR:

$$\Delta X_t = \alpha\beta' X_{t-2} + \Gamma_1^* \Delta X_{t-1} + \Phi U_{t-1} + \mu_1 t + \mu D_t + \epsilon_t. \tag{6.5}$$

where $\Gamma_1^* := (\Gamma_1 + \Pi)$, $\Phi_1 := \Gamma_2$, $\Phi_2 := (\Gamma_3 : \ldots : \Gamma_{k-1})$, $\Phi := (\Phi_1 : \Phi_2)$ and $U_{t-1} := (\Delta X'_{t-2} : \ldots : \Delta X'_{t-k+1})'$ is $m \times 1$, $m := p(k-2)$, and $\mu := (\mu_0 : \mu_d)$, $D_t := (1 : d'_t)'$.

This EC form presents the level term measured in $t-2$; this can always be accomplished by adding and subtracting appropriate terms, even in the case of $k = 1$, see Johansen (1995b). This representation is chosen in order to simplify calculations in the following, and it is completely general, because results for any other EC formulation can be deduced from it, see the following section 6.4.3.

We assume that the VAR process satisfies the following condition:

**I(1) Assumption**

I(1)_$a$ : Every root $z$ of the characteristic polynomial of $X_t$ satisfies $z = 1$ or $|z| > 1$.
I(1)_$b$ : $\Pi := -A(1) = \alpha\beta'$, where $\alpha$ and $\beta$ are $p \times r$ matrices of full rank $r < p$.
I(1)_$c$ : $\mu_1 = \alpha\beta_0'$ with $\beta_0'$ a $r \times 1$ vector.
I(1)_$d$ : $\alpha'_\perp \Gamma \beta_\perp$ has full rank $p - r$, where $\Gamma := -I + \sum_{i=1}^{k-1} \Gamma_i$.

These assumptions guarantee that $\Delta X_t$ and $\beta' X_t + \beta_0' t$ are stationary processes, apart from the influence of initial values, and that $X_t$ has at most a linear trend in all directions, see Johansen (1995b). In the following all results do not depend on how $\beta$ is identified. Hence we do not explicitly describe any normalizing restriction on $\beta$, but simply assume that some (possibly over-)identifying restriction is imposed in estimation.

The associated state space representation is $\widetilde{X}_t = A\widetilde{X}_{t-1} + u_t$ with $u_t := J(\mu^* D_t^* + \epsilon_t)$, $J := (I_p : 0)'$, and

$$\widetilde{X}_t \;:=\; \begin{pmatrix} \Delta X_t \\ \beta' X_{t-1} \\ U_t \end{pmatrix} \begin{matrix} p \\ r \\ m \end{matrix}$$

$$A \;:=\; \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} := \begin{matrix} & \begin{matrix} p & r & p & m-p \end{matrix} & \\ & \left(\begin{array}{cc|cc} \Gamma_1^* & \alpha & \Phi_1 & \Phi_2 \\ \beta' & I_r & & \\ \hline I_p & & & \\ & & I_{m-p} & \end{array}\right) & \begin{matrix} p \\ r \\ p \\ m-p \end{matrix} \end{matrix} \tag{6.6}$$

where we have reported dimensions alongside blocks of the the state vector and of the companion matrix. Note that for brevity the $A_{22}$ block in (6.6) is partitioned in blocks of $p$ and $m-p$ rows times $m-p$ and $p$ columns, unlike the other blocks. Zero entries are not reported unless when needed for clarity.

Let $i_j$ denote a $j \times 1$ vector of ones. The following proposition applies.

PROPOSITION 1 (IF in I(1) systems). *Consider state space form* (6.6) *under the* I(1) *assumption; then all eigenvalues of A are within the unit circle and the impact factor* $F := (I - A)^{-1} - I$ *has the following form: let*

$$B := \begin{pmatrix} C & (C\Gamma^\circ - I)\bar{\beta} \\ \bar{\alpha}'(\Gamma^\circ C - I) & \bar{\alpha}'(\Gamma^\circ C \Gamma^\circ - \Gamma^\circ)\bar{\beta} \end{pmatrix}$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

*$c_1 := c_2 \otimes I_p$, with $c_2$ a lower triangular matrix with ones on and below the main diagonal, $\Gamma^\circ := -\Gamma$, $C = \beta_\perp (\alpha_\perp' \Gamma^\circ \beta_\perp)^{-1} \alpha_\perp'$, $\psi := (\psi_2 : ... : \psi_{k-1})$, $\psi_i = \sum_{j=i}^{k-1} \Gamma_j$; then*

$$F + I = \begin{pmatrix} B & B \begin{pmatrix} \psi \\ 0 \end{pmatrix} \\ (i_{k-2} \otimes I : 0)B & c_1 + i_{k-2} \otimes C\psi \end{pmatrix}$$

$$= \begin{pmatrix} C & (C\Gamma^\circ - I)\bar{\beta} & C\psi \\ \bar{\alpha}'(\Gamma^\circ C - I) & \bar{\alpha}'(\Gamma^\circ C \Gamma^\circ - \Gamma^\circ)\bar{\beta} & \bar{\alpha}'(\Gamma^\circ C - I)\psi \\ i_{k-2} \otimes C & i_{k-2} \otimes (C\Gamma^\circ - I)\bar{\beta} & c_1 + i_{k-2} \otimes C\psi \end{pmatrix}.$$

*From this expression one can read the impact factors; in particular $F_{y_t,x_t}$ equals*

(1) $C - I$ *for* $y_t = x_t := \Delta X_t$
(2) $(C\Gamma^\circ - I)\bar{\beta}$ *for* $y_t := \Delta X_t$, $x_t := \beta' X_{t-1}$
(3) $\bar{\alpha}'(\Gamma^\circ C - I)$ *for* $y_t := \beta' X_{t-1}$, $x_t := \Delta X_t$
(4) $\bar{\alpha}'(\Gamma^\circ C \Gamma^\circ - \Gamma^\circ)\bar{\beta}$ *for* $y_t = x_t := \beta' X_{t-1}$.

A special interpretation applies to the I(1) case. Consider $F_{y,x}$ for $y_t := \Delta X_t$, $x_t := \widetilde{X}_t$. The cumulated forecasts of the differences $\sum_{i=1}^{\ell} \Delta X_{t+i|t} = X_{t+\ell|t} - X_t$ give the forecast on the levels minus the initial value. Hence the total effect of a change in $x_t$ is given by $\text{TE} = X_{\infty|t}^c - X_{\infty|t}$, where $X_{\infty|t}^c$ indicates the forecast on the level of $X_\infty$ based on $\widetilde{X}_t^c$. Thus TE measures the change in the long-run forecast on the *levels*, and IF is a sensitivity measure of the *level* forecast with respect to changes in the FS variables.

This interpretation has been emphasized in Bedini and Mosconi (2000). In particular they focus on $F_{\Delta X_t, \beta' X_{t-1}} = (C\Gamma^\circ - I)\bar{\beta}$, which they call the *long-run adjustment coefficients* to disequilibrium errors. The approach of the present chapter gives a forecasting interpretation of the long-run adjustment coefficients, as well as of other IF.

Note also that case 3 involves lag-lead effects, because $y_t$ is chosen as $\beta' X_{t-1}$ while $x_t$ is $\Delta X_t$. The lag-lead effect was introduced in section 6.3.4. The problem can be solved subtracting $\beta'$ from the corresponding IF.

**6.4.2. Cointegrated I(2) VAR.** Consider the equilibrium correction (EC) representation of the VAR suggested in Paruolo and Rahbek (1999) for I(2) systems:

$$\Delta^2 X_t = \alpha(\beta' X_{t-1} + \delta \beta_2' \Delta X_{t-1}) + (\zeta_1 : \zeta_2)(\beta : \beta_1)' \Delta X_{t-1} + + \Upsilon_1 \Delta^2 X_{t-1}' + \Phi W_{t-1} + \mu^* D_t^* + \epsilon_t \quad (6.7)$$

where $W_{t-1} := (\Delta^2 X_{t-2}' : ... : \Delta^2 X_{t-k+2}')'$, of dimension $m \times 1$, $m := p(k-3)$, $\Phi := (\Upsilon_2 : ... : \Upsilon_{k-2})$. $\mu^* := (\mu_1 : \mu_0 : \mu_d)$, $D_t^* := (t : 1 : d_t')'$.

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

**I(2) Assumption**

I(2)_$a$ : Assumptions I(1)_$a$, I(1)_$b$, I(1)_$c$ hold.

I(2)_$b$ : $P_{\alpha_\perp} \Gamma P_{\beta_\perp} = \alpha_1 \beta_1'$ where $\alpha_1$ and $\beta_1$ are $p \times s$ matrices of full rank $s < p - r$, or, equivalently, $\alpha_\perp' \Gamma \beta_\perp = \xi \eta'$ where $\xi = \alpha_\perp' \alpha_1$ and $\eta = \beta_\perp' \beta_1$ are $p - r \times s$ matrices of full rank $s < p - r$.

I(2)_$c$ : $\alpha_2' \theta \beta_2$ has full rank $p_2 := p - r - s$, where $\alpha_2 = (\alpha : \alpha_1)_\perp$, $\beta_2 = (\beta : \beta_1)_\perp$ and $\theta$ is defined as

$$\theta := \Gamma \bar{\beta} \bar{\alpha}' \Gamma + \phi. \quad (6.8)$$

I(2)_$d$ : $\alpha_\perp' \mu_0 = \xi \eta_0' + \alpha_\perp' \Gamma \bar{\beta} \beta_0'$, with $\eta_0'$ a $s \times 1$ vector.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

| symbol | definition | dim | symbol | definition | dim |
|---|---|---|---|---|---|
| $\zeta_1$ | $:= \Gamma\beta$ | $p \times r$ | $\zeta_2$ | $:= \Gamma\beta_1$ | $p \times s$ |
| $\Gamma$ | $:= -I + \sum_{i=1}^{k-1} \Gamma_i$ | $p \times p$ | $\Phi_1$ | $:= \Upsilon_2$ | $p \times p$ |
| | $= \alpha\delta\beta_2' + \zeta_1\beta' + \zeta_2\beta_1'$ | | $\Phi_2$ | $:= (\Upsilon_3 : ... : \Upsilon_{k-2})$ | $p \times p(k-4)$ |
| $\zeta_1^*$ | $:= \zeta_1 + 2\alpha$ | $p \times r$ | $\Upsilon_1^*$ | $:= (\Upsilon_1 + \Gamma + \alpha\beta')$ | $p \times p$ |
| $\phi$ | $:= I - \sum_{i=1}^{k-2} \Upsilon_i$ | $p \times p$ | $\phi^*$ | $:= \phi - \Gamma - \alpha\beta'$ | $p \times p$ |
| $\theta$ | $:= \zeta_1\bar{\alpha}'\Gamma + \phi$ | $p \times p$ | $\theta^*$ | $:= \zeta_1^*\bar{\alpha}'\Gamma + \phi^*$ | $p \times p$ |
| $C_2$ | $:= \beta_2 \left(\alpha_2'\theta\beta_2\right)^{-1} \alpha_2'$ | $p \times p$ | $\tau$ | $:= (\beta : \beta_1)$ | $p \times (r+s)$ |
| $h$ | $:= I - \theta^* C_2$ | $p \times p$ | $\varrho$ | $:= \bar{\alpha}'(I - \zeta_2\bar{\alpha}_1')h$ | $r \times p$ |
| $\psi_i$ | $:= \sum_{j=i}^{k-2} \Upsilon_j$ | $p \times p$ | $\psi$ | $:= (\psi_2 : ... : \psi_{k-2})$ | $p \times p(k-3)$ |

Table 6.1: Symbol definitions for the expression of the IF in the I(2) systems.

Johansen's I(2) representation theorem, see Johansen (1992) or Johansen (1995b, theorem 4.6), establishes that, under I(1)_$a$, necessary and sufficient conditions for

$$\Delta^2 X_t, \quad \beta' X_t + \delta\beta_2'\Delta X_t + \beta_0' t, \quad \beta_1'\Delta X_t \tag{6.9}$$

to be stationary, apart from initial values, and for $X_t$ to have at most a linear trend in all directions are the conditions I(2)_$a$ to $d$. This result is reported in Rahbek et al. (1999).[4] In the following 'I(2) assumption' and 'I(2) conditions' are used as synonyms.

The EC formulation in (6.7) imposes some of the I(2) restrictions; we refer to Paruolo and Rahbek (1999) for complete definitions of coefficients and background. As for the I(1) case, we choose a specific timing of the EC terms in order to simplify calculations. Again this is done without loss of generality, since results for any other EC formulation can be deduced from it, see again section 6.4.3.

Proposition 5 in section 6.A shows that one of the many possible equivalent EC formulation of this system is

$$\begin{aligned}\Delta^2 X_t &= \alpha(\beta' X_{t-3} + \delta\beta_2'\Delta X_{t-2} + \beta_0' t) + (\zeta_1^* : \zeta_2)(\beta : \beta_1)'\Delta X_{t-2} + \\ &\quad + \Upsilon^* \Delta^2 X_{t-1}' + \Phi W_{t-1} + \mu D_t + \epsilon_t,\end{aligned} \tag{6.10}$$

where we have imposed $\mu_1 = \alpha\beta_0'$, condition I(1)_$c$. The timing of the EC terms $(\beta' X_{t-3} + \delta\beta_2'\Delta X_{t-2})$, $(\beta : \beta_1)'\Delta X_{t-2}$ is different from the one in (6.7) and $\zeta_1^* := \zeta_1 + 2\alpha$ and $\Upsilon_1^* := (\Upsilon_1 + \Gamma + \Pi)$. Note that this affects the definition only of $\zeta_1^*$ and $\Upsilon_1^*$ and not of the remaining coefficients. This timing can always be achieved, also for $k = 2$.[5] We summarize notation in table 6.1.

The present derivations do not depend on how $\beta$, $\beta_1$ and $\delta$ are identified; see Paruolo and Rahbek (1999) for a discussion of identification in (6.7). Thus we do not explicitly describe any normalizing restriction on the CI parameters, but we will simply assume that some (possibly over-)identifying restriction is imposed in estimation.

The system can be cast in the state space form $\widetilde{X}_t = A\widetilde{X}_{t-1} + u_t$ with $u_t := J(\mu^* D_t^* + \epsilon_t)$ and

$$\widetilde{X}_t := \begin{pmatrix} \Delta^2 X_t \\ \beta'\Delta X_{t-1} \\ \beta_1'\Delta X_{t-1} \\ \beta' X_{t-2} + \delta\beta_2'\Delta X_{t-1} \\ W_t \end{pmatrix} \begin{matrix} p \\ r \\ s \\ r \\ m \end{matrix},$$

[4]Note that the stationarity of the variables in (6.9) implies that also $\beta'\Delta X_t$ is stationary.

[5]Following the literature, we do not consider $k = 1$ in the I(2) case.

 DOI: 10.2870/52609

$$
(6.11) \qquad A := \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{array}{c} \\ \left(\begin{array}{cccc|cc} p & r & s & r & p & m-p \\ \Upsilon_1^* & \zeta_1^* & \zeta_2 & \alpha & \Phi_1 & \Phi_2 \\ \beta' & I_r & & & & \\ \beta_1' & & I_s & & & \\ \delta\beta_2' & I_r & & I_r & & \\ \hline I_p & & & & & \\ & & & & I_{m-p} & \end{array}\right) \end{array} \begin{array}{c} \\ p \\ r \\ s \\ r \\ p \\ m-p \end{array}
$$

where we have reported dimensions. As in (6.6), the $A_{22}$ block in (6.11) is partitioned in blocks of $p$ and $m-p$ rows times $m-p$ and $p$ columns, differently from the remaining blocks. The following proposition applies.

PROPOSITION 2 (IF in I(2) systems). *Consider the state space form (6.11) under the I(2) assumption; then all eigenvalues of $A$ are within the unit circle and the impact factor $F := (I-A)^{-1} - I$ has the following form: let*

$$
B := \begin{pmatrix} p & r+s & r \\ C_2 & (C_2\phi^* - I)\bar{\tau} & -C_2\zeta_1 \\ -\delta\beta_2' C_2 & -\delta\beta_2' C_2 \phi^* \bar{\tau} & \delta\beta_2' C_2 \zeta_1 - I \\ -\bar{\alpha}_1' h & -\bar{\alpha}_1' h \phi^* \bar{\tau} & \bar{\alpha}_1' h \zeta_1^* \\ -\varrho & -\varrho\phi^*\bar{\tau} & \varrho\zeta_1^* \end{pmatrix} \begin{array}{c} \\ p \\ r \\ s \\ r \end{array}
$$

*where symbols are defined in table* 6.1 *and* $c_1 := c_2 \otimes I_p$, *where* $c_2$ *is a lower triangular matrix with ones on and below the main diagonal; then* $F + I$ *equals*

$$
(6.12) \quad \begin{pmatrix} B & B\begin{pmatrix} \psi \\ 0 \end{pmatrix} \\ (i_{k-2} \otimes I : 0)B & c_1 + i_{k-2} \otimes C_2\psi \end{pmatrix}
$$

$$
= \begin{pmatrix} C_2 & (C_2\phi^* - I)\bar{\tau} & -C_2\zeta_1 & C_2\psi \\ -\delta\beta_2' C_2 & -\delta\beta_2' C_2\phi^*\bar{\tau} & -I + \delta\beta_2' C_2 \zeta_1 & -\delta\beta_2' C_2 \psi \\ -\bar{\alpha}_1' h & -\bar{\alpha}_1' h\phi^*\bar{\tau} & \bar{\alpha}_1' h \zeta_1 & -\bar{\alpha}_1' h \psi \\ -\varrho & -\varrho\phi^*\bar{\tau} & \varrho\zeta_1^* & -\varrho\psi \\ -i_{k-2} \otimes C_2\zeta_1 & & c_1 + (i \otimes I_p) C_2 \psi & \end{pmatrix}.
$$

*From this expression one can read the impact factors; in particular* $F_{y_t, x_t}$ *equals*

(1) $C_2 - I$ *for* $y_t = x_t := \Delta^2 X_t$;
(2) $-\delta\beta_2' C_2$ *for* $y_t := \beta' \Delta X_t$, $x_t := \Delta^2 X_t$
(3) $\bar{\alpha}_1'(\theta^* C_2 - I)$ *for* $y_t := \beta_1' \Delta X_t$, $x_t := \Delta^2 X_t$
(4) $\bar{\alpha}'(I - \zeta_2 \bar{\alpha}_1')(\theta^* C_2 - I)$ *for* $y_t := \beta' X_{t-2} + \delta\beta_2' \Delta X_{t-1}$, $x_t := \Delta^2 X_t$.

Again we note that IF of the type $F_{b'\Delta X, x}$ present the level interpretation given for I(1) systems: they measure the change in the long-run forecast of $b'X_t$ induced by a change in $x_t$. We observe that there are several long-run adjustment coefficients to various disequilibrium errors; they appear in the second and third column in formula (6.12). One can note that timing of the EC terms used in (6.10) is perhaps not the most natural. The following subsection discusses the relation among IF obtained for the various choices of timing of the EC terms, both for the I(1) and the I(2) cases.

Finally we note that Case 4 presents lag-lead effects, because $y_t := \beta' X_{t-2} + \delta\beta_2' \Delta X_{t-1}$ and $x_t := \Delta^2 X_t$. The IF (6.12) is calculated in the empirical application in section 6.7, where the lag-lead effect is also illustrated.


DOI: 10.2870/52609

**6.4.3. Timing of the EC terms.** The choice of timing of the EC terms in an EC formulation like (6.10) is arbitrary. It is well known that in the I(1) case the level term $\beta' X_{t-j}$ can be shifted to any lag $j$, $1 \leq j \leq k$, by changing the definition of the coefficients to the variables $\Delta X_{t-1}$, ..., $\Delta X_{t-k+1}$. The same applies to the EC terms $(\beta, \beta_1)'\Delta X_{t-1}$ and $(\beta' X_{t-1} + \delta\beta_2'\Delta X_{t-1})$ in the I(2) systems: the level term $X_{t-j}$ can be shifted to any lag $j$, $1 \leq j \leq k$ and the differences $\Delta X_{t-j}$ to any lag $j$, $1 \leq j \leq k-1$. The choices made in the previous sections were only motivated by ease of calculations.

Let $Z_t$ and $V_t$ be two possible choices of the state vector $\widetilde{X}_t$ corresponding to a specific timings of the EC terms. It is simple to see that they are connected by a linear map $Z_t = NV_t$, where $N$ is square and non-singular, see examples in section 6.B. The two state vectors satisfy recursions $Z_t = A_Z Z_{t-1} + u_t$, and $V_t = A_V V_{t-1} + u_t$. Substituting $Z_t = NV_t$ in the first equation one sees that $NV_t = A_Z NV_{t-1} + u_t$ or $V_t = N^{-1}A_Z NV_{t-1} + N^{-1}u_t$, i.e. the companion matrices are related by $A_V = N^{-1}A_Z N$, or $NA_V N^{-1} = A_Z$. This implies a similar relation between the corresponding IF, which is a special case of the basic property (6.2), with $N_X = N_Y = N$.

Let $F_Z$ and $F_V$ indicate the IF calculated for state vectors $Z_t$ and $V_t$. The following proposition applies.

PROPOSITION 3 (Timing and $IF$). *One has* $F_Z = NF_V N^{-1}$ *for* $Z_t := NV_t$.

The previous proposition shows that one can transform IFs just as easily as one can redefine the timing of EC terms. A few leading examples of transformation $N$ are described in section 6.B, which collects also proofs of this subsection. Two remarks emerge from the analysis of these cases.

- The choice of timing of the EC term involves a transformation matrix $N$ that contains either known elements (0 and 1s) or cointegrating parameters, $\beta$ in the I(1) case and $\beta$, $\beta_1$, $\beta_2$ and $\delta$ in the I(2) case.
- The inverse $N^{-1}$ of $N$ is easily calculated, and often corresponds to a matrix with the same entries of $N$ with same sign on the main diagonal and opposite sign in the rest of the matrix.

It is thus possible to calculate a single set of IF and deduce other possible choices from this set. The following proposition states which of the IF are invariant with respect to the choice of lag of the EC terms.

PROPOSITION 4 (Invariance of some IF w.r.t timing of EC terms). (1) *In the* $I(1)$ *case, for any state space vector of the form*

$$\left(\Delta X_t' : X_{t-j}'\beta : U_t'\right)', \qquad j = 1, \ldots, k$$

*the IF* $F_{y_t,x_t}$ *are invariant for* $y_t = \Delta X_t$, $U_t$ *and* $x_t = \Delta X_t$, $\beta' X_{t-j}$.

(2) *In the* $I(2)$ *case for any state space vector of the form*

$$\left(\Delta^2 X_t' : \Delta X_{t-i}'\beta : \Delta X_{t-j}'\beta_1 : X_{t-l}'\beta + \Delta X_{t-m}'\beta_2\delta' : W_t'\right)', \\ i, j, m = 1, \ldots, k-2, \ l = 1, \ldots, k-1$$

*the IF* $F_{y_t,x_t}$ *are invariant for* $y_t = \Delta^2 X_t$, $W_t$ *and* $x_t = \Delta^2 X_t$, $\beta_1'\Delta X_{t-j}$, $\beta' X_{t-l} + \delta\beta_2'\Delta X_{t-m}$.

This shows that some IFs are invariant w.r.t choice of lags. Other IFs are not. Note that in the I(1) case the long-run adjustment coefficient $F_{\Delta X_t, \beta X_{t-j}}$ is invariant. In the I(2) case the long-run adjustment coefficient for the multicointegration relation $F_{\Delta^2 X_t, \beta' X_{t-l} + \delta\beta_2'\Delta X_{t-m}}$ is also invariant. Note that one other long-run adjustment coefficient $F_{\Delta^2 X_t, \beta_1'\Delta X_{t-j}}$ is invariant, whereas the last one $F_{\Delta^2 X_t, \beta'\Delta X_{t-j}}$ is not.

 DOI: 10.2870/52609

## 6.5. Selected areas of application

This section reports two possible areas of application of the notion of IF. They regard the effectiveness of economic policy in the long run and the impact of data revisions on forecasts.

**6.5.1. Policy effectiveness.** The analysis of IF can be used in the context of policy analysis, when the FS statistic $\widetilde{X}_t$ includes variables that may be influenced, directly or indirectly, by economic policy.

Sometimes the policy maker (PM) may be able to set the value of some instrument variable, as in the case of government expenditure or tax rates. This case may be called the one of an *ideal* instrument. More often, the PM may be in the position to influence the value of some economic variable, at the margin. In the case of short-term interest rates, for instance, open market operation on the market of short-term bills by the PM marginally influence the value of the market interest rate. In this case one may think that the PM cannot set the value of the variable, but it can add a small positive or negative perturbation to its value. We call this case the one of a *partial* instrument.

If perturbations of the FS statistic $\widetilde{X}_t$ can be hypothetically induced by policy interventions, both in the case of ideal and partial instruments, the IF captures the long-run response of the forecasted variables to this type of intervention. If some perturbation induced by policy action does not affect the cumulated forecast on some 'target' variable included in the forecast variables $Y_{t+h}$, this means that the policy is ineffective in the long run.

Therefore it appears of importance to test if some IF are significantly different from zero. In this interpretation, insignificant IF correspond to ineffectiveness of policies. If the system is I(1) and the target variable is the growth rate of some non-stationary variable, policy ineffectiveness is measured with respect to the long-run forecast of the level of the target variable, see section 6.4.1.

The superposition principle for linear forecast functions applied here implies that one can restrict attention to single perturbations $\widetilde{v}$. We observe that the perturbations $\widetilde{v}$ may involve variables at different points in time: for the policy intervention interpretation to apply, one needs to restrict attention to perturbations $\widetilde{v}$ that regard the most recent time subscript, i.e. of the form $\widetilde{v} = Jv$, where $J := (I_p : 0)'$. This type of perturbation corresponds to a factual experiment, in which some variables (instruments) are affected by policy. We hence call this type of perturbation *factual*.

On the contrary all perturbations $\widetilde{v}$ that are not of the form $Jv$ are *counterfactual*, in the sense that they cannot possibly be obtained by single policy actions, which affect variables at a single point in time. The counterfactual perturbations correspond to a thought experiment where variables at different lags are perturbed simultaneously. Given the superposition property, see section 6.3.6, these counterfactual experiments correspond to multi-period policy actions.

Some economists may question the possibility to choose *any* perturbation $\widetilde{v}$, and in particular the ones that just select one input variable at the time, without perturbing the other input variables with the same time subscript. This question regards the ceteris paribus condition, and prompts an analysis of the simultaneity structure of the system.

In some cases it is possible that, except for the instrument, all other variables have a delayed response, so that the PM may assume the ceteris paribus condition, and simply consider a perturbation to the instrument variable. This may be the case, for instance, in monthly money demand systems which include one interest rate and quantity variables like money, prices and GDP, if the re-balance in the money demand schedule and in the other quantity variables induced by marginal changes in the interest rate take longer than one month to materialize.

In some other instances, the simultaneity structure of the economy may be non-trivial, and must be taken into proper account for policy analysis. Consider e.g. a factual perturbation $\widetilde{v} = Jv$.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

Assume that the simultaneity in the system can be represented by $v = H^*u$, where $u$ are the perturbations for which the ceteris paribus condition is satisfied, and $H^*$ reflects the simultaneous reaction of the variables in the system. Then the effectiveness of the associated policy action may be measured through $F\widetilde{v} = FJH^*u$. Hence failure to reject that an appropriate entry in $FJH^*$ is zero is an indication of ineffectiveness of the associated policy in the long run. Note that $FJH^*$ is a simple function of the IF $F$. Similar remarks apply to the case of counterfactual perturbations.

Summarizing, the significance analysis of the IF, or of simple transformations of the IF, can serve to evaluate the effectiveness of policy interventions. Inference on the IF is treated in section 6.6, where we also cover the case of simultaneity coefficients as in $FJH^*$. In section 6.7 we also report a discussion of the policy interpretation of the empirical results.

**6.5.2. Data revisions.** In this section we discuss two different interpretations of the IF in case of data revision. The first one is of numerical type, and regards the numerical variation of long-run forecasts due to data revisions. The second one is probabilistic, and regards the distribution of long-run forecasts induced by uncertainty on the data used in forecasting.[6] We present them in turn.

Several macroeconomic indicators are first published in preliminary form, and next adjusted, e.g. on the basis of national accounts available at the end of the year. The perturbations $\widetilde{v}$ may be interpreted as induced by data revisions. Because the FS statistic $\widetilde{X}_t$ containing preliminary data is used in the forecast function as input in order to produce preliminary forecasts of major macroeconomic aggregates, IF can be interpreted as a sensitivity measure of the cumulated forecast profile to revisions in the data.

In this interpretation, the value of the IF is the multiplier applied to data changes to obtain changes in the long run forecasts. When a single IF is greater (respectively less) than one in absolute value, data revisions have amplified (respectively damped) effects on the long-run forecasts. This interpretation may be combined with the fact that, when $y_t$ consists of variables in first differences, IFs measure the sensitivity of the long-run forecast of the corresponding levels. A similar comment applies to growth rates in I(2) systems.

We next discuss how the IF can be used to evaluate long-run forecast uncertainty induced by data revisions. We start from a simple example, which is then generalized. Assume that the econometrician is interested in the forecast of the growth rate of GDP, $y_{t+i} := \Delta DGP_{t+i} = b'Y_{t+i}$. Let the forecast be based on a VAR specification, which has been estimated on the final data for the previous years, while the forecast is based on preliminary data of the current year.

For simplicity assume that a single variable $x_t$ is affected by data revisions, $x_t := a'\widetilde{X}_t$, where $a$ is a selection vector. The relevant entry of the TE, indicated as $c := b'\text{TE}(a'\widetilde{v}) = b'\text{TE}(\widetilde{v})\bar{a}$, is equal to $DGP^c_{\infty|t} - DGP_{\infty|t}$, the change in the long-run forecast of $DGP$ due to the data revision process. Before the actual data revision, $h := a'\widetilde{v}$ is a random variable, and so is $c$, the induced long-run forecast uncertainty. From (6.1) one finds $c = b'\text{TE}(a'\widetilde{v}) = b'F\bar{a} \cdot a'\widetilde{v} = F_{y_t,x_t} \cdot h$, which is a linear function of $h := a'\widetilde{v}$.

Since the data revision process is systematic, the econometrician may assume that $h := a'\widetilde{v}$ follows some distribution, like $N(0, \sigma^2)$. It then follows that $c \sim N(0, F^2_{y_t,x_t}\sigma^2)$. Note that $F_{y_t,x_t}$ is the relevant entry of the IF, which acts as a multiplier of the standard deviation: the standard deviation of data-revision uncertainty is multiplied by $F_{y_t,x_t}$ to obtain the standard deviation of long-run forecast uncertainty about the level of DGP.

This illustrates how the IF also conveys information on the impact of data revisions on the long-run forecast uncertainty associated with the data revision process. This example can be generalized as shown below, where we do not strive for maximal generality, but rather wish to illustrate some of the possible extensions.

[6] The latter topic was suggested by one of the referees.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

Let $b'Y_t$ and $h = a'\widetilde{v}$ be vectors of the same dimensions, and assume a normal distribution for data revisions, $h \sim N(0, \Sigma_h)$. Following the same steps as above, one finds a normal distribution for forecast uncertainty $c = Fh \sim N(0, F\Sigma_h F')$, where $F$ is the IF.

More generally let $h$ present density $f_h(v)$, $v \in D$, w.r.t. Lebesque measure, not necessarily Gaussian. Let also $c := H(h) = b'\text{TE}(h)$ be an injective function of $h$, such that the absolute Jacobian of $H$, denoted $J_H(v) := |\det(F(v))|$, is positive for all $v$. Note that $J_H(v)$ is a function of the IF $F$. In the case of constant IF, like in the VAR case, the condition of non-singular Jacobian for all $v$ reduces to a rank condition on the IF $F$.

Let $f_c$ indicate the induced density of $c$. Standard change of variable formula, see e.g. Hoffmann-Jørgensen (1994) eq. (8.2.2), yields

$$f_c(u) = \frac{f_h(H^{-1}(u))}{J_H(H^{-1}(u))} 1(u \in H(D)).$$

When $H$ is not injective over all $D$ but it is so over a partition of $D$, a similar formula applies with summation over the partition. The discussion can be extended to the case of transformations $H$ with different number of elements in $h$ and $c$, with countable numbers of points of non-differentiability and/or of singularity of $F(v)$, along the lines of chapter 8.6 in Hoffmann-Jørgensen (1994). We do not report details here for space constraints.

Hence the IF summarizes the relevant information to describe the relation between the data uncertainty distribution $f_h$ and the forecast uncertainty distribution $f_c$. Overall, there appears to be a vast potential for applications of IF to the field of forecast uncertainty evaluation.

## 6.6. Inference on the IF

In this section we consider inference on IF in a unified framework for stationary, I(1) and I(2) systems. The approach is based on the observation that CI parameters are estimated super-consistently. This implies that the inclusion of estimated CI parameters in the definition of regressors does not affect the limit distribution of the IF. Inference on the IF is associated with the one on the companion matrix $A$. This matrix is estimated below through a specific regression system, which is specified in the next subsection for the I(0), I(1) and I(2) cases. In subsection 6.6.2, we then address the issue of inference on the IF $F$, which is calculated as $(I - A)^{-1} - I$.

**6.6.1. Regression setup.** In order to estimate the IF, one needs to estimate the companion matrix $A$. We define $G^* := J'A$ and $L := J_\perp' A$, where $J := (I_p : 0)'$ and $J_\perp = (0 : I)$. The matrix $G^*$ contains the adjustment coefficients, while $L$ contains only known values, 0 or 1, and CI parameters in the integrated cases. The matrix $A$ is then reconstructed as $A = (G^{*\prime} : L')'$.

In the stationary case let $X_{0t} := J'\widetilde{X}_t = X_t$ be the regression dependent variable and $X_{1t} := (X_{t-1}' : ... : X_{t-k}')'$ be the matrix of stochastic regressors. For homogeneity with the integrated cases we assume that $\mu_1 = 0$, so that the system equations can be written as

$$X_{0t} = GX_{1t} + \mu D_t + \epsilon_t, \tag{6.13}$$

where $G := (A_1 : ... : A_k)$. The likelihood analysis of the stationary VAR in (6.13) is simply performed by OLS. For later reference we also set $H := I$, $G^* := G$, $\widehat{X}_{1t} := X_{1t}$.

Consider now the integrated cases. The I(1) cointegration analysis with the deterministic specification used above is described in Johansen (1995b), while the corresponding one for the I(2) model is described in Rahbek et al. (1999)[7].

Consider the I(1) case. Let $X_{0t} := J'\widetilde{X}_t = \Delta X_t$ be the regression dependent variable. The I(1) analysis permits to determine the CI rank $r$ and to estimate $\beta^* := (\beta' : \beta_0')'$. These estimates

[7]The estimation of the cointegrating coefficients can be accomplished via likelihood techniques in I(1) and I(2) systems or via the 2SI2 procedure in I(2) systems, see Johansen (1995c), Paruolo (1996), Rahbek et al. (1999).

 DOI: 10.2870/52609

are (at least) $T$-consistent, see Johansen (1995b). The estimate of $\beta^*$ permits to calculate the regressor vector $\widehat{X}_{1t}^* := (\Delta X_{t-1}' : (\widehat{\beta}' X_{t-2} + \widehat{\beta}_0' t)' : U_{t-1}')$, and eq. (6.5) can be rewritten as

$$X_{0t} = G^* \widehat{X}_{1t}^* + \mu D_t + \widehat{\epsilon}_t \tag{6.14}$$

where $G^* = (\Gamma_1^* : \alpha : \Phi)$ and $\widehat{\epsilon}_t := \epsilon_t - \alpha((\widehat{\beta} - \beta)' X_{t-2} + (\widehat{\beta}_0 - \beta_0)' t)'$ is the error term. Here and in the following we indicate with $\widehat{}$ quantities where the CI coefficients have been substituted with their estimators.

In the special case $k = 1$ listed in section 6.C, $G^*$ has reduced rank because of the reduced rank of $A = \widetilde{A} H'$, $G^* := J' A = J' \widetilde{A} H'$. In this case define $\widehat{X}_{1t} := \widehat{H}' \widehat{X}_{1t}^*$, $G := J' \widetilde{A}$; otherwise we let $H = I$ and define $G := G^*$, $\widehat{X}_{1t} := \widehat{X}_{1t}^*$. Eq (6.14) then reads

$$X_{0t} = G \widehat{X}_{1t} + \mu D_t + \widehat{\epsilon}_t. \tag{6.15}$$

Consider the I(2) case. We define $G^* := J' A$ and let $X_{0t} := J' \widetilde{X}_t = \Delta^2 X_t$ be the regression dependent variable. The I(2) analysis permits to determine the integration indices $r$ and $s$ and to estimate $\beta^* := (\beta' : \beta_0')'$ and $\delta$, $\beta_1$, $\beta_2$. These estimates are (at least) $T$-consistent, see Johansen (1997) and Paruolo (2000). Substituting these estimates, one obtains the regressor vector $\widehat{X}_{1t}^* := (\Delta^2 X_{t-1}' : \Delta X_{t-2}' (\widehat{\beta} : \widehat{\beta}_1) : (\widehat{\beta}' X_{t-3} + \widehat{\beta}_0' t + \widehat{\delta} \widehat{\beta}_2' \Delta X_{t-2})' : W_{t-1}')$, and eq. (6.10) can be rewritten as (6.14) where $G^* = (\Upsilon_1^* : \zeta_1^* : \zeta_2 : \alpha : \Phi)$ and the error term $\widehat{\epsilon}_t$ depends on $\epsilon_t$ and on the estimation error of the CI parameters.

In the special case $k = 2$ listed in section 6.C, $G^*$ has reduced rank because of the reduced rank of $A = \widetilde{A} H'$, $G^* := J' A = J' \widetilde{A} H'$. In this case define $\widehat{X}_{1t} := \widehat{H}' \widehat{X}_{1t}^*$, $G := J' \widetilde{A}$; otherwise we let $H = I$ and define $G := G^*$, $\widehat{X}_{1t} := \widehat{X}_{1t}^*$. Eq (6.14) can then be transformed in (6.15) as in the I(1) case.

In all cases the matrix $A$ is then reconstructed as

$$A = \begin{pmatrix} G^* \\ L \end{pmatrix} = \begin{pmatrix} GH' \\ L \end{pmatrix}.$$

**6.6.2. Inference.** Equation (6.15) is the regression equation on which we base inference on the IF. For fixed CI coefficients, the ML estimates of $G$ and $\Omega$ are computed by OLS,

$$\widehat{G} = \widehat{S}_{01} \widehat{S}_{11}^{-1} \qquad \widehat{\Omega} = \widehat{S}_{00.1} := \widehat{S}_{00} - \widehat{S}_{01} \widehat{S}_{11}^{-1} \widehat{S}_{10},$$

where $S_{ij} := T^{-1} \sum_{t=1}^T R_{it} R_{jt}'$, $R_{it} := X_{it} - M_{iD} M_{DD}^{-1} D_t$, $M_{iD} := T^{-1} \sum_{t=1}^T X_{it} D_t'$, $M_{DD} := T^{-1} \sum_{t=1}^T D_t D_t'$, and $\widehat{}$ indicates quantities where the CI coefficients have been substituted with their estimators. Similarly $\widehat{H}$ and $\widehat{L}$ indicate the $H$ and $L$ matrices with CI coefficients have been substituted with their estimators.

The expressions of the regression estimators for the stationary case in (6.13) are identical, but obviously do not involve moments with pre-estimated CI coefficients. An analogous comment applies to the $H$ and $L$ matrices.

The corresponding estimate of $A$ is

$$\widehat{A} = \begin{pmatrix} \widehat{G} \widehat{H}' \\ \widehat{L} \end{pmatrix}$$

and $\widehat{F} = (I - \widehat{A})^{-1} - I$. We next introduce some notation. Let $Z_{1t} := H' X_{1t}^*$ and $\Sigma := E((Z_{1t} - E(Z_{1t}))((Z_{1t} - E(Z_{1t}))'$

The following theorem states the relevant limit distributions for inference on the impact factors.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

THEOREM 13 (limit distribution of IF). *In the $I(1)$ and $I(2)$ cases the estimator $\widehat{H}$ and $\widehat{L}$ are superconsistent, i.e. $\widehat{H}-H$, $\widehat{L}-L \in O_p(T^{-1})$. In the $I(0)$, $I(1)$ and $I(2)$ cases the estimator of the adjustment coefficients $\widehat{G}$ is $T^{1/2}$-consistent and has a Gaussian limit distribution*

$$T^{1/2} vec(\widehat{G}' - G') \xrightarrow{w} N(0, \Omega \otimes \Sigma^{-1}). \tag{6.16}$$

*Moreover*

$$T^{1/2} vec(\widehat{F}' - F') \xrightarrow{w} N(0, \Psi)$$

*with*

$$\Psi := KJ\Omega J'K' \otimes K'H\Sigma^{-1}H'K, \tag{6.17}$$

*where $K := (I-A)^{-1}$. The asymptotic covariance matrix of the impact factors can be estimated consistently by substituting parameter matrices with their regression-based consistent estimators, $\widehat{\Sigma} := \widehat{S}_{11}$, $\widehat{K} = \widehat{F} + I$, $\widehat{\Omega} = \widehat{S}_{00.1}$ within* (6.17).

We observe that the asymptotic covariance matrix of $F$ is singular. This singularity is due to several factors. The first source of singularity is due to the fact that $L$ is known in the I(0) case and it is estimated superconsistently in the integrated cases. This singularity is reflected in the matrix $J := (I : 0)'$ in the expression of the asymptotic covariance matrix. A similar phenomenon appears in connection with $H$ for the special cases where $H$ is not the identity matrix.

Other singularities are associated with the singularities of the matrix $C$ in the I(1) case and of $C_2$ in the I(2) cases. Instead of focusing on these cases we refer to Paruolo (1997a,b) for inference on $C$ and to Paruolo (2002a) for inference on $C_2$.

The results in the theorem allow to define Wald-type statistics for individual IF. For simple hypotheses of the type $F_{ij} = c$, for instance, if the corresponding asymptotic variance $\sigma^2$ is non-zero, one can define an asymptotically $\chi^2(1)$ statistic $(\widehat{F}_{ij} - c)^2/\widehat{\sigma}^2$, or the corresponding asymptotic $N(0,1)$ statistic $(\widehat{F}_{ij} - c)/\widehat{\sigma}$. These statistics are illustrated with an application in the section 6.7.

**6.6.3. Modification for IR.** In this subsection we briefly sketch how the present results can be modified and applied to the case of cumulated IR with IF $FJH^*$, see section 6.3.5. We assume that $\widehat{H}^*$ is a function of the unrestricted variance estimator $\widehat{\Omega} := M_{\widehat{\epsilon}\widehat{\epsilon}}$, which is used to decompose shocks. Let $\Sigma_*$ be its asymptotic variance, $T^{1/2}(\widehat{H}^* - H^*) \xrightarrow{w} N(0, \Sigma_*)$.

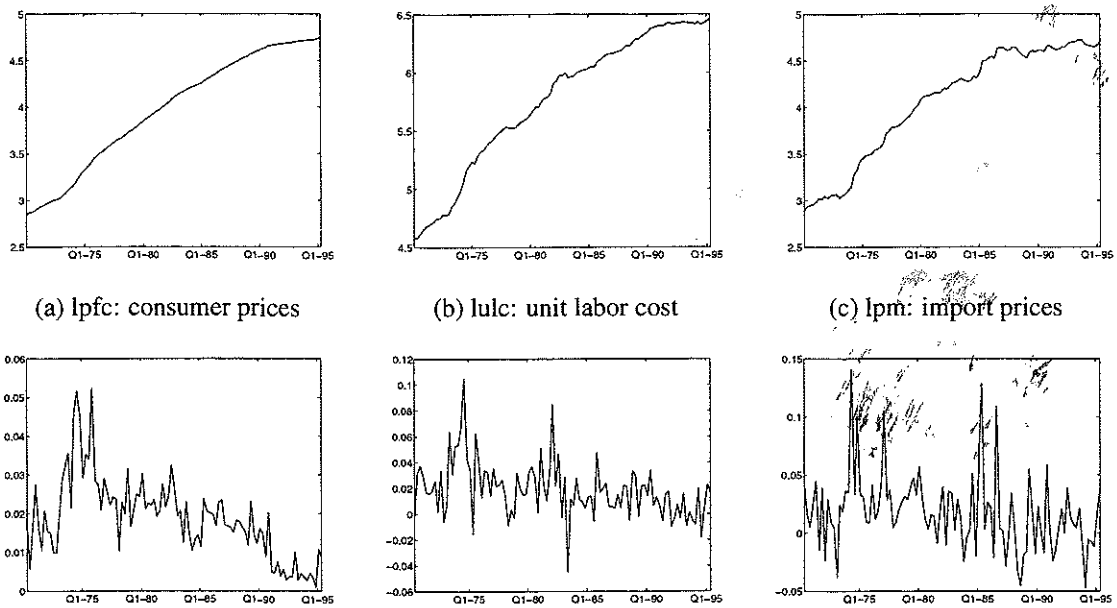
As noted e.g. in Paruolo (1997b), one has

$$\begin{aligned} T^{1/2}(\widehat{F}J\widehat{H}^* - FJH^*) &= T^{1/2}(\widehat{F} - F)JH^* + T^{1/2}FJ(\widehat{H}^* - H^*) + o_p(1) \\ &= a + b + o_p(1), \text{ say} \end{aligned}$$

where $a$ and $b$ are asymptotically independent. It is simple to see that the asymptotic variance of the r.h.s. equals $Q_1'\Psi Q_1 + Q_2'\Sigma_* Q_2$ where $Q_1' := (H^{*\prime}J' \otimes I)$, $Q_2' := (I \otimes FJ)$, which can be consistently estimated substituting consistent estimators for the parameters.

Hence the asymptotic standard error of a cumulated IR depends on the asymptotic variance of the IF, which can be calculated as detailed in theorem 13. The second term depends on $\Sigma_*$, which varies according to the definition of $H^*$ i.e. the type of IR. This is not treated further here, but we refer to Pesaran and Shin (1998), Koop et al. (1996), Amisano and Giannini (1997).

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609



(a) lpfc: consumer prices (b) lulc: unit labor cost (c) lpm: import prices

(d) Δlpfc (e) Δlulc (f) Δlpm

Figure 6.1: Australian data in levels and differences

## 6.7. An illustration: price mark-up in Australia

As an illustrative example of IFs, we consider the data set analyzed by Banerjee et al. (2001).[8] It consists of three Australian macroeconomic data series: the consumer price deflator at factor cost (lpfc), unit labor costs in the non-farm sector (lulc) and import prices (lpm). All three variables are quarterly data measured in natural logs, and run from 1970Q1 to 1995Q2 for a total of 102 observations. The variables are graphed in levels and first differences in Fig. 6.1. The levels of the variables appear non-stationary, and also the differences show signs of possible non-stationarity. No apparent break in the deterministic terms is visible in Fig. 6.1.

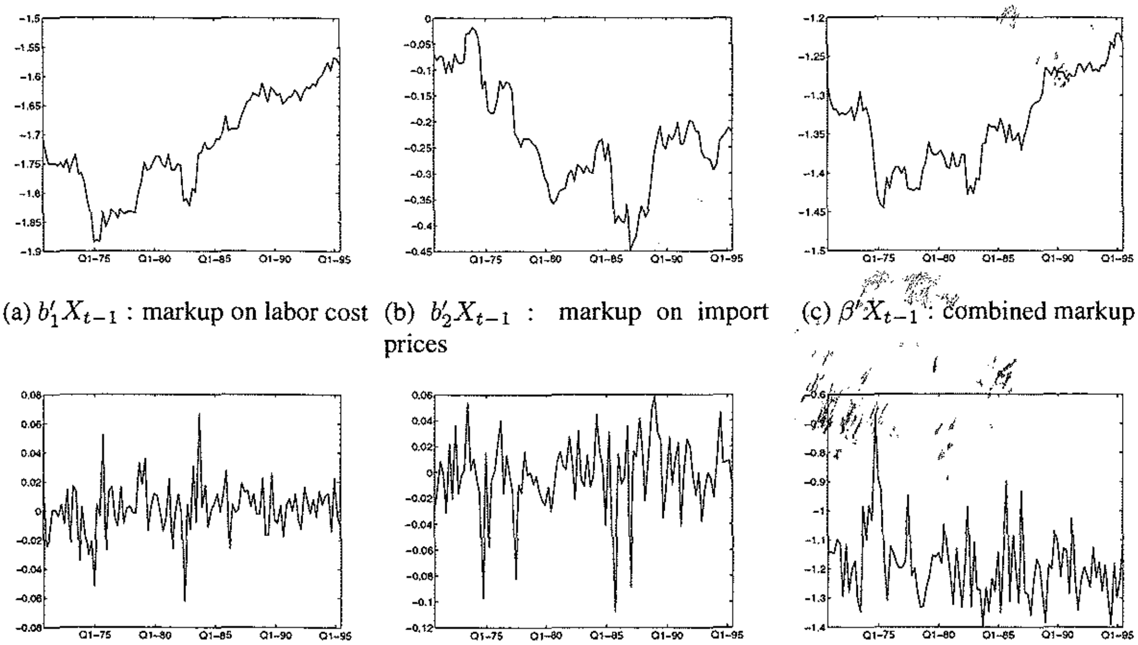
We include dummy variables to take account of a number of shocks to the economy, like the oil shocks. The dummies take value 1 in one quarter and zero otherwise; the quarters are 1974Q2, 1974Q3, 1975Q2, 1982Q1, 1983Q2, 1985Q2 and 1986Q3.[9] We fit an unrestricted VAR in levels with $k = 2$ lags, centered seasonal dummies, a constant and a trend. We employ the package Me2, enclosed with this thesis, which performs maximum likelihood (ML) analysis for the I(2) models.

We next perform some mis-specification tests for normality and autocorrelation of the errors proposed by Doornik and Hansen (1994) and Doornik (1996). The normality test statistic is equal to 8.25 with a $p$-value of 0.22; the AR1 and AR4 test statistics are equal to 2.72 and 35.99, with $p$-values equal to 0.97 and 0.47. ARCH tests on residuals do not reject the null of homoskedasticity. These results indicate that the model appears to be well specified. We next test for the degree of integration of the system, allowing for the possibility of I(2). The resulting cointegration analysis is presented in subsection 6.7.1.

[8]The data set is available at the data archive of the Journal of Applied Econometrics: http://qed.econ.queensu.ca/jae.

[9]Banerjee et al. (2001) conditioned on a number of stationary variables we do not consider here. Their selection of integration indices is the same as the one reached here; moreover we do not reject the nominal-to-real transformation, as in their paper.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609




(a) $b_1' X_{t-1}$ : markup on labor cost (b) $b_2' X_{t-1}$ : markup on import prices (c) $\beta' X_{t-1}$ : combined markup

(d) $b_1' \Delta X_{t-1}$ (e) $b_2' \Delta X_{t-1}$ (f) $mec_{t-1}$, multicointegration relation

Figure 6.2: Cointegration and multicointegration relationships

During the period under study, the Australian economy moved from a fixed to a floating exchange rate regime and from a national-award-based wage system to a localized system. In order to check for the possible breaks in the model, we perform the Andrews (1993) tests imposing the estimated error correction terms from the CI analysis. Results are reported at the end of subsection 6.7.1; they imply the non-rejection of the hypothesis of no breaks. The IF are then calculated. The estimates are reported in subsection 6.7.2.

**6.7.1. Cointegration analysis.** Since I(1) behavior of the growth rates implies that the levels are I(2), see Fig. 6.1, we leave open the possibility to select an I(2) model for the data. We first test for the number of unit roots allowing both I(1) and I(2) behavior, by selecting the integration indices of the system. This analysis considers all I(1) and I(2) submodels of the unrestricted VAR.

The selection of the integration indices is based on the 2SI2 estimator (Johansen, 1995c; Paruolo, 1996; Rahbek et al., 1999); the test statistics for the specification $\mu_1 = \alpha\beta_0'$ of Rahbek et al. (1999) are reported in table 6.2. Below each entry we report the 95% quantile of the asymptotic distribution, taken from Rahbek et al. (1999). We select $(r, s) = (1, 1)$, which corresponds to one I(1) trend and one I(2) trend. The roots of the characteristics polynomial are 1, 1, 1, 0.38, $-0.21$ and 0.11; there is no evidence of additional non-stationary trends.[10] The same integration indices were selected by Banerjee et al. (2001).

We test the nominal-to-real transformation (Kongsted, 1998, 2002), i.e that lpfc$-$lulc (the markup of internal prices on unit labor cost) and lpfc$-$lpm (the markup of price over import prices) are at most I(1). We use the likelihood ratio statistic; under the null the test has an asymptotic $\chi^2(2)$-distribution, see Johansen (2002c). The test statistic takes the value 0.935, with a $p$-value of 0.63, giving ample support to the transformation. This implies that $\beta = b\rho$, and

[10]The roots of the unrestricted polynomial are $1.00$, $0.89 \pm 0.02i$, $0.41$, $-0.22$ and $0.14$.

 DOI: 10.2870/52609

| $p-r$ | $r$ | | | | |
|---|---|---|---|---|---|
| 3 | 0 | 279.3 (87.6) | 158.6 (68.2) | 74.7 (53.2) | 53.0 (42.7) |
| 2 | 1 | | 117.8 (47.6) | **31.8** (34.4) | 23.7 (25.4) |
| 1 | 2 | | | 17.5 (19.9) | 10.8 (12.5) |
| $p_2$ | | 3 | 2 | 1 | 0 |

Table 6.2: 2SI2 inference on the integration indices $r, s$. $p_2 := p - r - s$. The sequence of tests starts from the upper left corner to the lower right corner, proceeding row-wise from left to right. The first unrejected model is shown in boldface.

| | | | lpfc | lupc | lpm | $t$ |
|---|---|---|---|---|---|---|
| $\rho$ | 0.7423 | $b'$ | 1 | $-1$ | 0 | 0.0013 |
| | 0.2577 | | 1 | 0 | $-1$ | -0.0029 |
| $\delta$ | 2.6760 | $\beta_2'$ | 1 | 1 | 1 | |

Table 6.3: Estimates of the cointegration parameters under the nominal to real transformation; $b$ is a basis of $sp(\tau)$, $\beta = b\rho$.

$\beta_2 = b_\perp = (1:1:1)'$, where

$$b := \begin{pmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The maximum likelihood estimates of the cointegration parameters are reported in table 6.3. The $CI(2,1)$ relations, that is the cointegration relations from $I(2)$ to $I(1)$, are the two markups, pictured in figure 6.2; they appear I(1). The combined mark-up on price $\beta$, obtained as a linear combination of the two, $\widehat{\beta} = b\widehat{\rho}$, is also $I(1)$:

$$\widehat{\beta}'X_{t-2} + \widehat{\beta}_0 t = \text{lpfc}_{t-2} - 0.74\text{lulc}_{t-2} - 0.26\text{lpm}_{t-2} + 0.0013t.$$

The remaining $CI(2,1)$ relationship $\widehat{\beta}_1 = \bar{b}\widehat{\rho}_\perp$ is also I(1), where

$$\widehat{\beta}_1'X_{t-2} = -0.28\text{lpfc}_{t-2} - 0.72\text{lulc}_{t-2} + \text{lpm}_{t-2}.$$

The fact that the combined mark-up, $\widehat{\beta}'X_t$, is still I(1) by itself is consistent with imperfect competition theories, which predict that a high mark-up is associated with low inflation.[11] The combined markup $\widehat{\beta}'X_t$ next cointegrates with the I(1) trend in the first differences, represented by $\widehat{\beta}_2'\Delta X_t = (1:1:1)\Delta X_t$, proportional to the average inflation in the 3 series. This gives the following stationary multicointegration relationship

$$(6.18)\qquad mec_t = \widehat{\beta}'X_{t-2} + \widehat{\delta}\widehat{\beta}_2'\Delta X_{t-1} + \widehat{\beta}_0't = \text{lpfc}_{t-2} - 0.74\text{lulc}_{t-2} - 0.26\text{lpm}_{t-2} + 2.68\left(\Delta\text{lpfc}_{t-1} + \Delta\text{lulc}_{t-1} + \Delta\text{lpm}_{t-1}\right) + 0.0013t.$$

This multicointegration relation represents a *compensated* markup relation, where the markup of internal prices over labor cost and imports depends negatively on the average inflation in the three series: high average inflation is associated with low markups and vice versa.

[11]For a full overview of the economic theory, we refer to Banerjee et al. (2001).

 DOI: 10.2870/52609

| | $\Delta^2$lpfc$_t$ | $\Delta^2$lulc$_t$ | $\Delta^2$lpm$_t$ | $\beta'\Delta X_{t-1}$ | $\beta_1'\Delta X_{t-1}$ | $mec_t$ |
|---|---|---|---|---|---|---|
| $\Delta^2$lpfc$_t$ | 0.00 (0.10) | 0.08 (13.02) | 0.02 (3.57) | −0.77 (−29.57) | 0.15 (17.58) | −0.11 (−28.08) |
| $\Delta^2$lulc$_t$ | 0.00 (0.10) | 0.08 (13.02) | 0.02 (3.57) | 0.31 (11.92) | 0.43 (50.42) | −0.11 (−28.08) |
| $\Delta^2$lpm$_t$ | 0.00 (0.10) | 0.08 (13.02) | 0.02 (3.57) | 0.01 (0.38) | −0.65 (−77.02) | −0.11 (−28.08) |
| $\beta'\Delta X_{t-1}$ | −0.03 (−0.11) | −0.65 (−13.02) | −0.15 (−3.57) | 1.20 (5.75) | 0.20 (3.01) | −0.10 (−3.21) |
| $\beta_1'\Delta X_{t-1}$ | −0.18 (−0.26) | −0.57 (−4.50) | 1.06 (9.93) | 0.02 (0.04) | 1.95 (11.26) | 0.04 (0.46) |
| $mec_t$ | 14.56 (5.02) | 2.67 (5.12) | 1.19 (2.73) | 9.88 (4.52) | −2.99 (−4.20) | 3.29 (9.83) |

Table 6.4: $K := F + I$: Impact factors $(+I)$ in the Australian mark-up model. $F$ is calculated using the restricted estimates in eq. (6.18) and in the preceding two displays. $t$-values are reported in brackets.

We check for the possible presence of model breaks. We test whether the structural changes (exchange rate, labor market) have changed the speed of adjustment to equilibrium. We calculate the estimated error correction terms mec$_t$, $\widehat{\beta}'X_{t-2}$, $\widehat{\beta}_1'X_{t-2}$ using the estimates of the cointegration parameters; we then perform Andrews (1993) stability test on the adjustment coefficients $(\alpha : \zeta)$. The unknown sample-fraction break-point was chosen in the range $[\pi_0, 1 - \pi_0] = [0.20, 0.80]$. The sup-LR test for breaks gives a test statistics of 23.36. The 5% critical value in table 1 in Estrella (2003)[12] for $\pi_0 = 0.2$ and dimension 9 is 25.16, which implies a non-rejection. We thus conclude that there is no evidence of breaks in the model.

**6.7.2. Impact Factors.** This subsection presents the IF calculated in the model specified in the previous subsection. We first present the interpretation of the IF as a sensitivity measure of long-run forecasts of the stationary variables in the system. This discussion also illustrates the lag-lead effects. We then present the relation of IFs with the impulse responses, and finally conclude with comments on policy effectiveness and the influence of data revisions in the present example. In the following we omit $\widehat{}$ over estimated CI coefficients.

table 6.4 reports the impact factors in the I(2) model. The first three columns in table 6.4 are the impact factors which correspond to a factual experiment. The last three columns correspond to counterfactual experiments. The lower left panel contains lag-lead effects.

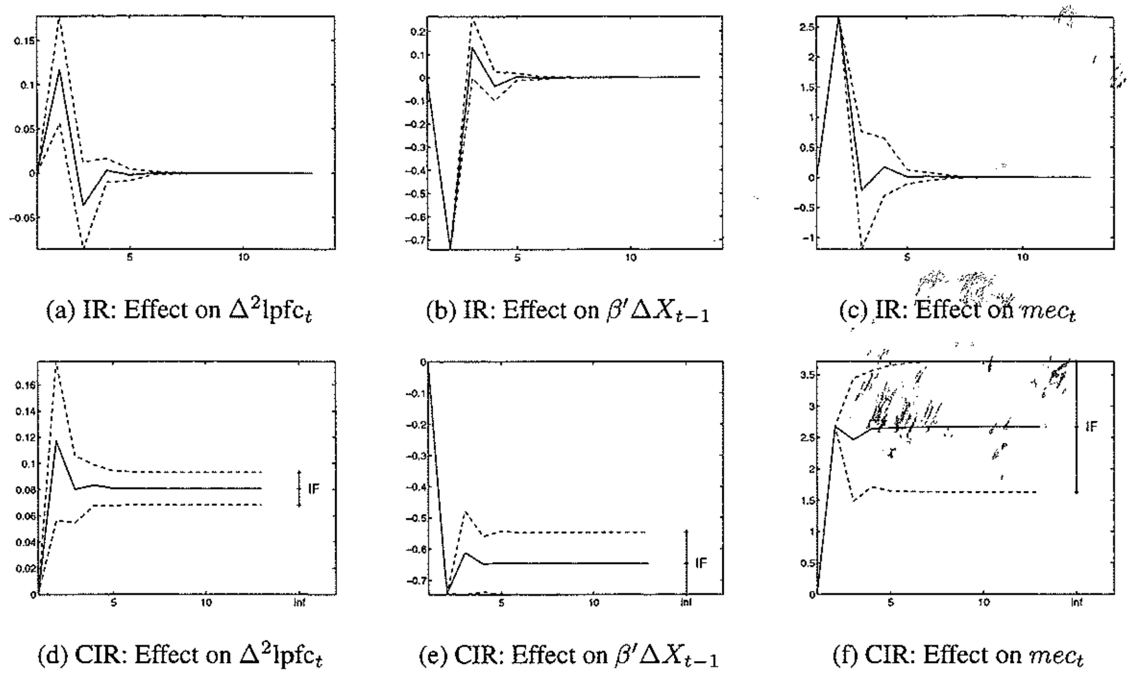
The consequences of a perturbation to the lpfc$_t$ can be read off from the first column in table 6.4. Such a perturbation does not lead to significant changes in the forecast of the growth rates of all three price series. The effect on the long run forecast of the level of the markups $\beta'X_{t-1}$ and $\beta_1'X_{t-1}$ is also insignificant.

Note however that the latter two IFs contain a lag-lead effect. In this case it can be shown that the lag-lead effects have zero sum, because the forecasted variables $\beta'\Delta X_{t-1}$ and $\beta_1'\Delta X_{t-1}$ are in first differences. As an illustration consider $y_t := \beta'\Delta X_{t-1}$ and $x_t := \Delta^2$lpfc$_t$, and note that $y_{t+1} := \beta'\Delta X_t$. The forecast $E(y_{t+1}|X_t)$ is a.s. equal to $\beta'\Delta X_t$ itself, so that $e_1$ contains a contribution equal to 1 not associated with forecasting, due to the coefficient of 1 of $\Delta$lpfc$_t$ in $\beta'\Delta X_t$. At forecast horizon 2, $y_{t+2} := \beta'\Delta X_{t+1}$ includes a negative effect of the same magnitude through the presence of $-\beta'X_t$ in $y_{t+2}$. Hence the net lag-lead effects is zero.

The IF of $\Delta^2$lpfc$_t$ on the compensated markup (6.18) is positive and significant, in line with economic expectations. Also in this case one can discount lag-lead effects. Let $y_{t+1} := \beta'X_{t-1} + 2.68 \cdot i_3'\Delta X_t$. At forecast horizon 1, a perturbation to $\Delta^2$lpfc$_t$ induces an equal change within

[12]See also Andrews (2003).

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609



(a) IR: Effect on $\Delta^2$lpfc$_t$ (b) IR: Effect on $\beta'\Delta X_{t-1}$ (c) IR: Effect on $mec_t$

(d) CIR: Effect on $\Delta^2$lpfc$_t$ (e) CIR: Effect on $\beta'\Delta X_{t-1}$ (f) CIR: Effect on $mec_t$

Figure 6.3: Effect of perturbation to unit labor cost (lulc): Impulse Response functions (IR:top) and Cumulated Impulse Response functions (CIR: bottom) with 95% confidence intervals. Impact Factors with 95% confidence interval are given along CIR at horizon $\infty$, indicated as 'inf'. (b) (c) (e) (f) present the lag-lead effect.

$i_3'\Delta X_t$, and this gives a direct contribution of $+2.68$ to $e_1$. At forecast horizon 2 one has $y_{t+2} := \beta' X_t + 2.68 \cdot i_3'\Delta X_{t+1}$, the same perturbation induces a change equal to $+1$ to lpfc$_t$ within $\beta' X_t$ and a change equal to $-2.68$ due to the coefficient 2.68 and the $-1$ change induced through $i_3'\Delta X_{t+1}$.

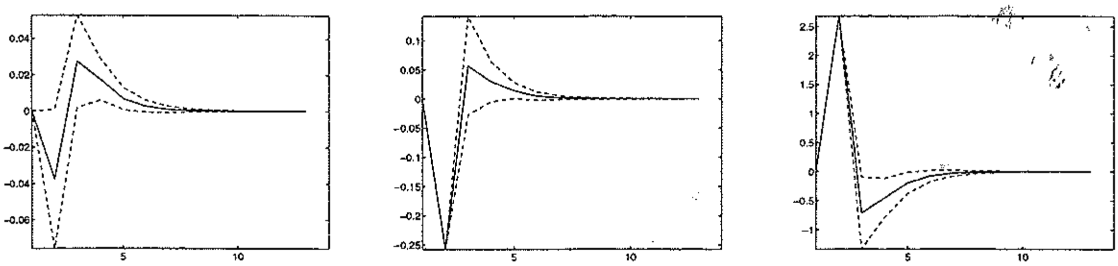
This gives a contribution of $-1.68$ to $e_2$ due to lag-lead effects. Summing one finds $+2.68 - 1.68 = 1$. Subtracting 1 from the relevant IF, one obtains a sensitivity measure of 13.56 net of the lag-lead effects, with a significant $t$-ratio of 4.68. Hence, also when correcting for lag-lead effects, the sensitivity of forecasts of the compensated mark-up to changes to lpfc$_t$ is marked and positive. A unit change to lpfc$_t$ is amplified by 13.56 over all future forecasts of the compensated mark-up.

If the forecast changes are due to data revisions, this high coefficients measures the impact of the use of imprecise measurements of lpfc$_t$ in the forecasts of mec$_t$. More accurate input data on lpfc$_t$ would thus induce a high improvement in the forecasts of mec$_t$.

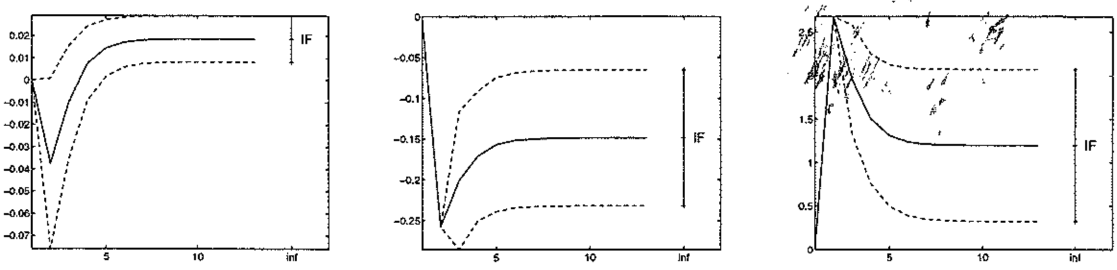
We next comment the impact factors collected in the second and third column in table 6.4, and their association with the IR and the Cumulated IR, CIR, graphed in Fig. 6.3 and 6.4. Standard errors for IR are calculated as in Lütkephol (1991). IFs appear in these graphs as the limit of the CIR at horizon $\infty$, indicated as 'inf'. Note that some of the IR have 0 standard errors for the first lead, because of the different timing of the variables in the state vector, i.e. the lag-lead effect.

Figure 3(a) shows the effect on the second difference of the price level, that is the acceleration rate of inflation. The initial impact is positive and followed by a small decline. The cumulated impulse response function shows the effect on the inflation rate. This effect converges rapidly to 0.08, the impact factor; this corresponds to a permanent increase in the annual inflation rate of 0.32% (due to a one percent perturbation of unit labor costs).

 DOI: 10.2870/52609



(a) IR: Effect on $\Delta^2\text{lpfc}_t$ (b) IR: Effect on $\beta'\Delta X_{t-1}$ (c) IR: Effect on $mec_t$




(d) CIR: Effect on $\Delta^2\text{lpfc}_t$ (e) CIR: Effect on $\beta'\Delta X_{t-1}$ (f) CIR: Effect on $mec_t$

Figure 6.4: Effect of perturbation of import prices (lpm): Impulse Response functions (IR:top) and Cumulated Impulse Response functions (CIR: bottom) with 95% confidence intervals. Impact Factors with 95% confidence interval are given along CIR at horizon $\infty$, indicated as 'inf'. (b) (c) (e) (f) present lag-lead effects.

Graphs (b) and (e) show that an increase in unit labor costs leads a decline in the combined mark-up. Note that the combination of an increase in inflation and a decrease in this mark-up is completely in the line with the prediction of imperfect competition models. Graphs (c) and (f) report effects on the compensated mark-up.

Figure 6.4 reports the effect of perturbation to import prices. The adjustment to the new equilibrium of the multicointegrating relation takes longer than for unit labor costs. Apart from the effect on relation $\beta_1'\Delta X_{t-1}$, the impact factors have the same sign as the impact factors above, but are 2 to 5 times smaller in magnitude. Overall, labor costs have a greater impact on the forecast of price inflation than import prices, in line with expectations.

When reading the above findings from the perspective of data-revisions, one concludes that more timely and accurate data on labor costs would provide the most dramatic improvement in the uncertainty associated with long run forecasts of inflation.

Finally, we comment on the possible policy interpretation of these findings. Unit labor cost may be influenced by the policy maker by collective wage bargaining, the social benefit system and the setting of wages in the public sector. Hence $\text{lulc}_t$ can be considered a partial instrument. If one assumes the coeteris paribus condition, the IF associated with perturbation in $\text{lulc}_t$ can be interpreted as indicator of policy effectiveness in the long run. The empirical analysis suggests that policy actions aimed at influencing labor costs are effective.

The ceteris paribus condition for this instrument, however, may be questionable. Wage increases may simultaneously affect internal prices, hence violating the ceteris paribus condition. No simultaneous effect is expected on import prices. These observations require an analysis of the simultaneity structure of the system.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

| | $\Delta^2\text{lpfc}_t$ | $\Delta^2\text{lulc}_t$ | $\Delta^2\text{lpm}_t$ |
|---|---|---|---|
| $\Delta^2\text{lpfc}_t$ | 0.005137 | | |
| $\Delta^2\text{lulc}_t$ | $0.39803^a$ | 0.013794 | |
| $\Delta^2\text{lpm}_t$ | $0.20201^b$ | $0.110714^c$ | 0.02393 |

Table 6.5: Estimates of $\widehat{\Omega}$. Diagonal elements are the estimated standard deviations, off-diagonal elements report estimated correlations. $^a$: significant at 5% level, $^b$: significant at 1% level, $^c$: insignificant at 5% level.

Let us assume that the PM has conducted an institutional investigation on its instruments, and expects a unit increase in $\text{lulc}_t$ to be associated with 0.148 increase in $\text{lpfc}_t$ and no change in $\text{lpm}_t$. The sensitivity of the policy action aimed at influencing $\text{lulc}_t$ can then be determined by considering IFs with respect to $a'\widetilde{X}_t := a'(\Delta^2\text{lpfc}_t\text{: } \Delta^2\text{lulc}_t\text{: } \Delta^2\text{lpm}_t\text{: } \widehat{\beta}'\Delta X_{t-1}\text{: } \widehat{\beta}_1'\Delta X_{t-1}\text{: } \text{mec}_t)'$ where $a := (0.148\text{: } 1\text{: } 0\text{: } ... \text{ : } 0)'$. By the results in section 6.3.2, this amounts to multiplying $F$ by $\bar{a} = (0.112\text{: } 0.758\text{: } 0\text{: } ... \text{ : } 0)'$, i.e. by summing the first two columns of table 6.4 with weights 0.112 and 0.758. This example shows that, for any given simultaneity structure, an appropriate linear combination of the IFs presents a policy effectiveness interpretation.

There are many ways to analyze the simultaneity structure of the system. One possible analysis is based on the historical variance-covariance matrix of the systems innovations, $\widehat{\Omega}$, which we report in table 6.5. In the present case the estimated correlation between innovations in $\text{lpfc}_t$ and $\text{lulc}_t$ is equal to 0.398. This correlation is significant on the basis of the asymptotic standard errors for $\widehat{\Omega}$. Note that the correlation between innovations in $\text{lpm}_t$ and $\text{lulc}_t$ is not significant, also in line with expectations.

Multiplying 0.398 by the ratio 0.005137/0.013794 of the standard deviations of the innovations in the two equations, one finds a 0.148 increase in $\text{lpfc}_t$ associated with a unit increase in $\text{lulc}_t$. This reasoning follows the idea of GIR of Pesaran and Shin (1998) and Koop et al. (1996). We note here that when estimating the standard error of the modified IF one needs to account for the fact that $\Omega$ is estimated, as shown in section 6.6.3.

Import prices are denominated in local currency and thus reflect fluctuations in exchange rates. Economic policy aimed at influencing the exchange rate may be responsible for fluctuations in import prices; hence also $\text{lpm}_t$ can be considered a partial instrument. When one assumes the ceteris paribus condition, one observes significant long-run effects on internal prices and on all mark-ups. Exchange-rate policies may thus be effective in influencing internal prices through import prices.

A more detailed analysis of the simultaneity structure is however needed also in this case, as for the case of $\text{ulc}_t$ discussed above. In particular one can postulate a simultaneous effect of $\text{lpm}_t$ on $\text{lpfc}_t$, which agrees with the findings in table 6.5. This would lead to a linear combination of the first and third column of table 6.4. This analysis is not reported for space constraints.

## 6.8. Conclusions

In this chapter we have defined impact factors as a sensitivity measure of long-run forecasts, and discussed their properties. We have applied the definition to vector autoregressive processes, in the stationary, I(1) and I(2) cases. Not surprisingly, the impact factors are functions of the moving average total impact matrix of the stationary representation of the systems, which is singular in cointegrated processes.

An application to price mark-up in Australia shows, among other things, how perturbations to labor cost can have a permanent positive effect on inflation and a permanent negative effect on the mark-up. This is in line with imperfect competition models.

## 6.A. Derivation of the IF

In this appendix we report proofs of the propositions in the sections 6.4.1 and 6.4.2. The first lemma gives a well known result on the inversion of a partitioned matrix, see also Faliva and Zoia (2000).

LEMMA 6. *Given the $p \times s$ matrices $a$, $b$ of full column rank $s < p$, and $Q$ any square $p \times p$ matrix, then a necessary and sufficient condition for the matrix*

$$S := \begin{pmatrix} Q & a \\ b' & 0 \end{pmatrix}$$

*to be invertible is that $a'_\perp Q b_\perp$ be of full rank $p - s$; in this case*

$$S^{-1} := \begin{pmatrix} Q & a \\ b' & 0 \end{pmatrix}^{-1} = \begin{pmatrix} R & (I - RQ)\bar{b} \\ \bar{a}'(I - QR) & \bar{a}'(QRQ - Q)\bar{b} \end{pmatrix}$$

*where $R := b_\perp (a'_\perp Q b_\perp)^{-1} a'_\perp$.*

PROOF. We observe that $S$ has the same rank as $K := J_1 S J_2$ for $J_i$ invertible square matrices. Choose $J_1$ and $J_2$ as follows and calculate the resulting product $K := J_1 S J_2$

$$J_1 := \begin{pmatrix} & I_s \\ (\bar{a}, \bar{a}_\perp)' & \end{pmatrix}, \qquad J_2 := \begin{pmatrix} (\bar{b}, \bar{b}_\perp) & \\ & I_s \end{pmatrix}$$

$$K := J_1 S J_2 = \begin{pmatrix} & & I_s \\ Q_{a_\perp b} & Q_{a_\perp b_\perp} & \\ Q_{ab} & Q_{ab_\perp} & I_s \end{pmatrix}$$

where we have used the notation $Q_{cd} := \bar{c}' Q \bar{d}$, $c, d = a, b, a_\perp, b_\perp$. $J_1 S J_2$ is block triangular and it is invertible iff $Q_{a_\perp b_\perp}$, or equivalently if $a'_\perp Q b_\perp$ is invertible. If this is the case, the inverse $S^{-1}$ can be calculated as $S^{-1} = J_2 (J_1 S J_2)^{-1} J_1 = J_2 K^{-1} J_1$. By straightforward application of the partitioned inverse formula to $K$, one finds

$$K^{-1} := (J_1 S J_2)^{-1} = \begin{pmatrix} & I_s & \\ -Q^{-1}_{a_\perp b_\perp} Q_{a_\perp b} & Q^{-1}_{a_\perp b_\perp} & \\ -(Q_{ab} - Q_{ab_\perp} Q^{-1}_{a_\perp b_\perp} Q_{a_\perp b}) & -Q_{ab_\perp} Q^{-1}_{a_\perp b_\perp} & I_s \end{pmatrix}.$$

Finally calculating $S^{-1} = J_2 K^{-1} J_1$ one finds the results in the statement. □

PROOF. of proposition 1. We apply partitioned inverses to the matrix $(I - A)$ partitioned conformably to the $A_{ij}$ blocks in (6.6), using lemma 6. Let

$$K := \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} := (I - A) = \begin{pmatrix} I - A_{11} & -A_{12} \\ -A_{21} & I - A_{22} \end{pmatrix}$$

and indicate by $K^{ij}$ blocks of $K^{-1}$ conformable with $A_{ij}$. Note that $K^{11} = K^{-1}_{11.2}$, where $K_{11.2} := K_{11} - K_{12} K^{-1}_{22} K_{21} = I - (A_{11} + A_{12}(I - A_{22})^{-1} A_{21})$, where

$$A_{12}(I - A_{22})^{-1} A_{21} = diag\left(\sum_{i=2}^{k-1} \Gamma_i, 0\right),$$

so that

$$K_{11.2} = I - (A_{11} + A_{12}(I - A_{22})^{-1} A_{21}) = \begin{pmatrix} I - \Gamma^\circ + \alpha\beta' & -\alpha \\ -\beta' & 0 \end{pmatrix}.$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

where $\Gamma^\circ := -\Gamma = I - \sum_{i=1}^{k-1} \Gamma_i$. Applying lemma 6

$$K_{11.2}^{-1} = \begin{pmatrix} C & (C\Gamma^\circ - I)\bar{\beta} \\ \bar{\alpha}'(\Gamma^\circ C - I) & \bar{\alpha}'(\Gamma^\circ C \Gamma^\circ - \Gamma^\circ)\bar{\beta} \end{pmatrix},$$

where $C := \beta_\perp (\alpha_\perp' \Gamma^\circ \beta_\perp)^{-1} \alpha_\perp'$. The remaining blocks of $K^{-1}$ can be expressed as $K^{21} = -K_{22}^{-1} K_{21} K_{11.2}^{-1}$, $K^{12} = -K_{11.2}^{-1} K_{12} K_{22}^{-1}$ and $K^{22} = K_{22}^{-1} + K_{22}^{-1} K_{21} K_{11.2}^{-1} K_{12} K_{22}^{-1}$, where $K_{11.2}^{-1}$ has already been calculated and $K_{22}^{-1} = c_2 \otimes I$. Substituting one obtains the expression in the proposition. □

The following EC formulation is convenient in the I(2) case.

PROPOSITION 5. *An equivalent EC formulation of* (6.7) *is*

$$\begin{aligned} \Delta^2 X_t &= \alpha(\beta' X_{t-3} + \delta \beta_2' \Delta X_{t-2}) + (\zeta_1^* : \zeta_2)(\beta : \beta_1)' \Delta X_{t-2} + \\ &\quad + \Upsilon_1^* \Delta^2 X_{t-1} + \Phi W_{t-1} + \mu^* D_t^* + \epsilon_t \end{aligned} \tag{6.19}$$

*where* $\zeta_1^* := \zeta_1 + 2\alpha$ *and* $\Upsilon_1^* := (\Upsilon_1 + \Gamma + \Pi)$.

PROOF. Adding and subtracting $\Pi(X_{t-1} - X_{t-3}) = \Pi \Delta X_{t-1} + \Pi \Delta X_{t-2}$ on the r.h.s. of (6.7) one obtains

$$\Delta^2 X_t = \Pi X_{t-3} + (\Gamma + \Pi)\Delta X_{t-1} + \Pi \Delta X_{t-2} + \Upsilon_1 \Delta^2 X_{t-1} + \Phi W_t + \epsilon_t.$$

Further adding and subtracting $(\Gamma + \Pi)\Delta X_{t-2}$ on the r.h.s. yields

$$\begin{aligned} \Delta^2 X_t &= \Pi X_{t-3} + (\Gamma + 2\Pi)\Delta X_{t-2} + (\Upsilon_1 + \Gamma + \Pi)\Delta^2 X_{t-1} + \Phi W_t + \epsilon_t \\ &= \Pi X_{t-3} + \Gamma^* \Delta X_{t-2} + \Upsilon_1^* \Delta^2 X_{t-1} + \Phi W_t + \epsilon_t, \end{aligned} \tag{6.20}$$

where $\Gamma^* := \Gamma + 2\Pi$, $\Upsilon_1^* := \Upsilon_1 + \Gamma + \Pi$. In order to recover the EC terms within (6.20) note that $\Gamma^* \bar{\beta}_2 = (\Gamma + 2\Pi)\bar{\beta}_2 = \Gamma \bar{\beta}_2 = \alpha\delta$ and hence

$$\Gamma^* = \Gamma^*(P_\tau + P_{\beta_2}) = (\Gamma^* \bar{\tau})\tau' + (\Gamma^* \bar{\beta}_2)\beta_2' = \zeta^* \tau' + \alpha\delta\beta_2',$$

where $\zeta^* := \Gamma^* \bar{\tau}$, $\tau := (\beta, \beta_1)$ and we observe that $\zeta_2^* := \Gamma^* \bar{\beta}_1 = \Gamma \bar{\beta}_1 =: \zeta_2$. Substituting within (6.20) one finds (6.19). □

PROOF. of proposition 2. Let $m := p(k-3)$ be the dimension of $W_t$. In order to compute $(I - A)^{-1}$, we apply partitioned inverses to the matrix $(I - A)$ partitioned conformably to $A_{ij}$ blocks in (6.6), using lemma 6. As in the I(1) case let

$$K := \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} := (I - A) = \begin{pmatrix} I - A_{11} & -A_{12} \\ -A_{21} & I - A_{22} \end{pmatrix}$$

and indicate by $K^{ij}$ blocks of $K^{-1}$ conformable with $A_{ij}$. Note that $K^{11} = K_{11.2}^{-1}$, where $K_{11.2} := K_{11} - K_{12} K_{22}^{-1} K_{21} = I - (A_{11} + A_{12}(I - A_{22})^{-1} A_{21})$, where

$$A_{12}(I - A_{22})^{-1} A_{21} = diag\left( \sum_{i=2}^{k-1} \Upsilon_i, 0 \right),$$

so that

$$K_{11.2} = I - (A_{11} + A_{12}(I - A_{22})^{-1} A_{21}) = \left( \begin{array}{c|ccc} \phi - \Gamma - \alpha\beta' & -\zeta_1^* & -\zeta_2 & -\alpha \\ -\beta' & & & \\ -\beta_1' & & & \\ \hline -\delta\beta_2' & -I_r & & \end{array} \right).$$

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

where $\phi := I - \sum_{i=1}^{k-2}\Upsilon_i$. In order to calculate $K_{11.2}^{-1}$ we express it as $K_{11.2}^{-1} = (J_3K_{11.2})^{-1}J_3$ where

$$J_3 := \begin{pmatrix} I_p & & & \\ & & I_r & \\ & I_r & & \\ & & & I_s \end{pmatrix}$$

$$J_3K_{11.2} = \left(\begin{array}{cc|cc} \phi^* & -\zeta_1^* & -\zeta_2 & -\alpha \\ -\delta\beta_2' & -I_r & & \\ \hline -\beta' & & & \\ -\beta_1' & & & \end{array}\right) = \begin{pmatrix} Q & a \\ b' & 0 \end{pmatrix},$$

where $\phi^* := \phi - \Gamma - \alpha\beta'$ and $Q$ is $(p+r)\times(p+r)$. We now wish to apply lemma 6, observing that $b_\perp = diag(\beta_2, I_r)$ and $a_\perp = diag(\alpha_2, I_r)$, because $\zeta_2 = \alpha\mathrm{F}_{\alpha\beta_1} + \alpha_1' \in sp(\alpha : \alpha_1)$. Let $\theta^* := \phi^* + \zeta_1^*\bar{\alpha}'\Gamma$, $h := I - \theta^* C_2$ and recall that $\phi^* = \phi - \Gamma - \alpha\beta'$, $\zeta_1^* = \zeta_1 + 2\alpha$, $\bar{\tau} = (\bar{\beta} : \bar{\beta}_1)$. One finds

$$(J_3K_{11.2})^{-1} = \begin{pmatrix} R & (I-RQ)\bar{b} \\ \bar{a}'(I-QR) & \bar{a}'(QRQ-Q)\bar{b} \end{pmatrix} =$$

$$= \begin{pmatrix} C_2 & -C_2\zeta_1 & (C_2\phi^* - I)\bar{\tau} \\ -\delta\beta_2'C_2 & \delta\beta_2'C_2\zeta_1 - I & -\delta\beta_2'C_2\phi^*\bar{\tau} \\ -\bar{\alpha}_1'h & \bar{\alpha}_1'h\zeta_1^* & -\bar{\alpha}_1'h\phi^*\bar{\tau} \\ -\bar{\alpha}'(I-\zeta_2\bar{\alpha}_1')h & \bar{\alpha}'(I-\zeta_2\bar{\alpha}_1')h\zeta_1^* & -\bar{\alpha}'(I-\zeta_2\bar{\alpha}_1')h\phi^*\bar{\tau} \end{pmatrix}$$

where

$$\bar{a}' = \begin{pmatrix} -\bar{\alpha}_1' & 0 \\ -\bar{\alpha}'(I-\zeta_2\bar{\alpha}_1') & 0 \end{pmatrix} \qquad \bar{b} = \begin{pmatrix} -\bar{\tau} \\ 0 \end{pmatrix}.$$

Thus $B := K_{11.2}^{-1} = (J_3K_{11.2})^{-1}J_3$ corresponds to the expression found above for $(J_3K_{11.2})^{-1}$ with the last 2 blocks of columns interchanged. The rest of the calculations are exactly the same as in the proof of Proposition 1; this completes the proof. □

## 6.B. Timing and IF

In this appendix we illustrate various possible choices of lag for the EC terms, and report proofs for results in section 6.4.3. In all cases below we adopt the following convention: the various subvectors of the state vector $Z_t := NV_t$ or $V_t$ are numbered consecutively. Consider the $i$-th subvector of $Z_t$ and the $j$-th subvector of $V_t$, of dimension $n_i$ and $n_j$ respectively; the elements of the transformation matrix $N$ corresponding to these subvectors are indicated with the subscript $ij$, $N_{ij}$, of dimension $n_i \times n_j$. When not otherwise specified, elements of the $N$ matrix are assumed to be zero.

(1) I(1) case, EC in lag 1. Let $V_t$ be the choice of state vector used above, $V_t := (V_{1t}' : V_{2t}' : V_{3t}')' = (\Delta X_t' : X_{t-1}'\beta : U_t')'$, and consider the following possible alternative choice of state vector $Z_t := (Z_{1t}' : Z_{2t}' : Z_{3t}')' = (\Delta X_t' : X_t'\beta : U_t')'$. It is simple to see that $Z_t = NV_t$ with $N_{ii} = I$, $i = 1, 2, 3$ and $N_{21} = \beta'$.

(2) I(1) case, EC in lag $j$, where $1 < j \leq k$. Let $V_t$ be the choice of state vector used above, and let $Z_{ht} := V_{ht}$, $h = 1, 3$ and $Z_{2t} := \beta' X_{t-j}$. It is simple to see that $Z_t = NV_t$ with $N_{ii} = I$, $i = 1, 2, 3$ and $N_{23} = (-i_{j-1}' \otimes \beta', 0)$, where $i_j$ is an $j \times 1$ vector of ones.

(3) I(2) case, level term in lag 1. Let $V_t$ be the choice of state vector used above, $V_t := (V_{1t}' : \ldots : V_{5t}')' = (\Delta^2 X_t' : \Delta X_{t-1}'\beta : \Delta X_{t-1}'\beta_1 : X_{t-2}'\beta + \Delta X_{t-1}'\beta_2\delta' : W_t')'$ and consider the following possible alternative choice of state vector $Z_t := (Z_{1t}' : \ldots : Z_{5t}')'$ where $Z_{ht} := V_{ht}$, $h = 1, 2, 3, 5$ and $Z_{4t} := \beta' X_{t-1} + \delta\beta_2'\Delta X_{t-1}$. The only term that has been

shifted is $X_s'\beta$ form $s = t-2$ to $s = t-1$. It is simple to see that $Z_t = NV_t$ with $N_{ii} = I, i = 1, ..., 5$ and $N_{42} = I_r$.

(4) I(2) case, level term in lag $s$, where $2 < s \leq k$. Let $V_t$ be the choice of state vector used above and consider the following possible alternative choice of state vector $Z_t := (Z_{1t}' : ... : Z_{5t}')'$ where $Z_{ht} := V_{ht}$, $h = 1, 2, 3, 5$ and $Z_{4t} := \beta' X_{t-s} + \delta\beta_2' \Delta X_{t-1}$ where the only term that has been shifted is $\beta' X_i$ form $i = t-2$ to $i = t-s$. It can be checked that $Z_t = NV_t$ with $N_{ii} = I, i = 1, ..., 5$, $N_{42} = -(s-2)I_r$, $N_{45} = (j' \otimes \beta')$, $j := (s-3, s-4, ..., 1, 0, ..., 0)'$.

(5) I(2) case, differenced term in lag $s$, where $2 \leq s \leq k$. Let $V_t$ be the choice of state vector used above and consider the following possible alternative choice of state vector $Z_t := (Z_{1t}' : ... : Z_{5t}')'$ where $Z_{ht} := V_{ht}$, $h = 1, 2, 3, 5$ and $Z_{4t} := \beta' X_{t-2} + \delta\beta_2' \Delta X_{t-s}$ where the only term that has been shifted is $\beta_2' \Delta X_i$ form $i = t-1$ to $i = t-s$. It can be checked that $Z_t = NV_t$ with $N_{ii} = I, i = 1, ..., 5$, $N_{45} = (-i_{s-2}' \otimes \delta\beta_2', 0)$.

PROOF. of proposition 3. By definition

$$\begin{aligned} F_Z := (I - A_Z)^{-1} - I = & \ (N(I - A_S)N^{-1})^{-1} - I = N(I - A_S)^{-1}N^{-1} - I = \\ = & \ N((I - A_S)^{-1} - I)N^{-1} =: NF_S N^{-1}. \end{aligned}$$

□

PROOF. of Proposition 4. Let $Z_t = NV_t$ indicate the change of state vector, and let $F_Z$ and $F := F_S$ indicate the corresponding IF. From proposition 3 it follows that $F_Z = NFN^{-1}$. Hence $(F_Z)_{y,x} = \sum_i \sum_j N_{yi} F_{ij} N^{jx}$, where we use subscripts to indicate blocks. Blocks of $N^{-1}$ are indicated with $N^{ij} := (N^{-1})_{ij}$. Thus if $N_{yi} = 0$, $N_{yy} = I$, $N_{jx} = 0$, $N_{xx} = I$, for $i \neq y$, $j \neq x$ one finds that $(F_Z)_{y,x} = F_{y,x}$, and that the IF are invariant.

For the first result we take $V_t := (\Delta X_t' : X_{t-1}'\beta : U_t')'$ and $Z_t := (\Delta X_t : X_{t-j}'\beta : U_t')'$, $j = 1, \ldots, k-1$ and note that $Z_t = NV_t$ with

$$N = \begin{pmatrix} I_p & & \\ & I_r & (-i_{j-1}' \otimes \beta' : 0) \\ & & I_m \end{pmatrix}, \qquad N^{-1} = \begin{pmatrix} I_p & & \\ & I_r & (i_{j-1}' \otimes \beta' : 0) \\ & & I_m \end{pmatrix}$$

It is thus immediate to note that $F_{y_t, x_t}$ is invariant for $y_t = \Delta X_t$, $U_t$ and $x_t = \Delta X_t$, $\beta' X_{t-j}$. When $Z_t$ includes $\beta' X_t$, case $j = 0$ above, then the transformation matrix has a similar shape, but $N_{21} = \beta'$, $N_{23} = 0$, $N^{21} = -\beta'$, $N^{23} = 0$. The same conclusion thus applies.

For the I(2) results, we take $Z_t = NV_t$ with

$$V_t := V_t(i, j, l, m) := (\Delta^2 X_t' : \Delta X_{t-i}'\beta : \Delta X_{t-j}'\beta_1 : X_{t-l}'\beta + \Delta X_{t-m}'\beta_2\delta' : W_t')'$$

and $Z_t := V_t(1, 1, 2, 1)$. One finds

$$N = \begin{pmatrix} I & & & & \\ & I & & & N_{25} \\ & & I & & N_{35} \\ & N_{42} & & I & N_{45} \\ & & & & I \end{pmatrix}, \qquad N^{-1} = \begin{pmatrix} I & & & & \\ & I & & & -N_{25} \\ & & I & & -N_{35} \\ & -N_{42} & & I & Q \\ & & & & I \end{pmatrix},$$

where $Q := -N_{45} + N_{42}N_{25}$, where $N_{45} := N_{45a} + N_{45b}$,

$$N_{25} = (-i_{i-1}' \otimes \beta' : 0), N_{35} = (-i_{j-1}' \otimes \beta_1' : 0), N_{45a} = (-i_{m-1}' \otimes \delta\beta_2' : 0).$$

If $l = 1$, one has $N_{42} = I_r$, $N_{45b} = 0$ whereas if $l \geq 3$, $N_{42} = -(l-2)I_r$, $N_{45b} = (g \otimes \beta' : 0)$, $g := (l-3 : l-4 : \ldots : 1 : 0 : \ldots : 0)$.


DOI: 10.2870/52609

From the expressions on $N$ and $N^{-1}$ we find that $F_{y_t,x_t}$ is invariant for $y_t = \Delta^2 X_t$, $W_t$ and $x_t = \Delta^2 X_t$, $\beta_1' X_{t-j}$, $\beta' X_{t-l} + \delta\beta_2'\Delta X_{t-m}$. We also observe that $F_{\Delta^2 X_t, \beta_1' X_{t-j}}$ can be simplified as follows

$$(C_2\phi^* - I)\bar{\beta}_1 = (C_2\phi - I)\bar{\beta}_1 - C_2\Gamma\bar{\beta}_1 - C_2\alpha\beta'\bar{\beta}_1 = (C_2\phi - I)\bar{\beta}_1,$$

where we have used that $C_2\Gamma\bar{\beta}_1$ contains the term $\alpha_2'\Gamma\beta_1$, which equals zero by assumption I(2)_$b$. □

## 6.C. Inference on the IF

In this appendix we give proofs of section 6.6. We start by illustrating how the state space representations used for $k = 1$ in the I(1) and $k = 2$ for the I(2) cases are not minimal.

The non-minimality of the state space vectors does not affect the derivations of IF, although it is relevant for inference. We thus show how the companion matrices $A$ can be rank-decomposed in $A = \widetilde{A}H'$. In case of no rank reduction of the matrix $A$, we take $H = I$ in the decomposition $A = \widetilde{A}H$, i.e. $A = \widetilde{A}$.

In the I(1) state space formulation, when $k = 1$ the companion matrix $A$ reduces to the block $A_{11}$ in formula (6.6), where, moreover, $\Gamma_1 = 0$, i.e. $\Gamma_1^* = \Pi = \alpha\beta'$. It is simple to see that $A$ has in this case reduced rank, since

$$A := \begin{pmatrix} \alpha\beta' & \alpha \\ \beta' & I \end{pmatrix} = \begin{pmatrix} \alpha \\ I \end{pmatrix} \begin{pmatrix} \beta' & I \end{pmatrix} =: \widetilde{A}H', \tag{6.21}$$

where $\widetilde{A} := (\alpha' : I)'$, $H := (\beta' : I)'$ are $p + r \times r$ matrices with full column rank $r$.

For the state space representation of I(2) systems, when $k = 2$ the companion matrix $A$ reduces to the block $A_{11}$ in formula (6.11) above, where, moreover, $\Upsilon_1 = 0$, i.e. $\Upsilon_1^* = \Gamma + \Pi = \Gamma + \alpha\beta'$. It can be checked that, similarly to the I(1) case, $A$ has in this case reduced rank:

$$\begin{aligned} A &= \begin{pmatrix} \Gamma + \alpha\beta' & \Gamma\bar{\beta} + 2\alpha & \Gamma\bar{\beta}_1 & \alpha \\ \beta' & I_r & & \\ \beta_1' & & I_s & \\ \delta\beta_2' & I_r & & I_r \end{pmatrix} = \\ &= \begin{pmatrix} \Gamma\bar{\beta} + \alpha & \Gamma\bar{\beta}_1 & \alpha \\ I_r & & \\ & I_s & \\ & & I_r \end{pmatrix} \begin{pmatrix} \beta' & I_r & & \\ \beta_1' & & I_s & \\ \delta\beta_2' & I_r & & I_r \end{pmatrix} =: \widetilde{A}H', \end{aligned} \tag{6.22}$$

where

$$\widetilde{A} := \begin{pmatrix} \Gamma\bar{\beta} + \alpha & \Gamma\bar{\beta}_1 & \alpha \\ I_r & & \\ & I_s & \\ & & I_r \end{pmatrix} = \begin{pmatrix} \zeta_1 + \alpha & \zeta_2 & \alpha \\ I_r & & \\ & I_s & \\ & & I_r \end{pmatrix}, \quad H := \begin{pmatrix} \beta & \beta_1 & \beta_2\delta' \\ I_r & & I_r \\ & I_s & \\ & & I_r \end{pmatrix}$$

are $(p + 2r + s) \times (2r + s)$ matrices with full column rank $(2r + s)$.

PROOF. of theorem 13. $\widehat{H} - H$, $\widehat{L} - L \in O_p(T^{-1})$ because they are functions of the cointegrating coefficients, which are at least $T$-consistent. Result (6.16) follows by standard regression arguments, after observing that, due to superconsistency, one can substitute the estimated cointegration coefficients with their true values, see Paruolo (2002b) for a detailed proof of $\widehat{S}_{ij} - S_{ij} = O_p(T^{-1})$. In fact one has $\widehat{G} = \widehat{S}_{01}\widehat{S}_{11}^{-1} = S_{01}S_{11}^{-1} + O_p(T^{-1}) = G + S_{\epsilon 1}S_{11}^{-1} + O_p(T^{-1})$, from which $T^{1/2}(\widehat{G} - G) = T^{1/2}S_{\epsilon 1}S_{11}^{-1} + o_p(1)$, and $T^{1/2}vec(\widehat{G} - G)' \xrightarrow{w} N(0, \Omega \otimes \Sigma^{-1})$.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

In order to show (6.17) note that differentiating $F$ one has $\mathrm{d}F = K\mathrm{d}AK$, so that

$$T^{1/2}(\widehat{F} - F) = T^{1/2}K(\widehat{A} - A)K + o_p(1). \tag{6.23}$$

Because $\widehat{H} - H, \widehat{L} - L \in O_p(T^{-1})$ one has

$$T^{1/2}(\widehat{A} - A) = \begin{pmatrix} T^{1/2}(\widehat{G} - G)H' \\ 0 \end{pmatrix} + o_p(1) = JT^{1/2}(\widehat{G} - G)H' + o_p(1)$$

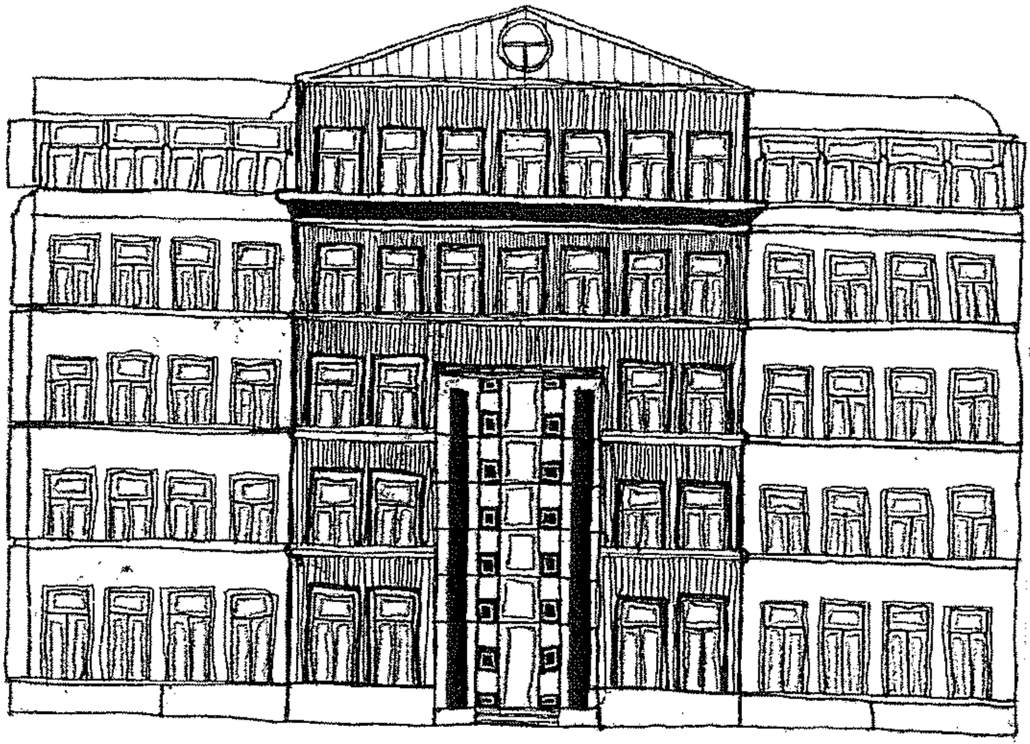
Substituting in (6.23) one finds $T^{1/2}(\widehat{F} - F) = T^{1/2}KJ(\widehat{G} - G)H'K + o_p(1)$. Transposing and vectorizing one obtains (6.17) from (6.16). □

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

# Part 4

# Dutch Money Demand


DOI: 10.2870/52609



'S GRAVENHAGE

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute

DOI: 10.2870/52609

CHAPTER 7

# Money Demand in the Netherlands

## 7.1. Introduction

The introduction[1] of the Euro in 12 of the 15 members of the European Union, has led to a large debate, scholarly, political and popular, on the costs and benefits of a monetary union. Discussion still takes place on the costs and benefits of joining for the three old EU members, that have so far stayed out, and the ten new members, who will join the EU in 2004. The most frequently cited cost is probably the inability to react differently to asymmetric shocks. Different transmission mechanisms in the different countries are a second source of asymmetry: due to institutional differences, like a prevalence of fixed rate mortgages (the Netherlands) or floating rate mortgages (the UK) and differences in structural parameters of the economy can cause a monetary intervention to have different effects even to economies hit by exactly the same shock.

The Netherlands has in the last few years often been praised for its economic successes of the last two decades: a rapidly declining unemployment rate, increasing participation rate and a massive public deficit, which had turned into a small surplus before the latest economic slowdown. This success took place, while the Netherlands had a de facto monetary union with Germany, by far its most important trade partner. The Dutch central bank had rendered control of its monetary polity to the German Bundesbank. Within the Euro-area, countries relinquish virtually all their monetary autonomy to the European Central Bank in exchange for a very small say in the actual running of monetary policy.

Studying Dutch money demand and monetary transmission mechanism may therefore provide a valuable insight in this success story and provide understanding as to how a monetary union can work or fail, a question which is still relevant.

In the next section, the Dutch macro-economic and political situation are described together with the monetary policy pursued in the period under study, 1979 1st quarter until 1998, 4th quarter, when the Euro was introduced. From 1999 there are no publicly available statistics on the money supply in the individual countries inside the Euro area, such that the study cannot be extended.

Then follows section 7.3 with a short overview of the literature and a description of the data. Section 7.4 contains an overview of the relevant economic literature. In section 7.5 we discuss the I(2) and I(1) methodology, the nominal to real transformation and automated model selection as well as small sample properties of the estimators used. The analysis of the Dutch data follows in section 7.6 before final conclusions are drawn.

A technical result concerning bootstrapping linear within equation restrictions in the cointegration space is given in section 7.A.

Throughout this paper we shall argue that the bootstrap (of tests on the cointegration parameters) should be based on the unrestricted estimate. We then need to find alternative null hypotheses

[1] I am grateful to Katarina Juselius for her help and patience in the econometric modelling of the data. She is not responsible for any errors or policy statements.

Gijs de Bruin (postbank), Annelies van Rhenen (spaarbeleg), IJda Geerts (NVM), Dhr. Elferich (De Nederlandsche Bank) kindly provided data for this paper.

for the bootstrap samples. In the appendix to this paper an alternative null hypothesis for linear restrictions within each cointegration vector is given.

## 7.2. The Netherlands

In this long section we give an overview of the economic and political situation in the Netherlands in the period we study. We then proceed with stating the monetary objectives, the instruments that were used to achieve them and the exchange rate policy followed. Then follows a short account of the liberalization process.

**7.2.1. Political and Economic situation.** During the second World war, the Netherlands had undergone extensive economic damage. After the war, a policy of national reconstruction was started, strongly led by the central government, which channelled credit to specific sectors, set explicit targets for the number of houses to be built each year, and in agreement with the social partners, implemented a policy of moderate wage increases until 1963. After that year a wage-price spiral erupted, which was fuelled by the revenue of newly discovered natural gas reserves in the North of the country. The 1960s and 1970s saw a rapid expansion of the Welfare State and even the two oil shocks did not immediately hit the Netherlands as hard as it did other countries: natural gas income provided a rapidly increasing source of income. In 1979 a housing market asset bubble burst and by 1981 the labour share of national income had risen to over 95%. Unemployment started to rise rapidly as the Netherlands entered its severest post-war depression in 1982, during which national output fell to 1978 levels, and unemployment tripled.

From 1977 to 1982, 3 increasingly unstable coalitions governed the Netherlands. Although the social democrats were the big winners of the 1982 general election, a coalition between the Christian Democrats (center) and Conservative Liberals was formed in November 1982, under prime minister Ruud Lubbers, which was to govern until 1989. Under pressure from the government and the unfavourable economic situation, the leaders of the Dutch employers, C. van der Lede and employees, Wim Kok, agreed on the so-called "Agreement of Wassenaar", which contained wage restraint, a shorter working week and redivision of existing jobs. The government itself pledged to reduce the taxation on labour, once the deficit, which was heading for double digit figure, would allow so. It also cut nominal wages in the public sector, benefits and minimum wages and embraced the market: privatization processes were started, markets were liberalized and the public sector was gradually sized down.

At the same time, the Netherlands entered the hard ERM in 1983, after an unexpected devaluation in 1983. All in all, the macroeconomic performance of the Netherlands improved remarkably from the second half of the 1980s onwards, when inflation was well below German levels, the unemployment rate declined as did the deficit of the government.

In 1989 a new coalition of the Christian Democrats and Social Democrats, who had decidedly moved to the center, took over with a promise of a more social policy. The fall of the German wall and German unification provided a positive demand shock, but when the business cycle swung down, the government, which was still running a deficit, felt that it was necessary to implement a hugely unpopular austerity package of roughly 3% of GDP in the first few months of 1991. This was accompanied by a large increase in the current account balance. The package was also necessary in the light of the Maastricht treaty, which was being drafted by the Dutch government, and contained debt and deficit criteria, to which the Netherlands would not be able to stick without the package. Furthermore the high nominal and real interest rates on the capital market were aggravating the problem for a country with a debt/GDP ratio of 80%. The treaty was approved by parliament, without much discussion and without a popular vote, meaning that the Netherlands would be in the first wave entrants to Euroland.


DOI: 10.2870/52609

During the ERM crisis of 1992-93, the Guilder-DM mark parity was never seriously tested by the market and when the system collapsed the Dutch and German authorities entered into a bilateral agreement to maintain the old parities. As Belgium and France, two other important trade partners, also recovered to their old parities quite quickly, the crisis probably had less of an impact on the Dutch economy than on other European economies.

A new government of Social Democrats, Conservative Liberals and Liberals superseded the old government in 1994 and continued with a socially tinted, neo-liberal economic policy. The Dutch economy flourished more than ever, as its official unemployment was the lowest in the Euro-area, and in 1999 the government was able to record its first surplus in 25 years.

**7.2.2. Monetary Policy.** The Dutch Bank Law[2] of 1948 states that:

> "It shall be the task of the Dutch Central Bank to regulate the value of the Dutch currency in such a way as is most conductive for the prosperity of the nation and in doing so, to stabilize that value as much as possible."[3]

Apparently there are three objectives: a welfare objective, internal price stability and external, exchange rate stability. In fact the last part of the phrase was added by parliament to interpret the first part, see De Jong (1960). The law already curbed the independence of the Central Bank with respect to the pre-war law in two important ways: all its shares were bought by the Dutch state (probably at a price slightly below market value) and an important new article was added, which stated that the Minister of Finance could give binding "directions" to the Bank. They have never been given and the article has been removed from the law in 1998 in preparation for the Monetary Union, but in the frequent informal meetings between the President of the Bank and the Minister of Finance the sheer possibility of them gave considerable leverage to the Minister. So the Dutch central bank was somewhat less independent than the German Bundesbank.

In practice Dutch monetary policy, often referred to as "Moderate Monetarism", aimed at exchange rate stability, deemed very important for a trading nation, first in the Bretton Woods system, later in the snake, ERM and EMU. A broad liquidity ratio was used as the key indicator for monetary policy. Whenever money growth was perceived to be too high, the Dutch monetary authorities hit the break with an over time evolving array of policy instruments, which will be described below, as will the exchange rate arrangements. A very complete overview is given in De Greef et al. (1998).

The banks own view is that path of moderate monetarism was a fairly constant one. Still in the 1970s the Dutch government took recourse to monetary financing of its deficit (1975-1983), inflation was high and the Dutch currency was devalued several times, so in fact an expansionary Keynesian policy was followed between 1972 and 1983. Before and after that period, moderate monetarism is an adequate description of monetary policy.

In 1998 a new bank law superseded the old one, in preparation for monetary union. Under the new law, price stability is the main objective and as long as that is not endangered, contributing to reaching the goals of article 2 of the treaty of Rome. The right of the Minister of Finance to overrule the Bank was abolished.

**7.2.3. Liquidity ratio and monetary instruments.** From the 1970s onwards, the liquidity ratio was the key variable watched by the Central Bank in the belief that excess liquidity would ultimately lead to inflation. Whenever the Bank felt that M2-growth was excessive, it used its instruments to bring it down. Over time we see two broad developments in the instruments used:

[2]Bankwet 1948, published in het Staatsblad I-166 on May 14th 1948. Another unofficial translation in English can be found in the Annual Report of De Nederlandsche Bank over 1948.

[3]Article 9(1) of De Bankwet 1948: "De Bank heeft tot taak de waarde van de Nederlandse geldeenheid te reguleren op zodanige wijze als voor 's lands welvaart het meest dienstig is, en daarbij die waarde zoveel mogelijk te stabiliseren."

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

the transition from direct to indirect instruments and a gradual orientation to more market based instruments. Both processes were intrinsically linked to the capital and credit market liberalization, which is described below.

From July 1973 until November 1979, the Central Bank imposed a liquidity reserve requirement system on banks. As the liquidity ratio increased nonetheless and the aforementioned measure had undesirable side effects, the Bank decided to apply a net credit restriction between May 1977 and June 1981. The percentage growth rate allowed varied over time and also differed somewhat between banks. Consumptive credit was restricted between April 1979 and March 1980.

In 1986 and 1987 there was a gentlemen's agreement with commercial banks to limit the growth of credit. Especially the limit in 1987, 2%, was tight. Note that at various other times, there were discussions between the Central Bank and commercial banks. Due to the small number of banks (four and after a merger three large banks had a market share of well over 80%), a formal agreement was not always necessary to limit credit expansion.

Between 1987 and 1993, the Bank held a small portfolio of government bonds for open market operations: it was used only once in March 1989. Between July 1989 and April 1990, a monetary cash reserve (non-interest bearing) applied.

During the 1980s the liquidity ratio was gradually abandoned and during the 1990s the bank only targeted exchange rates, not so much as a policy choice, but a necessity after the capital market liberalizations of the1980s. It was simply impossible to continue targeting both a fixed exchange rate with Germany and the liquidity ratio, but the bank for a long time paid lip service to targeting liquidity (several issues of the annual report of DNB). The liquidity ratio is reported in figure 1(j) and illustrates this point.

**7.2.4. Exchange rate.** The Netherlands have a long history of aiming for exchange rate stability: the gold standard was only abandoned after all other countries had left the gold block in 1936, even though unemployment was over 20% at the time. The first bank president after the war, Holtrop held an almost dogmatic aversion to realignments in the Bretton woods system, in which the Guilder was effectively anchored to the dollar: revaluations of the German mark were only partially followed.

When the system broke down, the Dutch government took the initiative and entered into an agreement with Belgium and Luxembourg, fixing exchange rates. A few month later these countries entered the snake and later the Netherlands became one of the founding members of the European Monetary System. In all these arrangements, the Dutch Guilder was restricted to the smallest possible band. In the EMS, the Dutch Guilder was devalued with respect to the German mark on 18 October 1976, 16 October 1978, 24 September 1979 and 21 March 1983. At each time it was devalued by 2% vis-a-vis the German mark. Especially the last devaluation came as a surprise: despite the massive government deficit at the time and consistently higher inflation than in Germany, the interest rate differential with Germany had been closed. The surprise devaluation of the new center-right government re-opened the interest rate gap with Germany for about five years, in which the government ran expensive deficits.

Cumulative inflation since 1983 has been lower in the Netherlands than in Germany and the 1983-peg was maintained afterwards without problems. After the signature of the Maastricht treaty in 1991, drawn up by the Dutch government, which foresaw the creation of a singly European currency on January 1st 1999 and in which the participation of both Netherlands and Germany was never in doubt, turmoil broke out in the EMS: After two waves of speculative attacks in 1992 and 1993 it effectively collapsed: all bands were widened to 15%. Still the Dutch and German authorities immediately entered into a bilateral agreement in August 1993 to maintain the old bands, which were never challenged by the markets. Just before fixing the Euro-exchange rates, the Guilder was markedly stronger than the Mark on speculation that it might be revalued. Wellink, the president of the central bank since 1998, admitted afterwards that they had considered the idea of a revaluation.[4]

**7.2.5. Liberalization process.** Already in 1961, current account transactions had been fully liberalized and foreign direct investment was allowed virtually without restrictions. There never was a period of really strict capital controls. Other capital account restrictions remained strict, but were considerably simplified in 1977. In 1983 restrictions on capital inflows were abolished.

On January 1 1986, October 1 1986 and January 1 1988, the domestic capital market was almost completely deregulated: among other things bullet loans, commercial paper, floating rate notes and bank issues of certificate were allowed, as were deep discount and zero-coupon bonds. Foreigners were also allowed to tap the market. In 1991 the prohibition of indexed loans was finally abrogated.[5]

The abolition of a strict separation between banks and insurance companies on January 1 1990 quickly led to a number of mergers and take-overs, which profoundly changed the market structure: by the mid-1990s, three big financial groups, ABN-Amro, ING and the Rabobank controlled almost the entire banking market in the Netherlands. The first two are now rapidly expanding abroad. On the other hand a few insurance companies entered the banking market, offering postal savings account at very competitive rates. Due to the increased competition, banks and these insurance companies started offering savings accounts with interest rates well above the money market rates. At the time of writing this article, April 2003, large banks are offering 4% on instant saving accounts and some smaller ones even 4.8%. Yet the yield curve on the money market (to one year) is well below 2,75%, whereas the yield on 10 year government bonds is 4,2%. Interest rates on saving accounts are thus effectively used to attract and maintain clients and to entice them to buy other profitable services from the bank.

## 7.3. Money demand, data and the Netherlands

In this section we shall discuss the literature on money demand in the Netherlands and the data used in this paper.

**7.3.1. Money demand in the Netherlands.** The central question in the Dutch money demand literature of the last decade has been: what is the cause of the rise in the liquidity ratio over the last twenty years.

From a monetarist point of view, excess money should ultimately lead to inflation, which given the fixed exchange rate over the period (and even more now in Euroland) should be a real worry to policy makers.

A number of explanations have been put forward in the empirical literature: profit hoarding by firms, who are uncertain about the future (Kuipers and Boertje, 1988), an increased importance of the financial transactions motive (Sterken, 1992) and increased wealth in the Dutch economy (Fase and Winder, 1990). All of these studies leave large misspecification in their estimated equations in the form of extremely low Durbin-Watson statistics, a sign that the non-stationarity of the data has not been completely taken care of.

The Dutch Central Bank decided to do a survey of firms in the late eighties to find out why they held more money. Only from the mid 1980s did the central bank record where the liquidity increase took place. Large part of it was in the business sector and from the survey they conclude

[4]At the presentation of the annual report of the Dutch Central Bank over 1998, on May 25th 1999, he admitted that the Dutch central bank had entertained the idea of a revaluation. They decided not to revalue for fear of turbulence on the exchange rate market (see NRC Handelsblad, May 26 1999)

[5]Unfortunately no indexed bonds have been issued by the Dutch state so far, which means that real interest rates will have to be approximated in the rest of this paper.


DOI: 10.2870/52609

that this is due to the preference for internal financing over external financing in firms, which given the increased profitability of the business sector, became easier in the 1980s.

Jacobs and Van der Horst (1996) are from a methodological point of view closest to this paper: they consider a small VAR model with the log of real GDP, annual inflation, short and long term interest rates (but not money) and find that the real long term interest rate in the Netherlands in stationary over the period 1977(1)-1992(4).

**7.3.2. Data vector.** The data vector of this study is very similar to that used in a number of other studies, namely Juselius (1998, 2001); Juselius and Toro (1999); Beyer (1998). It consists of quarterly data on national income, money, a short term interest rate, a long term interest rate and the quarterly inflation rate. The last four variables are daily averages (monthly average in the case of money supply) over the whole quarter (and thus not end of quarter data).

The log of real gross domestic product(according to ESA 1995 definitions and with base year 1995) $y_t$ is supplied by the Dutch central statistical agency (CBS, available on-line at statline: www.cbs.nl).

The log of nominal M3, $m3_t$ as supplied by the Dutch central bank was chosen, because it was not targeted over the period (a special Dutch definition of M2 was in the beginning of the period under study) and it is the only broad measure for which data are readily available for the whole period, as measurement of M2 stopped a few years earlier. This measure was not targeted, but it appears subject to a large number of data revisions.

The log of the CPI with base year 1975 $p_t$ is supplied by the Dutch central statistical agency and has been taken from datastream. Unfortunately no chain weighted price index is available and the GDP deflator for the statistical agency suffers from large data problems[6].

$Lti_t$ and $sti_t$ are measures of long and short term interest rates respectively, each divided by 400 to make them comparable to the quarterly inflation rate $\Delta p_t = p_t - p_{t-1}$.

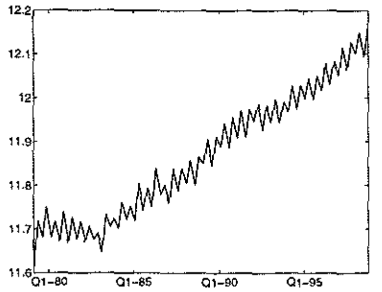
$Lti_t$ is the interest rate on liquid[7] Dutch government bonds with a remaining time to maturity between 5 and 8 years and has been taken from statline.

$Sti_t$ is a measure of interest rates available on saving accounts and deposits of less than two years: it has been constructed on purpose for this study by taking the daily maximum over all such savings products of the postbank/rijkspost-girospaarbank/ING and Spaarbeleg. The first used to be the state-owned giro service at the post office, was privatized in the 1980s and became part of the ING group in the early 1990s. Throughout the period it had a fairly constant and consistent market share. The latter is part of an insurance company and aggressively entered the market with a postal savings account in the 1990s. This interest is markedly different from short term interest rates on the money markets. It is a good proxy measure for the savings rate available to households (and some small companies) over the period. A gradual increase in competition lead to banks offering interest rates to private consumers which were well above the money market rate towards the end of the period. They did so to attract other more profitable business. Clearly not all companies did and do not have access to these saving rates.

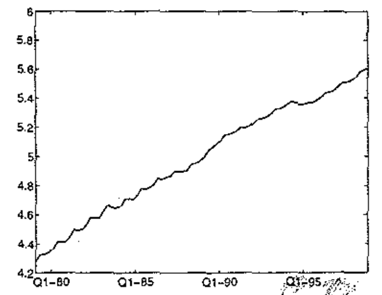
The 5 data series, $y_t, m3_t, p_t, lti_t$ and $sti_t$ together with inflation $\Delta p_t$ are plotted in figures 1(a)-1(f). In figures 1(g)-1(j) we report 4 derived data series, namely the interest rate differential $id_t = lti_t - sti_t$, the real short run interest rate $rsti_t = sti_t - \Delta p_t$, the real supply of M3, $m3r_t = m3_t - p_t$ and the log liquidity ratio $lr_t = m3r_t - y_t$. All these four last variables play an important role in the theories, that will be tested in this paper.

[6]The GDP deflator supplied by the Central Statistical Office(CBS) fluctuates more than 5% per quarter in the early 1980s. In other parts of the series no similar behaviour exists. The CBS explained they were aware of the problem, caused by linking series, but had no idea of the causes or indeed how to fix it.

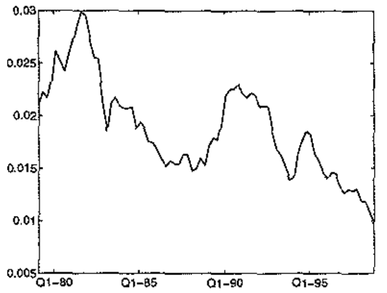
[7]Trade has to take place in a certain bond and there is a minimal amount outstanding (source: CBS statline)



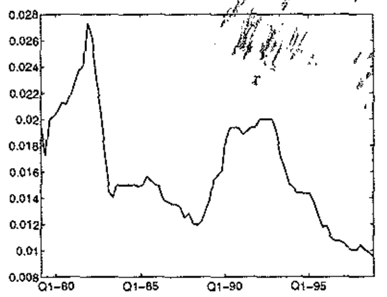

(a) $y_t$ real gdp



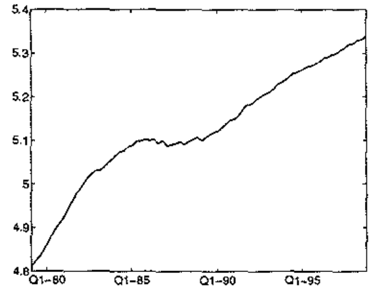

(b) $m3_t$ nominal M3



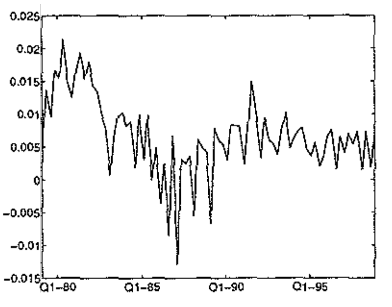

(c) $lri_t$, long run interest rate



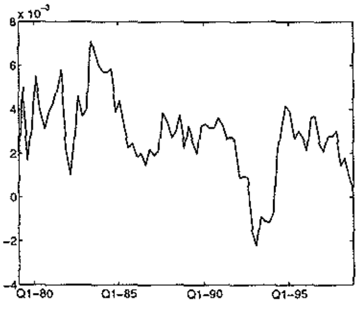

(d) $sri_t$, short run interest rate



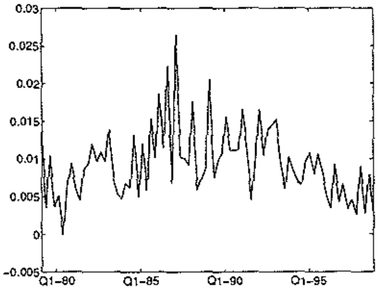

(e) $p_t$ CPI, base year 1975



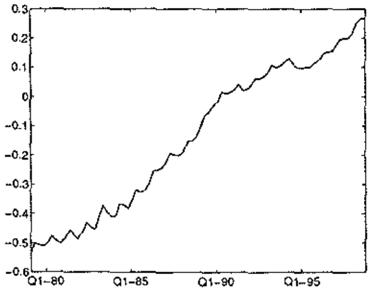

(f) $\Delta p_t$ inflation



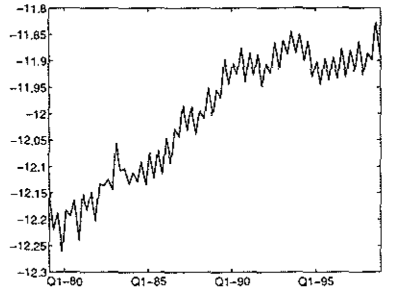

(g) $id_t$ interest rate differential




(h) $rsti_t$ real short term interest rate




(i) $m3r_t$ real M3 ($m3_t - p_t$)




(j) Log liquidity ratio, $lr_t = m3r_t - y_t$

Figure 7.1: The data series

Graph 1(g) shows that there is a decline in the interest rate spread over the period and that there is one episode in 1993 of an inverted yield curve. The real short term interest rate also declined in the last years of the sample. The growth in real M3 has been phenomenal an has far outstripped the growth in real income, such that the liquidity ratio increased considerable over the period under study.

## 7.4. Economic Theory

A host of economic theories predict constant relationships between the above mentioned variables. The following discussion also purports to show which ones have found empirical support in the empirical literature, which uses cointegrated VAR-models.

*Money demand*, $m^d$ in its most general form is given by:

$$m_t^d = b_1 y_t + b_2 p_t + b_3 sti_t - b_4 lti_t - b_5 \Delta p_t + b_6 t + u_t \tag{7.1}$$

where all parameters, with the exception of $b_6$, are assumed to be non-negative and $u_t$ is stationary. Often unit price $b_2$ and income $b_1$ elasticities are imposed. These are of course testable restrictions and will be treated as such. Furthermore $b_3 = b_4$ is often believed to be necessary, as the differential should give a measure of the opportunity cost of holding money. $b_6$ is a fairly crude way of including a long liberalization process or alternatively technological innovation.

*Aggregate income*. The standard $IS$ relationship predicts that trend-adjusted real aggregate income is negatively related to the real long term interest rate.

Alternatively trend-adjusted real income may cointegrate with inflation to yield a short-run Phillips curve as in Hendry and Mizon (1993) or Juselius (1996). Both alternatives are captured in the following relationship

$$y_t = b_1 t - b_2 lti_t + b_3 \Delta p_t + u_t \tag{7.2}$$

where $b_1 \geq 0,$ , $b_2 = b_3$ and $u_t \sim I(0)$ is consistent with an $IS$ curve and $b_2 = 0$ and $b_3 > 0$ specifies a short-run Phillips curve.

*Interest rate relations*. According to the Fisher parity, the expected real interest rate is a stationary process:

$$sti_t = \mathcal{E}_t(\Delta_8 p_{t+8})/8 + u_{1t} \tag{7.3}$$

(Here the yield curve is supposed to be increasing over the first two years, such that M3 yields the highest interest rate over 8 quarters). Unfortunately we cannot measure expectations with the present data set, so we have to rely on the outcome. If we make the auxiliary assumptions that:

$$u_{2t} = \mathcal{E}_t(\Delta_8 p_{t+8})/8 - \Delta_8 p_{t+8}/8 \tag{7.4}$$

and :

$$u_{3t} = \mathcal{E}_t(\Delta_8 p_{t+8})/8 - \Delta p_t \tag{7.5}$$

where $u_{2t}$ and $u_{3t}$ are stationary, just as $u_{1t}$ is, then we get:

$$rsti_t = sti_t - \Delta p_t \tag{7.6}$$

$$= u_{1t} + u_{2t} + u_{3t} \tag{7.7}$$

which is easily testable, but of course heavily dependent on the two auxiliary assumptions. Therefore a rejection does not imply a refutation of the Fisher parity.

The expectations hypothesis, augmented with the two auxiliary assumptions states that the interest differential between long and short term interest rates is stationary:

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

(7.8) $$id_t = lti_t - sti_t = u_{4t}$$

where $u_t$ is once more a stationary process.

*Central bank policy rules.* As the Central Bank targeted the exchange rate in a small open economy with already very liberal capital restrictions in the beginning of the period, it was able to influence neither the (short term) interest rate (which was set by Germany), nor the money supply. We thus do not expect to find a central bank policy rule.

Policy rules are found to yield stationary relationships by Beyer (1998) and Juselius (2001), but in these two cases, the respective countries executed their independent monetary policy.

*Monetarist theories* Textbook treatment of monetarist theories and also the Dutch moderate monetarism framework assume that the liquidity ratio (or alternatively money velocity, which is its inverse) is stationary:

(7.9) $$lr_t = m3_t - p_t - y_t = u_{5t}$$

Furthermore excess money will lead to inflation in the medium run and the central bank is assumed to be able to control inflation by increasing the short term interest rates. The last two statements are testable in the moving average representation of the $I(1)$ model.

## 7.5. The statistical model

In this section we discuss the statistical models used together with some extra remarks on particular outstanding issues.

**7.5.1. The I(2) model.** One representation of the $p$-dimensional I(2) model (Johansen, 1992) with 2 lags is given by:

(7.10) $$\Delta^2 X_t = \Pi X_{t-1} - \Gamma \Delta X_{t-1} + \mu_0 + \mu_1 t + \varepsilon_t$$

where $\varepsilon_t$ is distributed normally with mean zero and variance-covariance matrix $\Omega$

Define the characteristic polynomial of this process as:

(7.11) $$F(\lambda) = \lambda^2 I - (\Pi + 2I - \Gamma)\lambda - (\Gamma - I)$$

and let $\lambda_1, \ldots, \lambda_{2p}$ be the roots of $|F(\lambda)| = 0$

The following assumptions apply:

**I(2) a:** $\Pi = \alpha\beta'$ where $\alpha$ and $\beta$ are $p \times r$ matrices of full column rank. $r < p$

**I(2) b:** $2p - 2r - s$ roots $\lambda$ of the characteristic polynomial (7.11) equal one $\lambda = 1$. The other $2r + s$ roots are smaller than one in absolute value $|\lambda| < 1$. Let $\{\lambda_i^*\}, i = 1, \ldots, 2r + s$ indicate the roots of the second group. It then follows that $\alpha'_\perp \Gamma \beta_\perp = \xi\eta'$ where $\xi$ and $\eta$ are full rank matrices of dimension $(p - r) \times s$. Another equivalent way of stating this result is $\bar{\alpha}_\perp \alpha'_\perp \Gamma \bar{\beta}_\perp \beta'_\perp = \alpha_1 \beta'_1$ where $\alpha_1$ and $\beta_1$ are full rank matrices of dimension $p \times s$, $s < p - r$.

**I(2) c:** $\alpha'_2 \Theta \beta_2$ where $\Theta = \Gamma\bar{\beta}\bar{\alpha}'\Gamma + I$, $\alpha_2 = (\alpha : \alpha_1)_\perp$ and $\beta_2 = (\beta : \beta_1)_\perp$ is a matrix of full rank $(p - r - s)$.

On the deterministics, we put restrictions to make sure that all variables have a trend in the levels, but no quadratic or cubic trend. This implies that all the variables can be decomposed in a stochastic part $Y_t$ and deterministic part as:

(7.12) $$X_t = Y_t + m_1 + m_2 t$$

The following particular specification for the deterministic part of the I(2) model was originally proposed by (Rahbek et al., 1999) .

**I(2) t1:** $\mu_1 = \alpha\beta'_0$ where $\beta_0$ is a vector of length $r$.


DOI: 10.2870/52609

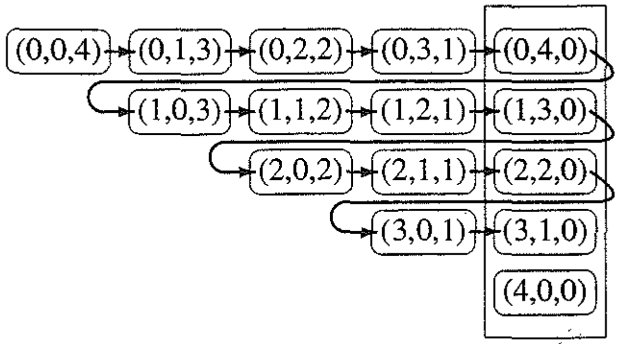



Figure 7.2: Selection of ranks in the I(2) model

**I(2) t2:** $\alpha_{\perp}'\mu_0 = \xi\eta_0' + \alpha_{\perp}'\Gamma\bar{\beta}\beta_0'$, where $\eta_0$ is a vector of length $p-r-s$.

No restrictions are placed on $\alpha'\mu_0$.

A few comments on these conditions are warranted: If $s = p - r$ we are back at the well known I(1) model, which is given below. Furthermore all unit roots are at 1 and nowhere else: seasonal unit roots are not considered in this chapter.

The number of I(1) trends equals $s$, whereas the number of I(2) trends is $p - r - s$.

The advantage of this representation is that all the restrictions are explicitly introduced in the standard VAR-model, one of the main workhorses of modern econometrics. A major disadvantage is that the restrictions are very complicated and non-linear. Furthermore the stationary relations are not obvious in this representation, so this representation does not offer any direct economic interpretation.

Another problem for the statistician is that no direct methods to explicitly maximize the likelihood function have been derived for this representation: it is difficult to maximize a likelihood under these complicated linear restrictions. So following Paruolo and Rahbek (1999) we reparametrize the model, such that conditions I(2)a, I(2)t1 and I(2)t2 become embedded in the parametrization:

$$\Delta^2 X_t = \alpha\left(\rho'\left(\tau' X_{t-1} + \tau_2 t\right) + \delta\tau_{\perp}'\Delta X_{t-1}' + \tau_1\right) + \zeta'\left(\tau'\Delta X_{t-1} + \tau_2\right) + \varepsilon_t \tag{7.13}$$

7.5.1.1. *Determination of the rank in the I(2) model.* In the I(2) model, we need to determine two ranks, namely $r$ and $s$. We can maximize the likelihood under the restrictions $H_0 : r = r_0, s = s_0$ to obtain $L_{\max}\left(H\left(r,s,p-r-s\right)\right)$ and form the likelihood ratio

$$Q\left(H\left(r,s,p-r-s\right)|H\left(p,0,0\right)\right) = \frac{L_{\max}\left(H\left(r,s,p-r-s\right)\right)}{L_{\max}\left(H\left(p,0,0\right)\right)}$$

Under the null $r = r_0, s = s_0$ the test statistic $-2\ln Q\left(H\left(r,s,p-r-s\right)|H\left(p,0,0\right)\right)$ asymptotically converges to a functional of Brownian motions, which for the deterministic specification in this paper has been tabulated by Rahbek et al. (1999). We start by testing $H\left(0,0,p\right)$. If rejected we test $H\left(0,1,p-1\right)$ to $H\left(0,p,0\right)$ and then test $H\left(1,0,p-1\right)$ as in figure 7.2. The grey column coincides with the standard I(1) rank tests. The number of I(2) trends, p-r-s equals zero in that column.

Paruolo (1996) shows that this procedure has an asymptotic rejection probability of 5% if tests performed are 5% tests.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

7.5.1.2. *The two step procedure and maximum likelihood estimation.* Two approaches have been proposed for the estimation of an I(2) system. Historically the so-called two step procedure (Johansen, 1995c) preceded the full maximum likelihood procedure (Johansen, 1997). Even though most applications so far have used the two step procedure, we shall apply the maximum likelihood procedure, except for the estimation of the rank, where we use both. Me2 (Omtzigt, 2003) is a computer package executing the maximum likelihood procedure.

For the determination of the cointegration ranks, we use the two step estimator as Paruolo (1996) proves that the procedure in figure 7.2, based on the two step estimator selects the correct integration indices with probability 95% (if 5% tests are used). To our knowledge a similar proof for the maximum likelihood estimator is not available. We shall however calculate the statistics for both procedures.

We shall only need to do inference on the parameters $\tau$ and $\tau_2$ in equation (7.13). Johansen (2002c) derives conditions under which likelihood ratio tests on $\tau$ are asymptotically $\chi^2$ distributed[8]. As little is known about the small sample properties of these tests (and nothing is known on the asymptotic distribution of tests on $\tau_2$), we shall also resort to bootstrapping these tests. Small sample properties will be more fully discussed in subsection 7.5.4.

**7.5.2. The I(1) model.** The $p$-dimensional vector autoregressive model with 2 lags can be represented in its reduced form as:

$$\Delta X_t = \alpha\beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + \mu_0 + \mu_1 t + \varepsilon_t \tag{7.14}$$

where $\varepsilon_t$ are distributed $N(0,\Omega)$.

Define the characteristic polynomial of this process as:

$$F(\lambda) = \lambda^2 I - (\Pi + I + \Gamma)\lambda - \Gamma \tag{7.15}$$

and let $\lambda_1, \ldots, \lambda_{2p}$ be the roots of $|F(\lambda)| = 0$

The assumptions, which assure that the model is I(1) are:

**I(1) a:** $p-r$ roots $\lambda$ of the characteristic polynomial (7.15) equal one: $\lambda = 1$. The other $p+r$ roots are smaller than one in absolute value $|\lambda| < 1$. Let $\{\lambda_i^*\}, i = 1, \ldots, p+r$ indicate the roots of the second group.

**I(1) b:** $\alpha$ and $\beta$ are full rank $p \times r$ matrices, $r < p$.

The following assumption is placed on the trend variable to assure that no quadratic trend is generated in the data, but that all variables have a trend in their levels is:

**I(1) t1:** $\alpha'_\perp \mu_1 = 0$ or equivalently $\mu_1 = \alpha\rho'$.

Inference in the I(1) model is thoroughly described in the monograph by Johansen (1995b).

**7.5.3. The nominal to real transformation (NtRT).** The I(2) model is a submodel of the I(1) model: it has the extra rank restriction I(2)c imposed on it. Inference in the I(2) model involves a number of still unknown distributions and maximum likelihood still has not been implemented for general restrictions. Hence a transformation from the I(2) to the I(1) model is highly desirable. Kongsted (2002) proposes a testable transformation, in which no information is lost and calls it the nominal to real transformation. The name suggest that the only nominal variables exhibit I(2) behaviour and that it can be removed by subtracting an appropriate price index. This holds true in many cases, but is not an absolute prerequisite, such that the name is slightly misleading.

The transformation starts from the observation that $(\tau' X_t, \tau'_\perp \Delta X) \sim I(1)$ or in fact any transformation $(\tau' X_t, s'_\perp \Delta X) \sim I(1)$ for which $|s'_\perp \tau_\perp| \neq 0$. So the proposed transformation is

[8] The I(2) model of Johansen (2002c)does not have any deterministics. We shall assume that the results will continue to hold true if these are included.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

to analyze $(\tau' X_t, s'_\perp \Delta X)$, which will be shown to have an autoregressive structure and a reduced rank, such that it is an I(1) model.

We shall now derive the autoregressive representation of the new process after the transformation. Let us first define $q$ and $x$ from the following equality, which we shall apply repeatedly:

$$\begin{aligned} I &= s\left(\tau' s\right)^{-1}\tau' + \tau_\perp\left(s'_\perp \tau_\perp\right)^{-1} s'_\perp \\ &= \left(q\tau' + x s'_\perp\right) \end{aligned}$$

If we multiply (7.13) by $\tau'$ and ignore the deterministic part for a moment, we obtain

$$\begin{aligned} \tau'\Delta X_t &= \tau'\alpha\rho'\left(\tau' X_{t-1}\right) + \tau'\alpha\delta\tau'_\perp q\left(\tau'\Delta X_{t-1}\right) + \tau'\alpha\delta\tau'_\perp x\left(s'_\perp \Delta X_{t-1}\right) + \tau'\zeta'\left(\tau'\Delta X_{t-1}\right) \\ &\quad + \tau'\Delta X_{t-1} + \tau'\varepsilon_t \end{aligned}$$

and pre-multiplying by $s'_\perp$ we get

$$\begin{aligned} s'_\perp \Delta^2 X_t &= s'_\perp \alpha\rho'\left(\tau' X_{t-1}\right) + s'_\perp \alpha\delta\tau'_\perp q\left(\tau'\Delta X_{t-1}\right) + s'_\perp \alpha\delta\tau'_\perp x\left(s'_\perp \Delta X_{t-1}\right) \\ &\quad + s'_\perp \zeta'\left(\tau'\Delta X_{t-1}\right) + s'_\perp \varepsilon_t \end{aligned}$$

Now collecting terms we get the transformed model, which is I(1) with rank $r$:

$$\begin{aligned} \begin{bmatrix} \tau'\Delta X_t \\ s'_\perp \Delta^2 X_t \end{bmatrix} &= \begin{bmatrix} \tau'\alpha \\ s'_\perp \alpha \end{bmatrix} \begin{bmatrix} \rho' & \delta\tau'_\perp x \end{bmatrix} \begin{bmatrix} \tau' X_t \\ s'_\perp \Delta X_t \end{bmatrix} \\ &\quad + \begin{bmatrix} \tau'\alpha\delta\tau'_\perp q + \tau'\zeta' + I & 0 \\ s'_\perp \alpha\delta\tau'_\perp q + s'_\perp \zeta' & 0 \end{bmatrix} \begin{bmatrix} \tau'\Delta X_{t-1} \\ s'_\perp \Delta^2 X_{t-1} \end{bmatrix} \\ &\quad + \begin{bmatrix} \tau' \\ s'_\perp \end{bmatrix} \varepsilon_t \end{aligned}$$

As there is only a trend in the levels of the variables, we see that the new variable $s'_\perp \Delta X_t = s'_\perp \Delta Y_t + s'_\perp m_2$ should only contain a level intercept, but no trend, whereas of course $\tau' X_t = \tau' Y_t + \tau' m_1 + \tau' m_2 t$ should still contain a trend.

So to keep the model exactly the same after the transformation, we should take account of the following two facts:

(1) The coefficient to the second lag of $s'_\perp \Delta^2 X_{t-1}$ are zero.
(2) $s'_\perp \Delta X_t$ should only contain an intercept, but no trend, whereas $\tau' X_t$ should contain a trend.

Kongsted and Nielsen (2002) study the effect of ignoring the restrictions 1 and 2 in a model which does not contain dummies, seasonal or otherwise. They do so by means of an application on real data and a simulation study, both a 3-dimensional VAR and find that "unrestricted reduced rank regression is shown to yield only a minor loss of efficiency compared to imposing the restrictions in the simulation experiment." They thus argue to transform the model to an I(1) model, ignoring the additional restrictions. They thus transform the I(2) model (7.10) with restrictions I(2)a-c, I(2)t1-2 to the the I(1) model (7.10) with restrictions I(1)a-b, I(1)t1.

**7.5.4. Small sample properties of cointegrated VAR models.** The small sample properties of most, if not all tests in the cointegrated VAR models (7.14) and (7.10) have from the beginning given cause for concern. Most attention has focused on the restrictions on the cointegration parameters $\beta$ in the I(1) model (7.14) (see Gonzalo (1994)). Two methods to overcome severe size distortion in small samples have been proposed and applied in the literature, namely Bartlett corrections and bootstrap methods. We shall discuss the implementation of these methods in inference on the cointegration parameters in the I(1) model and then give some comments on small sample properties of the other tests.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

Define $\beta^* = (\beta, \rho)'$ and let us consider the general I(1) model (7.14) and the following hypotheses on $\beta$ and $\rho$:

(1) $\beta^* = (\beta_0^*, \psi)$ that is $q \leq r$ out of the cointegration relationships are known entirely, including their trend. The other cointegration relations are unknown.
(2) $\beta = H\varphi$, that is the same restriction on all cointegration vectors (but no restrictions on the trend parameters).
(3) $\beta^* = (H_1\varphi, \ldots, H_r\varphi)$ that is generically identifying, linear restrictions on each of the vectors $\beta^*$. If the restrictions are not generically identifying, the algorithm by Omtzigt (2002b) can be used to render them identifying.

Johansen (2000a) derives the Bartlett correction for tests 1 and 2 under an assumption on the dummies, which in the current deterministic set-up implies that seasonal dummies can be taken account of, but other dummies not. By means of a Monte Carlo study he shows that the Bartlett corrected test has a size close to 5%.

Gredenhoff and Jacobson (2001) propose to bootstrap all three kinds of tests and do so in a small Monte Carlo study. This method can also be applied if besides seasonal dummies, other dummies are present in the model. They base their bootstrap on the estimate of $\beta^*$ under the null hypothesis and show by means of a simulation study that the size properties of the bootstrapped test are adequate.

Omtzigt and Fachin (2002) show that both methods can fail in terms of power. They argue that if the null hypothesis is false, then the estimated model under the (false) null hypothesis does not contain $p - r$ unit roots, but (at least) $p - r + 1$ unit roots, as (at least) one of the cointegration relations becomes non-stationary. If the estimated $(p - r + 1)$th unit root is close to unity, the Bartlett correction factor grows without bound (and becomes undefined when it is unity). Consequently the Bartlett corrected likelihood ratio test statistic becomes very small and the null hypothesis is wrongly accepted. By means of simulations they show that the corrected likelihood ratio can well become a biased test, as can the bootstrapped test statistic.

Let $\hat{\theta}_r = \left(\hat{\beta}_r, \hat{\rho}_r, \hat{\alpha}_r, \hat{\Gamma}_{1r}, \hat{\Omega}_r\right)$ be the estimates under $\mathcal{H}_0$ and $\hat{\theta}_u = \left(\hat{\beta}_u, \hat{\rho}_u, \hat{\alpha}_u, \hat{\Gamma}_{1u}, \hat{\Omega}_u\right)$ the estimates under the alternative. Then Omtzigt and Fachin (2002) propose to base the Bartlett correction on $\hat{\theta}_u$. They also suggest resampling from the DGP based on $\hat{\theta}_u$ (and not $\hat{\theta}_r$) when applying the bootstrap. $\mathcal{H}_0$ does not necessarily hold in the bootstrap sample, so propose to take a new null hypothesis $\mathcal{H}_0^\natural$ which holds in the bootstrap sample and equals $\mathcal{H}_0$ if $\mathcal{H}_0$ were to hold true in that sample. They propose $\mathcal{H}_0^\natural$ for the cases 1 and 2, but not for the more frequently applied case 3. In the appendix to this chapter, section 7.A we give one proposal for $\mathcal{H}_0^\natural$ in case 3.

As a general point we shall report $|\lambda^*_{\max}|$, that is the the largest root, that is not restricted to be unity by the model. If $|\lambda^*_{\max}|$ is substantially larger in the restricted model (and close to unity) than in the unrestricted model, then we consider that a sign that $\mathcal{H}_0$ should be rejected. The points raised by Omtzigt and Fachin (2002) are then amply illustrated, as both the bootstrap and the Bartlett correction cease to function, when $|\lambda^*_{\max}|$ is close to unity. If this is not the case there are no substantial differences between basing the Bartlett correction or bootstrap on the unrestricted estimates.

The asymptotic theory for parameters in the I(2) model has only just been developed and not even for all the tests on $\tau$ and $\tau_2$ we perform do we know the asymptotic distribution. So the small sample properties of the estimators are still very much unknown. We therefore bootstrap these tests. The point raised above on the extra unit root(s) is equally valid, but we still bootstrap using the restricted parameter estimates, as we do not have any equivalent null hypotheses for the bootstrap. We do however report the largest root in both the restricted and unrestricted model and note that in the test we perform they are extremely close, such that even with an equivalent hypothesis we are confident that we would obtain very much the same results.


DOI: 10.2870/52609

The Bartlett correction of Johansen (2002b) for the trace test in the I(1) model is not valid in the presence of seasonal dummies, which are present in our application. Subsequently we cannot use the correction. An alternative would be to block-bootstrap the residuals and resample under the null as proposed by Van Giersbergen (1996). Yet the length of the block makes a large difference and no clear guidance as to how to select the block length is available.

Two issues are of interest in the rank selection procedure. In the I(2) model the differences between rank selection based on maximum likelihood and rank selection based on the 2 step method has not been commented upon, so we compare the two.

Secondly if we were to apply a nominal-to-real transformation and impose the restrictions 1, 2 in subsection 7.5.4,if no dummies were present and if we were to use the unrestricted estimate of $\tau$, then the trace statistic for the rank in the I(2) model $\frac{L_{\max}(H(r,0,p-r))}{L_{\max}(H(p,0,0))}$ and the trace test in the I(1) model $\frac{L_{\max}(H(r))}{L_{\max}(H(p))}$ would be the same. We shall find that ignoring these conditions will cause them to be quite different: we use a restricted estimate of $\tau$, seasonal dummies are present in the model and we do not impose the two restrictions in subsection 7.5.4.

**7.5.5. Automated Model Selection.** The identification and restriction of the cointegration space in the I(1) model is a long and fairly arduous process. Following Davidson (1998a) who first automates the search for restrictions, Omtzigt (2002a) proposes a procedure for restriction and identification, which mimics the way Juselius (2002) searches for cointegration vectors.

If there is only one cointegration vector, the search procedure is as follows:

(1) The program creates $p+1$ unit vectors $h_1, \ldots, h_{p+1}$, corresponding to the $p$ variables and the trend in $\beta^*$. The user can specify additional vectors $e_i, i = 1, \ldots, u$. If the first two variables are money and income, then the user may define $e = [1, -1, 0, \ldots, 0]'$ which can be seen as the 'new' variable $m - y$. The program takes any possible combination of maximal $p$ of the vectors $h_1, \ldots, h_{p+1}, e, \ldots, e_u$ and forms the matrices $H_1$ through to $H_K$ all of full column rank. For each $i \neq j, sp(H_i) \neq sp(H_j)$ and the matrix which only consists of column $h_{p+1}$ (and would thus correspond to testing whether the trend is a stationary relation) is excluded.
(2) The programs tests restrictions of the kind $\beta^* = H_v \phi_v$. All those accepted at the 5% level are listed. First the accepted tests with the highest number of over-identifying restrictions (that is with the lowest number of columns in $H$) are reported. In case of an equal number of over-identifying restrictions, the test with the highest $p$-value is reported first.

If the rank of the cointegration matrix is 2, then the matrices $H$ in step 1 contain at most $p-1$ columns. In step 2, we first test all the individual restrictions of the form $\beta^* = (H_v \phi_v, \psi)$. Those accepted at the 1% level are then combined in a further step. Let $C_1 = \{1, \ldots c_1\}$ denote the set of accepted restrictions. We then test each combination $i, j \in C_1, i \neq j$ for which the restrictions are generically identifying. We thus test $\beta^* = (H_i \varphi, H_j \varphi)$ Let $C_2 = \{\{i,j\}_l, l = 1, \ldots, c_2\}$ define the set of combined restrictions that are accepted at the 5% level. Then order all the restrictions in $C_1$, which are accepted at the 5% level and those in $C_2$ according to the criteria above.

A more detailed account of the procedure can be found in Omtzigt (2002a), who also performs a Monte Carlo study to test the effectiveness of the procedure. He argues that the researcher should chose between the top-5 models selected and shows that under-selection of the lag length leads to a higher probability of recovering the true model.

Even though his simulations show that there is a sizeable size distortion in the procedures, he does not use any corrections (Bartlett or bootstrap) in his Monte Carlo simulations. In this paper we combine automated model search with small sample corrections to gauge whether the resulting procedure is useful for the data set at hand.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

| Misspecification tests in the unrestricted VAR | | | | | | |
|---|---|---|---|---|---|---|
| | Univariate | | | | | Multivariate |
| Variable | $y99_t$ | $p_t$ | $m3_t$ | $rsri_t$ | $id_t$ | |
| Normality | 1.96(0.38) | 2.64(0.27) | 0.25(0.88) | 4.22(0.12) | 1.51(0.47) | 11.63(0.31) |
| AR1 | 0.03(0.87) | 0.03(0.85) | 0.01(0.91) | 0.06(0.81) | 0.55(0.46) | 25.08(0.46) |
| skewness | 0.36 | −0.25 | −0.02 | 0.37 | −0.08 | |
| kurtosis | 3.18 | 3.50 | 2.95 | 3.79 | 3.30 | |

Table 7.1: Misspecification tests of the unrestricted VAR with two lags

## 7.6. The empirical analysis

Even though data is available from 1977 quarter 1 (the start of the GDP data series), we effectively use data from 1979 first quarter onwards, as 1977 and 1978 contain a series of outliers, which would require a number of dummies.

Instead of modelling straight away the five data series, we choose to model the following transformation: $y_t$, $p_t$, $m3_t, rsti_t$ and $id_t$. All these variables are plotted in figure 1. We needed exactly two lags and centered seasonal dummies, but no other dummies. The residuals of the unrestricted VAR with two lags show no sign of misspecification, see table 7.1. The normality and AR1 tests have been taken from Doornik and Hansen (1994) and Doornik (1996) respectively. The $p$-values between brackets indicate that the model is well specified.

**7.6.1. Determination of rank in the I(2) model.** The determination of the ranks in the I(2) model is done as described in section 7.5.1.1. We have calculated the relevant likelihood ratio statistics based on the 2 step estimator and those based on the maximum likelihood estimator. They are reported in tables 7.2 and 7.3. The 95% percentile of the asymptotic distribution is given in brackets below the test statistics: we remark once more that we have no formal proof that these percentiles are valid for the maximum likelihood estimator.

From a theoretical point of view we expect there to be just one I(2) trend, which is the nominal trend in $m3_t$ and $p_t$. Using table 7.2 the first accepted rank is $(r, s, p-r-s) = (1, 3, 1)$. According to Paruolo (1996) we should now stop and accept this rank.

We do however continue and find that the next hypothesis accepted is $(r, s, p-r-s) = (2, 2, 1)$: from a theoretical perspective we expect at least two cointegration relations to be present in the data set. Furthermore the rank $(r, s, p-r-s) = (1, 3, 1)$ was only just accepted.

We note that 2 step estimation and maximum likelihood give identical results when $r = 0$ and when $p-r-s = 0$: these are the top row and the right hand column in the rank tables. In all other cases the statistic based on maximum likelihood is remarkably lower than the statistic based on the two step estimator.

For two reasons shall we continue to do the rest of the analysis for $(r, s, p-r-s) = (1, 3, 1)$. Firstly it is the first rank accepted in both procedures. Secondly and more importantly, it turns out that using inference from $r = 1$ is beneficial for $r = 2$.

**7.6.2. The nominal to real transformation(s).** The aim of this section is to fully determine $\tau$, for once we have done so, we can apply the nominal to real transformation and continue our analysis in the I(1) model. $\tau' X_t$ contains all the variables and combinations of variables that are at most I(1). The variables are $X_t = (y_t, m_t, p_t, rsti_t, id_t)$: real GDP, m3, the consumer price index, the real short term interest rate and the interest rate differential between the nominal long


DOI: 10.2870/52609

| 2 Step Inference | | | | | | | |
|---|---|---|---|---|---|---|---|
| $p-r$ | $r$ | | | | | | |
| 5 | 0 | 593.2 (198.2) | 308.0 (167.9) | 217.1 (142.2) | 151.8 (119.8) | 114.8 (101.5) | 109.4 (87.2) |
| 4 | 1 | | 414.9 (137.0) | 169.0 (113.0) | 103.2 (92.2) | 68.8 (75.3) | 63.8 (62.8) |
| 3 | 2 | | | 201.7 (87.6) | 77.0 (68.2) | 42.6 (53.2) | 27.1 (42.7) |
| 2 | 3 | | | | 118.3 (47.6) | 44.8 (34.4) | 13.3 (25.4) |
| 1 | 4 | | | | | 46.0 (19.9) | 2.1 (12.5) |
| $p-r-s$ | | 5 | 4 | 3 | 2 | 1 | 0 |

Table 7.2: Test statistics of the rank selection procedure, based on two step estimation

| Maximum Likelihood Inference | | | | | | | |
|---|---|---|---|---|---|---|---|
| $p-r$ | $r$ | | | | | | |
| 5 | 0 | 593.2 (198.2) | 308.0 (167.9) | 217.1 (142.2) | 151.8 (119.8) | 114.8 (101.5) | 109.4 (87.2) |
| 4 | 1 | | 198.2 (137.0) | 132.0 (113.0) | 90.8 (92.2) | 68.1 (75.3) | 63.8 (62.8) |
| 3 | 2 | | | 85.0 (87.6) | 59.0 (68.2) | 38.4 (53.2) | 27.1 (42.7) |
| 2 | 3 | | | | 36.8 (47.6) | 23.1 (34.4) | 13.3 (25.4) |
| 1 | 4 | | | | | 12.3 (19.9) | 2.1 (12.5) |
| $p-r-s$ | | 5 | 4 | 3 | 2 | 1 | 0 |

Table 7.3: Test statistics of the rank selection procedure, based on maximum likelihood estimation

term interest rate and the nominal short term interest rate. Let $h_1$ be the five-dimensional unit vector with 1 as the first element and $e_1 = [0, 1, -1, 0, 0]$.

In the unrestricted I(2) model with ranks $(r, s, p-r-s) = (1, 3, 1)$ the largest root of the characteristic polynomial that is not restricted to 1 equals 0.55 in absolute value, see table 7.4, where we report the outcome of the tests on $\tau$. The test that $\tau = [h_1, h_4, h_5, \psi]$, where $\psi$ varies freely. This hypothesis implies that that $y_t$, $rsti_t$ and $id_t$ are (at most) I(1) variables with a linear trend, and that $m_t$ and $p_t$ share the same I(2) trend, but that this trend does not feed proportionally into both variables. It takes a value of 9.82. Since under the null hypothesis this test has a $\chi^2$-distribution with 3 degrees of freedom, its $p$-value is 0.03 and the test is rejected. We note that the maximum root of the characteristic polynomial is 0.63 and thus close to 0.56 and bootstrap the test statistic to find that the hypothesis is accepted with a $p$-value of 0.11. The bootstrap procedure has been based on the restricted estimate (and not on the unrestricted estimate as argued before) of the parameters. Yet we shall see later, when testing in the I(1) model, that there may not be a large difference when $|\lambda^*_{\max}|$ does not increase too much.

The next hypothesis $\tau = [h_1, h_4, h_5, e_1]$, which implies that the I(2)-trend feeds proportionally into money and prices, is just rejected with a bootstrapped p-value of 0.03.

The test that $\tau = [h_1, h_4, h_5, e_1], \tau_2 = [*, 0, 0, *]$, which implies that $id_t$ and $rsti_t$ are at most I(1), but do not contain a linear trend is accepted with a p-value of 0.08. Note that Johansen (2002c) does not report the asymptotic distribution for this latest test, such that we can only report the $p$-value of the bootstrapped test statistic.

 DOI: 10.2870/52609

| | The nominal to real transformation | | | | |
|---|---|---|---|---|---|
| $(r, s, p-r-s)$ | I(1) with trend | I(1) without trend | $\lvert\lambda^*_{\max}\rvert$ | LR test | BS $p$-value |
| $(1,3,1)$ | | | 0.55 | | |
| | $y_t, m3_t - 0.1p_t, id_t, rsti_t$ | | 0.63 | 9.29 (0.03) | 0.11 |
| | $y_t, m3_t - p_t, id_t, rsti_t$ | | 0.58 | 16.26 (0.00) | 0.03 |
| | $y_t, m3_t - p_t$ | $id_t, rsti_t$ | 0.60 | 17.07 | 0.08 |
| $(2,2,1)$ | | | 0.76 | | |
| | $y_t, m3_t - 0.8p_t, id_t, rsti_t$ | | 0.79 | 7.67 (0.05) | 0.18 |
| | $y_t, m3_t - p_t, id_t, rsti_t$ | | 0.80 | 7.97 (0.09) | 0.29 |
| | $y_t, m3_t - p_t$ | $id_t, rsti_t$ | 0.80 | 8.88 | 0.51 |

Table 7.4: Testing the nominal to real transformation for r=1 (top) and r=2 (bottom)

| | After transformation | | | | I(2) model | |
|---|---|---|---|---|---|---|
| | Trace | 95% | Lmax | 95% | Trace | Lmax |
| $r=0$ | 90.3 | 87.0 | 39.8 | 24.8 | 109.4 | 45.6 |
| $r=1$ | **50.5** | 62.2 | 21.8 | 20.0 | 63.8 | 36.7 |
| $r=2$ | 28.6 | 42.2 | **16.1** | 16.7 | **27.1** | **13.8** |
| $r=3$ | 12.6 | 25.5 | 10.1 | 13.1 | 13.3 | 11.2 |
| $r=4$ | 2.5 | 12.4 | 2.5 | 12.4 | 2.1 | 2.1 |

Table 7.5: Comparison of the rank tests in the I(1) and the I(2) model

For the choice $(r, s, p-r-s) = (2, 2, 1)$ all three hypotheses are accepted with relatively large bootstrapped $p$-values and $|\lambda^*_{\max}|$ does not increase much in the restricted models.

We thus conclude that the nominal to real transformation, where the I(2) trend feeds proportionally into money and prices is accepted for both choices of ranks.

We accept the transformation for both choices of ranks and proceed with an I(1)-analysis of the transformed data vector $X_t = [y_t, m3r_t, id_t, rsti_t, \Delta p_t]$, where $m3r_t = m3_t - p_t$.

**7.6.3. Rank tables after NtR transformation.** After the transformation, where following Kongsted and Nielsen (2002) we do not impose the restrictions 1 and 2 on page 146 on the transformed systems, we obtain the trace and Lmax statistics in table 7.5.(Note that additional differences are caused by the inclusion of seasonal dummies in both the I(2) and I(1) model). For comparison we also report the original statistics from the untransformed I(2) model from table 7.5. We see that if we use the trace statistic, then we accept $r = 1$, whereas in the I(2) model we would have accepted $r = 2$. Yet is were to use the Lmax statistic, both before and after the transformation, the choice would be $r = 2$. We thus continue with both choices of rank. In general the differences between the test statistics are substantial and thus ignoring the additional restrictions does make a large difference in this data set.

**7.6.4. Hypothesis testing on individual vectors.**

7.6.4.1. *Hypothesis testing for r=1.* Under the assumption that $r = 1$, we test a number of hypotheses on the cointegration space $\beta$. Since the all the restrictions are of the type $\beta = H\phi$ and only seasonal dummies are present, we can apply the Bartlett correction to these tests. We report the uncorrected LR-test statistic (with $p$-value underneath between brackets) and two Bartlett corrected LR-tests: the one based on the unrestricted estimates (only $r = 1$ is imposed) and the

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

restricted estimates. We shall base our decisions on the Bartlett corrected LR-test, which is based on the unrestricted estimate.

$\mathcal{H}_1 - \mathcal{H}_5$ are tests whether individual variables are trend-stationary. All these are soundly rejected. In $\mathcal{H}_6 - \mathcal{H}_9$ we test whether respectively the log-velocity $(y_t - m3r_t)$, the real long term interest rate $(rlti_t = lti_t - \Delta p_t)$, the nominal short term interest rate or the nominal long term interest rate are stationary. Once more all these hypotheses are rejected. With $\mathcal{H}_{10}$ we reject that any combination between $y_t$ and $m3r_t$ is stationary, whereas $\mathcal{H}_{11}$ rejects any stationary relationship among the interest rates and inflation. We conclude that the stationary relation is thus a combination of $y_t$ and $m3r_t$ on the one hand and interest rates and inflation on the other. This means either a money demand relation or an aggregate income relation. $\mathcal{H}_{12} - \mathcal{H}_{14}$ are different forms of money demand. $\mathcal{H}_{13}$ is the accepted relation with the largest number of restrictions (3) and is a special case of the money demand equation (7.1) with $b_1 = b_2 = b_3 = b_5$ and $b_4 = 0$. In $\mathcal{H}_{15} - \mathcal{H}_{18}$ we test different forms of aggregate income relations (7.2) (in $\mathcal{H}_{15}$ the coefficient on $m3r$ is estimated freely, but equals zero) and find that $\mathcal{H}_{17}$, an IS curve (a relation between real income and the real long term interest rate) and $\mathcal{H}_{18}$, a short run Phillips curve are both accepted.

We conclude that of the hypotheses considered, we can accept either of $\mathcal{H}_{13}$, $\mathcal{H}_{17}$ and $\mathcal{H}_{18}$.

All Bartlett corrections are defined, whether they are based on the unrestricted or the restricted estimators, as the maximum unrestricted eigenvalue of the characteristic polynomial (7.15) never exceeds one in absolute value. Yet we note that the two hypotheses, that are most strongly rejected, $\mathcal{H}_2$ and $\mathcal{H}_6$ are accepted, when the Bartlett correction is based on the restricted estimators. Their $|\lambda^*_{\max}|$ is very close to unity, such that the Bartlett factor is extremely large. This further corroborates the point on power raised in this paper.

7.6.4.2. *Hypothesis testing for r=2.* For $r = 2$, we first consider hypotheses of the kind $\beta^* = (H_1\varphi, \psi)$. No Bartlett correction has yet been derived for them, so we bootstrap the test statistics. Once again we have the choice of basing the bootstrap on the unrestricted estimate and the restricted estimate.

To base the bootstrap on the unrestricted estimate, we need to formulate an alternative null hypothesis, which is satisfied by the bootstrap sample. The construction of the alternative hypothesis is given in the appendix of this paper.

We test 37 hypothesis on one of the two vectors (leaving the other unrestricted) and calculate the three corresponding $p$-values. The results are reported in table 7.7. The last row of the table contains $|\lambda^*_{\max}|$. The Bartlett correction is not defined, when $|\lambda^*_{\max}| \geq 1$, but from a computational point of view, the bootstrap can be applied. We base our conclusions on the bootstrapped $p$-value that is based on the unrestricted estimate.

$\mathcal{H}_1 - \mathcal{H}_9$ are tests that exactly one of the variables is stationary, whereas $\mathcal{H}_{10} - \mathcal{H}_{18}$ are the corresponding tests for trend-stationarity. $\mathcal{H}_{19}$ is a test for trend-stationarity of velocity without imposing equal coefficients on real money and prices. $\mathcal{H}_{20} - \mathcal{H}_{27}$ are tests on the stationarity of combinations of interest rates and inflation, $\mathcal{H}_{28} - \mathcal{H}_{31}$ are hypotheses, corresponding to an aggregate demand curve and $\mathcal{H}_{32} - \mathcal{H}_{37}$ hypotheses on a money demand relationship.

The bootstrap based on the restricted estimate accepts all 37 hypotheses, which means the procedure does not have any discriminatory power in this context. The uncorrected likelihood ratio test accepts 10 hypotheses, whereas the bootstrap based on the unrestricted estimate accepts 25 hypotheses, among whom five out of $\mathcal{H}_{10} - \mathcal{H}_{18}$. Under the assumption that $r = 2$, at most two of them can hold true as each of these 9 hypothesis concerned test that one variable or a predefined linear combination of two or three variables is stationary.

Combining two hypotheses at a time out of the 25 accepted, we find that 145 out of 228 possible combinations of restrictions, which are generically identifying, are accepted. The one with the largest number of over-identifying restrictions accepted is $\mathcal{H}_5 + \mathcal{H}_{10}$. Yet the large number

| | Restricted estimates of $\beta$ | | | | | | | | Unrestricted es | | Restricted es | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $y$ | $m3r$ | $rsti$ | $id$ | $\Delta p$ | $1000t$ | dof | LR test | BF | LR test | BF | LR test | $\lvert\lambda^*_{\max}\rvert$ |
| $\mathcal{H}_1$ | 1 | 0 | 0 | 0 | 0 | −6.67 | 4 | 23.72 (0.00) | 1.43 | 16.54 (0.00) | 2.30 | 10.30 (0.04) | 0.84 |
| $\mathcal{H}_2$ | 0 | 1 | 0 | 0 | 0 | −12.40 | 4 | 33.08 (0.00) | 1.43 | 23.06 (0.00) | 3.46 | 9.57 (0.05) | 0.96 |
| $\mathcal{H}_3$ | 0 | 0 | 1 | 0 | 0 | 0.04 | 4 | 24.56 (0.00) | 1.43 | 17.12 (0.00) | 1.96 | 12.50 (0.01) | 0.73 |
| $\mathcal{H}_4$ | 0 | 0 | 0 | 1 | 0 | 0.05 | 4 | 29.74 (0.00) | 1.43 | 20.73 (0.00) | 2.41 | 12.36 (0.01) | 0.78 |
| $\mathcal{H}_5$ | 0 | 0 | 0 | 0 | 1 | 0.04 | 4 | 22.55 (0.00) | 1.43 | 15.72 (0.00) | 1.88 | 11.99 (0.02) | 0.77 |
| $\mathcal{H}_6$ | 1 | −1 | 0 | 0 | 0 | 2.85 | 4 | 33.96 (0.00) | 1.43 | 23.67 (0.00) | 4.25 | 7.99 (0.09) | 0.95 |
| $\mathcal{H}_7$ | 0 | 0 | 1 | 1 | 0 | 0.08 | 4 | 25.04 (0.00) | 1.43 | 17.45 (0.00) | 1.99 | 12.59 (0.01) | 0.74 |
| $\mathcal{H}_8$ | 0 | 0 | 1 | 0 | 1 | 0.11 | 4 | 25.31 (0.00) | 1.43 | 17.65 (0.00) | 1.93 | 13.09 (0.01) | 0.72 |
| $\mathcal{H}_9$ | 0 | 0 | 1 | 1 | 1 | 0.06 | 4 | 28.21 (0.00) | 1.43 | 19.67 (0.00) | 2.14 | 13.20 (0.01) | 0.86 |
| $\mathcal{H}_{10}$ | 1 | 0.05 | 0 | 0 | 0 | −7.20 | 3 | 23.66 (0.00) | 1.46 | 16.20 (0.00) | 2.40 | 9.84 (0.02) | 0.84 |
| $\mathcal{H}_{11}$ | 0 | 0 | 0.12 | 1 | 0.83 | 0.08 | 2 | 21.19 (0.00) | 1.49 | 14.26 (0.00) | 1.92 | 11.02 (0.00) | 0.71 |
| $\mathcal{H}_{12}$ | 1 | −1 | 0 | −50.84 | 0 | 2.07 | 3 | 27.57 (0.00) | 1.46 | 18.88 (0.00) | 2.25 | 12.26 (0.01) | 0.74 |
| $\mathcal{H}_{13}$ | 1 | −1 | 15.06 | 0 | 0 | 5.02 | 3 | 7.03 (0.07) | 1.46 | 4.81 (0.19) | 1.54 | 4.58 (0.21) | 0.50 |
| $\mathcal{H}_{14}$ | 1 | −1 | 17.00 | 0 | 4.59 | 5.38 | 2 | 3.84 (0.15) | 1.49 | 2.58 (0.27) | 1.53 | 2.51 (0.29) | 0.49 |
| $\mathcal{H}_{15}$ | 1 | −0.00 | 0 | 0.95 | −5.71 | −6.69 | 1 | 8.25 (0.00) | 1.51 | 5.46 (0.02) | 1.65 | 5.00 (0.03) | 0.57 |
| $\mathcal{H}_{16}$ | 1 | 0 | 3.33 | 3.33 | −3.56 | −6.40 | 2 | 5.84 (0.05) | 1.49 | 3.93 (0.14) | 1.57 | 3.72 (0.16) | 0.55 |
| $\mathcal{H}_{17}$ | 1 | 0 | 7.12 | 7.12 | 0 | −5.94 | 3 | 9.78 (0.02) | 1.46 | 6.70 (0.08) | 1.59 | 6.14 (0.10) | 0.52 |
| $\mathcal{H}_{18}$ | 1 | 0 | 0 | 0 | −5.83 | −6.77 | 3 | 8.35 (0.04) | 1.46 | 5.72 (0.13) | 1.58 | 5.28 (0.15) | 0.57 |

Table 7.6: Tests in model with r=1 and 2 lags. All tests are Bartlett corrected in two ways: once the Bartlett factor is based on the unrestricted estimates and the second time it is based on the restricted estimates. For the unrestricted estimate $|\lambda^*_{\max}| = 0.51$.

of accepted hypotheses indicates that the bootstrapped test has got relatively little discriminatory power.

We have executed a model search as described in section 7.5.5 with one exception: we have selected only a handful of hypotheses and certainly not every combination of unit vectors $\mathcal{H}_1 - \mathcal{H}_5$ and additional vectors $\mathcal{H}_6 - \mathcal{H}_9$. Yet the large number of accepted hypotheses in table 7.7 and the computer intensity of the bootstrap would make a full bootstrapped search costly in terms of computing time. Furthermore even in this limited search large numbers of mutually exclusive models were accepted, leading to the conclusion that for the data set at hand, basing the final specification on a bootstrapped automated model search is not the way to proceed.

**7.6.5. Automated Model Selection.** Bearing in mind the apparent failure of (semi-)automated model selection, based on the bootstrapped test statistics, we proceed with a full automatic model search, based on $\mathcal{H}_1 - \mathcal{H}_9$ (four user specified variables). That is with the data vector $X_t = [y_t, m3r_t, id_t, rsti_t, \Delta p_t]$, we define the following search directions $e_1 = [1, -1, 0, 0, 0], e_2 = [0, 0, 1, 1, 0], e_3 = [0, 0, 1, 1, 1]$ and $e_4 = [0, 0, 1, 0, 1]$, which correspond to the log-velocity, the


DOI: 10.2870/52609

| | Restricted estimates of $\beta$ | | | | | | | | P-values | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | y | m3r | rsi | id | $\Delta p$ | 1000t | dof | LR test | LR | BS(un) | BS(res) | $\lvert\lambda^*_{max}\rvert$ |
| $\mathcal{H}_1$ | 1 | 0 | 0 | 0 | 0 | 0 | 4 | 19.24 | 0.00 | 0.01 | 0.46 | 1.009 |
| $\mathcal{H}_2$ | 0 | 1 | 0 | 0 | 0 | 0 | 4 | 18.27 | 0.00 | 0.02 | 0.41 | 1.000 |
| $\mathcal{H}_3$ | 0 | 0 | 1 | 0 | 0 | 0 | 4 | 16.65 | 0.00 | 0.02 | 0.26 | 0.903 |
| $\mathcal{H}_4$ | 0 | 0 | 0 | 1 | 0 | 0 | 4 | 14.92 | 0.00 | 0.05 | 0.36 | 0.824 |
| $\mathcal{H}_5$ | 0 | 0 | 0 | 0 | 1 | 0 | 4 | 7.30 | 0.12 | 0.48 | 0.72 | 0.870 |
| $\mathcal{H}_6$ | 1 | −1 | 0 | 0 | 0 | 0 | 4 | 17.12 | 0.00 | 0.02 | 0.38 | 0.983 |
| $\mathcal{H}_7$ | 0 | 0 | 1 | 1 | 0 | 0 | 4 | 19.62 | 0.00 | 0.00 | 0.20 | 0.930 |
| $\mathcal{H}_8$ | 0 | 0 | 1 | 1 | 1 | 0 | 4 | 11.64 | 0.02 | 0.15 | 0.57 | 0.941 |
| $\mathcal{H}_9$ | 0 | 0 | 1 | 0 | 1 | 0 | 4 | 11.55 | 0.02 | 0.26 | 0.53 | 0.929 |
| $\mathcal{H}_{10}$ | 1 | 0 | 0 | 0 | 0 | −6.73 | 3 | 8.23 | 0.04 | 0.22 | 0.54 | 0.898 |
| $\mathcal{H}_{11}$ | 0 | 1 | 0 | 0 | 0 | −12.34 | 3 | 15.23 | 0.00 | 0.04 | 0.23 | 0.954 |
| $\mathcal{H}_{12}$ | 0 | 0 | 1 | 0 | 0 | 0.11 | 3 | 13.21 | 0.00 | 0.03 | 0.29 | 0.918 |
| $\mathcal{H}_{13}$ | 0 | 0 | 0 | 1 | 0 | 0.05 | 3 | 12.46 | 0.01 | 0.07 | 0.29 | 0.760 |
| $\mathcal{H}_{14}$ | 0 | 0 | 0 | 0 | 1 | 0.02 | 3 | 7.13 | 0.07 | 0.39 | 0.59 | 0.849 |
| $\mathcal{H}_{15}$ | 1 | −1 | 0 | 0 | 0 | 1.69 | 3 | 17.05 | 0.00 | 0.02 | 0.18 | 0.972 |
| $\mathcal{H}_{16}$ | 0 | 0 | 1 | 1 | 0 | 0.14 | 3 | 12.57 | 0.01 | 0.03 | 0.29 | 0.916 |
| $\mathcal{H}_{17}$ | 0 | 0 | 1 | 1 | 1 | 0.12 | 3 | 7.46 | 0.06 | 0.24 | 0.49 | 0.722 |
| $\mathcal{H}_{18}$ | 0 | 0 | 1 | 0 | 1 | 0.06 | 3 | 10.63 | 0.01 | 0.21 | 0.41 | 0.865 |
| $\mathcal{H}_{19}$ | 1 | 0.12 | 0 | 0 | 0 | −8.01 | 2 | 7.95 | 0.02 | 0.11 | 0.32 | 0.899 |
| $\mathcal{H}_{20}$ | 0 | 0 | 1 | −2.59 | 0 | 0 | 3 | 13.54 | 0.00 | 0.02 | 0.23 | 0.826 |
| $\mathcal{H}_{21}$ | 0 | 0 | 1 | 3.52 | 0 | 0.25 | 2 | 11.77 | 0.00 | 0.01 | 0.12 | 0.828 |
| $\mathcal{H}_{22}$ | 0 | 0 | 1 | 0 | 23.03 | 0 | 3 | 7.27 | 0.06 | 0.38 | 0.56 | 0.872 |
| $\mathcal{H}_{23}$ | 0 | 0 | 1 | 0 | 4.31 | 0.15 | 2 | 6.91 | 0.03 | 0.24 | 0.33 | 0.845 |
| $\mathcal{H}_{24}$ | 0 | 0 | 0 | 1 | 3.35 | 0 | 3 | 7.20 | 0.07 | 0.32 | 0.56 | 0.872 |
| $\mathcal{H}_{25}$ | 0 | 0 | 0 | 1 | 1.01 | 0.07 | 2 | 6.42 | 0.04 | 0.24 | 0.40 | 0.798 |
| $\mathcal{H}_{26}$ | 0 | 0 | 1 | 3.55 | 12.21 | 0 | 2 | 7.18 | 0.03 | 0.24 | 0.34 | 0.875 |
| $\mathcal{H}_{27}$ | 0 | 0 | 1 | 2.30 | 1.65 | 0.20 | 1 | 4.87 | 0.03 | 0.19 | 0.21 | 0.704 |
| $\mathcal{H}_{28}$ | 1 | 0 | 1.15 | 1.15 | −5.18 | −6.65 | 1 | 0.30 | 0.59 | 0.70 | 0.66 | 0.662 |
| $\mathcal{H}_{29}$ | 1 | 0 | 4.75 | 4.75 | 0 | −6.20 | 2 | 6.51 | 0.04 | 0.14 | 0.29 | 0.858 |
| $\mathcal{H}_{30}$ | 1 | 0 | 0 | 0 | −6.06 | −6.79 | 2 | 0.49 | 0.78 | 0.89 | 0.89 | 0.676 |
| $\mathcal{H}_{31}$ | 1 | 0.02 | 0 | 0 | −6.10 | −7.00 | 1 | 0.46 | 0.50 | 0.68 | 0.63 | 0.674 |
| $\mathcal{H}_{32}$ | 1 | −1 | 0 | 0 | −60.83 | 2.18 | 2 | 7.16 | 0.03 | 0.24 | 0.34 | 0.839 |
| $\mathcal{H}_{33}$ | 1 | −0.49 | 8.58 | 0 | 0 | −0.78 | 1 | 0.80 | 0.37 | 0.51 | 0.48 | 0.651 |
| $\mathcal{H}_{34}$ | 1 | −1 | 15.24 | 0 | 0 | 5.18 | 2 | 5.04 | 0.08 | 0.21 | 0.30 | 0.818 |
| $\mathcal{H}_{35}$ | 1 | −1 | 17.44 | 0 | 6.13 | 5.55 | 1 | 0.24 | 0.62 | 0.75 | 0.73 | 0.633 |
| $\mathcal{H}_{36}$ | 1 | 0.40 | 0 | 16.61 | 0 | −10.39 | 1 | 4.98 | 0.03 | 0.09 | 0.23 | 0.827 |
| $\mathcal{H}_{37}$ | 1 | −1 | 0 | −69.44 | 0 | 1.20 | 2 | 11.42 | 0.00 | 0.02 | 0.15 | 0.744 |

Table 7.7: Tests on a single cointegration vector in model with $r = 2$ and 2 lags. All tests are Bartlett corrected in two ways: once the Bartlett factor is based on the unrestricted estimates and the second time it is based on the restricted estimates. For the unrestricted estimate $|\lambda^*_{max}| = 0.642$

long term real interest rate, the long term nominal interest rate and the short term nominal interest rate respectively.

We run the algorithm for both $r = 1$ and $r = 2$, and base our decisions on the asymptotic LR-tests, uncorrected for small sample properties.

7.6.5.1. *Rank = 1, 2 Lags.* We report the first five models $\mathcal{M}_1 - \mathcal{M}_5$ from the automated model selection in table 7.8. We also add the two models, which we accepted previously (after


DOI: 10.2870/52609

| Model | Nr Res | LR | P value | $y$ | $m3r$ | $rsti$ | $id$ | $\Delta p$ | $trend$ |
|---|---|---|---|---|---|---|---|---|---|
| $\mathcal{M}_1$ | 3 | 1.14 | 0.767 | 1 | −0.55 | 9.26 | 0 | 0 | 0 |
| $\mathcal{M}_2$ | 3 | 6.47 | 0.091 | 1 | −0.52 | 9.90 | 9.90 | 0 | 0 |
| $\mathcal{M}_3$ | 3 | 7.03 | 0.071 | 1 | −1 | 15.06 | 0 | 0 | 0.0050 |
| $\mathcal{M}_{3e1}$ | 3 | 8.35 | 0.039 | 1 | 0 | 0 | 0 | −5.83 | −0.0067 |
| $\mathcal{M}_{3e2}$ | 3 | 9.78 | 0.021 | 1 | 0 | 7.12 | 7.12 | 0 | −0.0059 |
| $\mathcal{M}_4$ | 2 | 0.62 | 0.733 | 1 | −0.54 | 10.03 | 1.15 | 1.15 | 0 |
| $\mathcal{M}_5$ | 2 | 0.82 | 0.662 | 1 | −0.48 | 8.53 | 0 | 0 | −0.0008 |

Table 7.8: Automated model selection with $r = 1$ and 2 lags

| Model | Nr Res | LR | P value | $y$ | $m3r$ | $rsti$ | $id$ | $\Delta p$ | $trend$ |
|---|---|---|---|---|---|---|---|---|---|
| $\mathcal{M}_1$ | 6 | 8.42 | 0.209 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | | | 1 | −0.55 | 9.72 | 0 | 9.72 | 0 |
| $\mathcal{M}_2$ | 6 | 10.76 | 0.096 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | | | 1 | −1 | 17.17 | 0 | 17.17 | 0.0054 |
| $\mathcal{M}_3$ | 6 | 12.08 | 0.060 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | | | 1 | −0.50 | 11.36 | 11.36 | 11.36 | 0 |
| $\mathcal{M}_4$ | 6 | 12.29 | 0.056 | 0 | 0 | 1 | 1 | 1 | 0 |
| | | | | 1 | −0.55 | 9.02 | 0 | 0 | 0 |
| $\mathcal{M}_5$ | 6 | 12.52 | 0.051 | 0 | 0 | 1 | 0 | 1 | 0 |
| | | | | 1 | −0.55 | 0 | 0 | −9.18 | 0 |
| $\mathcal{M}_{65}$ | 4 | 6.21 | 0.184 | 1 | −1 | 14.92 | 0 | 0 | 0.0051 |
| | | | | 1 | 0 | 0 | 0 | −6.16 | −0.0068 |
| $\mathcal{M}_{71}$ | 4 | 6.99 | 0.136 | 1 | −1 | 15.27 | 0 | 1 | 0.0053 |
| | | | | 1 | 0 | 3.66 | 3.66 | 0 | −0.0063 |

Table 7.9: Automated Model Selection with $r = 2$ and 2 lags

Bartlett correction) in subsection 7.6.4.1: they were the first two models rejected after $\mathcal{M}_3$ as their $p$-value is below 0.05.

The first two models are difficult to accept, so based on the automated model selection, we choose $\mathcal{M}_3$, a money demand relationship. This corresponds to $\mathcal{H}_{13}$ in table 7.6. $\mathcal{M}_{3e1}$ and $\mathcal{M}_{3e2}$ are $\mathcal{H}_{18}$ and $\mathcal{H}_{17}$ in the aforementioned table.

7.6.5.2. *Rank = 2 and two lags.* With rank=2 and two lags, we obtain the models in table 7.9. We note that the 71th accepted model is the combination of the restrictions implied by $\mathcal{M}_3$ and $\mathcal{M}_{3e2}$ in table 7.8. In general we see that models with a large number of over-identifying restrictions are accepted. Each accepted relation has more restrictions that those that are accepted in the model with rank 1.

7.6.5.3. *Rank = 2 and 1 lag.* Omtzigt (2002a) noted in a Monte Carlo study that when under-selecting the true lag length, one had a greater chance of recovering the right restrictions: we therefore run the automated model selection procedure with just one lag and note that while $\mathcal{M}_1$ and $\mathcal{M}_2$ are non-interpretable in the light of economic theory outlined earlier in the paper, $\mathcal{M}_{2e2}$ in table 7.10 corresponds to $\mathcal{M}_{71}$ in table 7.9. We select this as our final model.

**7.6.6. Selected model.** The estimated model is thus a combination of a money demand equation (which is exactly the equation selected, when r=1), depicted in figures 3(c)-3(d) and an

 DOI: 10.2870/52609

| Model | Nr Res | LR | P value | $y$ | $m3r$ | $rsti$ | $id$ | $\Delta p$ | $trend$ |
|---|---|---|---|---|---|---|---|---|---|
| $\mathcal{M}_1$ | 4 | 6.24 | 0.182 | 1 | −0.48 | 14.87 | 14.87 | 0 | 0 |
| | | | | 0 | 1 | −17.65 | 0 | 0 | −0.0117 |
| $\mathcal{M}_2$ | 4 | 8.73 | 0.068 | 1 | −0.52 | 17.03 | 0 | 0 | 0 |
| | | | | 1 | 0 | 9.04 | 9.04 | 0 | −0.0057 |
| $\mathcal{M}_{2e2}$ | 4 | 9.62 | 0.047 | 1 | −1 | 32.70 | 0 | 0 | 0.0058 |
| | | | | 1 | 0 | 11.40 | 11.40 | 0 | −0.0056 |
| $\mathcal{M}_{2e26}$ | 4 | 18.51 | 0.001 | 1 | −1 | 16.85 | 0 | 0 | 0.0053 |
| | | | | 1 | 0 | 0 | 0 | −5.03 | −0.0067 |
| $\mathcal{M}_3$ | 3 | 1.52 | 0.678 | 1 | −1 | 20.73 | 0 | 0 | −0.0054 |
| | | | | 1 | 0 | 5.15 | 8.21 | −3.06 | −0.0061 |
| $\mathcal{M}_4$ | 3 | 2.17 | 0.537 | 1 | −0.54 | 12.66 | 0 | 0 | 0 |
| | | | | 1 | 0 | 4.84 | 7.03 | −2.19 | −0.0061 |
| $\mathcal{M}_5$ | 3 | 2.30 | 0.513 | 1 | −0.49 | 13.52 | 13.52 | 0 | 0 |
| | | | | 0 | 1 | −16.95 | 0 | −3.95 | −0.0119 |

Table 7.10: Automated Model Selection with $r = 2$ and 1 lag

IS-curve depicted in figures 3(a)-3(b). The last one The IS curve clearly shows the large disequilibrium (recession) the Netherlands faced in the beginning of the period and a smaller one in the early 1990s. The money demand relations is far more stable. The only disequilibrium coincides with the only serious policy intervention in 1986-87, when there was a gentlemen's agreement with the Dutch banks to limit credit expansion. It is probably somewhat surprising that it only depends on the own return on money, namely the real short run interest rate. Different groups faced rather different interest rate (the market rate for firms, but a higher rate for small savers), which makes finding this stable relation even more remarkable. The trend is also very important (2.1 percent autonomous growth a year in the real money supply). It is most likely a consequence of the gradual liberalization process.

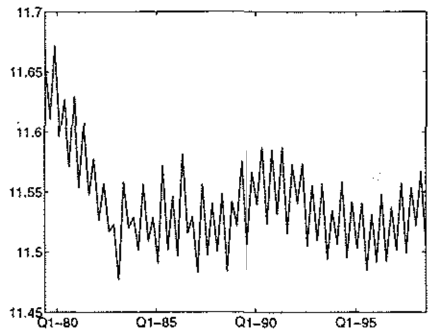
## 7.7. Conclusions

We have studied money demand in the Netherlands in the 1980s and 1990s and modelled it by means of a cointegrated VAR. We found a stable money demand function, which only depends on the own interest rate, not on rates of return on other financial assets considered. With all the shocks and fundamental changes in the economy over the period, this is fairly remarkable.

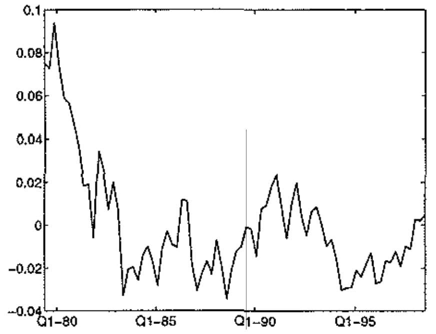
In modelling the VAR we have applied full maximum likelihood in the I(2) model and found that substantial differences exist between the two step procedure and maximum likelihood for the selection of the rank in the I(2) model. Furthermore ignoring the restrictions of the I(2) model in the I(1) after the nominal to real transformation, does make a large difference.

On the methodology used, we make two points. Firstly the largest stationary root in the characteristic polynomial is very important. If it grows large, then that is a sure sign of misspecification of the model, but contemporaneously makes the small sample corrections (bootstrap and Bartlett corrections) fail miserably, if they are based on the unrestricted estimate. Secondly in selecting the rank of the model, we have selected r=2 in the I(1) model, but for identification, we have used the information from automated model selection with r=1 and 2 lags as well as r=2 and 1 lag. There is no theory yet on the asymptotic distribution of the tests, when the rank and/or lag length are deliberately under-selected and this will be a fruitful alley for further research.

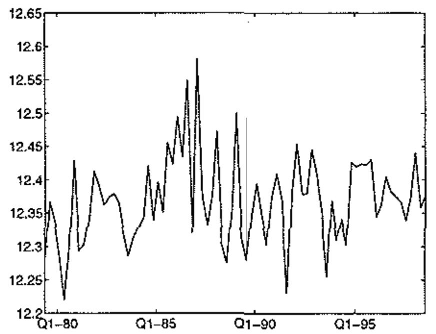
Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609



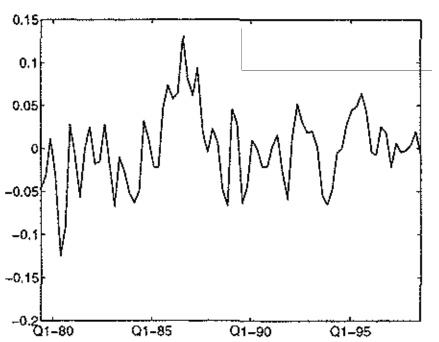

(a) IS curve $\beta_1' X_t$




(b) IS curve $\beta_1' R_t$




(c) Money Demand Relationship $\beta_2' X_t$




(d) Money Demand Relationship $\beta_2' R_t$

Figure 7.3: The cointegration relationships. $R_t$ is $X_t$ corrected for $\Delta X_t$ and $D_{t+1}$.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

## 7.A. Equivalent hypotheses

In this appendix we propose a solution for the following problem:

PROBLEM 1. *We have the null hypothesis of linear within-equation restrictions*[9]

$$\mathcal{H}_0 : \beta = (H_1\varphi_1, \ldots, H_r\varphi_r) \tag{7.16}$$

*where the matrices* $H_1, \ldots, H_r$ *are generically identifying in the sense defined by Johansen (1995a). (If they do not, render them generically identifying, see Omtzigt (2002b)). Let matrix* $H_i$ *possess* $s_i$ *columns.*

*We have an unrestricted estimate of* $\beta$, $\hat{\beta}_u$, *which does not (necissarily) satisfy the restrictions implied by (7.16) and need to find an alternative null hypothesis* $\mathcal{H}_0^\natural : \beta = \left(\tilde{H}_1^*\varphi_1, \ldots, \tilde{H}_r^*\varphi_r\right)$, *which is satisfied by* $\hat{\beta}_u$ *and equals the restrictions implied by (7.16) if* $\hat{\beta}_u$ *satisfies those restrictions.*

Subsequently in the bootstrap we can resample from $\hat{\beta}_u$ and impose the new null hypothesis $\mathcal{H}_0^\natural$, see Omtzigt and Fachin (2002).

SOLUTION 2. *We need to rotate the space such that each of the* $r$ *vectors is as close as possible to its restrictions* $H_i$. *We do so as in Johansen (1995b, page 110-111). Solve the following* $r$ *eigenvalue problems*

$$\left|\mu_i\hat{\beta}_u'\hat{\beta}_u - \hat{\beta}_u'H_i\left(H_i'H_i\right)^{-1}H_i'\hat{\beta}_u\right| \qquad i = 1, \ldots, r$$

*with* $r$ *ordered rows of* $r$ *eigenvalues each* $\mu_{i,1} \geq \mu_{i,2} \geq \ldots \geq \mu_{i,r} \geq 0$ *and corresponding* $r$ *sets of eigenvectors* $(v_{i,1}, \ldots, v_{i,r})$. *Then let*

$$\tilde{\beta}_i = \hat{\beta}_u v_{i,1} \qquad i = 1, \ldots, r$$

*such that* $\tilde{\beta} = \left(\tilde{\beta}_i, \ldots, \tilde{\beta}_r\right)$ *is the ordered unrestricted estimate.*

*Next find the part of* $H_i$ *which is a close to the null-space of* $\tilde{\gamma}_i$, *that is* $\tilde{\gamma}_{i\perp}$ *as possible:*

$$\left|\kappa_i H_i'H_i - H_i'\tilde{\beta}_{i\perp}\left(\tilde{\beta}_{i\perp}'\tilde{\beta}_{i\perp}\right)^{-1}\tilde{\beta}_{i\perp}'H_i\right| \qquad i = 1, \ldots, r$$

*with* $r$ *ordered rows of* $s_i$ *eigenvalues each* $\kappa_{i,1} \geq \kappa_{i,2} \geq \ldots \geq \kappa_{i,s_i} \geq 0$ *and corresponding* $r$ *sets of eigenvectors* $(w_{i,1}, \ldots, w_{i,s_i})$.

*The new restrictions matrices* $\tilde{H}_i$ *then read*

$$\tilde{H}_i = \left(\tilde{\beta}_i, H_i\left(w_{i,1}, \ldots, w_{i,s_{i-1}}\right)\right) \qquad i = 1, \ldots, r$$

*In the last step we have taken the the* $s - 1$ *columns of* $H_i$ *which are as close to the null space of* $\tilde{\beta}_i$ *as possible. The reason for which we take these is that if* $H_i$ *is exactly identifying on a column, then* $\tilde{\beta}_i \in sp(H_i)$ *and thus* $\tilde{\beta}_i = H_i w_{i,s_i}$. *If we were then to include the* $H_i w_{i,s_i}$ *we would have that* $\tilde{H}_i$ *is no longer of full column rank.*

*As a final (optional) step we normalize the new restrictions on the old ones:* $\tilde{H}_i^* = \tilde{H}_i\left(H_i'\tilde{H}_i\right)^{-1}$. *The new restrictions then read*

$$\tilde{H}_i^* \qquad i = 1, \ldots, r$$

*and are satisfied by the unrestricted estimate* $\tilde{\beta}$.

[9] We take away the star from $\beta^*$ in the main text to avoid clutter in notation.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

## Bibliography


Abadir, K., K. Hadri, and E. Tzavalis (1999). The influence of VAR dimensions on estimator biases. *Econometrica 67*, 163–181.

Amisano, G. and C. Giannini (1997). *Topics in Structural VAR Econometrics*. Berlin: Springer.

Andrews, D. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*, 821–856.

Andrews, D. (2003). Tests for parameter instability and structural change with unknown change points: A correction. *Econometrica 71*, 395–397.

Attfield, C. (1995). A Bartlett adjustment to the likelihood ratio test for a system of equations. *Journal of Econometrics 66*, 207–223.

Attfield, C. (1998). Bartlett adjustments for systems of linear equations with linear restrictions. *Economic Letters 60*, 277–283.

Banerjee, A., L. Cockeress, and R. B. (2001). An I(2) analysis of inflation and the markup. *Journal of Applied Econometrics 16*, 221–240.

Banerjee, A. N. and J. Magnus (1999). The sensitivity of OLS when the variance matrix is (partially) unknown. *Journal of Econometrics 92*, 295–323.

Banerjee, A. N. and J. Magnus (2000). On the sensitivity of the usual t- and F-tests to covariance misspecification. *Journal of Econometrics 95*, 157–176.

Barndorff-Nielsen, O. and P. Hall (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika 75*, 374–378.

Bartlett, M. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London Series A 160*, 268–282.

Bedini, C. and R. Mosconi (2000). New tools for the dynamic analysis of cointegrated VAR models. Technical report, Department of economics and production, Politecnico di Milano.

Beran, K. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association 83*, 687–697.

Berkowitz, J. and L. Kilian (2000). Recent developments in bootstrapping time series. *Econometric Reviews 19*, 1–48.

Bewley, R., D. Orden, M. Yang, and L. Fisher (1994). Comparison of Box-Tiao and Johansen canonical estimator of cointegration vectors in VEC(1) models. *Journal of Econometrics 64*, 3–27.

Beyer, A. (1998). Modelling money demand in Germany. *Journal of Applied Econometrics 13*, 57–76.

Boswijk, H. (1995). Identifiability of cointegrated systems. Technical Report 95-78, Tinbergen Institute, University of Amsterdam. Available at http://www1.fee.uva.nl/ke/boswijk/.

Brüggeman, R. and H. Lütkepohl (2001). Lag selection in subset VAR models with an application to a U.S. monetary system. In L. Knüppel and H. Lütkepohl (Eds.), *Econometric Studies: A Festschrift in Honour of Joachim Frohn*, Münster, pp. 107–128. LIT-Verlag.

Chan, N. and C. Wei (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Mathematical Statistics 16*, 367–410.

Chang, Y., R. Sickles, and W. Song (2001). Bootstrapping unit root tests with covariates. Technical Report 2001/07, Department of Economics, Rice University. Available at




Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609

http://www.ruf.rice.edu/~econ/papers/.

Clements, M. and D. Hendry (2002). *A Companion to Economic Forecasting*. Oxford: Blackwell Publishers.

Cribari-Neto, F. and G. Cordeiro (1996). On Bartlett and Bartlett-type corrections. *Econometric Reviews 15*, 339–367.

Davidson, J. (1998a). Structural relations, cointegration and identification: Some simple results and their applications. *Journal of Econometrics 87*, 87–113.

Davidson, J. (1998b). A Wald test of restrictions on the cointegrating space based on Johansen's estimator. *Economics Letters 59*, 183–187.

Davidson, J. (2002). Alternative bootstrap procedures for testing cointegration in fractionally integrated processes. Technical report, Cardiff University. Available at http://www.cf.ac.uk/carbs/econ/davidsonje.

Davidson, R. and J. MacKinnon (2000). Improving the reliability of bootstrap tests. Technical Report 995, Queen's University Institute for Economic Research.

De Greef, I., P. Hilbers, and L. Hoogduin (1998). Moderate monetarism: A brief survey of Dutch monetary policy in the post-war period. Technical Report 28, Dutch Central Bank.

De Jong, A. (1960). *De Wetgeving Nopens de Nederlandsche Bank*. Den Haag: Nijhoff.

De Nederlandsche Bank (1946-2002). Annual report, various issues.

Doornik, J. (1995). Testing general restrictions on the cointegration space. Technical report, Nuffield College, Oxford. Available at http://www.nuff.ox.ac.uk/Users/Doornik/.

Doornik, J. (1996). Testing vector autocorrelation and heteroscedasticity. Technical report, Nuffield College, Oxford. Available at http://www.nuff.ox.ac.uk/Users/Doornik/.

Doornik, J. and H. Hansen (1994). An omnibus test for univariate and multivariate normality. Technical report, Nuffield College, Oxford.

Doornik, J. and D. Hendry (1997). *Modelling Dynamic Systems Using PcFiml 9 for Windows*. Timberlake, London.

Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. London: Chapman & Hall.

Engle, R. and C. Granger (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica 55*, 251–276.

Estrella, A. (2003). Critical values and p values of Bessel process distributions: Computation and application to structural break tests. Technical report, Federal Reserve Bank of New York.

Fachin, S. (2000). Bootstrap and asymptotic tests of long-run relationships in cointegrated systems. *Oxford Bulletin of Economics and Statistics 62*, 577–585.

Fachin, S. and L. Bravetti (1996). Asymptotic normal and bootstrap inference in structural VAR analysis. *Journal of Forecasting 15*, 329–341.

Faliva, M. and M. Zoia (2000). On a partitioned inversion formula having useful applications in econometrics. *Econometric theory 18*(2), 525–530.

Fase, M. and C. Winder (1990). The demand for money in the Netherlands revisited. *De Economist 38*, 276–301.

Fisher, F. (1966). *The Identification Problem in Econometrics*. New York: McGraw-Hill.

Gonzalo, J. (1994). Five alternative methods of estimating long-run equilibrium relationships. *Journal of Econometrics 64*, 203–233.

Gourerioux, C. and A. Monfort (1995). *Statistics and Econometric Models*. Cambridge: Cambridge University Press.

Gredenhoff, M. and T. Jacobson (2001). Bootstrap testing linear restrictions on cointegrating vectors. *Journal of Business and Economic Statistics 19*, 63–72.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer.

Hall, P. (1995). Methodology and theory for the bootstrap. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Vol. IV*, Amsterdam. North-Holland.

 DOI: 10.2870/52609

Hansen, H. and K. Juselius (1994). *CATS for RATS4: Manual to Coinegration Analysis to Time Series*. Estima.

Harbo, I., S. Johansen, B. Nielsen, and A. Rahbek (1998). Test for cointegrating rank in partial systems. *Journal of Business and Economic Statistics 16*, 388–399.

Haug, A. (1996). Tests for cointegration - a Monte Carlo comparison. *Journal of Econometrics 71*, 89–115.

Hendry, D. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

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

Hendry, D. and H. Krolzig (2003). New developments in automatic general-to-specific modelling. In B. Stigum (Ed.), *Econometrics and the Philosophy of Economics*. MIT Press.

Hendry, D. and G. Mizon (1993). Evaluating econometric models by encompassing the VAR. In P. Phillips (Ed.), *Models, Methods and Applications of Econometrics*. Blackwell.

Hoffmann-Jørgensen, J. (1994). *Probability with a View Toward Statistics, Vol. II*. New York: Chapman and Hall, Probability Series.

Hoover, K. and S. Perez (1999). Data mining reconsidered: Encompassing and the general -to-specific approach to specification search. *Econometrics Journal 2*, 167–191.

Horowitz, J. (2002). The bootstrap. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics, Vol. V*, Amsterdam. North-Holland.

Hsiao, C. (1983). Identification. In Z. Grilliches and M. Intriligator (Eds.), *Handbook of Econometrics, Vol. I*, Amsterdam. North Holland.

Inoue, A. and L. Kilian (2002). Bootstrapping autoregressive processes with possible unit roots. *Econometrica 70*, 377–391.

Jacobs, J. and A. Van der Horst (1996). VAR-ing the economy of the Netherlands. Technical Report 24, Economics Department of the University of Groningen.

Jensen, J. (1993). A historical sketch and some new results on the improved log likelihood ratio statistic. *Scandinavian Journal of Statistics 20*, 1–15.

Jensen, J. and A. Wood (1997). On the non-existence of a Bartlett correction for unit root tests. *Statistics and Probability Letters 35*, 181–187.

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 processes. *Econometrica 59*, 1551–1580.

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

Johansen, S. (1995a). Identifying restrictions of linear equations: With applications to simultaneous equation and cointegration. *Journal of Econometrics 69*, 111–132.

Johansen, S. (1995b). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

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

Johansen, S. (1997). Likelihood analysis of the I(2) model. *Scandinavian Journal of Statistics 24*, 433–462.

Johansen, S. (2000a). A Bartlett correction factor for tests on the cointegrating relations. *Econometric Theory 16*, 740–778.

Johansen, S. (2000b). Testing hypotheses in the I(2) model. Technical report, Economics Department, European University Institute, Florence.

Johansen, S. (2002a). A small sample correction for tests of hypotheses on the cointegrating vectors. *Journal of Econometrics 111*, 195–221.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

Johansen, S. (2002b). A small sample correction for the test of cointegrating rank in the vector autoregressive model. *Econometrica 70*, 1929–1961.

Johansen, S. (2002c). Statistical analysis of hypotheses on the cointegrating relations in the I(2) model. Technical Report 13, Department of Theoretical Statistics, University of Copenhagen. Available at http://www.math.ku.dk/~sjo.

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

Juselius, K. (1996). An empirical analysis of the changing role of the German Bundesbank after 1983. *Oxford Bulletin of Economics and Statistics 58*, 791–819.

Juselius, K. (1998). A structured VAR for Denmark under changing monetary regimes. *Journal of Business and Economics Statistics 16*, 400–411.

Juselius, K. (2001). European integration and monetary transmission mechanisms: The case of Italy. *Journal of Applied Econometrics 16*, 341–358.

Juselius, K. (2002). *The Cointegrated VAR Model: Econometric Methodology and Macroeconomic Applications*. Unpublished manuscript. Available at http://www.econ.ku.dk/okokj/.

Juselius, K. and J. Toro (1999). The effects of joining the EMS: Monetary transmission mechanisms in Spain. Technical Report 99-22, University of Copenhagen, Institute of Economics. Available at http://www.econ.ku.dk/wpa/.

Kailath, T. (1980). *Linear Systems*. Englewood Cliffs, Prentice-Hall.

Kongsted, H. (1998). An I(2) cointegration analysis of small country import price determination. Technical Report 98-22, Institute of Economics, University of Copenhagen. Available at http://www.econ.ku.dk/wpa/.

Kongsted, H. (2002). Testing the nominal-to-real transformation. Technical Report 02/06, Institute of Economics, University of Copenhagen. Available at http://www.econ.ku.dk/wpa/.

Kongsted, H. and H. Nielsen (2002). Analyzing I(2) systems by transformed vector autoregressions. Technical Report 02-20, Institute of Economics, University of Copenhagen. Available at http://www.econ.ku.dk/wpa/.

Koop, G., M. Pesaran, and S. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics 74*, 119–147.

Koopmans, T., H. Rubin, and R. Leipnik (1950). *Statistical Inference in Dynamic Economic Models*. New York: Cowles Commission Monograph 10, J.Wiley.

Kuipers, S. and B. Boertje (1988). On the causes of the rise in the liquidity ratio in the Netherlands in the early eighties. *De Economist 136*, 50–90.

Lawley, D. (1956). A general method for approximating the distribution for the likelihood ratio criteria. *Biometrika 43*, 293–303.

Li, H. and G. Maddala (1996). Bootstrapping time series models. *Econometric Reviews 15*, 115–158.

Li, H. and G. Maddala (1997). Bootstrapping cointegrating regressions. *Journal of Econometrics 80*, 297–318.

Lovell, M. (1983). Data mining. *The Review of Economics and Statistics 65*, 1–12.

Lütkephol, H. (1991). *Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.

Nielsen, B. (1997). Bartlett correction of the unit root test in autoregressive models. *Biometrika 84*, 500–504.

Nielsen, B. and A. Rahbek (2000). Similarity issues in cointegration analysis. *Oxford Bulletin of Economics and Statistics 62*(1), 5–22.

Omtzigt, P. (2001). Selezione automatica dei modelli autoregressivi cointegrati. In P. Giudice, C. Tarantola, and F. Verrecchia (Eds.), *Modelli Statistici per le Applicazioni di Data Mining*, pp. 109–114. Università di Pavia.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute DOI: 10.2870/52609

Omtzigt, P. (2002a). Automatic identification and restriction of the cointegration space. Technical Report 2002/16, UvA-Econometrics, University of Amsterdam. Available at http://www.fee.uva.nl/ke/Uva-Econometrics. (Chapter two of this thesis).

Omtzigt, P. (2002b). Automatic identification of simultaneous equations models. Technical Report 2002/14, UvA-Econometrics, University of Amsterdam. Available at http://www.fee.uva.nl/ke/Uva-Econometrics. (Chapter one of this thesis).

Omtzigt, P. (2003). Me2: A computer package for the maximum likelihood estimation of I(2) systems. Technical report, University of Amsterdam.

Omtzigt, P. and S. Fachin (2002). Bootstrapping and Bartlett corrections in the cointegrated VAR model. Technical Report 2002/15, UvA-Econometrics, University of Amsterdam, Amsterdam. Available at http://www.fee.uva.nl/ke/Uva-Econometrics. (Chapter four of this thesis).

Paparoditis, E. and D. Politis (2001). Unit-roots testing via the continuous-path block bootstrap. Technical Report 2001-06, University of California at San Diego, Department of Economics. Available at http://www.econ.ucsd.edu/papers/.

Paruolo, P. (1996). On the determination of integration indices in I(2) systems. *Journal of Econometrics 72*, 313–356.

Paruolo, P. (1997a). Asymptotic inference on the moving average impact matrix in cointegrated I(1) VAR systems. *Econometric Theory 13*, 79–118.

Paruolo, P. (1997b). Standard errors for the long-run variance matrix. *Econometric Theory 13*, 305 ( Problem 97.1.1). Solution in vol 14, 152-153.

Paruolo, P. (2000). Asymptotic efficiency of the two stage estimator in I(2) systems. *Econometric Theory 16*, 524–550.

Paruolo, P. (2001). LR cointegration tests when some cointegrating relations are known. *Statistical Methods and Applications - Journal of the Italian Statistical Society 10*, 123–137.

Paruolo, P. (2002a). Asymptotic inference on the moving average impact matrix in cointegrated I(2) VAR systems. *Econometric Theory 18*, 673–690.

Paruolo, P. (2002b). Common features and common I(2) trends in VAR systems. Technical Report 2002/29, Department of Economics, University of Insubria, Italy. Available at http://eco.uninsubria.it/dipeco/Quaderni/.

Paruolo, P. and A. Rahbek (1999). Weak exogeneity in I(2) VAR systems. *Journal of Econometrics 93*, 281–308.

Pesaran, H. and Y. Shin (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters 58*, 17–29.

Phillips, P. (1998). Impulse response and forecast error variance asymptotics in nonstationary VARs. *Journal of Econometrics 83*, 21–56.

Phillips, P. (2001). Bootstrapping spurious regression. Technical Report 1330, Cowles Foundation. Available at http://cowles.econ.yale.edu/P/cd/cfdpmain.htm.

Rahbek, A., H. Kongsted, and C. Jørgensen (1999). Trend stationarity in the I(2) cointegration model. *Journal of Econometrics 90*, 265–289.

Sargan, J. (1988). *Lecture Notes on Advanced Econometric Theory*. Oxford: Basil Blackwell.

Sterken, E. (1992). The financial transactions motive in the Dutch money demand function. *Kredit und Kapital 25*, 386–405.

Taniguchi, M. (1988). Asymptotic expansions of the distributions of some test for Gaussian ARMA processes. *Journal of Multivariate Analysis 27*, 494–511.

Taniguchi, M. (1991). *Higher Order Asymptotic Theory for Time Series Analysis*. Berlin: Springer-Verlag.

Van Giersbergen, N. (1996). Bootstrapping the trace statistic in VAR models: Monte Carlo results and applications. *Oxford Bulletin of Economics and Statistics 58*, 391–408.

Omtzigt, Pieter (2003), Essays on Cointegration Analysis
European University Institute
DOI: 10.2870/52609