# INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

# UMI®

UNIVERSITY OF CALIFORNIA AT SAN DIEGO

**Structural Changes in Cointegrated Processes**

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Economics

by

Peter Reinhard Hansen

Committee in charge:

Professor James D. Hamilton, Chair
Professor Graham Elliott
Professor Bruce E. Lehmann
Professor Dimitris Politis
Professor Halbert White

2000

# UMI®
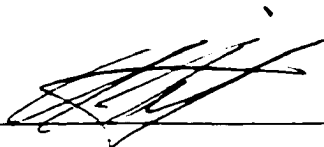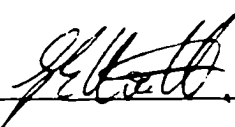
The dissertation of Peter Reinhard Hansen is approved. and it

is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

Chair

University of California at San Diego

2000

iii

# Table of Contents

# List of Figures and Tables

vi

# ACKNOWLEDGEMENTS

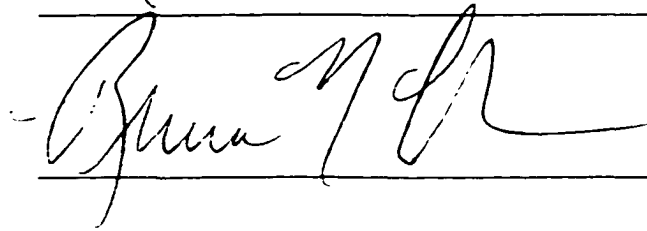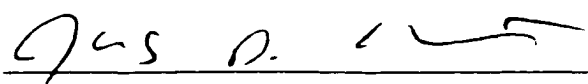As is the case for anyone involved in research. I am deeply indebted to many people who inspired me. both before and during my time as a graduate student at the University of California at San Diego.

I would like to thank Soren Johansen who initiated my interest in econometrics. He was an inspiring and motivating teacher while I studied at the University of Copenhagen. and had I not met him, I probably would not have pursued an academic career.

It has been a great privilege to study econometrics at the University of California at San Diego. where my mentors have been many. The guidance by James Hamilton through the past three years has been invaluable. He helped me stay focused. and offered his time. advice. and encouragement beyond the call of duty. No one could have been a better supervisor for me. My debt to Halbert White is also enormous. Many questions that had kept my econometric soul unrestful were resolved in his classes, and I learned much while working for him on econometric projects and helping with the revision of his book *Asymptotic Theory for Econometricians.* Graham Elliott offered several good suggestions and raised questions that forced me to think harder about the problems at hand. which led me to important discoveries. Other important inputs came from discussions with Clive Granger. Wouter den Haan. and Allan Timmermann.

I also want to thank Katarina Juselius. who invited me to conferences in Trondheim and Copenhagen. and to the European University Insitute in Florence. On these occasions I received inspiration and valuable comments from her and other conference and seminar participants.

Finally. I want to thank my fellow graduate students who made my years in San Diego a happy and productive time. and in particular thanks to Sivan Ritz for her special friendship and her contributions in making this dissertation readable (I hope) for others than myself.

vii

# Vita

| | |
|---|---|
| 1994 | B.A., University of Copenhagen |
| 1994 - 1996 | Research Assistant for Professor Soren Johansen |
| 1995 | M.Sc. (cand.scient.oecon.) University of Copenhagen |
| 1996 - 1997 | Graduate Student at University of Copenhagen |
| 1997 - 2000 | Graduate Student at University of California, San Diego |
| 1998 - 2000 | Teaching Assistant at University of California, San Diego |
| 1998 - 2000 | Research Assistant for Professor Halbert White |
| 2000 | Ph.D., University of California, San Diego |

# Publications

*Workbook on Cointegration*, with Søren Johansen. Oxford University Press, Oxford, 1998.

"Solutions", with Halbert White. In *Asymptotic Theory for Econometricians*, 2nd edn. Academic Press, San Diego, 2000, pp. 213–259, forthcoming.

viii

# ABSTRACT OF THE DISSERTATION

Structural Changes in Cointegrated Processes

by

Peter Reinhard Hansen

Doctor of Philosophy in Economics

University of California at San Diego, 2000

Professor James D. Hamilton, Chair

In my dissertation, I show how structural changes in cointegrated processes can be formulated in the vector autoregressive model, how parameters can be estimated, and how one can test for structural changes in the cases where the change points are known or unknown. My dissertation also contains new results about the Granger representation for $I(1)$ processes and a general estimation technique.

Chapter one contains a new proof of the Johansen-Granger representation theorem and derives an explicit expression of the Granger representation. This representation is useful for impulse response analysis and for the asymptotic analysis of cointegrated processes with structural changes.

Chapter two develops the case where potential change points and the number of cointegrating relations are known. The number of cointegrating relations may vary over the sample. I show how a large class of structural changes can be formulated in a unified framework, and that parameters can be estimated with a new estimation technique. This technique is called the *generalized reduced rank regression* (GRRR) technique, and is described in more detail in Chapter five.

ix

Tests for structural changes, and hypotheses that can be expressed by linear parameter restrictions, are shown to have an asymptotic $\chi^2$ distribution. The chapter includes an empirical application to the US term structure of interest rates.

Chapter three considers the case where a change point is unknown. Various tests for parameter constancy are studied. These tests are constructed from a set of likelihood ratio (LR) statistics that test for a structural change in the cointegrating relations over a pre-specified interval. Some tables with critical values are provided along with a study of the power of the different tests.

Chapter four derives a test to determine the number of cointegrating relations in processes with one or multiple structural changes. When the potential change points are known, the asymptotic distribution of the LR test turns out to be a convex combination of squared Dickey-Fuller distributions.

Chapter five presents the most general version of the GRRR technique. Several applications of the estimation technique are presented.

x

# Chapter 1

## The Johansen-Granger Representation Theorem:

## An Explicit Expression for $I(1)$ Processes[†]

**Abstract**

The Johansen-Granger representation theorem for the cointegrated vector autoregressive process is derived using the companion form. This approach yields an explicit representation of all coefficients and initial values.

This result is useful for impulse response analysis, common feature analysis and asymptotic analysis of cointegrated processes.

1

## 1.1. Introduction

The Johansen-Granger representation theorem[1] states that a vector autoregressive process $A(L)X_t$ $= \varepsilon_t$, integrated of order one, has the representation $X_t = C \sum_{i=1}^{t} \varepsilon_i + C(L)\varepsilon_t + A_0$, where $\{C(L)\varepsilon_t\}$ is stationary if $\{\varepsilon_t\}$ is stationary and where $A_0$ depends on initial values $(X_0, X_{-1}, \ldots)$. (see Johansen (1991, 1996)). Johansen's result gives explicit values of $C$ whereas the coefficients of the lag polynomial, $C(L)$, and the initial value, $A_0$, are given implicitly.

This representation of cointegrated processes is known as the Granger representation and is synonymous with the Wold representation for stationary processes. Because the representation divides $X_t$ into a random walk and a stationary process, it can be viewed as multivariate Beveridge-Nelson decomposition where the labels are permanent and transitory components. (see Beveridge and Nelson (1981)).

The Granger representation is valuable in the asymptotic analysis of cointegrated processes, where typically only an explicit expression for $C$ is needed. Explicit values for the coefficients in $C(L)$ are useful in common feature analysis. (see Engle and Kozicki (1993)), and in impulse response analysis. (see Lütkepohl and Reimers (1992), Warne (1993), and Lütkepohl and Saikkonen (1997)), where the coefficients of $C(L)$ are interpreted as the transitory effects of the shocks $\varepsilon_t$. Similarly, in asymptotic analysis of the model with structural breaks, it is valuable to have an explicit value for $A_0$.

In this paper, explicit values of coefficients as well as initial values are found using the companion form, making use of the algebraic structure that characterizes this model.

From Johansen (1996) we adopt the following definitions: for an $m \times n$ matrix $a$ with full column rank $n$, we define $\bar{a} = a(a'a)^{-1}$ and let the orthogonal complement of $a$, be the full rank $m \times (m - n)$ matrix $a_\perp$ that has $a'_\perp a = 0$.

---

[1] The original Granger representation theorem, given by Engle and Granger (1987), asserts the existence of an error correction representation of $X_t$, under the assumptions that $\Delta X_t$ and $\beta'X_t$ have stationary and invertible VARMA representations, for some matrix $\beta$. The Johansen-Granger representation theorem, of Johansen (1991, 1996), makes assumptions on the autoregressive parameters, that precisely characterizes $I(1)$ processes, and states results on the moving average representation of $X_t$.

In Section 2 the explicit representation is derived. In Section 3 we consider deterministic aspects of the representation. Section 4 contains concluding remarks and the appendix contains relevant algebra.

## 1.2. The Granger Representation for Autoregressive Processes Integrated of Order One

We consider the $p$-dimensional vector autoregressive process of order $k$

$$X_t = \Pi_1 X_{t-1} + \Pi_2 X_{t-2} + \cdots + \Pi_k X_{t-k} + \Phi D_t + \varepsilon_t, \quad t = 1, \ldots, T.$$

where the process' deterministic terms are contained in $D_t$ and where $\varepsilon_t, t = 1, \ldots, T$ is a sequence of independent identically distributed stochastic variables with mean zero[2].

The process can be re-written in error correction form:

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t, \quad t = 1, \ldots, T$$

where $\Pi = -I + \sum_{i=1}^{k} \Pi_i$ and $\Gamma_i = -\sum_{j=i+1}^{k} \Pi_j$. The conditions that ensure that $X_t$ is integrated of order one, referred to as $X_t$ being $I(1)$, are stated in the following assumption:

**Assumption 1.2.1.** *The assumptions of the Johansen-Granger representation theorem are:*

($i$) *The roots of the characteristic polynomial*

$$\det(A(z)) = \det(I - \Pi_1 z - \Pi_2 z^2 - \cdots \Pi_k z^k)$$

*are either outside the unit circle or equal to one.*

---

[2] The Granger representation is not relying on the assumptions on $\varepsilon_t$, since it is entirely an algebraic derivation. However the i.i.d. assumption is important for some of the interpretations of the representation.

(*ii*)  The matrix $\Pi$ has reduced rank $r < p$. and can therefore be expressed as the product $\Pi = \alpha\beta'$ where $\alpha$ and $\beta$ are $p \times r$ matrices of full column rank $r$.

(*iii*)  The matrix $\alpha'_\perp \Gamma \beta_\perp$ has full rank. where $\Gamma = I - \sum_{i=1}^{k-1} \Gamma_i$ and where $\alpha_\perp$ and $\beta_\perp$ are the orthogonal complements to $\alpha$ and $\beta$.

The first assumption ensures that the process is not explosive (roots in the unit circle) or seasonally cointegrated (roots on the boundary of the unit circle different from $z = 1$). (see Hylleberg, Engle, Granger, and Yoo (1990) or Johansen and Schaumburg (1998)). The second ensures that there are at least $p - r$ unit roots and induces cointegration whenever $r \geq 1$. The third assumption restricts the process from being $I(2)$. because (*iii*) together with (*ii*) ensures that the number of unit roots is exactly $p - r$.

Under these assumptions, Johansen (1991) showed that $X_t$ has the representation $X_t = C \sum_{i=1}^{t} (\varepsilon_i + \Phi D_i) + C(L)(\varepsilon_t + \Phi D_t) + A$. where $C = \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp$. By using the companion form of the process. it is possible to obtain explicit values for the coefficients of the lag polynomial $C(L) = C_0 + C_1 L + C_2 L^2 + \cdots$. and the initial values contained in $A$. as I show below.

The following lemma will be useful.

**Lemma 1.2.2.** *Let $a$ and $b$ be $m \times n$ matrices. $m \geq n$ with full column rank $n$. and let $a_\perp$ and $b_\perp$ be their orthogonal complements, respectively.*

*The following five statements are equivalent.*

(*i*)  The matrix $(I + b'a)$ does not have 1 as an eigenvalue.

(*ii*)  Let $v$ be a vector in $\mathbf{R}^m$. Then $(b'a)v = 0$ implies $v = 0$.

(*iii*)  The matrix $b'a$ has full rank.

(*iv*)  The $m \times m$ matrix $(b. a_\perp)$ has full rank.

(*v*)  The matrix $b'_\perp a_\perp$ has full rank.

**Proof.** The equivalence of $(i)$, $(ii)$ and $(iii)$ is straightforward, and the identity

$$|(a,a_\perp)|\,|(b,a_\perp)| = |(a,a_\perp)'(b,a_\perp)| = \left| \begin{pmatrix} a'b & 0 \\ a'_\perp b & a'_\perp a_\perp \end{pmatrix} \right| = |a'b|\,|a'_\perp a_\perp|$$

proves that $(iii)$ holds if and only if $(iv)$ holds. Finally, the identity

$$|(b,b_\perp)|\,|(b,a_\perp)| = |(b,b_\perp)'(b,a_\perp)| = \left| \begin{pmatrix} b'b & 0 \\ b'_\perp b & b'_\perp a_\perp \end{pmatrix} \right| = |b'b|\,|b'_\perp a_\perp|$$

completes the proof. ∎

### 1.2.1. The Companion Form

We transform the process into the companion form. by defining

$$X_t^* = \left( X_t',X_{t-1}',\ldots,X_{t-k+1}' \right)'$$

so that with suitable definitions

$$\begin{aligned} \Delta X_t^* &= \Pi^* X_{t-1}^* + \Phi_t^* + \varepsilon_t^* \\ &= \alpha^* \beta^{*\prime} X_{t-1}^* + \Phi_t^* + \varepsilon_t^*. \end{aligned}$$

which converts the process to a vector autoregressive process of order one. The needed definitions are

$$
\Pi^* = \begin{pmatrix} \alpha\beta' + \Gamma_1 & \Gamma_2 - \Gamma_1 & \cdots & \Gamma_{k-1} - \Gamma_{k-2} & -\Gamma_{k-1} \\ I & -I & & & 0 \\ & & \ddots & & \vdots \\ & & & -I & 0 \\ 0 & 0 & & I & -I \end{pmatrix}.
$$

$$
\alpha^* = \begin{pmatrix} \alpha & \Gamma_1 & \cdots & \Gamma_{k-1} \\ 0 & I & & 0 \\ \vdots & & \ddots & \\ 0 & & & I \end{pmatrix}, \beta^* = \begin{pmatrix} \beta & I & 0 & \cdots & 0 \\ 0 & -I & I & & \\ \vdots & & & \ddots & \\ & & & & I \\ 0 & \cdots & 0 & & -I \end{pmatrix}.
$$

$$
\varepsilon_t^* = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \Phi_t^* = \begin{pmatrix} \Phi D_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.
$$

It is easily verified that the orthogonal complements of $\alpha^*$ and $\beta^*$ are given by

$$
\alpha_\perp^* = \begin{pmatrix} \alpha_\perp \\ -\Gamma_1'\alpha_\perp \\ \vdots \\ -\Gamma_{k-1}'\alpha_\perp \end{pmatrix}, \beta_\perp^* = \begin{pmatrix} \beta_\perp \\ \vdots \\ \beta_\perp \end{pmatrix}.
$$

**Lemma 1.2.3.** *Let $\alpha$, $\beta$, $\alpha^*$ and $\beta^*$ be defined as above, and assume that Assumption 1.2.1 holds. Then the eigenvalues of the matrix $(I + \beta^{*'}\alpha^*)$ are all less than one in absolute value.*

**Proof.** By Assumption 1.2.1 (*iii*), the identity

$$\alpha_{\perp}^{*\prime}\beta_{\perp}^{*} = \alpha_{\perp}'(I - \Gamma_1 - \cdots - \Gamma_{k-1})\beta_{\perp}$$

shows that $\alpha_{\perp}^{*\prime}\beta_{\perp}^{*}$ has full rank, and by Lemma 1.2.2, we have that 1 is not an eigenvalue of $(I + \beta^{*\prime}\alpha^{*})$. However we need to show that the eigenvalues are smaller than one in absolute value. Therefore consider an eigenvector $v = (v_1', \ldots, v_k')' \neq 0$ of $(I + \beta^{*\prime}\alpha^{*})$, e.g. $(I + \beta^{*\prime}\alpha^{*})v = \lambda v$. The upper $r + p$ rows of $(I + \beta^{*\prime}\alpha^{*})v$ yields

$$v_1 + \beta'(\alpha v_1 + \Gamma_1 v_2 + \cdots + \Gamma_{k-1}v_k) = \lambda v_1$$

$$(\alpha v_1 + \Gamma_1 v_2 + \cdots + \Gamma_{k-1}v_k) = \lambda v_2$$

which implies $\lambda\beta' v_2 = (\lambda - 1)v_1$, and the remaining part implies $v_2 = \lambda v_3 = \cdots = \lambda^{k-2}v_k$. The case $\lambda = 0$ clearly fulfills $|\lambda| < 1$ so assume $\lambda \neq 0$. Multiply the second set of equations by $(\lambda - 1)/\lambda^k$ and substitute $z = 1/\lambda$ to obtain

$$[I(1-z) - \alpha\beta'z - \Gamma_1(1-z)z - \cdots - \Gamma_{k-1}(1-z)z^{k-1}]v_k = 0.$$

This is equivalent to

$$\left| I(1-z) - \alpha\beta'z - \sum_{i=1}^{k-1}\Gamma_i(1-z)z^i \right| = 0.$$

and since Assumption 1.2.1 has $|z| > 1$ we conclude that $|\lambda| < 1$. ∎

The result has the implication that under Assumption 1.2.1 the sum $\sum_{i=0}^{\infty}(1 + \beta^{*\prime}\alpha^{*})^i$ is convergent with limit $(\beta^{*\prime}\alpha^{*})^{-1}$, such that a process defined by $Y_t = \sum_{i=0}^{\infty}(1 + \beta^{*\prime}\alpha^{*})^i u_{t-i}$ is stationary whenever $u_t$ is stationary.

**Lemma 1.2.4.** *With the definitions above we have the identities:*

$$(I - C\Gamma) = (I - C\Gamma)\bar{\beta}\beta'$$

$$I = (I - C\Gamma)\bar{\beta}\beta' + C(\Gamma - I) + C\bar{\alpha}_\perp\alpha'_\perp.$$

**Proof.** Since $I = \beta(\beta'\beta)^{-1}\beta' + \beta_\perp(\beta'_\perp\beta_\perp)^{-1}\beta'_\perp = \bar{\beta}\beta' + \beta_\perp\bar{\beta}'_\perp$, the first identity follows from

$$(I - C\Gamma) = (I - C\Gamma)(\bar{\beta}\beta' + \beta_\perp\bar{\beta}'_\perp)$$

$$= (I - C\Gamma)\bar{\beta}\beta' + \beta_\perp\bar{\beta}'_\perp - \beta_\perp(\alpha'_\perp\Gamma\beta_\perp)^{-1}\alpha'_\perp\Gamma\beta_\perp\bar{\beta}'_\perp$$

$$= (I - C\Gamma)\bar{\beta}\beta'.$$

and the second follows by applying the first identity and that $C = C\bar{\alpha}_\perp\alpha'_\perp$. ■

We are now ready to formulate the main result.

**Theorem 1.2.5 (The Johansen-Granger representation theorem).** *Let a process be given by the equation*

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1}\Gamma_i\Delta X_{t-i} + \Phi D_t + \varepsilon_t.$$

*and assume that Assumption 1.2.1 holds. Then the process has the representation*

$$X_t = C\sum_{i=1}^{t}(\varepsilon_i + \Phi D_i) + C(L)(\varepsilon_t + \Phi D_t) + C(X_0 - \Gamma_1 X_{-1} - \cdots - \Gamma_{k-1}X_{-k+1})$$

*where $C = \beta_\perp(\alpha'_\perp\Gamma\beta_\perp)^{-1}\alpha'_\perp$ and where the coefficients of $C(L)$ are given by*

$$C_i = GQ^iE_{1,2}$$

*where*

$$G = ((I - C\Gamma), -C\Gamma_1, \ldots, -C\Gamma_{k-1})$$

$$Q = \begin{pmatrix} I + \Pi & \Gamma_1 & \cdots & \Gamma_{k-2} & \Gamma_{k-1} \\ \Pi & \Gamma_1 & \cdots & \Gamma_{k-2} & \Gamma_{k-1} \\ 0 & I & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & I & 0 \end{pmatrix}$$

$$E_{1,2} = (I_p, I_p, 0, \cdots, 0)'.$$

**Proof.** Under Assumption 1.2.1 the $pk \times pk$ matrix $(\beta^*, \alpha^*_\perp)$ has full rank. We can therefore obtain the Granger representation for $X^*_t$ by finding the moving average representation for the processes $\beta^{*\prime} X^*_t$ and $\alpha^{*\prime}_\perp X^*_t$ individually and then stacking them and multiplying by $(\beta^*, \alpha^*_\perp)^{\prime -1}$.

First, consider the process

$$\beta^{*\prime} X^*_t = (I + \beta^{*\prime}\alpha^*)\beta^{*\prime} X^*_{t-1} + \beta^{*\prime}(\varepsilon^*_t + \Phi^*_t).$$

Since all the eigenvalues of $(I + \beta^{*\prime}\alpha^*)$, according to Lemma 1.2.3, are smaller than one in absolute value, the process has the stationary representation

$$\beta^{*\prime} X^*_t = C^*(L)(\varepsilon^*_t + \Phi^*_t)$$

where $C^*_i = (I + \beta^{*\prime}\alpha^*)^i \beta^{*\prime}$, and where by stationary we mean that $\beta^{*\prime} X^*_t - E\left(\beta^{*\prime} X^*_t\right)$ is stationary.

Next consider the random walk

$$\begin{aligned} \alpha^{*\prime}_\perp X^*_t &= \alpha^{*\prime}_\perp X^*_{t-1} + \alpha^{*\prime}_\perp(\varepsilon^*_t + \Phi^*_t) \\ &= \alpha^{*\prime}_\perp X^*_0 + \sum_{i=1}^{t} \alpha^{*\prime}_\perp(\varepsilon^*_i + \Phi^*_i). \end{aligned}$$

A representation for $X_t^*$ is now obtained as

$$X_t^* = (\beta^*.\alpha_\perp^*)'^{-1} \begin{pmatrix} C^*(L)(\varepsilon_t^* + \Phi_t^*) \\ \sum_{i=1}^t \alpha_\perp^{*\prime}(\varepsilon_i^* + \Phi_i^*) + \alpha_\perp^{*\prime}X_0^* \end{pmatrix}.$$

The entire matrix $(\beta^*.\alpha_\perp^*)'^{-1}$ is given in the Appendix. for our purposes we only need its upper $p$ rows that define the equation for $X_t$. These rows are given by

$$((I - C\Gamma)\bar{\beta}.-C\Gamma_1^s.\dots.-C\Gamma_{k-1}^s.C\bar{\alpha}_\perp)$$

with the definition $\Gamma_i^s = \Gamma_i + \cdots + \Gamma_{k-1}$. For simplicity. we define

$$F = ((I - C\Gamma)\bar{\beta}.-C\Gamma_1^s.\dots.-C\Gamma_{k-1}^s)$$

and obtain the representation for $X_t$:

$$\begin{aligned} X_t &= (F.C\bar{\alpha}_\perp) \begin{pmatrix} C^*(L)(\varepsilon_t^* + \Phi_t^*) \\ \sum_{i=1}^t \alpha_\perp^{*\prime}(\varepsilon_i^* + \Phi_i^*) + \alpha_\perp^{*\prime}X_0^* \end{pmatrix} \\ &= FC^*(L)(\varepsilon_t^* + \Phi_t^*) + C\bar{\alpha}_\perp \sum_{i=1}^t \alpha_\perp^{*\prime}(\varepsilon_i^* + \Phi_i^*) + C\bar{\alpha}_\perp\alpha_\perp^{*\prime}X_0^* \\ &= C(L)(\varepsilon_t + \Phi D_t) + C\sum_{i=1}^t (\varepsilon_i + \Phi D_i) + A. \end{aligned}$$

where the initial value is explicitly given by

$$A = C\bar{\alpha}_\perp\alpha_\perp^{*\prime}X_0^* = C(X_0 - \Gamma_1 X_{-1} - \cdots - \Gamma_{k-1}X_{-(k-1)}).$$

and the coefficients of the polynomial $C(L)$ are given by

$$C_i = F(I + \beta^{*\prime}\alpha^*)^i\beta^{*\prime}E_1$$

$$= FD^i B' E_{1,2}.$$

with the additional definitions

$$D = (I + \beta^{*\prime} \alpha^*)$$

$$B = \begin{pmatrix} \beta & 0 & 0 & \cdots & 0 \\ 0 & I & 0 & & 0 \\ 0 & 0 & 0 & & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} . \quad E_1 = \begin{pmatrix} I_p \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{and } E_{1,2} = \begin{pmatrix} I_p \\ I_p \\ 0 \\ \vdots \\ 0 \end{pmatrix} .$$

Because $(I + \beta'\alpha)\beta' = \beta'(I + \alpha\beta')$ we have that

$$DB' = B'Q$$

where $Q$ is as given in the theorem. Thus. the coefficients can be written as

$$C_i = FD^i B' E_{1,2} = FB'Q^i E = GQ^i E_{1,2}$$

where

$$G = FB' = ((I - C\Gamma). -C\Gamma_1^s, \ldots . -C\Gamma_{k-1}^s) .$$

where we applied the identity $(I - C\Gamma)\bar{\beta}\beta' = (I - C\Gamma)$ of Lemma 1.2.4. This completes the proof. ∎

**Corollary 1.2.6.** *The coefficients of $C(L)$ can be obtained recursively from the formula*

$$C_i = C_{i-1} + \sum_{j=1}^{i} (\Pi + \Gamma_j)\Delta C_{i-j}. \quad i = 1.2. \ldots .$$

*where $C_0 = I - C$ and $\Delta C_0 = I$ and where we set $\Gamma_j = 0$ for $j \geq k$.*

**Proof.** From the proof of the Johansen-Granger representation theorem we have that $C_i = GQ^i E_{1,2}$. So by defining

$$A_i = \begin{pmatrix} A_{1,i} \\ A_{2,i} \\ \vdots \\ A_{k,i} \end{pmatrix} = QA_{i-1} = Q^i E_{1,2}.$$

tedious algebra (given in the Appendix) leads to the relation

$$C_i = C_{i-1} + A_{2,i}, \quad C_{-1} = -C. \quad i = 0.2. \dots$$

and the structure of $Q$ yields the equation

$$A_{2,i} = \sum_{j=1}^{i} (\Pi + \Gamma_j) A_{2,i-j}. \quad A_{2,0} = I \quad i = 1.2. \dots$$

By inserting $A_{2,i-j} = C_{i-j} - C_{i-j-1}$ we find the equation of the corollary. ∎

As a special case we formulate the representation for the vector autoregressive process of order one.

**Corollary 1.2.7.** Let $\Delta X_t = \alpha\beta' X_{t-1} + \varepsilon_t$ be a process fulfilling Assumption 1.2.1. Then we have the representation

$$X_t = C \sum_{i=1}^{t} \varepsilon_i + (1 - C) \sum_{i=0}^{\infty} (I + \alpha\beta')^i \varepsilon_{t-i} + CX_0$$

where $C = \beta_\perp (\alpha'_\perp \beta_\perp) \alpha'_\perp$.

The result of Corollary 1.2.7 is derived directly in Johansen (1996) by dividing the process into its stationary and non-stationary part with the identity $I = \alpha (\beta'\alpha) \beta' + \beta_\perp (\alpha'_\perp \beta_\perp) \alpha'_\perp$. The proof of Theorem 1.2.5 made use of the more general identity $I = (I - C\Gamma) \bar{\beta}\beta' + C(\Gamma - I) + C\bar{\alpha}_\perp \alpha'_\perp$ of

Lemma 1.2.4. which simplifies to the identity in Johansen (1996) when $\Gamma = I$. as is the case for a VAR(1) process.

## 1.3. Deterministic Terms

In this section we study the stationary polynomial's role for the deterministic term. The deterministic part plays an important role for the asymptotic analysis of this model, because the limits of some test statistics depend on the deterministic term. The literature has developed a notation for models with different deterministic terms which we shall adopt.

First we analyze the model $H_1$. This model contains only a constant $\Phi D_t = \mu_0$. which in general will give rise to a linear trend in the process $X_t$. Next. we also analyze its sub-model $H_1^*$. which has the deterministic term $\Phi D_t = \alpha \rho_0$. This is equivalent to the restriction on the constant $C\mu = 0$. which is precisely what is needed for $X_t$ not to have a linear trend. We also analyze the models $H$ and $H^*$. Model $H$ has a linear deterministic trend $\Phi D_t = \mu_0 + \mu_1 t$. which gives rise to a quadratic trend in the process $X_t$. and the sub-model $H^*$. has the deterministic trend restricted to $\Phi D_t = \mu_0 + \alpha \rho_1 t$. which prevents the $X_t$ from having a quadratic trend.

### 1.3.1. The Models $H_1$ and $H_1^*$

When the deterministic term is simply a constant $\mu_0 = \Phi D_t$. the Granger representation is given by

$$X_t = C \sum_{i=1}^{t} \varepsilon_{t-i} + C(L)\varepsilon_t + C(1)\mu_0 + C\mu_0 t + A.$$

So unless $C\mu_0 = 0$. the constant $\mu_0$ leads to a deterministic linear trend in the process $X_t$. The matrix $C(1)$ is calculated in the appendix and is found to be

$$
\begin{aligned}
C(1) &= -(I - C\Gamma)\bar{\beta}\bar{\alpha}'(I - \Gamma C) - C\left(\sum_{i=1}^{k-1} i\Gamma_i\right)C \\
&= -BA' - C\Psi C.
\end{aligned}
$$

where $B = (I - C\Gamma)\bar{\beta}$. $A' = \bar{\alpha}'(\Gamma C - I)$ and $\Psi = \sum_{i=1}^{k-1} i\Gamma_i = \sum_{i=1}^{k-1} \sum_{j=i}^{k-1} \Gamma_j$.

This result encompasses two findings from Hansen and Johansen (1998). The first is that

$$E(\beta' X_t) = \beta' C(1)\mu_0 = \bar{\alpha}'(\Gamma C - I)\mu_0.$$

and the second is that in $H_1^*$, where $\mu_0 = \alpha\rho_0$, the linear trend vanishes while the constant in the process is given by $C(1)\mu = -(I - C\Gamma)\bar{\beta}\rho$.

### 1.3.2. Models $H$ and $H^*$

When the deterministic term contains a linear trend, $\Phi D_t = \mu_0 + \mu_1 t$, the deterministic part of the Granger representation is given by

$$\tfrac{1}{2}C\mu_1 t^2 + C\left(\mu_0 + \tfrac{1}{2}\mu_1\right) t + C(L)\left(\mu_0 + \mu_1 t\right).$$

(see Hansen and Johansen (1998)). This can be re-written as

$$\tfrac{1}{2}C\mu_1 t^2 + \left(C\mu_0 + \left(\tfrac{1}{2}C + C(1)\right)\mu_1\right) t + \left(C(1)\mu_0 - \sum_{i=0}^{\infty} iC_i\mu_1\right). \tag{1.3.1}$$

So unless $\alpha'_\perp \mu_1 = 0$ the linear trend $\mu_1$ leads to a quadratic deterministic trend in the process $X_t$. The only term of (1.3.1) not derived previously, is $\sum_{i=0}^{\infty} iC_i$. This term is derived in the appendix and is given by

$$BA' + C\Psi C + BA'\Gamma BA' - BA'\Psi C - C\Psi BA' - C\Psi C\Psi C - \frac{k(k-1)}{2}C\Psi C.$$

In model $H^*$ where the linear trend is restricted to $\mu_1 = \alpha\rho_1$. (1.3.1) reduces to

$$\left(C\mu_0 - (I - C\Gamma)\bar{\beta}\rho_1\right) t + C(1)\mu_0 + \left((C\Gamma - I)\bar{\beta} - C(1)\right)\rho_1$$

which encompasses a result from Johansen (1996, equation 5.20), because the expression for $\tau_1$.

in Johansen (1996, equation 5.20), equals $(C\mu_0 - (I - C\Gamma)\bar{\beta}\rho_1)$.

## 1.4. Conclusion

We gave an explicit expression of the moving average representation for processes integrated of order one using the companion form for the process. The explicit expression is useful to have in studies of impulse response functions and in common features analysis. As a side benefit the approach gives a new proof of the Johansen-Granger representation theorem, a proof that some might find more intuitive and easy to follow than previous proofs.

## Appendix A: Proofs

### A.1. The Inverse of $(\beta^*.\alpha_\perp^*)$

In the proof of the Johansen-Granger representation theorem we need an explicit expression for the first $p$ rows of $(\beta^*.\alpha_\perp^*)^{\prime-1}$. The entire matrix is given by

$$(\beta^*.\alpha_\perp^*)^{\prime-1} = \begin{pmatrix} (I-C\Gamma)\bar{\beta} & -C\Gamma_1^s & \cdots & -C\Gamma_{k-1}^s & C\bar{\alpha}_\perp \\ (I-C\Gamma)\bar{\beta} & -C\Gamma_1^s - I & \ddots & -C\Gamma_{k-1}^s & C\bar{\alpha}_\perp \\ (I-C\Gamma)\bar{\beta} & -C\Gamma_1^s & \ddots & \ddots & -C\Gamma_{k-1}^s & C\bar{\alpha}_\perp \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ (I-C\Gamma)\bar{\beta} & -C\Gamma_1^s & & \ddots & -C\Gamma_{k-1}^s & C\bar{\alpha}_\perp \\ (I-C\Gamma)\bar{\beta} & -C\Gamma_1^s & & & -C\Gamma_{k-1}^s - I & C\bar{\alpha}_\perp \end{pmatrix}.$$

which is verified by multiplying it by $(\beta^*.\alpha_\perp^*)'$ and using the identity $(I - C\Gamma)\bar{\beta}\bar{\beta}' = (I - C\Gamma)$ from Lemma 1.2.4.

## A.2. The Expression for $C_i$

In Corollary 1.2.6 we asserted the relation $C_i = C_{i-1} + A_{2,i}$, $i = 1, 2, \ldots$. This relation is proved as follows. First, notice from the equation for $A_i$ given by

$$
A_i = \begin{pmatrix} A_{1,i} \\ A_{2,i} \\ \vdots \\ A_{k,i} \end{pmatrix} = \begin{pmatrix} I + \Pi & \Gamma_1 & \cdots & \Gamma_{k-2} & \Gamma_{k-1} \\ \Pi & \Gamma_1 & \cdots & \Gamma_{k-2} & \Gamma_{k-1} \\ 0 & I & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & I & 0 \end{pmatrix} A_{i-1}. \quad A_0 = E_{1,2}
$$

that $A_{k,i} = A_{2,i-k+2}$ $k \geq 2$, and that $A_{1,i} = A_{1,i-1} + A_{2,i}$, why $A_{1,i} = \sum_{j=1}^{i} A_{2,j}$. So that

$$
A_{2,i} = \sum_{j=1}^{i} (\Pi + \Gamma_j) A_{2,i-j}. \quad A_{2,0} = I \quad i = 1, 2, \ldots.
$$

and note that $CA_{2,i} = C\Gamma_1 A_{2,i-1} + \cdots + C\Gamma_{k-1} A_{k,i-1}$.

Next consider

$$
\begin{aligned}
C_i &= GA_i = (I - C\Gamma) A_{1,i} - C\Gamma_1^s A_{2,i} - \cdots - C\Gamma_{k-1}^s A_{k,i} \\
&= (I - C\Gamma)(A_{2,i} + A_{1,i-1}) + C(\Gamma - I) A_{2,i} - C\Gamma_2^s A_{3,i} - \cdots - C\Gamma_{k-1}^s A_{k,i} \\
&= (I - C) A_{2,i} + (I - C\Gamma) A_{1,i-1} - C\Gamma_2^s A_{3,i} - \cdots - C\Gamma_{k-1}^s A_{k,i} \\
&= (I - C) A_{2,i} + (I - C\Gamma) A_{1,i-1} \\
&\quad -C(\Gamma_1^s - \Gamma_1) A_{2,i-1} - \cdots - C(\Gamma_{k-2}^s - \Gamma_{k-2}) A_{k-1,i-1} - C(\Gamma_{k-1} - \Gamma_{k-1}) A_{k,i-1} \\
&= GA_{i-1} + A_{2,i} \\
&= C_{i-1} + A_{2,i}.
\end{aligned}
$$

Which completes the proof.

## A.3. An Expression for $C(1)$

In the analysis of the deterministic terms we need to calculate

$$C(1) = F \sum_{i=0}^{\infty} \left(I + \beta^{*\prime}\alpha^*\right)^i \beta^{*\prime}E_1 = -F \left(\beta^{*\prime}\alpha^*\right)^{-1} \beta^{*\prime}E_1.$$

The inverse of

$$\beta^{*\prime}\alpha^* = \begin{pmatrix} \beta'\alpha & \beta'\Gamma_1 & \beta'\Gamma_2 & \cdots & \beta'\Gamma_{k-2} & \beta'\Gamma_{k-1} \\ \alpha & \Gamma_1 - I & \Gamma_2 & \cdots & \Gamma_{k-2} & \Gamma_{k-1} \\ 0 & I & -I & & & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ & & & \ddots & -I & 0 \\ 0 & & & & I & -I \end{pmatrix}.$$

is given by

$$\left(\beta^{*\prime}\alpha^*\right)^{-1} = \begin{pmatrix} \bar{\alpha}'(I - \Gamma C)\Gamma\bar{\beta} & \bar{\alpha}'(I - \Gamma C)\Gamma_1^s & \cdots & \bar{\alpha}'(I - \Gamma C)\Gamma_{k-1}^s \\ (I - C\Gamma)\bar{\beta} & -C\Gamma_1^s - I & & -C\Gamma_{k-1}^s \\ \vdots & & \ddots & \\ (I - C\Gamma)\bar{\beta} & -C\Gamma_1^s - I & \cdots & -C\Gamma_{k-1}^s - I \end{pmatrix}.$$

So

$$\left(\beta^{*\prime}\alpha^*\right)^{-1}\beta^{*\prime} = \begin{pmatrix} \bar{\alpha}'(I - \Gamma C) & -\bar{\alpha}'(I - \Gamma C)\Gamma_1^s & \cdots & -\bar{\alpha}'(I - \Gamma C)\Gamma_{k-1}^s \\ -C & C\Gamma_1^s + I & & C\Gamma_{k-1}^s \\ \vdots & & \ddots & \\ -C & C\Gamma_1^s & & C\Gamma_{k-1}^s + I \end{pmatrix}$$

and therefore we find

$$
\left(\beta^{*\prime}\alpha^*\right)^{-1}\beta^{*\prime}E_1 =
\begin{pmatrix}
\bar{\alpha}'\left(I - \Gamma C\right) \\
-C \\
\vdots \\
-C
\end{pmatrix}.
$$
(A.1)

and finally that

$$
\begin{aligned}
C(1) &= -\left(\left(I - C\Gamma\right)\bar{\beta}, -C\Gamma_1^s, \ldots, -C\Gamma_{k-1}^s\right)
\begin{pmatrix}
\bar{\alpha}'\left(I - \Gamma C\right) \\
-C \\
-C \\
-C
\end{pmatrix} \\
&= \left(I - C\Gamma\right)\bar{\beta}\bar{\alpha}'\left(\Gamma C - I\right) - C \left(\sum_{i=1}^{k-1}\sum_{j=i}^{k-1}\Gamma_j\right) C. \\
&= BA' - C\Psi C
\end{aligned}
$$

where $B = \left(I - C\Gamma\right)\bar{\beta}$. $A' = \bar{\alpha}'\left(\Gamma C - I\right)$ and $\Psi = \sum_{i=1}^{k-1}\sum_{j=i}^{k-1}\Gamma_j = \sum_{i=1}^{k-1}i\Gamma_i$.

## A.4. An Expression for $\sum_{i=0}^{\infty}iC_i$

In the case where the deterministic term is given by $\Phi D_t = \mu_0 + \mu_1 t$ we make use of

$$
\sum_{i=0}^{\infty}\left(I + \beta^{*\prime}\alpha^*\right)^i i = \left(\left(\beta^{*\prime}\alpha^*\right)^{-2} + \left(\beta^{*\prime}\alpha^*\right)^{-1}\right).
$$

The second term is calculated in the case with $\Phi D_t = \mu$, and the term we need to add is given by

$$
\left(\beta^{*\prime}\alpha^*\right)^{-2}\beta^{*\prime}E_1 =
\begin{pmatrix}
A'\Gamma BA' - A'\Psi C \\
BA' + C\Psi C + C \\
\vdots \\
BA' + C\Psi C + (k-1)C
\end{pmatrix}
$$

thus

$$F \left(\beta^{*\prime}\alpha^*\right)^{-2} \beta^{*\prime} E_1 \;=\; BA'\Gamma BA' - BA'\Psi C$$

$$-C\Psi BA' - C\Psi C\Psi C - \frac{k(k-1)}{2} C\Psi C$$

so that

$$\sum_{i=0}^{\infty} C_i i \;=\; F\left(\left(\beta^{*\prime}\alpha^*\right)^{-1} + \left(\beta^{*\prime}\alpha^*\right)^{-2}\right)\beta^{*\prime} E_1$$

$$= \; BA' + C\Psi C + BA'\Gamma BA' - BA'\Psi C$$

$$-C\Psi BA' - C\Psi C\Psi C - \frac{k(k-1)}{2} C\Psi C.$$

## References

BEVERIDGE, S., AND C. R. NELSON (1981): "A New Approach to Decompositions of Time Series Into Permanent and Transitory Components with Particular Attentions to Measurement of the 'Business Cycle'." *Journal of Monetary Economics*. 7. 151–174.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Co-Integration and Error Correction: Representation, Estimation and Testing," *Econometrica*. 55. 251–276.

ENGLE, R. F., AND S. KOZICKI (1993): "Testing for Common Features." *Journal of Business and Economic Statistics*, 11. 369–380.

HANSEN, P., AND S. JOHANSEN (1998): *Workbook on Cointegration*. Oxford University Press. Oxford.

HYLLEBERG, S., R. F. ENGLE, C. W. J. GRANGER, AND S. YOO (1990): "Seasonal Integration and Cointegration," *Journal of Econometrics*. 44. 215–238.

JOHANSEN, S. (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–1580.

——— (1996): *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford, 2nd edn.

JOHANSEN, S., AND E. SCHAUMBURG (1998): "Likelihood Analysis of Seasonal Cointegration." *Journal of Econometrics*, 88, 301–339.

LÜTKEPOHL, H., AND H.-E. REIMERS (1992): "Impulse Response Analysis of Cointegrated Systems." *Journal of Enonomic Dynamics and Control*, 16, 53–78.

LÜTKEPOHL, H., AND P. SAIKKONEN (1997): "Impulse Response Analysis in Infinite Order Cointegrated Vector Autoregressive Processes." *Journal of Econometrics*, 81. 127–157.

WARNE, A. (1993): "A Common Trends Model: Identification. Estimation and Inference." *Seminar Paper No. 555, IIES, Stockholm University*.

# Chapter 2

# Structural Changes in the Cointegrated Vector Autoregressive Model[†]

## Abstract

I generalize the cointegrated vector autoregressive model of Johansen (1988) to allow for structural changes. I derive the likelihood ratio test for structural changes occurring at fixed points in time, and show that it is asymptotically $\chi^2$. Moreover, I show how inference can be made when the null hypothesis is presence of structural changes.

The estimation technique derived for this purpose can be applied to other generalizations of the standard model, beyond the structural changes treated here. For example, the new technique can be applied to estimate models with heteroskedasticity.

I apply the generalized model to US term structure data, accounting for structural changes that coincide with the changes in the Fed's policy in September 1979 and October 1982. Contrary to previous findings I cannot reject the long-run implications of the expectations hypothesis.

---

## 2.1. Introduction

The modelling of structural changes in cointegrated processes has been addressed by several authors. In the vector autoregressive framework, Seo (1998) derived the Lagrange multiplier (LM) test for structural changes in cointegration relations and adjustment coefficients, and Inoue (1999) derived a rank test for cointegrated processes with a broken trend. Other approaches to modelling structural changes in cointegrated processes are the recursive estimation to identify structural changes by H. Hansen and Johansen (1999), the combination of cointegration and Markov switching by Krolzig (1996), the co-breaking theory by Hendry (1995), and a test for a cointegrating relation with a structural change against an $I(1)$ alternative was given by Gregory and B. E. Hansen (1996).

One of the main contributions of this paper is the development of a flexible framework in which structural changes can be formulated. The most related paper is the one of Seo (1998), who considered structural changes in cointegration relations and adjustment coefficients, under i.i.d. assumptions. The framework proposed here can handle a class of changes in integrated processes that are more general than previously treated. Partial structural changes[1] such as, a structural change in a particular cointegration relation or its mean can be handled, leaving other relations unchanged. In addition, the framework is applicable under weaker assumptions than the i.i.d. assumption. The test statistic invoked in this paper is the likelihood ratio (LR) test and it is shown that its asymptotic distribution is standard $\chi^2$ when the change points are taken as given[2]. Another contribution of this paper is that it enables hypotheses testing under the maintained hypothesis that the underlying process exhibits structural changes. The asymptotic $\chi^2$ results remain valid in this situation.

Another main contribution of this paper is the introduction of a new estimation technique, the *generalized reduced rank regression* (GRRR) technique. This technique has an applicability

---

[1] Partial structural changes in stationary processes has been analysed by Bai and Perron (1998) and Bai (1999).
[2] The case of an unknown change point leads to a non-standard asymptotic distribution. See Seo (1998) or Andrews and Ploberger (1994). I treat this aspect in Chapter 3.

beyond the estimation problems that arise from structural changes.

Estimation of the cointegrated vector autoregressive model was solved by Johansen (1988) as an eigenvalue problem, also known as reduced rank regression. This technique is directly applicable to estimation under simple linear restrictions on cointegration relations, $\beta$, and adjustment coefficients, $\alpha$. Johansen and Juselius (1992) proposed a switching algorithm for estimation under slightly more general restrictions. Boswijk (1995) derived a general estimation technique that can handle linear restrictions on $\text{vec}(\alpha)$ and $\text{vec}(\beta)$, where $\text{vec}(\cdot)$ is the vectorization operator.

The estimation technique of Boswijk (1995) is applicable to several estimation problems we face with structural changes in the cointegrated VAR. The GRRR technique introduced in this paper is a generalization of his technique in two directions. First of all, the GRRR technique allows for linear restrictions on all parameters apart from the variance parameter, by which it achieves a generality similar to the minimum distance approach by Elliott (1997, 1998a), since the generalization to nonlinear restrictions expressed by functions that are well-behaved[3] is straightforward. Secondly, the GRRR technique allows for a general covariance structure and is therefore applicable to models with heteroskedasticity.

The result of this paper is applied to the US term structure of interest rates. The results are that the long-run implications of the expectations hypothesis cannot be rejected once structural changes have been accounted for.

The paper is organized as follows. Section 2 contains the statistical formulation of various structural changes in the cointegration model. The estimation problems are treated in Section 3, and Section 4 contains the asymptotic analysis. Section 5 contains an empirical analysis of the expectations hypothesis applied to the US term structure of interest rates. Section 6 concludes, and the appendix contains proofs.

---

[3] Functions that are continuously differentiable.

## 2.2. The Statistical Model

In this section we give some of the details of the cointegrated vector autoregressive model by Johansen (1988). The model is generalized to allow for various structural changes and it is shown how these changes can be formulated as parameter restrictions in a unified framework.

### 2.2.1. The Cointegrated Vector Autoregressive Model

I take the $p$-dimensional vector autoregressive model $X_t = \Pi_1 X_{t-1} + \cdots + \Pi_k X_{t-k} + \Phi D_t + \varepsilon_t$ as my point of origin, where $\varepsilon_t$ is assumed to be independent and Gaussian distributed with mean zero and variance $\Omega$. The variable $D_t$ contains deterministic terms such as a constant, a linear trend and seasonal dummies. The error correction form for the model is

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t.$$

and it is well known that if the characteristic polynomial, here given by $A(z) = I(1-z) - \Pi z - \sum_{i=1}^{k-1} \Gamma_i(1-z)z^i$, has all its roots outside the unit-disk, then $X_t$ is stationary. If the polynomial has one or more unit roots, then $X_t$ is an integrated process as defined by Johansen (1996). A unit root implies that $\Pi$ has reduced rank $r < p$ and if the number of unit roots equals $p - r$, then the process $X_t$ is integrated of order one, denoted $I(1)$. When $\Pi$ has reduced rank, it can be written as a product of two $p \times r$ matrices $\Pi = \alpha\beta'$, such that the model can be expressed in the form

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t. \tag{2.2.1}$$

This process can be inverted to an infinite moving average representation, also known as the Granger representation, (see Chapter 1 or P. R. Hansen (2000b)). The representation shows ($i$) how the adjustment coefficient, $\alpha$, relates to the common stochastic trends in the process and ($ii$) that $\beta$ defines the cointegration relations.

It is convenient to rewrite the model as

$$Z_{0t} = \alpha \beta' Z_{1t} + \Psi Z_{2t} + \varepsilon_t. \tag{2.2.2}$$

where $Z_{0t} = \Delta X_t$, $Z_{1t} = X_{t-1}$, $Z_{2t} = (\Delta X_{t-1}', \ldots, \Delta X_{t-k+1}', D_t')'$ and $\Psi = (\Gamma_1, \ldots, \Gamma_{k-1}, \Phi)$, so we separate the regressors with reduced rank parameters from the regressors with unrestricted parameters. In some situations we want to add variables to the cointegration space, such as exogenous variables or simply a linear trend or a constant. In such cases we redefine $Z_{1t}$ to include these variables and denotes its dimension by $p_1$ rather than $p$, which denotes the dimension of $Z_{0t}$. The regression problem in equation (2.2.2), with no additional restrictions on the parameters, is referred to as a reduced rank regression (RRR).

We define a generalized reduced rank regression. as the following regression problem:

$$Z_{0t} = AB' Z_{1t} + C Z_{2t} + \varepsilon_t. \tag{2.2.3}$$

$$\text{s.t. } \text{vec}(A, C) = G\upsilon.$$

$$\text{vec}(B) = H\varphi.$$

where $G$ and $H$ are known matrices with full column rank. and $\{\varepsilon_t\}$ obeys the following assumption.

**Assumption 2.2.1.** $\{\varepsilon_t\}$ is a sequence of independent $p$-dimensional Gaussian variables. where $\varepsilon_t$ is independent of $Z_{1t}$ and $Z_{2t}$ and has the marginal distribution $N(0, \Omega(t))$.

By this formulation the i.i.d. assumption on $\{\varepsilon_t\}$ is relaxed. by no longer requiring an identical distribution. We leave the exact structure of $\Omega(t)$. $t = 1, \ldots, T$ to be determined from model-specific assumptions on heteroskedasticity. The assumption still implies independence of $\{\varepsilon_t\}$. Estimation and inference under a weaker assumption than Assumption 2.2.1 is treated in Chapter 5. see also P. R. Hansen (2000a).

Obviously. the estimation problems that can be solved by a RRR can also be solved by a

GRRR. by setting $G$ and $H$ as identity matrices, and with $\Omega(t) = \Omega$.

As shown by Boswijk (1995), the following assumption is necessary for generic identification of the parameters.

**Assumption 2.2.2.** *The matrices $H$ and $G$ in (2.2.3) have full column rank and are such that $A$ and $B$ have full column rank for all $(\psi', \varphi')' \in \mathbb{R}^n$ except on a set with Lebesgue measure zero, ($n$ denotes the number of column in $(H.G)$).*

Let the covariance parameters be expressed as $\Omega(t) = \Omega_t(\theta)$. $\theta \in \Theta_\theta$. $t = 1, \ldots, T$. This formulation does not necessarily impose any restrictions on the parameters.

**Assumption 2.2.3.** *The parameters $\psi$, $\varphi$ and $\theta$ are variation free. that is*

$$(\psi. \varphi. \theta) \in \Theta_\psi \times \Theta_\varphi \times \Theta_\theta.$$

This assumption is convenient for the parameter estimation. Suppose that Assumption 2.2.3 holds. and consider the procedure that iterates on the following three equations:

$$\psi^{(n)} = \arg\max_{\psi \in \Theta_\psi} L(\psi. \varphi^{(n-1)}. \theta^{(n-1)}).$$

$$\varphi^{(n)} = \arg\max_{\varphi \in \Theta_\varphi} L(\psi^{(n)}. \varphi. \theta^{(n-1)}).$$

$$\theta^{(n)} = \arg\max_{\theta \in \Theta_\theta} L(\psi^{(n)}. \varphi^{(n)}. \theta).$$

$n \geq 1$ until convergence of the likelihood function $L$. starting from some initial values of the parameters $(\psi^{(0)}. \varphi^{(0)}. \theta^{(0)})$. This procedure has the nice property that the value of the likelihood function is increased in every iteration; the ordering of the three parameters is irrelevant. Since the likelihood function is bounded by its global maximum, the procedure will eventually converge. Since finding a stationary point of the three equations is equivalent to solving the normal equations. a convergence point. say $(\hat{\psi}.\hat{\varphi}.\hat{\theta})$. will satisfy the normal equations. So whenever the normal equations uniquely define the global maximum of $L$. maximum likelihood estimation is achieved

with this procedure. The information matrix is asymptotically block diagonal which eliminates existence of local maxima asymptotically. However, in practise local maxima may exist in finite samples, so one should start the algorithm with different initial values of the parameters, and see if the algorithm converges to the same value of the likelihood function.

All the models we consider in this paper satisfy Assumption 2.2.3. An example of a model that does not satisfy this assumption is the GARCH model. This model has a dependence between the parameter space of the covariance matrix, typically denoted by $H_t$, and the other parameters, due to the dependence of $\hat{H}_t$ on the estimated residuals such as $\hat{\varepsilon}_{t-1}$.

We need to calculate the degrees of freedom in the parameter $\alpha(t)\beta(t)'$. The following lemma, taken from Johansen (1996), is useful for this purpose.

**Lemma 2.2.4.** *The function $f(x,y) = xy'$, where $x$ is $p \times r$ $(r \leq p)$ and $y$ is $p_1 \times r$ $(r \leq p_1)$, is differentiable at all points, with a differential given by*

$$Df(x,y) = x(dy)' + (dx)y'$$

*where $dy$ is $p \times r$ and $dx$ is $p_1 \times r$. If $x$ and $y$ have full rank $r$ then the tangent space at $(x,y)$, given by $\{x(dy)' + (dx)y' : dx \in \mathbb{R}^{p_1 \times r}, dy \in \mathbb{R}^{p \times r}\}$ has dimension $(p + p_1 - r)r$.*

So, in the case of a reduced rank regression, with $x = \alpha$ and $y = \beta$, the parameter space in which $\Pi = \alpha\beta'$ can vary has dimension $(p + p_1 - r)r$.

### 2.2.2. Structural Changes in the Cointegrated Vector Autoregressive Model

We now show that structural changes in model (2.2.1) can be viewed as a particular form of (2.2.3). Without loss of generality, we can focus just on changes in $\alpha$ and $\beta$, because changes in the parameters $\Gamma_1, \ldots, \Gamma_{k-1}$ or $\Phi$ in (2.2.1) are easily handled by redefining $Z_{2t}$ and $\Psi$. For now we keep the covariance matrix, $\Omega$, constant, but later we also generalize the model to allow for structural changes in this parameter. Estimation, when all parameters change their value is

easily done by estimating each subsample with the RRR technique, however in most applications it is desirable to keep some parameters fixed to avoid that the dimension of the parameter space increases too dramatically.

So, the generalization of model (2.2.1) that we consider is

$$Z_{0t} = \alpha(t)\beta(t)'Z_{1t} + \Psi Z_{2t} + \varepsilon_t. \tag{2.2.4}$$

We shall consider different choices of the time-dependent parameters $\alpha(t)$ and $\beta(t)$. More specifically, we consider various situations where $\alpha(t)$ and $\beta(t)$ are piecewise constant, which can be expressed as

$$\alpha(t)\beta(t)' = \alpha_1\beta_1'I_{1t} + \cdots + \alpha_q\beta_q'I_{qt} \tag{2.2.5}$$

where $I_{jt}, j = 1, \ldots, q$ are indicator functions that determine which $\alpha_j$ and $\beta_j$ are active. This formulation does not require $\alpha_i$ and $\alpha_j$ to have the same number of columns $i \neq j$, as long as $\alpha_j$ and $\beta_j$ have the same number of columns. So the formulation allows for changes in the number of cointegration relations as well as scenarios where some relations are constant over several subsamples while other relations change.

By defining $Z_{1jt} = I_{jt}Z_{1t}, j = 1, \ldots, q$, and $\tilde{Z}_{1t} = (Z_{11t}', \ldots, Z_{1qt}')'$, we obtain the regression problem

$$Z_{0t} = (\alpha_1, \ldots, \alpha_q)\begin{pmatrix} \beta_1 & 0 & \cdots & 0 & 0 \\ 0 & \beta_2 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & \beta_{q-1} & 0 \\ 0 & 0 & \cdots & 0 & \beta_q \end{pmatrix}' \tilde{Z}_{1t} + \Psi Z_{2t} + \varepsilon_t.$$

with a block diagonal structure of the matrix containing the cointegration relations, denoted by $B$. This structure can be expressed as a linear restriction on $\text{vec}(B) = H\varphi$, and the regression is therefore a special case of equation 2.2.3.

## Structural Changes in $\alpha$ and $\beta$

Consider a situation with $q - 1$ structural changes that occur at time $T_1, \ldots, T_{q-1}$, so that $\alpha_t$ and $\beta_t$ can take on $q$ different values. This can be formulated as

$$\beta_t = \begin{cases} \beta_1 & t = 1, \ldots, T_1 \\ \beta_2 & t = T_1 + 1, \ldots, T_2 \\ \quad \vdots \\ \beta_q & t = T_{q-1} + 1, \ldots, T. \end{cases}$$

and

$$\alpha_t = \begin{cases} \alpha_1 & t = 1, \ldots, T_1 \\ \alpha_2 & t = T_1 + 1, \ldots, T_2 \\ \quad \vdots \\ \alpha_q & t = T_{q-1} + 1, \ldots, T. \end{cases}$$

So in this case we define $Z_{11t} = Z_{1t}I(t \leq T_1)$, $Z_{12t} = Z_{1t}I(T_1 + 1 \leq t \leq T_2)$, $\ldots$, $Z_{1qt} = Z_{1t}I(T_{q-1} + 1 \leq t \leq T)$ and $\tilde{Z}_{1t} = (Z'_{11t}, \ldots, Z'_{1qt})'$, and obtain a model with the form of equation (2.2.3). This formulation allows for a change in the number of cointegration relations. Let $r_i$ denote the cointegration rank in subsample $i$, $i = 1, \ldots, q$. Then the dimension of the parameter space of $\Pi(t) = \alpha(t)\beta(t)'$ is by Lemma 2.2.4 found to be $\sum_{i=1}^{q}(p + p_1 - r_i)r_i$ where $r_i$ is the rank of $\alpha_i\beta'_i$, $i = 1, \ldots, q$. If the rank is constant over the entire sample, the expression for the degrees of freedom simplifies to $q(p + p_1 - r)r$.

## Structural Changes in the Adjustment Coefficients: $\alpha$

If the structural changes only affect the adjustment coefficients. $\alpha$, whereas the cointegration relations remain constant, we can express the model as

$$Z_{0t} = (\alpha_1, \ldots, \alpha_q) \begin{pmatrix} \beta & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \beta \end{pmatrix}' \tilde{Z}_{1t} + \Psi Z_{2t} + \varepsilon_t.$$

where $\tilde{Z}_{1t}$ is as defined above. Since $\beta$ is constant over the sample, so is the cointegration rank $r$. and the dimension of the parameter space for $\Pi(t)$ is simply given by $(qp + p_1 - r)r$.

## Structural Changes in the Cointegrating Relations: $\beta$

When the structural change is solely due to changes in the cointegration relations $\beta(t)$ while $\alpha(t)$ remains constant. the model simplifies to

$$\begin{aligned} Z_{0t} &= \alpha\beta_1' I_{1t} Z_{1t} + \cdots + \alpha\beta_q' I_{qt} Z_{1t} + \Psi Z_{2t} + \varepsilon_t \\ &= \alpha \left( \beta_1', \ldots, \beta_q' \right) \tilde{Z}_{1t} + \Psi Z_{2t} + \varepsilon_t. \end{aligned}$$

where $\tilde{Z}_{1t}$ is as defined previously. Here we again obtain an equation of the form of (2.2.3). but in this case without additional restrictions on $A$. $B$. and $C$. i.e. $G = I$. and $H = I_{p_1 rq}$. In this situation only a constant cointegration rank. $r$. is meaningful and the dimension of the parameter space for $\Pi(t)$ is given by $(p + qp_1 - r)r$.

The relations between the different structural changes are displayed in Figure 2.2.1. along with the relevant asymptotic distribution and degrees of freedom. The asymptotic distribution is derived below. and it is not surprisingly found to be asymptotically $\chi^2$.

General structural change

$$Z_{0t} = \alpha_1 \beta_1' Z_{11t} + \alpha_2 \beta_2' Z_{12t} + \Psi Z_{2t} + \varepsilon_t$$

$\chi^2((p_1 - r)r)$     $\chi^2((p - r)r)$

Structural change in $\alpha$

$$Z_{0t} = \alpha_1 \beta' Z_{11t} + \alpha_2 \beta' Z_{12t} + \Psi Z_{2t} + \varepsilon_t$$

Structural change in $\beta$

$$Z_{0t} = \alpha(\beta_1, \beta_2)'(Z_{11t}', Z_{12t}')' + \Psi Z_{2t} + \varepsilon_t$$

$\chi^2(pr)$     $\chi^2(p_1 r)$

Model without changes

$$Z_{0t} = \alpha \beta' Z_{1t} + \Psi Z_{2t} + \varepsilon_t$$
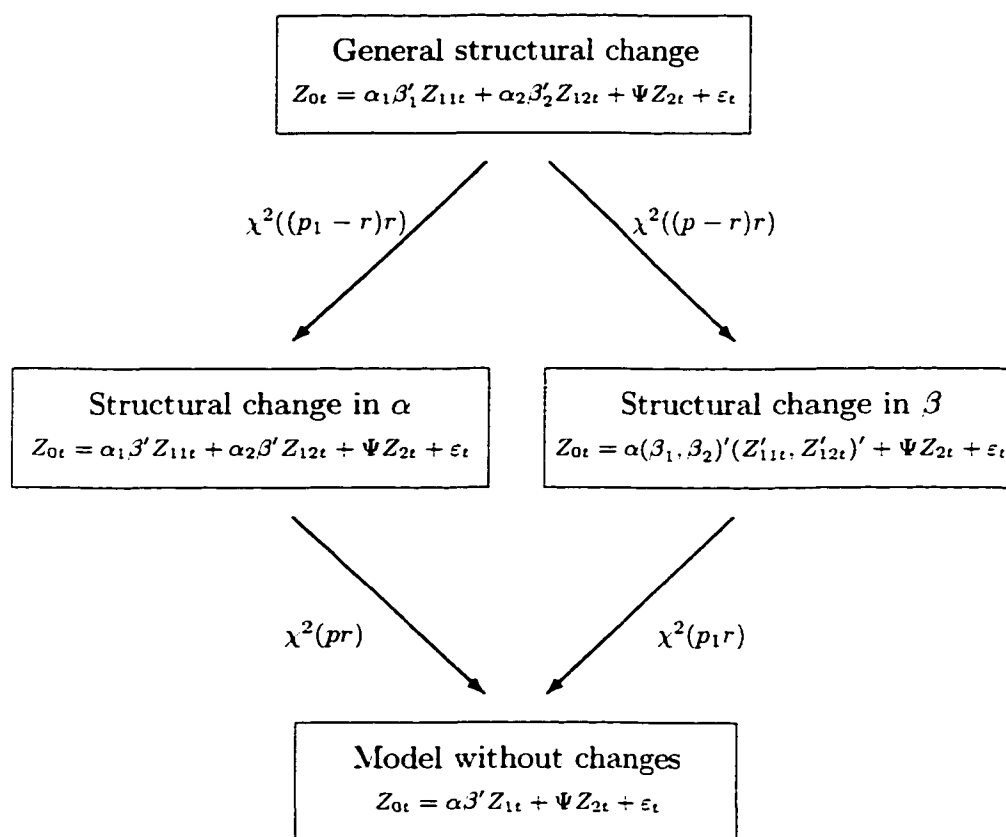
Figure 2.2.1: The relations between the different types of structural changes. The asymptotic distribution of the individual LR test is $\chi^2$ in all cases, with the degrees of freedom reported in the brackets.

## Temporary and Permanent Cointegration Relations

The scenario where some cointegration relations are present in the entire sample, whereas others are only present in a subsample can also be expressed in the form of equation (2.2.3). The simplest situation is where there are $r_1$ permanent cointegration relations, say $\beta_1$, and for $t \geq T_1 + 1$ there are an additional $r_2 - r_1$ temporary cointegration relations, say $\beta_e$ (linearly independent of $\beta_1$). This situation leads to two different cases – one where the adjustment coefficients corresponding to $\beta_1$ remain constant, and one where they may differ in the two subsamples. The latter is likely to be the most relevant, since the introduction of an extra adjustment from the added cointegration

relations might affect how the process adjusts to the permanent cointegration relations.

First we consider the case where $\alpha_1$ remains constant. This model is formulated as

$$Z_{0t} = \alpha_1\beta_1'Z_{1t} + \alpha_e\beta_e'Z_{1t}I_{(t>T_1)} + \Psi Z_{2t} + \varepsilon_t$$

$$= (\alpha_1.\alpha_e)\begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_e \end{pmatrix}'\begin{pmatrix} Z_{1t} \\ Z_{1t}I_{(t>T_1)} \end{pmatrix} + \Psi Z_{2t} + \varepsilon_t,$$

and the dimension of the parameter space for $\Pi(t)$ is slightly more complicated to derive. The degrees of freedom in $\Pi_1$ are given by $(p + p_1 - r_1)r_1$. and since $\Pi_2 = \Pi_1 + \alpha_e\beta_e'$ the additional contribution from $\Pi_2$ is given by $[p + (p_1 - r_1) - (r_2 - r_1)](r_2 - r_1)$. Adding the two terms gives the degrees of freedom in $\Pi(t)$ to be $(p + p_1 - r_2)r_2 + (r_2 - r_1)r_1$.

The model where the adjustment coefficients to the permanent cointegration relations may change. is formulated as

$$Z_{0t} = \alpha_{11}\beta_1'Z_{1t}I_{(t\leq T_1)} + (\alpha_{21}.\alpha_e)(\beta_1.\beta_e)'Z_{1t}I_{(t>T_1)} + \Psi Z_{2t} + \varepsilon_t$$

$$= (\alpha_{11}.\alpha_{12}.\alpha_e)\begin{pmatrix} \beta_1 & 0 & 0 \\ 0 & \beta_1 & \beta_e \end{pmatrix}'\begin{pmatrix} Z_{1t}I_{(t\leq T_1)} \\ Z_{1t}I_{(t>T_1)} \end{pmatrix} + \Psi Z_{2t} + \varepsilon_t.$$

which is also of the form of equation (2.2.3). but with a more complicated structure of $H$. due to the cross restrictions we have on $B$. The degrees of freedom are found by adding up the contributions from $\Pi_1$. $\alpha_{21}\beta_1'$ and $\alpha_e\beta_e'$. These are given by $(p+p_1-r_1)r_1$. $pr_1$ and $[p+(p_1-r_1)-(r_2-r_1)](r_2-r_1)$ respectively. where we used that $\beta_e$ may be chosen orthogonal to $\beta_1$. Adding the three terms up. gives the dimension of $\Pi(t)$ to be $(p + p_1 - r_2)r_2 + (p + r_2 - r_1)r_1$.

The former model is obviously nested in the latter. and both models are nested in the model where there are not necessarily any relations between the cointegration relations in the two samples. This model has a structure as given above with $r_1$ cointegration relations in the first subsample and $r_2$ in the second. So the model has $(p + p_1 - r_1)r_1 + (p + p_1 - r_2)r_2$ free parameters in

$\Pi(t)$. The relations between these three models are displayed in Figure 2.2.2. Below we prove that the likelihood ratio test for this hypothesis is asymptotically $\chi^2$ with degrees of freedom that correspond to the difference in dimensionality of $\Pi(t)$, as one would expect.

The extension to models with multiple sets of temporary cointegration relations in individual and overlapping subsamples is straightforward, only the calculation of degrees of freedom can be somewhat tricky.

---

**General structural change model**
$r_1$ cointegration relations for $t \leq T_1$
$r_2$ cointegration relations for $t \geq T_1 + 1$

$$Z_{0t} = \alpha_1 \beta_1' Z_{11t} + \alpha_2 \beta_2' Z_{12t} + \Psi Z_{2t} + \varepsilon_t$$

$\downarrow \quad \chi^2((p_1 - r_2)r_1)$

**Permanent cointegration relations:** $\quad r_1$
**Constant adjustment coefficients:** $\quad \div$
**Temporary cointegration relations:** $\quad r_2 - r_1$

$$Z_{0t} = \alpha_{11} \beta_1' Z_{11t} + (\alpha_{21}, \alpha_e)(\beta_1, \beta_e)' Z_{12t} + \Psi Z_{2t} + \varepsilon_t$$

$\downarrow \quad \chi^2(pr_1)$

**Permanent cointegration relations:** $\quad r_1$
**Constant adjustment coefficients:** $\quad \sqrt{}$
**Temporary cointegration relations:** $\quad r_2 - r_1$

$$Z_{0t} = \alpha_1 \beta_1' Z_{1t} + \alpha_e \beta_e' Z_{12t} + \Psi Z_{2t} + \varepsilon_t$$

Figure 2.2.2: The relations between the different models with structural changes and a change in the number of cointegration relations. The distribution of the LR test statistic between two of the models is asymptotically $\chi^2$ with the degrees of freedom given in the figure.

## Structural Changes in the Covariance Matrix

Structural changes in the covariance matrix also leads to a GRRR. The simplest case is a single structural change in the covariance matrix at time $T_1$. So $\text{var}(\varepsilon_t) = \Omega_1$ for $t \leq T_1$ and $\text{var}(\varepsilon_t) = \Omega_2$ for $t > T_1$, which implies the following structure on the covariance matrix

$$\Sigma = \begin{pmatrix} I_{T_1} \otimes \Omega_1 & 0 \\ 0 & I_{T-T_1} \otimes \Omega_2 \end{pmatrix}.$$

The combination of structural changes in the covariance matrix as well as other parameters, will also lead to a GRRR.

## Linear Restriction on Adjustment Coefficients and Cointegration Relations

Combining hypotheses of structural changes with linear restrictions on the cointegration relations will not complicate the estimation problem, because the two parameter restrictions can jointly be formulated as a linear restriction $\text{vec}(B) = H\varphi$ for a known matrix $H$ and some parameters $\varphi$. Adding linear restrictions to the adjustment coefficients, $\alpha_1, \ldots, \alpha_q$ can be formulated as $\text{vec}(A) = G\upsilon$, and is therefore also a GRRR.

## 2.3. Estimation

Estimation of the cointegrated vector autoregressive model, and other models that have the structure of equation (2.2.2), can be explicitly solved as an eigenvalue problem by reduced rank regression techniques. The method of reduced rank regression was developed by Anderson (1951) and applied to the $I(1)$ model by Johansen (1988).

The advantage of reduced rank estimation is that an explicit solution is obtained without iterations. Fortunately this method is applicable to estimation under simple linear restrictions on the reduced rank parameters. However, in most of the structural change models we face restrictions

that are beyond what the technique can handle. So a more general estimation technique is needed.

A few of the problems can be formulated as regression problems that can be handled by the *switching algorithm* of Johansen and Juselius (1992). This algorithm is an iterative procedure that in every iteration simplifies the problem to a reduced rank regression by keeping a subset of the parameters fixed. This method has the nice property that it increases the value of the likelihood function in every iteration, but unfortunately applications have shown that convergence can be very slow. More problematic is that general convergence to the global optimum cannot be proven: indeed it is easy to construct examples where the method will not converge.

A more general estimation technique was proposed by Boswijk (1995). This method is similar to the switching algorithm, in the sense that it increases the likelihood function in every iteration. It is more general because it can handle estimation problems with linear restrictions on $\text{vec}(B)$ and $\text{vec}(A)$. This method is therefore sufficient for most of the estimation problems that arise from structural change models. Applications of the method have shown that convergence is obtained in few iterations, and that it does converge to the global optimum. The fast convergence is not surprising because the information matrix is asymptotically block diagonal.

More general yet is the minimum distances approach by Elliott (1997, 1998a), which can estimate parameters under the general restriction $g(\theta) = c$, where $\theta$ is the vector of parameters, $c$ is a constant and $g$ is a well-behaved function. This method minimizes $\theta' \hat{V}_{\hat{\theta}} \theta$ subject to the constraints $g(\theta) = c$, where $\hat{V}_{\hat{\theta}}$ is an estimate of the asymptotic covariance matrix. This method is asymptotically equivalent to the maximum likelihood estimation, and with suitable choice of $\hat{V}_{\hat{\theta}}$ and if applied iteratively, (by recursive reestimation of $\hat{V}_{\hat{\theta}}$ as the estimate of $\theta$ changes), the minimum distance methods leads to the same estimator as the maximum likelihood method.

As we shall see below, it is possible to estimate under more general restrictions than those considered by Boswijk (1995) and Elliott (1997, 1998a). By handling restrictions as formulated in model (2.2.3) we obtain the same generality as the minimum distance method by Elliott (1997, 1998a), and can in addition estimate models with heteroskedasticity.

In the following we consider the reduced rank regression model

$$Z_{0t} = AB'Z_{1t} + CZ_{2t} + \varepsilon_t. \tag{2.3.1}$$

with various restrictions on the parameters, under Assumptions 2.2.1 and 2.2.2. We denote the dimension of $Z_{0t}$, $Z_{1t}$ and $Z_{2t}$ by $p$, $p_1$ and $p_2$ respectively, and for notional convenience we define the moment matrices $M_{ij} = \frac{1}{T}\sum_{t=1}^{T} Z_{it}Z'_{jt}$, $i, j = 0, 1, 2$, the residuals $R_{0t} = Z_{0t} - M_{02}M_{22}^{-1}Z_{2t}$, $R_{1t} = Z_{1t} - M_{12}M_{22}^{-1}Z_{2t}$. and the moment matrices of the residuals $S_{ij} = \frac{1}{T}\sum_{t=1}^{T} R_{it}R'_{jt}$. $i, j = 0, 1$.

## 2.3.1. Reduced Rank Regression

Estimation of reduced rank regressions is described in the following theorem.

**Theorem 2.3.1 (Reduced Rank Regression).** *The unrestricted estimators of Model (2.2.3) are given by*

$$\hat{B} = (\hat{v}_1, \ldots, \hat{v}_r)\varphi \tag{2.3.2}$$

$$\hat{A}(B) = S_{01}B(B'S_{11}B)^{-1} \tag{2.3.3}$$

$$\hat{\Omega} = S_{00} - S_{01}BA' + AB'S_{11}BA' - AB'S_{10}. \tag{2.3.4}$$

$$\hat{C} = M_{02}M_{22}^{-1} - \hat{A}\hat{B}'M_{12}M_{22}^{-1}. \tag{2.3.5}$$

*where $(\hat{v}_1, \ldots, \hat{v}_r)$ are the eigenvectors corresponding to the r largest eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_r$ of the eigenvalue problem*

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0.$$

*and where $\varphi$ is any $r \times r$ full rank matrix, by which $\hat{B}$ can be normalized. The maximum value of*

the (conditional) likelihood function is given by

$$L_{\max}^{-2/T}(\hat{A}.\hat{B}.\hat{C}.\hat{\Omega}) = (2\pi e)^p \, |S_{00}| \prod_{i=1}^{r}(1 - \lambda_i).$$

An algebraic proof that uncovers the structure of the problem is given in the appendix whereas the original proof can be found in Johansen (1996).

This theorem is directly applicable to the cointegrated vector autoregressive model given by equation 2.2.1. The maximum likelihood estimate is obtained by defining $Z_{0t} = \Delta X_t$. $Z_{1t} = X_{t-1}$ and $Z_{2t} = (\Delta X'_{t-1}.\ldots.\Delta X'_{t-k+1}.D'_t)'$.

## 2.3.2. Generalized Reduced Rank Regression

**Theorem 2.3.2.** *Let the parameter $A$. $B$. and $C$ be restricted by $\mathrm{vec}(A.C) = Gv$ and $\mathrm{vec}(B) = Ho$ and suppose that Assumptions 2.2.1-3 hold.*

*The maximum likelihood estimates $\hat{A}$. $\hat{B}$. $\hat{C}$. and $\hat{\Omega}(t)$ of $A$. $B$. $C$. and $\Omega(t)$ will satisfy*

$$\mathrm{vec}(\hat{A}.\hat{C}) = G\left[G'\sum_{t=1}^{T}\left[\begin{pmatrix} \hat{B}'Z_{1t}Z'_{1t}\hat{B} & \hat{B}'Z_{1t}Z'_{2t} \\ Z_{2t}Z'_{1t}\hat{B} & Z_{2t}Z'_{2t} \end{pmatrix} \times \hat{\Omega}(t)^{-1}\right]G\right]^{-1} \quad (2.3.6)$$

$$\times G'\sum_{t=1}^{T}\mathrm{vec}\left(\hat{\Omega}(t)^{-1}Z_{0t}(Z'_{1t}\hat{B}.Z'_{2t})\right).$$

$$\mathrm{vec}(\hat{B}) = H\left[H'\sum_{t=1}^{T}\left[\hat{A}'\hat{\Omega}(t)^{-1}\hat{A} \otimes Z_{1t}Z'_{1t}\right]H\right]^{-1} \quad (2.3.7)$$

$$\times H'\sum_{t=1}^{T}\mathrm{vec}\left(Z_{1t}(Z_{0t} - \hat{C}Z_{2t})'\hat{\Omega}(t)^{-1'}\hat{A}\right)$$

and $\hat{\Omega}(t) = \Omega_t(\hat{\theta})$. where $\hat{\theta}$ is given from the (model specific) equation

$$\hat{\theta} = \arg\max_{\theta \in \Theta_\theta} L(\hat{A}.\hat{B}.\hat{C}.\theta.Z_0.Z_1.Z_2). \quad (2.3.8)$$

The maximum value of the likelihood function is given by

$$L_{\max}(\hat{A}.\hat{B}.\hat{C}.\hat{\Omega}) = (2\pi)^{-\frac{Tp}{2}} \prod_{t=1}^{T} |\hat{\Omega}(t)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^{T} \hat{\varepsilon}_t' \hat{\Omega}(t)^{-1}\hat{\varepsilon}_t\right).$$

where $\hat{\varepsilon}_t = Z_{0t} - \hat{A}\hat{B}'Z_{1t} - \hat{C}Z_{2t}$.

The proof exploits that the estimation problem reduces to a GLS problem, when $(A.C.\Omega(t))$ or $(B.\Omega(t))$ are hold constant. The proof is given in the Appendix.

The theorem yields a procedure for parameter estimation, in the sense that the parameter estimates can be obtained by iterating on the three equations until convergence, from some initial values of the parameters. As described in the paragraph following Assumption 2.2.3, this procedure will converge to parameter values that satisfy the normal equations.

We now treat situations with fewer parameter restrictions.

**Corollary 2.3.3.** *Let the parameter $A$. $B$ and $C$ be restricted by* $\text{vec}(A.C) = Gv$ *and* $\text{vec}(B) = Ho$ *and suppose that* $\{\varepsilon_t\}$ *is i.i.d. Gaussian $N(0.\Omega)$.*

*The maximum likelihood estimates of $A$. $B$. $C$ and $\Omega$ satisfy the equations*

$$\text{vec}(\hat{A}.\hat{C}) = G\left[G'(\begin{pmatrix} \hat{B}'M_{11}B & \hat{B}'M_{12} \\ M_{21}\hat{B} & M_{22} \end{pmatrix} \times \hat{\Omega}^{-1})G\right]^{-1} G'\text{vec}\left(\hat{\Omega}^{-1}(M_{01}\hat{B}.M_{02})\right).$$

$$\text{vec}(\hat{B}) = H\left[H'\left(\hat{A}'\hat{\Omega}^{-1}\hat{A} \times M_{11}\right)H\right]^{-1} H'\text{vec}\left(M_{10} - M_{02}\hat{C}'\right)\hat{\Omega}^{-1}\hat{A}\right).$$

$$\hat{\Omega} = T^{-1}(Z_0 - \hat{A}\hat{B}'Z_1 - \hat{C}Z_2)(Z_0 - \hat{A}\hat{B}'Z_1 - \hat{C}Z_2)'.$$

*The maximum value of the likelihood function is given by*

$$L_{\max}^{-2/T}(\hat{A}.\hat{B}.\hat{C}.\hat{\Omega}) = (2\pi e)^p |\hat{\Omega}|.$$

If $C$ is unrestricted we obtain the following result of Boswijk (1995).

**Corollary 2.3.4.** *Let* $A$ *and* $B$ *be restricted by* $\text{vec}(A) = G\zeta$ *and* $\text{vec}(B) = H\phi$. *for known matrices* $G$ *and* $H$. *Then the maximum likelihood estimates satisfy the equations*

$$\text{vec}(\hat{B}) = H \left[ H' \left( \hat{A}'\hat{\Omega}^{-1}A \otimes S_{11} \right) H \right]^{-1} H' \left( \hat{A}' \otimes S_{10} \right) \text{vec} \left( \hat{\Omega}^{-1} \right) \tag{2.3.9}$$

*and*

$$\hat{A} = G \left[ G' \left( \hat{B}'S_{11}\hat{B} \otimes I_p \right) G \right]^{-1} G' \left( \hat{B}' \otimes \hat{\Omega}^{-1} \right) \text{vec}(S_{01}) \tag{2.3.10}$$

$$\hat{\Omega} = S_{00} - S_{01}\hat{B}\hat{A}' + \hat{A}\hat{B}'S_{11}\hat{B}\hat{A}' - \hat{A}\hat{B}'S_{10}$$

$$\hat{C} = M_{02}M_{22}^{-1} - \hat{A}\hat{B}'M_{12}M_{22}^{-1}.$$

*The maximum value of the likelihood function is given by*

$$L_{\max}^{-2/T}(\hat{A}. \hat{B}. \hat{C}. \hat{\Omega}) = (2\pi e)^P |\hat{\Omega}|.$$

**Corollary 2.3.5.** *Let* $B$ *be restricted by* $\text{vec}(B) = H\phi$. *Then the maximum likelihood estimates satisfy the equations*

$$\text{vec}(\hat{B}(A.\Omega)) = H \left[ H' \left( \hat{A}'\hat{\Omega}^{-1}\hat{A} \otimes S_{11} \right) H \right]^{-1} H' \left( \hat{A}' \otimes S_{10} \right) \text{vec} \left( \hat{\Omega}^{-1} \right)$$

$$\hat{A}(B) = S_{01}\hat{B} \left( \hat{B}'S_{11}\hat{B} \right)^{-1}.$$

$$\hat{\Omega}(B) = S_{00} - S_{01}\hat{B} \left( \hat{B}'S_{11}\hat{B} \right)^{-1} \hat{B}'S_{10}.$$

$$\hat{C} = M_{02}M_{22}^{-1} - \hat{A}\hat{B}'M_{12}M_{22}^{-1}.$$

*The maximum value of the likelihood function is given by*

$$L_{\max}^{-2/T}(\hat{A}. \hat{B}.\hat{C}.\hat{\Omega}) = (2\pi e)^P |\hat{\Omega}|.$$

With these results we have the tools available to estimate the parameters in the cointegrated vector autoregressive model under all the various structural changes considered in the previous section. However, the theorems presented here have a broader applicability, and can be used to estimate models with parameter restrictions that need not be related to structural changes, for example models with heteroskedasticity.

### 2.3.3. Applicability

**Example 2.3.6 (Structural changes in the covariance matrix).** *Consider the cointegrated vector autoregressive model (equation (2.2.1)), with a structural change at time $T_1$, in the sense that $\alpha(t) = \alpha_1$, $\beta(t) = \beta_1$ and $\Omega(t) = \Omega_1$ for $t \leq T_1$ and $\alpha(t) = \alpha_2$, $\beta(t) = \beta_2$ and $\Omega(t) = \Omega_2$ for $t \geq T_1 + 1$. This estimation problem can be written in the form of Model 2.2.3. The maximum likelihood estimators of $\Omega_1$ and $\Omega_2$ are given by*

$$\hat{\Omega}_1 = T_1^{-1} \sum_{t=1}^{T_1} \hat{\varepsilon}_t \hat{\varepsilon}_t'$$

$$\hat{\Omega}_2 = (T - T_1)^{-1} \sum_{t=T_1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'.$$

*So $\Omega(t)$, $t = 1, \ldots, T$ can be expressed in the functional form required by Theorem 2.3.2.*

**Example 2.3.7 (Heteroskedasticity).** *Models with the following type of heteroskedastic errors*

$$\text{var}(\varepsilon_t) = \Omega_t = f_\theta(\Omega_{t-1}, \Omega_{t-2}, \ldots, X_{t-1}, X_{t-2}, \ldots)$$

*can be expressed with the functional form in Theorem 2.3.2.*

## 2.4. Asymptotic Analysis

For simplicity, we derive the asymptotic results in the case of a single structural change at time $T_1$, and with the number of cointegrating relations being constant, $r$. However, it will be clear that the

results hold in the general situation with multiple changes, and varying number of cointegrating relations.

The process is described by

$$\Delta X_t = \alpha_1 \beta_1' X_{t-1} I_{(t \leq T_1)} + \alpha_2 \beta_2' X_{t-1} I_{(t > T_1)} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t.$$

where $\varepsilon_t$ is i.i.d.[1] $N(0, \Omega(t))$. $\Omega(t) = \Omega_1$ for $t \leq T_1$ and $\Omega(t) = \Omega_2$ for $t > T_1$.

In addition, we assume that the usual $I(1)$ assumptions hold in both subsamples. Specifically, that the roots of

$$\left| I(1-z) - \alpha_i \beta_i' z - \sum_{i=1}^{k-1} \Gamma_i (1-z) z^i \right| = 0$$

are outside the unit disc or equal to one, and that $\alpha_{i\perp}' (I - \Gamma_1 - \cdots - \Gamma_{k-1}) \beta_{i\perp}$ has full rank $p - r$. $i = 1, 2$.

### 2.4.1. The Granger Representation for Change Processes

In order to study the process's asymptotic properties. we need to derive the Granger representation for this process. The individual Granger representations for each of the sub-samples are given by

$$X_t = C \sum_{i=1}^{t} \varepsilon_i + C(L)\varepsilon_t + C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \quad t = 1, \ldots, T_1.$$

and

$$X_t = D \sum_{i=T_1+1}^{t} \varepsilon_i + D(L)\varepsilon_t + D(X_{T_1} - \sum_{i=1}^{k-1} \Gamma_i X_{T_1-i}) \quad t = T_1 + 1, \ldots, T.$$

where $C = \beta_{1\perp} (\alpha_{1\perp}' \Gamma \beta_{1\perp})^{-1} \alpha_{1\perp}'$, $D = \beta_{2\perp} (\alpha_{2\perp}' \Gamma \beta_{2\perp})^{-1} \alpha_{2\perp}'$ and $\Gamma = I - \Gamma_1 - \cdots - \Gamma_{k-1}$. (see P. R. Hansen (2000b)).

In order to get the representation in the appropriate form we need to express the second

---

[1] The asymptotic results will hold under more general conditions, though not always with the same asymptotic distribution. Both the Gaussian assumption and the i.i.d. assumption can be relaxed to $\{\varepsilon_t\}$ satisfying a Functional Central Limit Theorem, (see White (2000)).

representation with initial values depending only on $X_t$, $t = 0, -1, \ldots$, rather than $D(X_{T_1} - \sum_{i=1}^{k-1} \Gamma_i X_{T_1-i})$. This is obtained by the expression

$$
\begin{aligned}
D(X_{T_1} - \sum_{i=1}^{k-1} \Gamma_i X_{T_1-i}) &= D\left[ C\sum_{i=1}^{T_1} \varepsilon_i + C(L)\varepsilon_{T_1} + C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \right. \\
&\quad -\Gamma_1 \left( C\sum_{i=1}^{T_1-1} \varepsilon_i + C(L)\varepsilon_{T_1-1} + C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \right) \\
&\quad \vdots \\
&\quad \left. -\Gamma_{k-1} \left( C\sum_{i=1}^{T_1-k+1} \varepsilon_i + C(L)\varepsilon_{T_1-k+1} + C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \right) \right] \\
&= D\left[ \Gamma C\sum_{i=1}^{T_1} \varepsilon_i + C^*(L)\varepsilon_{T_1} + \Gamma C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \right] \\
&= D\Gamma C\sum_{i=1}^{T_1} \varepsilon_i + DC^*(L)\varepsilon_{T_1} + D\Gamma C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}).
\end{aligned}
$$

where

$$
C^*(L)\varepsilon_{T_1} = (I - (I - \Gamma_1)C)\varepsilon_{T_1} + (C_1 - \Gamma_1 C_0)\varepsilon_{T_1-1} + (C_2 - \Gamma_1 C_1)\varepsilon_{T_1-2} + \cdots
$$

is a stationary process. So altogether we have the Granger representation

$$
X_t = C\sum_{i=1}^{t} \varepsilon_i + C(L)\varepsilon_t + C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \quad t = 1, \ldots, T_1 \tag{2.4.1}
$$

$$
X_t = D\sum_{i=T_1+1}^{t} \varepsilon_i + D\Gamma C\sum_{i=1}^{T_1} \varepsilon_i + D(L)\varepsilon_t + DC^*(L)\varepsilon_{T_1} \tag{2.4.2}
$$

$$
+ D\Gamma C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \quad t = T_1 + 1, \ldots, T.
$$

Note that we have the stationary cointegrating relations in the second sub-sample $\beta_2' X_t = \beta_2' D(L)\varepsilon_t$, which is identical to what it would have been in the case of a constant process. For the first sub-sample the results are trivially the same as in the standard case without changes.

## 2.4.2. The Continuous Time Limits

In an asymptotic study of the process, we shall, as $T$ approaches infinity, keep the proportion of observations in each sub-sample constant. So we define $\rho = \frac{T_1}{T}$, which denotes the fraction of observations in the first sub-sample.

Donsker's invariance principle gives

$$T^{-\frac{1}{2}} \sum_{t=1}^{[Tu]} \varepsilon_t \xrightarrow{w} W(u), \qquad u \in [0,1].$$

where $W(u)$ is a Brownian motion with covariance matrix $\Omega$, and where $\xrightarrow{w}$ denotes weak convergence. We can split this into two independent Brownian motions which gives us

$$T^{-\frac{1}{2}} \left( \sum_{t=1}^{T_1} \varepsilon_t + \sum_{t=T_1+1}^{[Tu]} \varepsilon_t \right) \xrightarrow{w} W(\rho) + W(u) - W(\rho), \quad u > \rho.$$

So the random walk element in $X_t$ in each of the sub-samples, has the continuous time limits:

$$T^{-\frac{1}{2}} C \sum_{t=1}^{[Tu]} \varepsilon_t \xrightarrow{w} CW(u), \qquad u \leq \rho$$

$$T^{-\frac{1}{2}} \left( D\Gamma C \sum_{t=1}^{T_1} \varepsilon_t + D \sum_{t=T_1+1}^{[Tu]} \varepsilon_t \right) \xrightarrow{w} D\Gamma CW(\rho) + D\left(W(u) - W(\rho)\right), \quad u > \rho. \quad (2.4.3)$$

Equation (2.4.3) has an important implication for unit root tests, in processes with structural changes. Standard Dickey-Fuller type distributions, such as $\int (dB)\, B' \left[\int BB' du\right]^{-1} \int B\, (dB)'$ do not define the asymptotic distribution in this situation, because the Gaussian term $D\Gamma CW_1(\rho)$, that comes from the initial values, does not disappear. A unit root test based on observation after a structural change will therefore involve a term such as $\int (dB)\, (B+Z)' \left[\int (B+Z)(B+Z)' du\right]^{-1} \int (B+Z)\, (dB)'$. However, this problem does not occur if a constant (restricted to the cointegration space or unrestricted) is included as regressor, see Chapter 4.

From the Granger representation we find that

$$
T^{-1/2}
\begin{pmatrix}
X_{[Tu]} 1_{(u \leq \rho)} \\
X_{[Tu]} 1_{(u > \rho)}
\end{pmatrix}
\xrightarrow{w}
\begin{pmatrix}
CW(u) 1_{(u \leq \rho)} \\
[D\Gamma C W_1(\rho) + D(W(u) - W(\rho))] 1_{(u > \rho)}
\end{pmatrix}.
$$

since the other terms vanish in probability. Let

$$
B =
\begin{pmatrix}
\beta_1 & 0 \\
0 & \beta_2
\end{pmatrix}.
$$

let $B_\perp$ be the orthogonal complement to $B$. i.e. $B'_\perp B = 0$ and let $\bar{B}_\perp = B_\perp (B'_\perp B_\perp)^{-1}$. We define

$$
G(u) =
\begin{pmatrix}
G_1(u) \\
G_2(u)
\end{pmatrix}
= \bar{B}'_\perp
\begin{pmatrix}
CW(u) 1_{(u \leq \rho)} \\
[D\Gamma C W_1(\rho) + D(W(u) - W(\rho))] 1_{(u > \rho)}
\end{pmatrix}
$$

and by the continuous mapping theorem we have with $u = t/T$ that

$$
T^{-2} \sum_{t=1}^{T} \bar{B}'_\perp
\begin{pmatrix}
X_{[Tu]} 1_{(u \leq \rho)} \\
X_{[Tu]} 1_{(u > \rho)}
\end{pmatrix}
\begin{pmatrix}
X_{[Tu]} 1_{(u \leq \rho)} \\
X_{[Tu]} 1_{(u > \rho)}
\end{pmatrix}'
\bar{B}_\perp
\xrightarrow{w}
\int_0^1 G(u) G(u) du
$$

$$
=
\begin{pmatrix}
\int_0^\rho G_1(u) G_1(u) du & 0 \\
0 & \int_\rho^1 G_2(u) G_2(u) du
\end{pmatrix}.
$$

With this notation, the asymptotic results for unrestricted parameter estimates ($G = I$ and $H = I$) of $A$. $B$. $C$. and $\Omega(t)$. say $\hat{A}_u$. $\hat{B}_u$. $\hat{C}_u$. and $\hat{\Omega}_u(t)$. follows from Johansen (1988. lemma 13.1. 13.2). The results are that (a normalized) $\hat{B}_u$ is super consistent. with a mixed Gaussian asymptotic distribution, and that $\hat{A}_u$ is asymptotically normal. Further it also follows that the LR test of some over identifying restrictions, have a $\chi^2$ asymptotic distribution.

Consistency is not affected by imposing valid restrictions, and the results for the restricted parameter estimates given by expanding the normal equations. Assume for simplicity that $\Omega(t)$ is

constant, then

$$
\begin{aligned}
\mathrm{vec}(\dot{B}) \;=\;& H\left[H'\left[\hat{A}'\hat{\Omega}^{-1}\hat{A}\otimes T^{-2}\sum_{t=1}^{T}Z_{1t}Z_{1t}'\right]H\right]^{-1}\\
&\times H'\sum_{t=1}^{T}\mathrm{vec}\left(T^{-1}Z_{1t}(AB'Z_{1t}+(C-\hat{C})Z_{2t}+\varepsilon_t)'\Omega^{-1\prime}A\right)\\
=\;& \mathrm{vec}(B)+H\left[H'\left[\hat{A}'\hat{\Omega}^{-1}\hat{A}\otimes\sum_{t=1}^{T}Z_{1t}Z_{1t}'\right]H\right]^{-1}\\
&\times H'\sum_{t=1}^{T}\mathrm{vec}\left(Z_{1t}\varepsilon_t'\Omega^{-1\prime}A\right)+o_p(1).
\end{aligned}
$$

which by the consistency of $\hat{A}$, $\hat{C}$ and $\hat{\Omega}$ shows that

$$
\begin{aligned}
T\,\mathrm{vec}(\dot{B}-B)\;\xrightarrow{w}\;& H\left[H'\left[A'\Omega^{-1}A\otimes B_\perp\int_0^1 G(u)G'(u)du\,B_\perp'\right]H\right]^{-1}\\
&\times H'\mathrm{vec}\left(\int G(u)dW\Omega^{-1\prime}A\right).
\end{aligned}
$$

which is a mixed Gaussian distribution. Similarly

$$
\begin{aligned}
T^{1/2}\mathrm{vec}(\hat{A}-A,\hat{C}-C)\;=\;& G\left[G'\left[T^{-1}\sum_{t=1}^{T}\begin{pmatrix}\dot{B}'Z_{1t}Z_{1t}'\dot{B} & \dot{B}'Z_{1t}Z_{2t}'\\[4pt] Z_{2t}Z_{1t}'\dot{B} & Z_{2t}Z_{2t}'\end{pmatrix}\otimes\hat{\Omega}^{-1}\right]G\right]^{-1}\\
&\times G'\mathrm{vec}\left(\hat{\Omega}^{-1}T^{-1/2}\sum_{t=1}^{T}\varepsilon_t(Z_{1t}'\dot{B},Z_{2t}')\right).
\end{aligned}
$$

which asymptotically has a Gaussian distribution. The case with a varying $\Omega(t)$ leads to the same results, although the expressions have a more complicated structure.

From these results it follows by arguments similar to the ones of Johansen (1996, theorems 13.7, 13.9), that the likelihood ratio test has an asymptotically $\chi^2$ distribution, for hypotheses that can be formulated as linear restrictions.

## 2.5. Empirical Analysis of the US Term Structure of Interest Rates

In this section we analyze the US term structure using the structural change model we developed in Section 2.

### 2.5.1. The Expectations Hypothesis

A version of the term structure of interest rates is that the expected future spot rates equals the future rate plus a time-invariant term premium. We adopt the notation from Campbell. Lo, and Mackinlay (1997) and let $p_{n,t}$ denote the log of the price of a unit-par-value discount bond at date $t$, with $n$ periods to maturity. The continuously compounded yield to maturity for an $n$ period bond is defined as $y_{n,t} = -\frac{1}{n}p_{n,t}$, and the one-period future rate of return, earned from period $t + n$ to $t + n + 1$, (known at time $t$) is given by $1 + F_{n,t} = P_{n,t}/P_{n+1,t}$, such that

$$f_{n,t} = \log(1 + F_{n,t}) = p_{n,t} - p_{n+1,t}.$$

The expectations hypothesis[5] states that

$$f_{n,t} = E_t(y_{1,t+n}) + \Lambda_n.$$

where $\Lambda_n$ is the term premium. The restriction imposed by the expectations hypothesis is that the term premium does not depend on $t$. From the Fisher-Hicks relation $y_{nt} = n^{-1}\sum_{j=0}^{n-1} f_{jt}$, $n = 1, 2, \ldots$, and the identity $E_t(y_{1,t+j}) = \sum_{i=1}^{j} E_t(\Delta y_{1,t+i}) + y_{1,t}$, we obtain

$$y_{nt} - y_{1t} = n^{-1}\sum_{j=1}^{n-1}\sum_{i=1}^{j} E_t(\Delta y_{1,t+i}) + L_n. \tag{2.5.1}$$

where $L_n = n^{-1}\sum_{j=0}^{n-1} \Lambda_j$. This equation shows that if $y_{1t}$ is $I(1)$, such that the terms $\Delta y_{1,t}$ and $n^{-1}\sum_{j=1}^{n-1}\sum_{i=1}^{j} E_t(\Delta y_{1,t+i})$ are stationary[6], then $y_{nt}$ must be integrated of order one and

---

[5] For an overview of the expectations hypothesis theory and empirical studies of interest rates, see Shiller (1990).
[6] The stationarity of $E_t(\Delta y_{1,t+j})$ does not hold in general, but will hold for time-homogeneous processes. In particular it will hold for the vector autoregressive process we consider in this paper.

$y_{nt}$ and $y_{1t}$ are cointegrated with cointegration vector $(1, -1)$ as first analyzed by Campbell and Shiller (1987). Since the relationship will hold for any integer $n$, any pair of yields to maturity will be cointegrated with cointegration vector $(1, -1)$. We shall call this implication the long-run implication of the expectations hypothesis. This is only one of several implications of the expectations hypothesis. Equation (2.5.1) is the motivation for modelling interest rates as cointegrated processes, and illustrates the convenience of using this framework to test the long-run implication.

The implications of the expectations hypothesis are commonly rejected when tested on US term structure data; this is also the case for the long-run implication as concluded by Hall, Anderson, and Granger (1992), Engsted and Tanggaard (1994), Johnson (1994), and Pagan, Hall, and Martin (1996). Hall, Anderson, and Granger (1992) and Engsted and Tanggaard (1994) attributed their rejection to the unstable period for interest rates between September 1979 and October 1982, when the Fed did not target short interest rates directly. This period is also known as the period with the *nonborrowed reserves operating procedure*. Pagan, Hall, and Martin (1996) gave another possible explanation for the rejection. They extended the cointegration model with a parameter, $\gamma$, for the elasticity of volatility with respect to the level of the shortest interest rate. With simulations, they showed that hypothesis tests on cointegration vectors over-reject as $\gamma$ increases, and found the effect to be substantial as $\gamma$ increases beyond 0.5.

Whereas the expectations hypothesis has been rejected by most studies of US data (see Shiller (1990) for an overview), the results from studies of the term structure in other countries are mixed. Hardouvelis (1994) rejected the expectations hypothesis in 5 of the G7 countries. Cuthbertson (1996) found some evidence in favor of the expectations hypothesis using UK interbank rates and Engsted and Tanggaard (1995) found the long-run implications to hold for Danish data in the period where the central bank targeted interest rates.

## 2.5.2. Structural Changes in the US Term Structure of Interest Rates

There are several studies that find evidence of a structural change in the US term structure of interest rates. Hamilton (1988) applied a Markov switching model to 3- and 12-month T-bills. and the model detected a period that precisely coincides with the period with the *nonborrowed reserves operating procedure* as a separate regime. H. Hansen and Johansen (1999) have developed a recursive estimation of the cointegrated vector autoregressive model to detect structural changes. Their application to US data also indicates structural changes around the fall of 1979 and the fall of 1982.

Structural changes of US interest rates have also been analyzed within the framework of continuous time models. Chan, Karolyi, Longstaff, and Sanders (1992) estimated a diffusion process for the short term interest rate and rejected a structural shift in October 1979, and then estimated the elasticity of volatility to be 1.5. However Bliss and Smith (1998) found significant structural changes when the possibility of a structural shift by the end of 1982 is included in the analysis. They found evidence of structural changes in both 1979 as well as in 1982 when the Fed reversed to target the Fed funds rate. After these changes are accounted for, an elasticity as low as 0.5 is consistent with their data.

These studies have shown that the US term structure has had structural changes, and it is not surprising that these changes affect point estimates and inference.

Elliott (1998b) showed how standard inference can be misleading when there is a root close to unity. Using this local-to-unity approach, Lanne (1999) rejected the expectation hypothesis for US data in the period 1952:1–1991:2. However, after accounting for a structural change in 1979:10 the hypothesis could not be rejected.

In this paper, interest rates are modelled as $I(1)$ variables[7]. The fact that nominal interest rates cannot be negative and other considerations are strong arguments against interest rates being $I(1)$

---

[7] Ait-Sahalia (1996) found the short interest rates to behave as an $I(1)$ process within the band [4%, 18%] and a theoretical model in which interest rates are similar to a random walk is given by DenHaan (1995).

forever. Nevertheless, interest rates may very well be $I(1)$ in a particular sample period. Whenever this is the case, modelling interest rates as $I(1)$ is equivalent to invoking asymptotic results to finite samples. The parallel is that the sample in which interest rates behaved as $I(1)$ need to be long enough for asymptotic results of the $I(1)$ model to be valid, and that any constraint that may prevent interest rates from being $I(1)$ has had no relevance in the sample period analyzed. See Pagan, Hall, and Martin (1996) for another argument on this matter.

## 2.5.3. Data

The term structure data were extracted from the Bliss data[8] that are interpolated by the McCulloch cubic-spline method. This is the same technique as the one used to create the widely used data sets from McCulloch (1990) and McCulloch and Kwon (1993). However the Bliss data differs by not being tax adjusted.

The data used in the empirical analysis are monthly US zero-coupon yields with maturities of 1, 3, 6, 9, 12, 60, and 84 months[9] within the sample period 1970:1 - 1995:12. The yields are stacked in the vector $X_t$, ordered such that the first element in $X_t$ is the 1-month interest rate at time $t$. The most general model can be expressed as

$$\Delta X_t = \alpha(t)\beta(t)'X_{t-1} + \sum_{i=1}^{k-1}\Gamma_i\Delta X_{t-i} + \mu(t) + \varepsilon_t.$$

where $\alpha(t)$, $\beta(t)$ and $\mu(t)$ are piecewise constant with two change points: in 1979:10 and in 1982:10. To avoid a deterministic trend in the yields, the constant is restricted by $\mu(t) = \alpha(t)\rho(t)$, so the model can be rewritten as

$$\Delta X_t = \alpha(t)\beta(t)^{*'}X_{t-1}^{*} + \sum_{i=1}^{k-1}\Gamma_i\Delta X_{t-i} + \varepsilon_t.$$

---

[8] The data were provided to me by David Marshall, (see Bekaert, Hodrick, and Marshall (1997)). Interested parties are referred to Robert R. Bliss: rbliss@gsbalum.uchicago.edu.
[9] Longer maturities were not selected because precise estimate of these are difficult to obtain by interpolation techniques. See Bliss (1997)

where $X_t^{*\prime} = (X_t', 1)$ and $\beta^{*\prime} = (\beta(t)', \rho(t))$.

We may normalize the cointegration relations by

$$
\beta(t)^* = \begin{pmatrix}
\beta_{11,t} & \beta_{12,t} & \cdots & \beta_{1r,t} \\
-1 & 0 & & 0 \\
0 & -1 & & \\
\vdots & & \ddots & \\
0 & & & -1 \\
\rho_{1,t} & \rho_{2,t} & \cdots & \rho_{r,t}
\end{pmatrix}.
\tag{2.5.2}
$$

Since these relations define the stationary relations. the long-run implications of the expectations hypothesis - that the spreads $y_{n,t} - y_{1,t}$ are stationary - can be formulated as the parameter restrictions $\beta_{11,t} = \cdots = \beta_{1r,t} = 1$.

The individual cointegration relations in equation (2.5.2) can be written as

$$
b_{n,t} y_{1,t} - y_{n,t} + \rho_{n,t}. \quad n = 3.6.9.12.60.84.
\tag{2.5.3}
$$

where the maturities $n = 3.\ 6.\ 9.\ 12.\ 60.\ 84$ and $b_{n,t}$ correspond to $i = 1 \ldots r$ and $\beta_{1i,t}$ in equation (2.5.2). The Granger representation shows that $E(b_{n,t} y_{1,t} - y_{n,t} + \rho_{n,t}) = 0$. so $\hat{\rho}_{n,t}$ can be interpreted as the estimated term premia when $b_{n,t}$ is set to unity.

## 2.5.4. Estimation Results

The lag length was set to two using Akaike's and Hannan-Quinn's information criteria. The cointegration rank is set at six ($r = 6$) as predicted by the expectations hypothesis and as the existing literature has supported. No formal test was applied for this selection.

Table 2.5.1 shows that the covariance matrix clearly differs between the three subsamples. The variance estimates from the three subsamples are given in Table 2.5.2.

$$\Delta X_t - \alpha(t)\beta^*(t)'X_{t-1}^* - \Gamma_1\Delta X_{t-1} \sim N(0.\Omega(t))$$

| Model | $\max \log L(\alpha(t),\beta^*(t),\Gamma_1,\Omega(t))$ | Degrees of freedom | LR($M_i|M_0$) (p-value) |
|---|---|---|---|
| $M_0$: $\Omega(t)$ | 2009.25 | 295 | – |
| $M_1$: $\Omega_1 = \Omega_3$ | 1824.94 | 270 | 368.61 (0.0000) |
| $M_2$: $\Omega_1 = \Omega_2 = \Omega_3$ | 1631.77 | 239 | 754.96 (0.0000) |

Table 2.5.1: The maximum value of the likelihood function for the model with changing reduced rank parameters. and changing covariance $\Omega_t$.

The Estimated Covariance Matrices. $\Omega(t)$

1970:3–1979:9   $\hat{\Omega}_1 =$

$$\begin{pmatrix}
0.30 & 0.28 & 0.25 & 0.22 & 0.14 & 0.10 & 0.09 \\
0.28 & 0.27 & 0.25 & 0.22 & 0.15 & 0.11 & 0.10 \\
0.25 & 0.25 & 0.25 & 0.23 & 0.17 & 0.12 & 0.11 \\
0.22 & 0.22 & 0.23 & 0.23 & 0.17 & 0.13 & 0.11 \\
0.14 & 0.15 & 0.17 & 0.17 & 0.15 & 0.12 & 0.11 \\
0.10 & 0.11 & 0.12 & 0.13 & 0.12 & 0.10 & 0.09 \\
0.09 & 0.10 & 0.11 & 0.11 & 0.11 & 0.09 & 0.08
\end{pmatrix}$$

1979:10–1982:10   $\hat{\Omega}_2 =$

$$\begin{pmatrix}
1.75 & 1.68 & 1.51 & 1.28 & 0.92 & 0.63 & 0.54 \\
1.68 & 1.70 & 1.58 & 1.33 & 0.97 & 0.68 & 0.59 \\
1.51 & 1.58 & 1.50 & 1.30 & 0.97 & 0.69 & 0.61 \\
1.28 & 1.33 & 1.30 & 1.18 & 0.90 & 0.65 & 0.57 \\
0.92 & 0.97 & 0.97 & 0.90 & 0.72 & 0.54 & 0.48 \\
0.63 & 0.68 & 0.69 & 0.65 & 0.54 & 0.43 & 0.39 \\
0.54 & 0.59 & 0.61 & 0.57 & 0.48 & 0.39 & 0.35
\end{pmatrix}$$

1982:11–1995:12   $\hat{\Omega}_3 =$

$$\begin{pmatrix}
0.10 & 0.09 & 0.08 & 0.07 & 0.07 & 0.06 & 0.05 \\
0.09 & 0.09 & 0.09 & 0.09 & 0.09 & 0.08 & 0.07 \\
0.08 & 0.09 & 0.10 & 0.11 & 0.11 & 0.10 & 0.09 \\
0.07 & 0.09 & 0.11 & 0.12 & 0.12 & 0.11 & 0.11 \\
0.07 & 0.09 & 0.11 & 0.12 & 0.13 & 0.13 & 0.12 \\
0.06 & 0.08 & 0.10 & 0.11 & 0.13 & 0.13 & 0.13 \\
0.05 & 0.07 & 0.09 & 0.11 & 0.12 & 0.13 & 0.12
\end{pmatrix}$$

Table 2.5.2: The estimated covariance matrices, $\hat{\Omega}_j$, $j = 1.2.3$. from the most general change model.

It is not surprising that the variance of interest rates (see Table 2.5.2) were much higher in the 1979–1982 subsample when the Fed did not target interest rates directly. One conclusion from Table 2.5.1 is that the difference between the variance of interest rates in the first and third subsample is significant. From Table 2.5.2 it can be seen that the major difference between the covariance matrix in the first and last subsample is the reduced volatility of the interest rates with shorter maturities. This phenomenon may be explained by the less frequent adjustments of the Fed's target of the Fed's fund rate in the most recent sample, along with fact that the Fed now publicly announces what their target is.

Six models with different parameter restrictions were estimated[10]. The estimations results are given in Tables 2.5.3 and 2.5.4.

Model 1 in Table 2.5.3 is the most general model, where the parameters are left unrestricted. This model can be represented by the equation

$$\Delta X_t = \alpha(t)\left[\beta(t)'X_{t-1} + \rho(t)\right] + \Gamma_1 \Delta X_{t-1} + \varepsilon_t. \quad t = 1,\ldots,T.$$

$$\varepsilon_t \sim N(0,\Omega(t)).$$

where the parameters are constant within each subsample. i.e. $\alpha(t) = \alpha_1$ for $t \leq$ 1979:09. $\alpha(t) = \alpha_2$ for 1979:10 $\leq t \leq$ 1982:10 and $\alpha(t) = \alpha_3$ for $t \geq$ 1982:11, and similarly for $\beta(t)$. $\rho(t)$ and $\Omega(t)$. The long-run implication of the expectations hypothesis requires $b_n = 1$ for $n = 3, 6, 9, 12, 60$. and 84. The point estimates differ from unity by being systematically too small in the two first subsamples and too large in the last subsample.

In Model 2 the long-run implication of the expectations hypothesis is imposed as the parameter restriction $b_n = 1$ for all $n$ in all subsamples, whereas term premia $(\rho_n)$ adjustment coefficients $(\alpha_i. i = 1, 2, 3)$ as well as the covariance may differ across subsamples. This model can be written

---

[10]The empirical analysis was performed in Gauss. Code and documentation can be obtained by contacting the author.

| 1: Unrestricted Change Model $\alpha(t)$, $\beta(t)$, $\rho(t)$, $\Omega(t)$ | | | | 2 log L 4018.49 | #f 295 | LR – | p-value – |
|---|---|---|---|---|---|---|---|
| | n | 3 | 6 | 9 | 12 | 60 | 84 |
| 1970:3–1979:9 | $b_n$ | 0.9831 | 0.9767 | 0.9162 | 0.7473 | 0.6154 | 0.5947 |
| | $\rho_n$ | 0.3634 | 0.6356 | 1.1666 | 2.4113 | 3.4517 | 3.6640 |
| 1979:10–1982:10 | $b_n$ | 0.9234 | 0.8455 | 0.7716 | 0.7378 | 0.7179 | 0.6765 |
| | $\rho_n$ | 1.4726 | 2.5655 | 3.5156 | 3.8391 | 3.9931 | 4.4702 |
| 1982:11–1995:12 | $b_n$ | 1.0746 | 1.1391 | 1.2596 | 1.5328 | 1.7390 | 1.7989 |
| | $\rho_n$ | -0.2384 | -0.4607 | -0.9011 | -2.0401 | -2.8354 | -3.0585 |

| 2: Expectations Hypothesis $\alpha(t)$, $\beta(t) = \beta$, $\rho(t)$, $\Omega(t)$ | | | | 2 log L 3989.58 | #f 277 | LR 28.91 | p-value 0.0495 |
|---|---|---|---|---|---|---|---|
| | n | 3 | 6 | 9 | 12 | 60 | 84 |
| 1970:3–1979:9 | $b_n$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | $\rho_n$ | 0.2620 | 0.4935 | 0.6592 | 0.8935 | 1.1475 | 1.2357 |
| 1979:10–1982:10 | $b_n$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | $\rho_n$ | 0.6309 | 0.8628 | 0.9917 | 0.9370 | 0.8637 | 0.8800 |
| 1982:11–1995:12 | $b_n$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | $\rho_n$ | 0.2106 | 0.3694 | 0.6307 | 1.0520 | 1.3919 | 1.5010 |

| 3: Constant $\alpha_\perp$ & Expectations Hypothesis $\alpha(t) = \alpha\phi(t)$, $\beta(t) = \beta$, $\rho(t)$, $\Omega(t)$ | | | | 2 log L 3978.44 | #f 265 | LR 40.05 | p-value 0.1038 |
|---|---|---|---|---|---|---|---|
| | n | 3 | 6 | 9 | 12 | 60 | 84 |
| 1970:3–1979:9 | $b_n$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | $\rho_n$ | 0.2644 | 0.4999 | 0.6748 | 0.9221 | 1.1861 | 1.2753 |
| 1979:10–1982:10 | $b_n$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | $\rho_n$ | 0.6529 | 0.9089 | 1.0495 | 0.9896 | 0.9065 | 0.9281 |
| 1982:11–1995:12 | $b_n$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | $\rho_n$ | 0.2123 | 0.3753 | 0.6523 | 1.1248 | 1.5229 | 1.6487 |

| 4: Constant $\alpha$ & $\beta$ & EH. $\rho(t)$ may change. $\alpha(t) = \alpha$, $\beta(t) = \beta$, $\rho(t)$, $\Omega(t)$ | | | | 2 log L 3784.01 | #f 199 | LR 234.48 | p-value 0.0000 |
|---|---|---|---|---|---|---|---|
| | n | 3 | 6 | 9 | 12 | 60 | 84 |
| 1970:3–1979:9 | $b_n$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | $\rho_n$ | 0.2701 | 0.5061 | 0.6798 | 0.9381 | 1.2343 | 1.3332 |
| 1979:10–1982:10 | $b_n$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | $\rho_n$ | 0.5850 | 0.8015 | 0.9598 | 1.2261 | 1.4309 | 1.5107 |
| 1982:11–1995:12 | $b_n$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | $\rho_n$ | 0.2182 | 0.3826 | 0.6621 | 1.1599 | 1.5995 | 1.7405 |

Table 2.5.3: Estimation results: For each model we report the maximum value of the likelihood function, the model's degrees of freedom and the test statistic (tested against the most general model) with the correspondings $p$-value. The cointegration parameters $b_n$ and term premia $\rho_n$ from the cointegration relations $b_n y_{1,t} - y_{n,t} + \rho_n$ are reported for each model and subsample.

as

$$\Delta X_t = \alpha(t)\left[\beta'X_{t-1} + \rho(t)\right] + \Gamma_1\Delta X_{t-1} + \varepsilon_t, \quad t = 1,\ldots,T,$$

$$\varepsilon_t \sim N(0,\Omega(t)).$$

where $\beta$ has the structure required by the long-run implications. The likelihood ratio test of Model 2 against Model 1, has a $p$-value of 4.95%. This shows that there is not strong evidence against the long-run implication once structural changes in the parameters are accounted for.

Model 3 is a more parsimonious model where in addition to the restrictions in Model 2, the adjustment coefficients are required to span the same subspace, $\alpha(t) = \alpha \cdot o(t)$, where $o(t)$ is a full rank $r \times r$ matrix. This model can be written as

$$\Delta X_t = \alpha\phi(t)\left[\beta'X_{t-1} + \rho(t)\right] + \Gamma_1\Delta X_{t-1} + \varepsilon_t, \quad t = 1,\ldots,T,$$

$$\varepsilon_t \sim N(0,\Omega(t)).$$

The restriction implies that the orthogonal complement to $\alpha$ is constant, i.e. $\alpha_\perp(t) = \alpha_\perp$. The different strength of the adjustments between the three subsamples are expressed in terms of the matrix $o(t)$.

Recall the Granger representation from equations (2.4.1) and (2.4.2), and here extended with a third subsample:

$$X_t = C\sum_{i=1}^{t}\varepsilon_i + O_p(1), \quad t = 1,\ldots,T_1.$$

$$X_t = D\sum_{i=T_1+1}^{t}\varepsilon_i + D\Gamma C\sum_{i=1}^{T_1}\varepsilon_i + O_p(1) \quad t = T_1 + 1,\ldots,T_2.$$

$$X_t = E\sum_{i=T_2+1}^{t}\varepsilon_i + E\Gamma D\sum_{i=T_1+1}^{T_2}\varepsilon_i + E\Gamma D\Gamma C\sum_{i=1}^{T_1}\varepsilon_i + O_p(1).$$

$$t = T_2 + 1,\ldots,T.$$

An implication of the constancy of $\alpha_\perp$ and $\beta$ and $\Gamma_1$ is that the loading matrix is constant, i.e. $C = D = E = \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp$. This simplifies the Granger representation to a single equation given by

$$X_t = C \sum_{i=1}^{t} \varepsilon_i + O_p(1), \quad t = 1, \ldots, T.$$

using the fact that $CTC = C$.

The term $\alpha'_\perp \sum_{i=1}^{t} \varepsilon_i$ is called the common stochastic trend in $X_t$, because it describes the random walk element of $X_t$, and $C\bar{\alpha}_\perp$ defines how the stochastic trend is loaded into the process $X_t$. (note $C\bar{\alpha}_\perp\alpha'_\perp = C$). Thus the non-rejection of Model 3 (a $p$-value of 10.38% when tested against Model 1) can be interpreted as follows: The long-run implications are consistent with the data and we cannot reject that the common stochastic trend has been a constant linear combination of $\varepsilon_t$, and we cannot reject that the loading of the common stochastic trend has been constant. The non-constancy of the common stochastic trend comes from the changing variance of $\varepsilon_t$.

The last model in Table 2.5.3, Model 4, can be expressed as

$$\Delta X_t = \alpha \left[\beta' X_{t-1} + \rho(t)\right] + \Gamma_1 \Delta X_{t-1} + \varepsilon_t, \quad t = 1, \ldots, T.$$

$$\varepsilon_t \sim N(0, \Omega(t)).$$

In this model the adjustment coefficients have the same strength in the three subsamples. This is equivalent to the additional restriction: $\rho(t) = \rho$ on Model 3. This model is clearly inconsistent with the term structure data. The fact that the strength of the adjustments are non-constant is not puzzling, since the changes appear along with changes in volatility and term premia.

Thus, we find the term structure to have had structural changes in the covariance $\Omega(t)$ and the term premia $\rho(t)$ along with changes in the strength of the adjustments to dis-equilibria in the cointegration relations. However fundamentals such as the common stochastic trend and stable relationships between interest rates have remained relatively unchanged in the sample analyzed.

These findings are consistent with many of the suggestions that have been offered to explain

the rejection of the expectations hypothesis. The monetary changes in the fall of 1979 and the fall of 1982 had an important impact on the stochastic properties of interest rates. If the structural changes are not accounted for, the result can be incorrect inference, and a possible rejection of a true hypothesis, as was suggested by Hall, Anderson, and Granger (1992) and Engsted and Tanggaard (1994). The suggestion by Tzavalis and Wickens (1997) of a time varying term premium, is also consistent with the results, since we find $\rho(t)$ to vary as the volatility of interest rates changes. Finally, my finding of a changing variance is likely to distort hypothesis testing if not accounted for, which is similar to the volatility effect found by Pagan, Hall, and Martin (1996).

| 5: No Structural Changes $\alpha(t) = \alpha$, $\beta(t) = \beta$, $\rho(t) = \rho$, $\Omega(t) = \Omega$ | | | | $2\log L$ $2852$ | $\#f$ $131$ | LR $-$ | $p$-value $-$ |
|---|---|---|---|---|---|---|---|
| | n | 3 | 6 | 9 | 12 | 60 | 84 |
| 1970:3–1995:12 $b_n$ | | 1.0390 | 1.0417 | 1.0520 | 1.0529 | 1.0239 | 1.0191 |
| $\rho_n$ | | 0.0011 | 0.1680 | 0.2951 | 0.6209 | 1.1478 | 1.2875 |

| 6: No Changes & Expectations Hypothesis $\alpha(t) = \alpha$, $\beta(t) = H\phi$, $\rho(t) = \rho$, $\Omega(t) = \Omega$ | | | | $2\log L$ $2825$ | $\#f$ $125$ | $\mathcal{LR}$ $26.84$ | $p$-value $0.0002$ |
|---|---|---|---|---|---|---|---|
| | n | 3 | 6 | 9 | 12 | 60 | 84 |
| 1970:3–1995:12 $b_n$ | | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $\rho_n$ | | 0.2719 | 0.4570 | 0.6561 | 0.9888 | 1.3148 | 1.4215 |

Table 2.5.4: Estimation results. Testing the expectations hypothesis in the cointegrated VAR without structural changes. Note that the $p$-value is invalid because model 5 is strongly rejected against model 1.

The fifth and sixth models in Table 2.5.4 replicate previous empirical studies of the US term structure, by having constant parameters. Model 5 is the unrestricted model (with constant parameters) and Model 6 is the submodel in which the long-run implication of the expectations hypothesis is imposed. A test of Model 6 against Model 5 would have lead to a weak rejection of the expectations hypothesis, exactly as previous studies have concluded. Of course, this inference is invalid because model 5 is inconsistent with the data. The LR test statistic of Model 5 against Model 1 is 1166. Its distribution is asymptotically $\chi^2$ with 164 degrees of freedom, and is therefore clearly rejected.

## 2.6. Conclusion

This paper shows how structural changes in cointegrated processes can be formulated in a unified framework, using the familiar vector autoregressive model. It is possible to formulate and test various structural changes as simple parameter restrictions in this framework. Moreover, the parameters can be estimated under these restrictions with the new generalized reduced rank regression technique. This technique is also applicable to estimation problems unrelated to structural changes.

I derived the likelihood ratio test for structural changes occurring at known points in time, and showed that it is asymptotically $\chi^2$. Moreover, it was shown how hypotheses can be tested, when the maintained hypothesis is presence of structural changes. I derived the asymptotic distributions of the parameter estimates and likelihood ratio tests. Similar to the standard model without structural changes, the estimate of the cointegration relations is super-consistent and asymptotically mixed Gaussian, and the LR statistic is asymptotically $\chi^2$.

This combination of cointegration and structural changes may provide a fruitful framework for many economic questions of interest. In this paper I analyzed the US term structure and found evidence of structural changes that coincide with the Fed's policy changes in September 1979 and October 1982. Contrary to previous studies (see Hall, Anderson, and Granger (1992), Engsted and Tanggaard (1994), or Pagan, Hall, and Martin (1996)) I cannot reject the long-run implications of the expectations hypothesis, once these structural changes are accounted for. In fact, a parsimonious model is consistent with the data. This model has a different covariance structure in the three monetary regimes, and along with changes in the covariance matrix, only the term premia and the strength of adjustment coefficients changes.

In this paper, the cointegration rank was taken as given. Although this is reasonable when interest rates are analyzed, this need not always be the case. A formal test to determine the rank of cointegrated processes is derived in Chapter 4..

# Appendix B: Proofs

## B.1. Algebraic Treatment of the Reduced Rank Regression

Before I give the proofs of Theorem 2.3.1 I derive some intermediate results. The following lemma is a consequence of Poincaré's theorem, however, a direct proof is presented here.

**Lemma B.1.** *The function* $g(y) = |y'\Lambda y| / |y'y|$ *where* $y$ *is a* $p \times r$ *matrix,* $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$ *and* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ *has maximum value* $\prod_{i=1}^{r} \lambda_i$ *which is attained with* $y$ *equal to the first* $r$ *unit vectors, that is* $y = (I_r . 0_{r \times p-r})'$.

**Proof.** Let $J$ be an index set $J \subset \{1, \ldots, p\}$ of cardinality $r$, and define the $r \times r$ matrices $y_J$ and $\Lambda_J$ by $y_J = \{y_{ij}\}_{i \in J, j=1,\ldots,r}$ and $\Lambda_J = \{\Lambda_{ij}\}_{i,j \in J}$. So if $p = 3$, $r = 2$ and $J = \{1, 2\}$ we would have $y_J = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}$ and $\Lambda_J = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$.

Next, let $\mathbb{D}_p^r$ denote the set of all subsets of $\{1, \ldots, p\}$ containing exactly $r$ different elements (cardinality $r$). Below, I prove that

$$|y'\Lambda y| = \sum_{J \in \mathbb{D}_p^r} |y_J' \Lambda_J y_J| = \sum_{J \in \mathbb{D}_p^r} |y_J' y_J| \prod_{i \in J} \lambda_i = \sum_{J \in \mathbb{D}_p^r} |y_J|^2 \prod_{i \in J} \lambda_i. \tag{B.1}$$

So $g(y) = |y'\Lambda y| / |y'y| = \sum_{J \in \mathbb{D}_p^r} |y_J|^2 \prod_{i \in J} \lambda_i / \sum_{J \in \mathbb{D}_p^r} |y_J|^2$ is a convex combination over the elements in $\mathbb{D}_p^r$ with values given by $\prod_{i \in J} \lambda_i$, with the largest element being $\prod_{i=1}^{r} \lambda_i$ corresponding to $J = \{1, \ldots, r\}$. This value can be obtained with $\hat{y} = (I_r . 0_{r \times p-r})'$ which therefore maximizes the function $g(y)$.

The identity (B.1) is proved as follows. The second and third equality follows trivially from $|AB| = |A| |B|$ for matrices of proper dimensions, whereas the first equality is showed by induction

below. The equality trivially holds for $r = 1$ or $p = r$. So the scheme

| $p\backslash r$ | 1 | 2 | 3 | 4 | $\cdots$ |
|---|---|---|---|---|---|
| 1 | ✓ | – | – | – | |
| 2 | ✓ | ✓ | – | – | |
| 3 | ✓ | ? | ✓ | – | |
| 4 | ✓ | ? | ? | ✓ | |
| $\vdots$ | $\vdots$ | | | | $\ddots$ |

shows that the equality can be proven by showing it holds for cell $(p, r)$ when it is assume that it holds for cell $(p - 1, r - 1)$, say assumption $(A1)$, and for cell $(p - 1, r)$, say assumption $(A2)$.

Define $\bar{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_{p-1})$ and consider first the case where the last row of $y$ is a zero-row $(y_{p1}, \ldots, y_{pr}) = (0, \ldots, 0)$. Define in this case $\bar{y} = \{y_{ij}\}_{i=1,\ldots,p-1}$, that is $y$ without the zero-row. By applying assumption $(A2)$ we have the relation

$$
\begin{aligned}
|y'\Lambda y| &= |\bar{y}'\bar{\Lambda}\bar{y}| = \sum_{J \in \mathfrak{D}_{p-1}^r} |y'_J y_J| \cdot \Pi_{i \in J} \lambda_i \\
&= \sum_{J \in \mathfrak{D}_p^r, p \notin J} |y'_J y_J| \cdot \Pi_{i \in J} \lambda_i + \underbrace{\sum_{J \in \mathfrak{D}_p^r, p \in J} |y'_J y_J| \cdot \Pi_{i \in J} \lambda_i}_{=0} = \sum_{J \in \mathfrak{D}_p^r} |y'_J y_J| \cdot \Pi_{i \in J} \lambda_i
\end{aligned}
$$

which proves the lemma in this case.

Next assume that $(y_{p1}, \ldots, y_{pr}) \neq 0$, and choose a full rank $r \times r$-matrix $Q$, so that $(y_{p1}, \ldots, y_{pr})Q = (0, \ldots, 0, 1)$ Then define the $p - 1 \times r - 1$ matrix $\bar{z}$ as the first $r - 1$ columns of $\bar{y}Q$. Then it holds that

$$
\begin{aligned}
|Q|^2 |y'\Lambda y| &= \left| Q'\bar{y}'\bar{\Lambda}\bar{y}Q + \begin{pmatrix} 0_{r-1 \times r-1} & 0 \\ 0 & \lambda_p \end{pmatrix} \right| \\
&= \left| Q'\bar{y}'\bar{\Lambda}\bar{y}Q \right| + \left| \bar{z}'\bar{\Lambda}\bar{z} \right| \lambda_p.
\end{aligned}
$$ (B.2)

Applying assumption $(A2)$ on the first term of (B.2) we get

$$\left|Q'\bar{y}'\bar{\Lambda}\bar{y}Q\right| = |Q|^2 \sum_{J\in\mathbb{D}^r_{p-1}} |\bar{y}'_J\bar{\Lambda}_J\bar{y}_J| = |Q|^2 \sum_{J\in\mathbb{D}^r_p, p\notin J} |y'_J\Lambda_J y_J|. \tag{B.3}$$

Note that for $J \in \mathbb{D}^{r-1}_{p-1}$ we have that

$$|\tilde{z}_J| = \left|\begin{pmatrix} \tilde{z}_J & 0 \\ 0 & 1 \end{pmatrix}\right| = |y_J Q|, \text{ and } \lambda_p \left|\bar{\Lambda}_J\right| = |\Lambda_{\bar{J}}|$$

where $\bar{J} = \{J \cup \{p\}\} \in \mathbb{D}^r_p$. So applying assumption $(A1)$ to the second term of (B.2) we have

$$\left|\bar{z}'\bar{\Lambda}\bar{z}\right|\lambda_p = \lambda_p = |Q|^2 \sum_{J\in\mathbb{D}^{r-1}_p, p\in J} |y'_J\Lambda_J y_J|. \tag{B.4}$$

Combining the identities (B.2), (B.3), and (B.4) we have shown

$$|Q|^2|y'\Lambda y| = |Q|^2 \sum_{J\in\mathbb{D}^r_p, p\notin J} |y'_J\Lambda_J y_J| + |Q|^2 \sum_{J\in\mathbb{D}^r_p, p\in J} |y'_J\Lambda_J y_J| = |Q|^2 \sum_{J\in\mathbb{D}^r_p} |y'_J\Lambda_J y_J|$$

which completes the proof. ∎

In the proof for Lemma B.1 we obtained a representation for $|y'\Lambda y|$ which we formulate as a separate corollary.

**Corollary B.2.** *Let $\Lambda$ be a real $p \times p$ diagonal matrix, and $y$ a real $p \times r$ matrix, where $r \le p$. Then with the definitions above, we have that*

$$|y'\Lambda y| = \sum_{J\in\mathbb{D}^r_p} |y'_J\Lambda_J y_J| = \sum_{J\in\mathbb{D}^r_p} |y'_J y_J| \Pi_{i\in J}\lambda_i = \sum_{J\in\mathbb{D}^r_p} |y_J|^2 \Pi_{i\in J}\lambda_i.$$

**Lemma B.3.** *Let $x$ be a $p \times r$ matrix, $M$ and $N$ be $p \times p$ symmetric matrices, $M$ positive semi-definite and $N$ positive definite.*

The function $f(x) = |x'Mx| / |x'Nx|$ has $\prod_{i=1}^{r} \lambda_i$ as its maximum with is obtained for $x = (v_1, \ldots, v_r)$ where $v_1, \ldots, v_r$ are eigenvectors corresponding to the $r$ largest eigenvalues, $\lambda_1, \ldots, \lambda_r$ from the eigenvalue problem $|\lambda N - M| = 0$.

**Proof.** The matrix $\left(N^{-\frac{1}{2}} M N^{-\frac{1}{2}}\right)$ is symmetric positive semi-definite, hence we can diagonalize it as $N^{-\frac{1}{2}} M N^{-\frac{1}{2}} = Q\Lambda Q'$ where $QQ' = I$, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$ and $\lambda_1 > \lambda_2 > \cdots > \lambda_p \geq 0$. By defining $V = N^{-\frac{1}{2}}Q$ and $y = V^{-1}x$, we have that $|x'Mx| / |x'Nx| = |y'\Lambda y| / |y'y|$. According to Lemma B.1 this is maximized by $\hat{y} = (I_r, 0)'$, so $f(x)$ is maximized by $\hat{x} = V\hat{y} = N^{-\frac{1}{2}}Q\hat{y}$. ∎

**Proof of Theorem 2.3.1.**

The likelihood function is given by

$$L(\alpha, \beta, \Psi, \Omega) = \prod_{t=1}^{T} ((2\pi)^p |\Omega|)^{-\frac{1}{2}}$$
$$\times \exp\left(-\frac{1}{2}(Z_{0t} - \alpha\beta'Z_{1t} - \Psi Z_{2t})'\Omega^{-1}(Z_{0t} - \alpha\beta'Z_{1t} - \Psi Z_{2t})\right).$$

The estimate of the parameters are found by maximization of the likelihood function, or equivalently by maximization of the logarithm of the likelihood function

$$\log L(\alpha, \beta, \Psi, \Omega) = -\frac{T}{2}|\Omega| - \frac{T}{2}\log(2\pi)^p$$
$$-\frac{1}{2}\sum_{t=1}^{T}(Z_{0t} - \alpha\beta'Z_{1t} - \Psi Z_{2t})'\Omega^{-1}(Z_{0t} - \alpha\beta'Z_{1t} - \Psi Z_{2t}).$$

The maximization is done in three steps. First, we maximize with respect to $\Psi$ taking $\alpha$ and $\beta$ as given, then with respect to $\alpha$ and $\Omega$ taking $\beta$ as given, and finally with respect to $\beta$.

The estimate of $\Psi$, given $\alpha$ and $\beta$, is found by regressing $(Z_{0t} - \alpha\beta'Z_{1t})$ on $Z_{2t}$, with the Gaussian error term, the estimate is found by OLS

$$\hat{C}(\alpha, \beta) = M_{02}M_{22}^{-1} - \alpha\beta'M_{12}M_{22}^{-1}. \tag{B.5}$$

where $M_{ij} = T^{-1}\sum_{t=1}^{T} Z_{it}Z'_{jt}$. The concentrated likelihood function is given by

$$\log L(\alpha, \beta, \Omega) = -\frac{T}{2}|\Omega| - \frac{T}{2}\log(2\pi)^P - \frac{1}{2}\sum_{t=1}^{T}(R_{0t} - \alpha\beta'R_{1t})'\Omega^{-1}(R_{0t} - \alpha\beta'R_{1t}).$$

where the auxiliary residuals ($Z_{0t}$ and $Z_{1t}$ corrected for $Z_{2t}$) are given by $R_{0t} = Z_{0t} - M_{02}M_{22}^{-1}Z_{2t}$ and $R_{1t} = Z_{1t} - M_{12}M_{22}^{-1}Z_{2t}$.

Taking $\beta$ as given, the estimates of $\alpha$ and $\Omega$ are given by

$$\hat{\alpha}(\beta) = S_{01}\beta(\beta'S_{11}\beta)^{-1} \tag{B.6}$$

$$\hat{\Omega}(\beta) = S_{00} - S_{01}\beta(\beta'S_{11}\beta)^{-1}\beta'S_{10}. \tag{B.7}$$

again using that the errors are Gaussian.

What remains is to maximize the concentrated likelihood function with respect to $\beta$. Since

$$T^{-1}\sum_{t=1}^{T}(R_{0t} - \hat{\alpha}(\beta)\beta'R_{1t})'\left(\Omega(\beta)\right)^{-1}(R_{0t} - \hat{\alpha}(\beta)\beta'R_{1t}) = I.$$

the concentrated likelihood is given by

$$L(\beta) = \left((2\pi)^P |\hat{\Omega}(\beta)|\right)^{-\frac{T}{2}}\exp\left(-\frac{1}{2}Tp\right) = \left((2\pi e)^P \left|\hat{\Omega}(\beta)\right|\right)^{-\frac{T}{2}}.$$

So maximizing the likelihood function is equivalent to minimizing

$$|\hat{\Omega}(\beta)| = |S_{00} - S_{01}\beta(\beta'S_{11}\beta)^{-1}\beta'S_{10}| = |S_{00}|\frac{|\beta'(S_{11} - S_{10}S_{00}^{-1}S_{01})\beta|}{|\beta'S_{11}\beta|}.$$

which is solved by choosing the $r$ smallest eigenvalues of $|S_{11}\rho - (S_{11} - S_{10}S_{00}^{-1}S_{01})|$, or be defining $\lambda = 1 - \rho$, choosing the $r$ largest eigenvalues of $|S_{11}\rho - S_{10}S_{00}^{-1}S_{01}|$, which is identical to solve

$\max_\beta \frac{|\beta'(S_{10}S_{00}^{-1}S_{01})\beta|}{|\beta'S_{11}\beta|}$. By Lemma B.3 the estimator is given by

$$\hat{\beta} = (\hat{v}_1, \ldots, \hat{v}_r).$$

where $\hat{\lambda}_i$ and $\hat{v}_i$ are the eigenvalues and eigenvectors to the problem

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0.$$

ordered such that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_p$, and we find

$$|\hat{\Omega}(\hat{\beta})| = |S_{00}| \prod_{i=1}^{r}(1 - \hat{\lambda}_i).$$

Since the eigenvectors are normalized by $(\hat{v}_1, \ldots, \hat{v}_p)'S_{11}(\hat{v}_1, \ldots, \hat{v}_p) = I$, we have $\hat{\beta}'S_{11}\hat{\beta} = I$, such that (B.6) and (B.7) reduces to (2.3.3) and (2.3.4). By inserting these estimates into (B.5) we find (2.3.5). ∎

## B.2. Algebraic Treatment of the Generalized Reduced Rank Regression

Before we can formulate the general estimation result we need some additional notation. Define $Z_0 = (Z_{01}, \ldots, Z_{0T})$, $Z_1 = (Z_{11}, \ldots, Z_{1T})$, $Z_2 = (Z_{21}, \ldots, Z_{2T})$, and $E = (\varepsilon_1, \ldots, \varepsilon_T)$, so that Model 2.2.3 can be expressed as

$$Z_0 = AB'Z_1 + CZ_2 + E. \tag{B.8}$$

Next define

$$\mathbf{Z}_{1B2} = ((Z_1'B, Z_2') \otimes I_p)).$$

$$\mathbf{Z}_{1A} = (Z_1' \otimes A)K_{p_1,r}.$$

where $K_{p_1,r}$ is the commutation matrix, uniquely defined by $K_{p_1,r}\mathrm{vec}(B) \equiv \mathrm{vec}(B')$ for any $p_1 \times r$

matrix $B$. Thus $K_{p_1,r}$ is a $p_1 r \times p_1 r$ matrix consisting of zeros and ones.

Finally let $\varepsilon = \mathrm{vec}(\varepsilon_1, \ldots, \varepsilon_T)$ and set

$$\Sigma = \mathrm{var}(\varepsilon),$$

which is block diagonal under Assumption 2.2.1. The $p \times p$ matrices in the diagonal of $\Sigma$ are given by $\Omega(t)$, $t = 1, \ldots, T$, formally $\Sigma_{p(t-1)+i,p(t-1)+j} = \Omega_{i,j}(t)$ for $i, j = 1, \ldots, p$ and $t = 1, \ldots, T$. Hence $\Sigma^{-1}$ is a block diagonal matrix with $\Omega(t)^{-1}$ as diagonal matrices, $t = 1, \ldots, T$.

**Lemma B.4.** *With the definitions above, we have the relations:*

$$\mathbf{Z}'_{1A}\Sigma^{-1}\mathbf{Z}_{1A} = \sum_{t=1}^{T} \left[ A'\Omega(t)^{-1}A \times Z_{1t}Z'_{1t} \right]. \tag{B.9}$$

$$\mathbf{Z}'_{1A}\Sigma^{-1}\mathrm{vec}(Z_0 - CZ_2) = \sum_{t=1}^{T} \mathrm{vec}\left( Z_{1t}(Z_{0t} - CZ_{2t})'\Omega(t)^{-1'}A \right). \tag{B.10}$$

$$\mathbf{Z}'_{1B2}\Sigma^{-1}\mathbf{Z}_{1B2} = \sum_{t=1}^{T} \left[ \begin{pmatrix} B'Z_{1t}Z'_{1t}B & B'Z_{1t}Z'_{2t} \\ Z_{2t}Z'_{1t}B & Z_{2t}Z'_{2t} \end{pmatrix} \times \Omega(t)^{-1} \right]. \tag{B.11}$$

$$\mathbf{Z}'_{1B2}\Sigma^{-1}\mathrm{vec}(Z_0) = \sum_{t=1}^{T} \mathrm{vec}\left( \Omega(t)^{-1}Z_{0t}(Z'_{1t}B, Z'_{2t}) \right). \tag{B.12}$$

*If $\{\varepsilon_t\}$ is i.i.d. Gaussian with covariance matrix $\Omega$, the expressions simplify to:*

$$\mathbf{Z}'_{1A}\Sigma^{-1}\mathbf{Z}_{1A} = T\left[ A'\Omega^{-1}A \times M_{11} \right].$$

$$\mathbf{Z}'_{1A}\Sigma^{-1}\mathrm{vec}(Z_0 - CZ_2) = T\mathrm{vec}\left( (M_{10} - M_{02}C')\Omega^{-1}A \right).$$

$$\mathbf{Z}'_{1B2}\Sigma^{-1}\mathbf{Z}_{1B2} = T\left[ \begin{pmatrix} B'M_{11}B & B'M_{12} \\ M_{21}B & M_{22} \end{pmatrix} \times \Omega^{-1} \right].$$

$$\mathbf{Z}'_{1B2}\Sigma^{-1}\mathrm{vec}(Z_0) = T\mathrm{vec}\left( \Omega^{-1}(M_{01}B, M_{02}) \right).$$

**Proof.** The identity

$$\mathbf{Z}'_{1A}\Sigma^{-1}\mathbf{Z}_{1A} = K'_{p_1,r}(Z_1 \otimes A')\Sigma^{-1}(Z'_1 \otimes A)K_{p_1,r}$$

$$= K_{r,p_1}\sum_{t=1}^{T}(Z_{1t} \otimes A')\Omega(t)^{-1}(Z'_{1t} \otimes A)K_{p_1,r}$$

$$= K_{r,p_1}\sum_{t=1}^{T}(Z_{1t} \otimes A'\Omega(t)^{-1})(Z'_{1t} \otimes A)K_{p_1,r}$$

$$= K_{r,p_1}\sum_{t=1}^{T}(Z_{1t}Z'_{1t} \otimes A'\Omega(t)^{-1}A)K_{p_1,r}$$

$$= \sum_{t=1}^{T}(A'\Omega(t)^{-1}A \otimes Z_{1t}Z'_{1t}).$$

which proves (B.9). We used that $Z_{1t}$ is a column vector, and that we can write $\Omega(t)^{-1} = 1 \otimes \Omega(t)^{-1}$ and the formula $(M_1 \otimes M_2)(M_3 \otimes M_4) = (M_1 M_3 \otimes M_2 M_4)$ for matrices where the product $M_1 M_3$ and $M_2 M_4$ are well defined.

Next consider

$$\mathbf{Z}'_{1A}\Sigma^{-1}\text{vec}(Z_0 - CZ_2) = K_{r,p_1}\sum_{t=1}^{T}(Z_{1t} \otimes A')\Omega(t)^{-1}(Z_{0t} - CZ_{2t})$$

$$= K_{r,p_1}\sum_{t,\tau=1}^{T}(Z_{1t} \otimes A'\Omega(t)^{-1})\text{vec}(Z_{0t} - CZ_{2t})$$

$$= K_{r,p_1}\sum_{t,\tau=1}^{T}\text{vec}(A'\Omega(t)^{-1}(Z_{0t} - CZ_{2t})Z'_{1t})$$

$$= \sum_{t=1}^{T}\text{vec}(Z_{1t}(Z_{0t} - CZ_{2t})'\Omega(t)^{-1}A).$$

which proves (B.10). Equations (B.11) and (B.12) are proven similarly.

In the situation where $\{\varepsilon_t\}$ is i.i.d.. we have $\Omega(t)^{-1} = \Omega^{-1}$, which proves the last four equations. ∎

**Proof of Theorem 2.3.2.** Applying the vec operation to equation (B.8) yields the equation

$$
\begin{aligned}
\text{vec}(Z_0) &= (Z_1'B \otimes I_p)\text{vec}(A) + (Z_2' \otimes I_p)\text{vec}(C) + \varepsilon \\
&= [(Z_1'B \cdot Z_2') \otimes I_p)]\,\text{vec}(A \cdot C) + \varepsilon \\
&= \mathbf{Z}_{1B2}G\psi + \varepsilon.
\end{aligned}
$$

For fixed values of $B$ and $\Sigma$ this is a restricted GLS problem with the well-known solution given by

$$
\text{vec}(\hat{A}.\hat{C}) = G\left[G'\mathbf{Z}_{1B2}'\Sigma^{-1}\mathbf{Z}_{1B2}G\right]^{-1}G'\mathbf{Z}_{1B2}'\Sigma^{-1}\text{vec}(Z_0).
$$

which by Lemma B.4 simplifies to (2.3.6).

Similarly for fixed $A$. $C$. and $\Sigma$. we have the equation

$$
\begin{aligned}
\text{vec}(Z_0 - CZ_2) &= \text{vec}(AB'Z_1) + \varepsilon \\
&= (Z_1' \otimes A)\text{vec}(B') + \varepsilon \\
&= (Z_1' \otimes A)K_{p_1,r}\text{vec}(B) + \varepsilon \\
&= \mathbf{Z}_{1A}\text{vec}(B) + \varepsilon.
\end{aligned}
$$

This is also a restricted GLS problem. with the solution given by

$$
\text{vec}(\hat{B}) = H\left[H'\mathbf{Z}_{1A}'\Sigma^{-1}\mathbf{Z}_{1A}H\right]^{-1}H'\mathbf{Z}_{1A}'\Sigma^{-1}\text{vec}(Z_0 - CZ_2).
$$

which by Lemma B.4 reduces to (2.3.7). ∎

**Proof of Corollary 2.3.3.** Follows from Theorem 2.3.2 and Lemma B.4. ∎

**Proof of Corollary 2.3.4.** From Theorem 2.3.1, we obtain the equations for $\hat{C}$ and $\hat{\Omega}$. Rather

that handling the remaining estimation for $A$ and $B$ as a GLS problem we can obtain the likelihood

equations directly. The concentrated log-likelihood function is (apart from a constant) given by

$$\log L(A, B) = -\frac{T}{2} \text{tr} \left\{ \Omega^{-1}(S_{00} - AB'S_{10} + AB'S_{11}BA' - S_{01}BA') \right\}$$

holding $\Omega$ fixed. So the derivatives of $A$ and $B$ in the directions $a$ and $b$ are given by

$$
\begin{aligned}
D_A \log L(A, B)(a) &= T \text{tr} \left\{ \Omega^{-1}(S_{01} - AB'S_{11}) Ba' \right\} \\
&= T \left[ \text{tr} \left\{ \Omega^{-1} S_{01} Ba' \right\} - \text{tr} \left\{ I_p A(B'S_{11}B)a' \right\} \right] \\
&= T \text{vec}(a)' \left[ (B' \otimes \Omega^{-1}) \text{vec}(S_{01}) - (B'S_{11}B \otimes I_p) \text{vec}(A) \right].
\end{aligned}
$$

and

$$
\begin{aligned}
D_B \log L(A, B)(b) &= T \text{tr} \left\{ \Omega^{-1}(S_{01} - AB'S_{11}) bA' \right\} \\
&= T \text{tr} \left\{ A'\Omega^{-1}(S_{01} - AB'S_{11}) b \right\} \\
&= T \text{vec}(b)' \left[ (A' \otimes S_{10}) \text{vec}(\Omega^{-1}) - (A'\Omega^{-1}A \otimes S_{11}) \text{vec}(B) \right].
\end{aligned}
$$

using Theorem 3 from Magnus and Neudecker (1988, Chapter 2). So equations (2.3.9) and (2.3.10) are the first order conditions. ∎

**Proof of Corollary 2.3.5.** The result follows directly from Theorem 2.3.1 and Corollary 2.3.4. ∎

## References

AIT-SAHALIA. Y. (1996): "Testing Continuous-Time Models of the Spot Interest Rate." *Review of Financial Studies*. 9, 385–426.

ANDERSON, T. W. (1951): "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *Annals of Mathematical Statistics*. 22, 327–351.

ANDREWS. D. W. K., AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter is Present Only under the Alternative," *Econometrica*. 62, 1383–1414.

BAI. J. (1999): "Likelihood Ratio Test for Multiple Structural Changes." *Journal of Econometrics*. 91. 299–323.

BAI, J.. AND P. PERRON (1998): "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica*, 66, 47-78.

BEKAERT. G., R. J. HODRICK, AND D. A. MARSHALL (1997): "On Biases in Tests of the Expectations Hypohesis of the Term Structure of Interest Rates," *Journal of Financial Economics*, 44, 309-348.

BLISS. R. R. (1997): "Testing Term Structure Estimation Methods," *Advances in Futures and Options Research*, 9, 197-231.

BLISS. R. R.. AND D. C. SMITH (1998): "The Elasticity of Interest Rate Volatility: Chan. Karolyi. Longstaff and Sanders Revisited," Federal Reserve Bank of Atlanta Working Paper. 1997-13a. (March 1998).

BOSWIJK. P. H. (1995): "Identifiability of Cointegrated Systems." *Working paper, Tinbergen Institute.*

CAMPBELL. J. Y., A. W. LO. AND A. C. MACKINLAY (1997): *The Econometrics of Financial Markets.* Princeton University Press. Princeton. New Jersey.

CAMPBELL. J. Y.. AND R. J. SHILLER (1987): "Cointegration and Tests of Present Value Models." *Journal of Political Economy*, 95, 1062-1088.

CHAN. K. C.. G. A. KAROLYI. F. A. LONGSTAFF. AND A. B. SANDERS (1992): "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate." *Journal of Finance.* 47. 1209-1227.

CUTHBERTSON. K. (1996): "The Expectations Hypothesis of the Term Structure: The UK Interbank Market," *Economic Journal*, 106, 578-92.

DENHAAN. W. J. (1995): "The Term Structure of Interest Rates in Real and Monetary Economies," *Journal of Economic Dynamics and Control.* 19, 909-940.

ELLIOTT. G. (1997): "Minimum Distance Methods and Cointegrated Models." *UCSD manuscript.*

——— (1998a): "Estimating Restricted Cointegrating Vectors." *UCSD manuscript.*

——— (1998b): "The Robustness of Cointegration Methods When Regressors Almost Have Unit Roots." *Econometrica*, 66, 149-158.

ENGSTED. T.. AND C. TANGGAARD (1994): "Cointegration and the US Term Structure." *Journal of Banking and Finance,* 18, 167-181.

——— (1995): "The Predictive Power of Yield Spreads for Future Interest Rates: Evidence from the Danish Term Structure." *Scandinavian Journal of Economics.* 97, 149-59.

GREGORY. A. W.. AND B. E. HANSEN (1996): "Residual-Based Test for Cointegration in Models with Regime Shifts," *Journal of Econometrics.* 70, 99-126.

HALL. A. D.. H. M. ANDERSON, AND C. W. J. GRANGER (1992): "A Cointegration Analysis of Treasure Bill Yields," *The Review of Economic And Statistics*, 74, 116-126.

HAMILTON. J. D. (1988): "Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates." *Journal of Economic Dynamics and Control.* 12, 385-423.

HANSEN, H., AND S. JOHANSEN (1999): "Some Tests for Parameter Constancy in Cointegrated VAR-Models." *Econometrics Journal*, 2, 306–333.

HANSEN, P. R. (2000a): "Estimation of Cointegration Models with Heteroskedasticity and Autocorrelation under Parameter Restrictions," *Forthcoming UCSD Department of Economics Discussion Papers.*

——— (2000b): "The Johansen-Granger Representation Theorem: An Explicit Expression for *I*(1) Processes," *Forthcoming UCSD Department of Economics Discussion Papers.*

HARDOUVELIS, G. (1994): "The Term Structure Spread and Future Changes in Long and Short Rates in the G7 Countries." *Journal of Monetary Economics*. 33. 255–83.

HENDRY, D. F. (1995): "A Theory of Co-Breaking," *Working Paper, Institute of Economics and Statistics, Oxford.*

INOUE, A. (1999): "Test for Cointegrating Rank with a Trend-Break," *Journal of Econometrics*, 90, 215–237.

JOHANSEN, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

——— (1996): *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford, 2nd edn.

JOHANSEN, S., AND K. JUSELIUS (1992): "Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK," *Journal of Econometrics*, 53, 211–244.

JOHNSON, P. A. (1994): "On the Number of Common Unit Roots in the Term Structure of Interest Rates," *Applied Economics*, 26, 815–820.

KROLZIG, H.-M. (1996): "Statistical Analysis of Cointegrated VAR Processes with Markovian Regime Shifts," *SFB 373 Discussion Paper 25/1996, Humboldt-University Berlin.*

LANNE, M. (1999): "Near Unit Roots and the Predictive Power of Yield Spreads for Changes in Long-Term Interest Rates," *The Review of Economics and Statistics*, 81, 393–398.

MAGNUS, J. R., AND H. NEUDECKER (1988): *Matrix Differential Calculus with Applications in Statistics and Economics*. Wiley, New York.

McCULLOCH, J. H. (1990): "US Term Structure Data. 1946-1987," in *Handbook of Monetary Economics*, ed. by B. Friedman, and F. Hahn. North-Holland, Amsterdam, vol. 1, pp. 672–715.

McCULLOCH, J. H., AND H. C. KWON (1993): "US Term Structure Data. 1947-1991," *Ohio State University, Working Paper 93-6.*

PAGAN, A. R., A. D. HALL, AND V. MARTIN (1996): "Modelling the Term Structure," *Handbook of Statistics*, 14, 91–118.

SEO, B. (1998): "Tests for Structural Change in Cointegrated Systems," *Econometric Theory*. 14. 222–259.

SHILLER, R. J. (1990): "The Term Structure of Interest Rates," in *Handbook of Monetary Economics*, ed. by B. Friedman, and F. Hahn. North-Holland, Amsterdam, vol. 1, pp. 627–722.

TZAVALIS. E.. AND M. WICKENS (1997): "Explaining the Failures of the Term Spread Models of the Rational Expectations Hypothesis of the Term Structure." *Journal of Money, Credit and Banking*, 29, 364-80.

WHITE, H. (2000): *Asymptotic Theory for Econometricians*. Forthcoming at Academic Press, San Diego, 2nd edn.

# Chapter 3

## Testing for Structural Changes in Cointegrating Relations[†]

### Abstract

This paper derives likelihood-ratio based tests for a structural change in the cointegrating relations. When the potential change point, $\tau$, is known, the likelihood ratio test, $LR_T(\tau)$, is shown to be asymptotically $\chi^2$. Simulations show that the $\chi^2$ distribution is a reasonably good approximation in small samples, except when a stable root is "close" to unity.

The case with an unknown change point can be handled by statistics based on the sequence of tests: $LR_T(\tau)$, $\tau = \tau_0, \ldots, \tau_1$. Tests of this kind include the SupQ test, the MeanQ test, and the ExpQ test. Andrews and Ploberger (1994) showed that the MeanQ and ExpQ tests belong to a class of optimal tests in a setting with stationary variables. In this setting with $I(1)$ variables, simulations show that these tests only dominate the SupQ test against alternatives in the directions where the cointegration parameters are $T^{1/2}$-consistent. In the directions where the cointegration parameters are $T$-consistent, none of the tests clearly dominates the others.

---

71

## 3.1. Introduction

It is now well known that structural changes affect tests to determine the extent of $I(0)$- and $I(1)$-ness in time series, (see Perron (1990), Campos, Ericsson, and Hendry (1996), and Gregory and B. E. Hansen (1996)). Because statistical inference and economic interpretations are associated with variables or relations being $I(0)$ or $I(1)$, it is particularly important to model and test for structural changes in $I(1)$ processes. For an ongoing discussion of this issue, see Nelson and Plosser (1982), Perron (1989), Zivot and Andrews (1992) and Lumsdaine and Papell (1997).

This paper investigates structural changes in the cointegrating relations formulated in the vector autoregressive model of Johansen (1988). This framework is convenient because the maximum likelihood estimators are easily derived under both the null (no structural changes) as well as under the alternative (a structural change at time $\tau$). When the potential change point is taken as given, the likelihood ratio test is shown to be asymptotically $\chi^2$. When the change point is unknown, the testing problem is nonstandard, and I consider the SupQ, MeanQ, and ExpQ statistics, which are the likelihood ratio (LR) version of the tests proposed by B. E. Hansen (1992a), Andrews (1993), and Andrews and Ploberger (1994).

A Monte Carlo study shows that the small sample distributions of these tests depends on nuisance parameters. Another Monte Carlo experiment shows that the MeanQ and the ExpQ tests have the best power properties against alternatives in the direction where the estimator of the cointegration parameter is $T^{1/2}$-consistent, similar to the optimality these tests have been proved to have in a setting with stationary regressors. (see Andrews and Ploberger (1994)). Surprisingly, none of the tests seem to dominate the others against alternatives in the directions where the estimator of the cointegration parameter is $T$-consistent.

The tests considered by Andrews (1993) and Andrews and Ploberger (1994) can be based on either the Wald test (SupW, MeanW, and ExpW), the Lagrange multiplier (LM) test (SupF, MeanF, and ExpF), or the likelihood ratio test, (SupQ, MeanQ, and ExpQ). B. E. Hansen (1992a) derived the SupF and MeanF statistics to test for parameter constancy in cointegrated regressions,

and Seo (1998) considered the SupF, MeanF, and ExpF statistics in the cointegrated VAR model. A related test is the $L$ test by Nyblom (1989), extended to linear models by B. E. Hansen (1992b), to regressions with $I(1)$ variables by B. E. Hansen (1992a), and to cointegrated processes by Quintos (1997), Kuo (1998), and H. Hansen and Johansen (1999). H. Hansen and Johansen (1999) also proposed tests based on recursive estimation of eigenvalues that are associated with cointegration parameters. This approach has the advantage of reducing the dimension of the testing problem, and is well suited for graphical presentation, as implemented in CATS in RATS, (see H. Hansen and Juselius (1995)).

All of these tests have the same null hypothesis, (no parameter changes), but differ in their alternative. The Sup-, Mean-, and Exp-statistics are designed to test against an alternative of one structural change, whereas the alternative of the $L$ test is that the vector of parameters is a martingale.

The SupQ, MeanQ, and ExpQ statistics considered in this paper are similar to SupF, MeanF, and ExpF test for changes in the cointegrating relations considered by Seo (1998)[1]. But the tests in this paper differ by having power against both rotations/rescaling of the cointegration parameter as well as actual changes of the cointegration space, although the power in the former case is quite low when the sample size is small. In Seo (1998) the statistics are accompanied by tables with (asymptotic) critical values. One of the conclusions in this paper is that the tests have poor small sample properties for values of nuisance parameters that one is likely to encounter in economic time series. So in practice one should account for this bias rather than using the asymptotic tables alone. Based on the simulations, I argue that the MeanQ test is the best test.

This paper is organized as follows. In Section 2, the statistical model is presented and the LR test for a known change point is derived and shown to have an asymptotic $\chi^2$ distribution. In Section 3 we evaluate the small sample properties with Monte Carlo simulations. Power properties

---

[1] One of the motivations for using the LM based tests was a claim that the LR test is cumbersome to compute, (see Seo (1998) page 226). However, as shown in the next section, the LR test for changes in the cointegrating relations is easy to compute, so in practice this is not an obstacle.

are studied in Section 4 and Section 5 contains concluding remarks. All proofs are contained in the appendix.

## 3.2. The Statistical Model

The extended model to be estimated is given by

$$\Delta X_t = \alpha\beta(t)'X_{t-1} + \sum_{i=1}^{k-1}\Gamma_i\Delta X_{t-i} + \Phi D_t + \varepsilon_t. \quad t = 1,\dots,T$$

where $X_t$ has dimension $p$, the error term $\varepsilon_t$ is a sequence of i.i.d. Gaussian variables with mean zero and variance $\Omega$, and the only modification to the standard model is that

$$\beta(t) = \begin{cases} \beta_1 & t = 1,\dots,\tau \\ \beta_2 & t = \tau+1,\dots,T. \end{cases}$$

where $\alpha$, $\beta_1$, and $\beta_2$ are $p \times r$ matrices with full column rank.

We shall work under the null hypothesis that there are no structural changes in the process $(\beta_1 = \beta_2 = \beta)$ and assume that the process is $I(1)$. Specifically, we assume: $(i)$ The characteristic polynomial $A(z) = I(1 - z) - \alpha\beta'z - \sum_{i=1}^{k-1}\Gamma_i(1 - z)z^i$ has unit roots ($|A(1)| = 0$) and all other roots are outside the unit circle: $(ii)$ The number of unit roots equals $p - r$.

We define the orthogonal complements $\alpha_\perp$ and $\beta_\perp$ to be $p \times (p - r)$ matrices of full column rank that satisfy $\alpha_\perp'\alpha = \beta_\perp'\beta = 0$. and note that the assumptions above ensure that $\alpha_\perp'\Gamma\beta_\perp$ has full rank. where $\Gamma \equiv I - \sum_{i=1}^{k-1}\Gamma_i$. (see Johansen (1996)).

### 3.2.1. Estimation

It is well known how to estimate the model without changes. The model is reformulated as $Z_{0t} = \alpha\beta'Z_{1t} + \Psi Z_{2t} + \varepsilon_t$ with the definitions $Z_{0t} = \Delta X_t$, $Z_{1t} = X_{t-1}$. $Z_{2t} = (\Delta X_{t-1}',\dots,\Delta X_{t-k+1}',D_t)'$ and $\Psi = (\Gamma_1,\dots,\Gamma_{k-1},\Phi)$, and parameter estimates are found by solving the eigenvalue problem

$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0$ where moment matrices and auxiliary residuals are defined by $S_{ij} =$

$\frac{1}{T}\sum_{t=1}^{T}R_{it}R'_{jt}$. $i,j = 0,1$, $R_{0t} = Z_{0t} - M_{02}M_{22}^{-1}Z_{2t}$ and $M_{ij} = \frac{1}{T}\sum_{t=1}^{T}Z_{it}Z'_{jt}$, $i,j = 0,1,2$,

(see Johansen (1996)). The generalized model with structural changes in the cointegrating relations

can be estimated with the same technique. The model is rewritten as

$$\Delta X_t = \alpha\beta'_1 I_{(t\leq\tau)}X_{t-1} + \alpha\beta'_2 I_{(t>\tau)}X_{t-1} + \sum_{i=1}^{k-1}\Gamma_i\Delta X_{t-i} + \Phi D_t + \varepsilon_t$$

$$= \alpha\left(\beta'_1,\beta'_2\right)\tilde{Z}_{1t} + \sum_{i=1}^{k-1}\Gamma_i\Delta X_{t-i} + \Phi D_t + \varepsilon_t. \tag{3.2.1}$$

where $I_{(.)}$ are indicator functions and $\tilde{Z}_{1t} = (I_{(t\leq\tau)}X'_{t-1}, I_{(t>\tau)}X'_{t-1})'$. In the compact form the

model is given by $Z_{0t} = \alpha B'\tilde{Z}_{1t} + \Psi Z_{2t} + \varepsilon_t$, where $B = (\beta'_1,\beta'_2)'$.

The maximum likelihood estimators for this model are also obtained by solving an eigenvalue

problem. Define the moment matrices $\tilde{M}_{11} = \frac{1}{T}\sum_{t=1}^{T}\tilde{Z}_{1t}\tilde{Z}'_{1t}$. and $\tilde{M}_{12} = \frac{1}{T}\sum_{t=1}^{T}\tilde{Z}_{1t}Z'_{2t}$, and

the auxiliary residuals $\tilde{R}_{0t} = R_{0t}$. $\tilde{R}_{1t} = \tilde{Z}_{1t} - \tilde{M}_{12}\tilde{M}_{22}^{-1}Z_{2t}$ and the moment matrices of these

residuals as $\tilde{S}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\tilde{R}_{it}\tilde{R}'_{jt}$, $i,j = 0,1$.

**Theorem 3.2.1.** *The maximum likelihood estimators of the model 3.2.1 are given by*

$$\hat{B} = (\tilde{v}_1,\ldots,\tilde{v}_r) \tag{3.2.2}$$

$$\hat{\alpha} = \tilde{S}_{01}\hat{B} \tag{3.2.3}$$

$$\hat{\Omega} = \tilde{S}_{00} - \hat{\alpha}\hat{\alpha}' \tag{3.2.4}$$

$$\hat{\Psi} = \tilde{M}_{02}\tilde{M}_{22}^{-1} - \hat{\alpha}\hat{B}'\tilde{M}_{12}\tilde{M}_{22}^{-1}. \tag{3.2.5}$$

*where* $(\tilde{v}_1,\ldots,\tilde{v}_r)$ *are the eigenvectors corresponding to the* $r$ *largest eigenvalues* $\tilde{\lambda}_1,\ldots,\tilde{\lambda}_r$ *of the*

*eigenvalue problem*

$$|\lambda\tilde{S}_{11} - \tilde{S}_{10}\tilde{S}_{00}^{-1}\tilde{S}_{01}| = 0.$$

*The maximum value of the likelihood function is given by*

$$L_{max}^{-2/T}(\hat{\alpha}, \hat{B}, \hat{\Psi}, \hat{\Omega}) = (2\pi e)^p |\hat{S}_{00}| \prod_{i=1}^{r}(1 - \hat{\lambda}_i).$$

A proof of the theorem can be found in Johansen (1988).

This enables us to get a simple expression for the likelihood ratio test of the hypothesis of a constant model. The first theorem is applicable in a situation where a particular value for the cointegrating relations needs to be tested.

**Theorem 3.2.2.** *The likelihood ratio test of $\beta_1 = \beta_2 = \beta_o$ where $\beta_o$ is a known matrix is asymptotically $\chi^2$ with $2pr - r^2$ degrees of freedom.*

In the more general case where no particular value for $\beta$ is given, the following theorem is applicable.

**Theorem 3.2.3.** *The $LR_T(\tau)$ test of $\beta_1 = \beta_2$. that is no structural change at time $\tau$. is given by*

$$LR_T(\tau) \equiv -2\log Q = T\left[\sum_{i=1}^{r}\log\left(1 - \hat{\lambda}_i\right) - \log\left(1 - \tilde{\lambda}_i\right)\right]$$

*where $\hat{\lambda}_i, i = 1, \ldots, r$ are the $r$ largest eigenvalues of $|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0$. The asymptotic distribution of the test is $\chi^2$ with $pr$ degrees of freedom.*

The proofs of the two theorems are given in the appendix.

## 3.3. Test for a Structural Change when the Potential Change Point is Unknown

When the timing of a potential structural change is unknown, a more complicated situation arises, and the testing problem is nonstandard.

The situation with a structural change after the unknown time $\tau$, can be expressed in two ways. Either as

$$\beta(t) = \beta + \sum_{\tau=1}^{T-1} b_\tau \cdot 1[t > \tau]$$

where at most one of the parameters $b_1, \ldots, b_{T-1}$ are non-zero, or as

$$\beta(t) = \beta + b_\tau \cdot 1[t > \tau],$$

where $\tau$ is an unknown parameter, $1 \leq \tau < T$, that is only identified under the alternative. Problems of this nature have been analyzed by Andrews (1993), Andrews and Ploberger (1994), and B. E. Hansen (1996). The approach by Andrews and Ploberger (1994) is semi-Bayesian in the sense that it is based on a weighting function (prior), $J(\rho)$, over the possible change points, expressed as a fraction of the sample size, $\rho = \tau/T$. For example, the uniform distribution on the interval $[\pi_0, \pi_1]$, $0 \leq \pi_0 \leq \pi_1 \leq 1$, where $\tau_0 = [\pi_0 T]$ ($\tau_1 = [\pi_1 T]$) corresponds to the smallest (largest) possible change point. Andrews and Ploberger show that the class of optimal tests is given by

$$Exp\text{-}L_{Tc} \equiv (1 + c)^{-q/2} \int_{\pi_0}^{\pi_1} \exp\left(\frac{1}{2}\frac{c}{1+c}L_T(\rho)\right) dJ(\rho), \quad c \in (0, \infty),$$

where $L_T(\rho)$ is either the Wald, Lagrange multiplier, or the likelihood ratio test, for a structural change at $\tau = [\rho T]$. The parameter $c$ denotes how much weight is given to alternatives near the null, and $q$ denotes the dimension of parameter space for the change parameter, (in this setting $q = pr$).

The limits for $c \to 0$ and $c \to \infty$ (with suitable normalization) are given by

$$\text{MeanL}_T \equiv \lim_{c \to 0} 2(Exp\text{-}L_{Tc} - 1)/c = \int_{\pi_0}^{\pi_1} L_T(\rho) dJ(\rho).$$

the "average $L_T$", and

$$\text{ExpL}_T \equiv \lim_{c \to \infty} \log((1 + c)^{q/2} Exp\text{-}L_{Tc}) = \log\left(\int_{\pi_0}^{\pi_1} \exp(L_T(\rho)/2) dJ(\rho)\right).$$

the "exponential average $L_T$". It is worthwhile to notice that the Sup-test,

$$\text{Sup-L}_T = \sup_{\pi_0 \le \rho \le \pi_1} L_T(\rho).$$

does not belong to this class of optimal tests, as was pointed out by Andrews and Ploberger (1994). However, since the conditions that led to the class of optimal tests in Andrews and Ploberger (1994) are not satisfied in this setting with $I(1)$ variables. we cannot, a priori, exclude the Sup-$L_T$ as an optimal test.

Asymptotic distributions of these tests can be tabulated for various choices of $\pi_0 = \tau_0/T$ and $\pi_1 = \tau_1/T$. that defines the fraction of the subsample for which change points are considered. Often one can derive analytical expressions for the asymptotic distributions. e.g. H. Hansen and Johansen (1999). and simulation can be based on these expressions, or one can simply generate a large number of time series based on some choice of parameter values. provided that these parameters are not nuisance parameters in the asymptotic distribution.

The 90%. 95%. and 99% quantile for the SupQ, MeanQ, and ExpQ statistics are tabulated in Tables 3.4.3–3.4.5.

## 3.4. Size Properties of the Tests

To evaluate the small sample properties of the tests given in the previous section. we perform a Monte Carlo experiment. The experiments were made with the computer package GAUSS,

generating time-series based on the two-dimensional VAR(1) with one cointegrating relation,

$$\begin{pmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{pmatrix} = a \begin{pmatrix} -1 \\ 1 \end{pmatrix} (1, -1) \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \begin{pmatrix} 0.10 & 0.05 \\ 0.05 & 0.10 \end{pmatrix}). \quad (3.4.1)$$

The generated time series had length $n = 200, 300, 600, 1100$ and initial value $X_0 = 0$. When the statistics were calculated, the first 100 observations were discarded, to reduce any influence the choice of initial value may have. So test statistics were based on sample sizes of $T = 100, 200, 500$, and $1000$. Simulations were made for $a = 0.025, 0.050, 0.075, 0.100, 0.500$, and $0.900$. The parameter $a$ turns out be a nuisance parameter in our finite sample distributions. This parameter has a one-to-one correspondence with the stable root of the process, $z_2$, or the stable eigenvalue, $\lambda_2 = 1/z_2$. The characteristic polynomial of a process generated by equation (3.4.1) is given by

$$|A(z)| = | \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} (1 - z) - a \begin{pmatrix} -1 \\ 1 \end{pmatrix} (1, -1)| = (1 - z)(1 - (1 - 2a)z).$$

which has the roots $z_1 = 1$ and $z_2 = 1/(1 - 2a)$, and hence, eigenvalues, $\lambda_1 = 1$ and $\lambda_2 = (1 - 2a)$. The $I(1)$ conditions require that $\lambda_2$ satisfies $|\lambda_2| < 1$, or equivalently $a \in (0, 1)$, which is the case for all the simulated processes. The values of $a$ in the experiment translates into $\lambda_2 = 0.95, 0.90, 0.85, 0.80, 0.00$ and $-0.80$.

Critical values for the test statistics, for sizes of 10%, 5%, and 1%, were based on 50,000 generated time series, for every pair of $(a, T)$. The critical values are reported in Tables 3.4.1 and 3.4.2, along with the asymptotic $(\chi^2)$ critical values.

From Table 3.4.1 it can be seen that the size distortion is not alarming for $a \geq 0.1$ ($\lambda_2 \leq 0.8$). However, for $a$ close to zero ($\lambda_2$ close to one) the size distortion is increasing. The timing of the change does not play a big role in the size distortions. The simulations based on $a \leq 0.1$, is most relevant for empirical applications. In economic application, there are typically roots close to one, besides the unit roots, and an eigenvalue $\lambda$ with real($\lambda$) $\geq 0.8$ is usually the case. Given

Critical values, LR$_T$ test for a change at $[0.1T]$.

| $T$ | $a = 0.1$ ($\lambda_2 = 0.80$) | | | $a = 0.5$ ($\lambda_2 = 0.00$) | | | $a = 0.9$ ($\lambda_2 = -0.80$) | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 6.1310 | 7.7738 | 11.4580 | 4.5527 | 5.9259 | 9.1743 | 4.3820 | 5.7268 | 8.8171 |
| 200 | 5.6526 | 7.2627 | 10.9768 | 4.5066 | 5.8870 | 9.0041 | 4.3712 | 5.7503 | 8.7474 |
| 500 | 5.1897 | 6.6460 | 10.0702 | 4.5827 | 5.9330 | 9.1439 | 4.4901 | 5.8196 | 9.0563 |
| 1000 | 4.9050 | 6.3918 | 9.8848 | 4.5930 | 5.9983 | 9.3027 | 4.5618 | 5.9231 | 9.2515 |
| $\chi^2_{(2)}$ | 4.6052 | 5.9915 | 9.2103 | 4.6052 | 5.9915 | 9.2103 | 4.6052 | 5.9915 | 9.2103 |

Critical values. LR$_T$ test for a change at $[0.25T]$.

| $T$ | $a = 0.1$ ($\lambda_2 = 0.80$) | | | $a = 0.5$ ($\lambda_2 = 0.00$) | | | $a = 0.9$ ($\lambda_2 = -0.80$) | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 6.0567 | 7.8256 | 11.5581 | 4.7252 | 6.1182 | 9.4126 | 4.5528 | 5.9904 | 9.1717 |
| 200 | 5.4837 | 7.1334 | 10.7639 | 4.6767 | 6.1242 | 9.4022 | 4.5078 | 5.8709 | 9.2257 |
| 500 | 4.9711 | 6.3831 | 9.9940 | 4.6139 | 6.0221 | 9.0431 | 4.5910 | 5.9890 | 9.0826 |
| 1000 | 4.7659 | 6.2427 | 9.5385 | 4.6004 | 5.9572 | 9.2823 | 4.5801 | 5.9579 | 9.2403 |
| $\chi^2_{(2)}$ | 4.6052 | 5.9915 | 9.2103 | 4.6052 | 5.9915 | 9.2103 | 4.6052 | 5.9915 | 9.2103 |

Critical values, LR$_T$ test for a change at $[0.50T]$.

| $T$ | $a = 0.1$ ($\lambda_2 = 0.80$) | | | $a = 0.5$ ($\lambda_2 = 0.00$) | | | $a = 0.9$ ($\lambda_2 = -0.80$) | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 6.0276 | 7.7648 | 11.6975 | 4.7704 | 6.1958 | 9.5186 | 4.6816 | 6.1226 | 9.2817 |
| 200 | 5.3005 | 6.8921 | 10.5191 | 4.6818 | 6.0812 | 9.3086 | 4.6415 | 6.0237 | 9.3392 |
| 500 | 4.9586 | 6.4166 | 9.9239 | 4.6597 | 6.0287 | 9.2418 | 4.6022 | 5.9802 | 9.3219 |
| 1000 | 4.7438 | 6.1611 | 9.6193 | 4.5870 | 5.9603 | 9.0965 | 4.5894 | 6.0107 | 9.1048 |
| $\chi^2_{(2)}$ | 4.6052 | 5.9915 | 9.2103 | 4.6052 | 5.9915 | 9.2103 | 4.6052 | 5.9915 | 9.2103 |

Table 3.4.1: Critical values for the likelihood ratio test for the case with $p = 2$ and $r = 1$ based on simulations with 50,000 replications.

Critical values, LR$_T$ test for a change at $[0.1T]$.

| $T$ | $a = 0.025$ ($\lambda_2 = 0.95$) | | | $a = 0.05$ ($\lambda_2 = 0.90$) | | | $a = 0.075$ ($\lambda_2 = 0.85$) | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 7.3128 | 9.1252 | 13.0206 | 6.8261 | 8.5235 | 12.4170 | 6.4392 | 8.0782 | 11.8322 |
| 200 | 7.1444 | 8.9082 | 12.7554 | 6.4603 | 8.1118 | 11.7736 | 5.9457 | 7.5266 | 11.2938 |
| 500 | 6.3492 | 8.0043 | 11.7104 | 5.7253 | 7.3641 | 11.0342 | 5.3829 | 7.0089 | 10.5588 |
| 1000 | 5.7294 | 7.3884 | 11.0951 | 5.2539 | 6.7995 | 10.3381 | 5.0351 | 6.5278 | 10.0011 |
| $\chi^2_{(2)}$ | 4.6052 | 5.9915 | 9.2103 | 4.6052 | 5.9915 | 9.2103 | 4.6052 | 5.9915 | 9.2103 |

Critical values. LR$_T$ test for a change at $[0.25T]$.

| $T$ | $a = 0.025$ ($\lambda_2 = 0.95$) | | | $a = 0.05$ ($\lambda_2 = 0.90$) | | | $a = 0.075$ ($\lambda_2 = 0.85$) | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 7.8987 | 9.7984 | 14.1787 | 7.1010 | 8.9106 | 12.9989 | 6.4892 | 8.3059 | 12.2069 |
| 200 | 7.2001 | 9.0525 | 13.2050 | 6.2066 | 7.9262 | 11.9992 | 5.6397 | 7.3271 | 11.1988 |
| 500 | 6.0137 | 7.7950 | 11.4599 | 5.3503 | 6.9006 | 10.5298 | 5.0873 | 6.5553 | 10.0950 |
| 1000 | 5.3901 | 6.9856 | 10.5706 | 5.0298 | 6.5151 | 9.8866 | 4.8677 | 6.3484 | 9.7145 |
| $\chi^2_{(2)}$ | 4.6052 | 5.9915 | 9.2103 | 4.6052 | 5.9915 | 9.2103 | 4.6052 | 5.9915 | 9.2103 |

Critical values. LR$_T$ test for a change at $[0.50T]$.

| $T$ | $a = 0.025$ ($\lambda_2 = 0.95$) | | | $a = 0.05$ ($\lambda_2 = 0.90$) | | | $a = 0.075$ ($\lambda_2 = 0.85$) | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 8.0507 | 9.9911 | 14.4071 | 7.0272 | 8.8446 | 13.0488 | 6.3809 | 8.1670 | 12.1642 |
| 200 | 7.0967 | 8.9222 | 13.0676 | 6.0975 | 7.8120 | 11.6087 | 5.5795 | 7.2283 | 11.0295 |
| 500 | 5.8366 | 7.4845 | 11.4455 | 5.2715 | 6.8329 | 10.6469 | 4.9657 | 6.4610 | 9.8895 |
| 1000 | 5.2490 | 6.9117 | 10.5112 | 4.9314 | 6.4484 | 9.8773 | 4.8517 | 6.3053 | 9.6678 |
| $\chi^2_{(2)}$ | 4.6052 | 5.9915 | 9.2103 | 4.6052 | 5.9915 | 9.2103 | 4.6052 | 5.9915 | 9.2103 |

Table 3.4.2: Critical values for the likelihood ratio test for the case with $p = 2$ and $r = 1$ based on simulations with 50.000 replications.

the observed size distortion one should use critical values based on the $\chi^2$-distributions with care, when the sample size is moderate. The test is somewhat conservative for small sample sizes, but as an easy rule of thumb, one could used 1% $\chi^2$ critical values when the sample size is less than, say, 250 and the largest stable eigenvalue is larger than 0.9, and use the 5% $\chi^2$ critical values otherwise. This would take care of some of the distortion and would lead to an actual size of about 4% − 9%.

The size distortion for $\lambda_2$ close to unity, is somewhat in contrast to the observations made by Gregory, Nason, and Watt (1996). They investigated the tests of B. E. Hansen (1992a), and reported only moderate size distortion, but found power to be poor unless the stable eigenvalue is small.

### 3.4.1. Size Distortion of the SupQ, MeanQ, and ExpQ tests

The critical values for the SupQ, MeanQ, and ExpQ were based on the same simulations as the ones made for the $LR_T(\rho)$ test, although only for $a = 0.1$. 0.5, and 0.9. The critical values are reported in Tables 3.4.3–3.4.5

The size distortions for the SupQ, MeanQ, and ExpQ statistics are similar to the one of the $LR_T(\rho)$ statistic. For $a = 0.1$, a 5% test based on asymptotic critical values seem to be rejected approximately 10% of the times, for a sample size of $T = 100$.

For the simulated values of $a$, the size distortion is similar and moderate for all the statistics, SupQ, MeanQ, and ExpQ. Based on this, there is no reason to prefer one test over the others based on size properties alone. Given the results in Table 3.4.2, one would expect increasing size distortion as $a$ approaches zero ($\lambda_2$ approaches one).

## 3.5. Power Properties

In this section we study the power properties. We evaluate the power for a sample size of 200, when the change occurs at $\rho = 0.1$. $\rho = 0.25$. $\rho = 0.45$. $\rho = 0.5$. and the case where $\rho$ is random

Critical values: SupQ test for the interval $[0.10T]$ : $[0.90T]$.

| T | $a = 0.1$ | | | $a = 0.5$ | | | $a = 0.9$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 13.0964 | 15.1224 | 19.3934 | 10.0415 | 11.7753 | 15.4593 | 9.8384 | 11.4438 | 15.0366 |
| 200 | 12.3791 | 14.3145 | 18.2885 | 10.1298 | 11.8296 | 15.6366 | 10.0918 | 11.7574 | 15.3920 |
| 500 | 11.6036 | 13.5318 | 17.4276 | 10.4120 | 12.1095 | 15.7714 | 10.3943 | 12.1022 | 15.8074 |
| 1000 | 11.3536 | 13.1525 | 17.2453 | 10.5675 | 12.2131 | 15.9337 | 10.5783 | 12.2106 | 15.9890 |

Critical values: SupQ test for the interval $[0.25T]$ : $[0.75T]$.

| T | $a = 0.1$ | | | $a = 0.5$ | | | $a = 0.9$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 11.5163 | 13.5806 | 18.0332 | 8.8493 | 10.5813 | 14.2715 | 8.7221 | 10.3590 | 14.0159 |
| 200 | 10.6773 | 12.6482 | 16.7293 | 8.9334 | 10.6191 | 14.4508 | 8.8658 | 10.5644 | 14.2996 |
| 500 | 9.9534 | 11.8281 | 15.9182 | 9.0728 | 10.7630 | 14.4420 | 9.1184 | 10.7993 | 14.5569 |
| 1000 | 9.6728 | 11.5205 | 15.5274 | 9.2111 | 10.9048 | 14.5491 | 9.1996 | 10.9720 | 14.6100 |

Critical values: SupQ test for the interval $[0.45T]$ : $[0.55T]$.

| T | $a = 0.1$ | | | $a = 0.5$ | | | $a = 0.9$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 8.2322 | 10.2060 | 14.6510 | 6.4282 | 8.0034 | 11.6786 | 6.3505 | 7.9074 | 11.3794 |
| 200 | 7.5132 | 9.3435 | 13.3732 | 6.4964 | 8.1012 | 11.6897 | 6.4611 | 8.0570 | 11.5812 |
| 500 | 7.1433 | 8.8407 | 12.8055 | 6.5895 | 8.2301 | 11.8754 | 6.6023 | 8.1962 | 11.9676 |
| 1000 | 6.9323 | 8.5582 | 12.3834 | 6.6840 | 8.2449 | 11.7043 | 6.6876 | 8.2456 | 11.8379 |

Table 3.4.3: Critical values for the SupQ test for the case with $p = 2$ and $r = 1$ based on simulations with 50.000 replications.

Critical values; MeanQ test for the interval $[0.10T] : [0.90T]$.

| T | $a = 0.1$ | | | $a = 0.5$ | | | $a = 0.9$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 4.6804 | 5.6105 | 7.8162 | 3.6782 | 4.4941 | 6.4539 | 3.5772 | 4.3829 | 6.2884 |
| 200 | 4.2271 | 5.1249 | 7.2328 | 3.6325 | 4.4703 | 6.4066 | 3.5779 | 4.3758 | 6.2836 |
| 500 | 3.8938 | 4.7481 | 6.6875 | 3.5998 | 4.4308 | 6.2706 | 3.6054 | 4.3829 | 6.3162 |
| 1000 | 3.7336 | 4.5518 | 6.5095 | 3.5891 | 4.3670 | 6.1787 | 3.5860 | 4.3792 | 6.2380 |

Critical values: MeanQ test for the interval $[0.25T] : [0.75T]$.

| T | $a = 0.1$ | | | $a = 0.5$ | | | $a = 0.9$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 5.0601 | 6.2148 | 8.9266 | 4.0246 | 5.0241 | 7.4074 | 3.9225 | 4.9275 | 7.1663 |
| 200 | 4.5334 | 5.6746 | 8.2190 | 3.9587 | 4.9937 | 7.3705 | 3.9109 | 4.8962 | 7.1970 |
| 500 | 4.2164 | 5.2519 | 7.6751 | 3.9186 | 4.9061 | 7.2962 | 3.9071 | 4.8799 | 7.2438 |
| 1000 | 4.0128 | 5.0549 | 7.4785 | 3.8891 | 4.8507 | 7.0597 | 3.8794 | 4.8989 | 7.1792 |

Critical values: MeanQ test for the interval $[0.45T] : [0.55T]$.

| T | $a = 0.1$ | | | $a = 0.5$ | | | $a = 0.9$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 5.7773 | 7.3100 | 10.9838 | 4.5536 | 5.8791 | 8.9554 | 4.4797 | 5.7903 | 8.6346 |
| 200 | 5.1167 | 6.5616 | 9.9398 | 4.4882 | 5.7815 | 8.8768 | 4.4663 | 5.7630 | 8.7857 |
| 500 | 4.7581 | 6.1283 | 9.3511 | 4.4308 | 5.7725 | 8.7417 | 4.4427 | 5.7164 | 8.7102 |
| 1000 | 4.5580 | 5.8784 | 9.0470 | 4.4025 | 5.6926 | 8.6438 | 4.4235 | 5.7039 | 8.6359 |

Table 3.4.4: Critical values for the MeanQ test for the case with $p = 2$ and $r = 1$ based on simulations with 50.000 replications.

Critical values: ExpQ test for the interval $[0.10T]$ : $[0.90T]$.

| T | a = 0.1 | | | a = 0.5 | | | a = 0.9 | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 3.6941 | 4.5083 | 6.3919 | 2.6652 | 3.3397 | 4.9002 | 2.5771 | 3.1987 | 4.6876 |
| 200 | 3.2559 | 4.0310 | 5.7258 | 2.6250 | 3.2893 | 4.9077 | 2.5694 | 3.2078 | 4.7276 |
| 500 | 2.9045 | 3.6250 | 5.2463 | 2.6054 | 3.2557 | 4.7639 | 2.5872 | 3.2176 | 4.7419 |
| 1000 | 2.7709 | 3.4406 | 5.1137 | 2.5906 | 3.2286 | 4.7863 | 2.5866 | 3.2316 | 4.7413 |

Critical values: ExpQ test for the interval $[0.25T]$ : $[0.75T]$.

| T | a = 0.1 | | | a = 0.5 | | | a = 0.9 | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 3.5182 | 4.3459 | 6.2892 | 2.6042 | 3.3021 | 4.9218 | 2.5299 | 3.1869 | 4.7072 |
| 200 | 3.0662 | 3.8855 | 5.6166 | 2.5587 | 3.2592 | 4.8729 | 2.5263 | 3.1662 | 4.7410 |
| 500 | 2.7670 | 3.5060 | 5.1728 | 2.5239 | 3.1779 | 4.7769 | 2.5221 | 3.1771 | 4.7740 |
| 1000 | 2.6324 | 3.3346 | 5.0206 | 2.5178 | 3.1710 | 4.6755 | 2.5131 | 3.1905 | 4.7010 |

Critical values: ExpQ test for the interval $[0.45T]$ : $[0.55T]$.

| T | a = 0.1 | | | a = 0.5 | | | a = 0.9 | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 3.1722 | 4.0398 | 6.0322 | 2.4521 | 3.1591 | 4.8142 | 2.4169 | 3.1155 | 4.6726 |
| 200 | 2.7819 | 3.5738 | 5.4108 | 2.4142 | 3.1062 | 4.7590 | 2.3986 | 3.0886 | 4.7489 |
| 500 | 2.5612 | 3.2996 | 5.0918 | 2.3782 | 3.0848 | 4.6971 | 2.3768 | 3.0772 | 4.6929 |
| 1000 | 2.4450 | 3.1533 | 4.8359 | 2.3660 | 3.0563 | 4.6157 | 2.3672 | 3.0604 | 4.6186 |

Table 3.4.5: Critical values for the ExpQ test for the case with $p = 2$ and $r = 1$ based on simulations with 50,000 replications.

and uniformly distributed over the interval [0.1, 0.9].

We consider alternatives in two directions: the case where there is an actual change of the cointegration space. and the case where there is a rotation or rescaling of the cointegrating relations, but the cointegration space is unchanged.

We modify equation (3.4.1) by replacing $\beta$ with

$$\beta(t) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 0 \\ -c \end{pmatrix} I(t > [\rho T]).$$

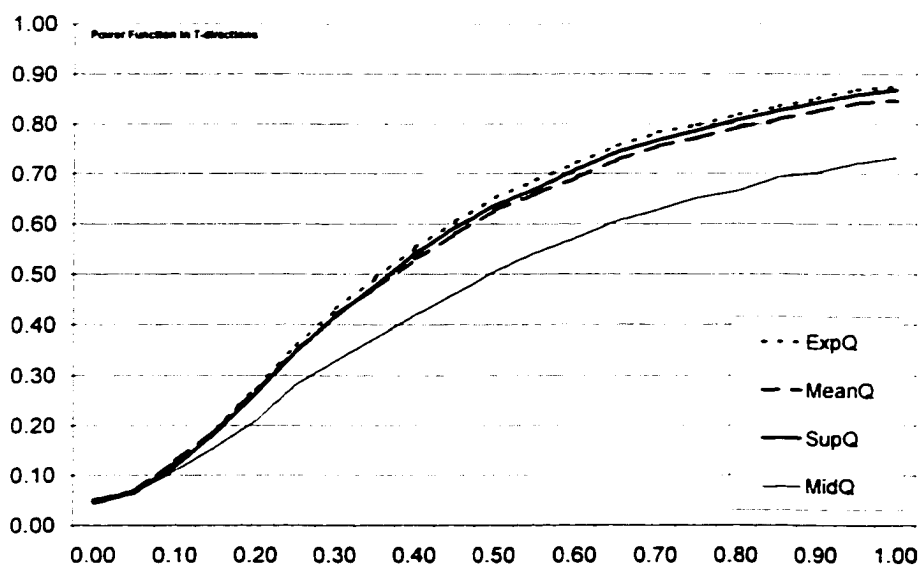This is the case where the cointegration space is changed after time $\tau = [\rho T]$.



Figure 3.5.1: The power functions in the $T$-convergent directions. based on simulations with 10.000 repetitions

Consider first the case where the change point is uniformly distributed on the points $[\pi_0 T]$. . . . . $[\pi_1 T]$. Figure 3.5.1 displays the power functions for SupQ, MeanQ, ExpQ, and MidQ $\equiv LR_T(T/2)$. under the alternative where $\tau$ is uniformly distributed over the points $[0.1T]$. . . . . $[0.9T]$. and where

the first three tests are based on the tests $LR_T(\tau)$, $\tau = [0.1T], \ldots, [0.9T]$, and thus exclude the first and last 10% of the LR statistics.

As can be seen from Figure 3.5.1, there is hardly any difference in the power of the first three statistics, and the difference may simply be due to sample variation. This indicates that the optimality of MeanQ and ExpQ over SupQ, shown by Andrews and Ploberger (1994), does not carry over to the situation with $I(1)$ variables. The naive MidQ test has, as expected, worse power properties. On the other hand the MidQ statistic dominates in terms of simplicity, because it only requires one estimation under the alternative, and has a $\chi^2$ distribution.
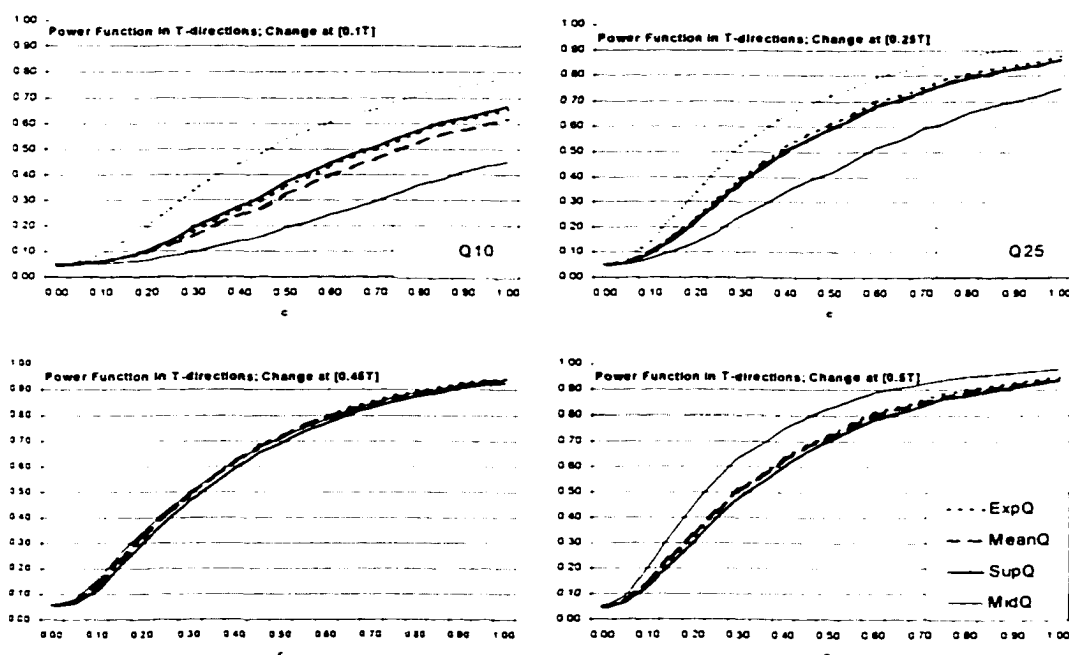


Figure 3.5.2: The power function for the test statistics, for a structural change in the $T$-consistent directions. Upper left, upper right, lower left, lower right panel are for a change at time $[0.1T]$, $[0.25T]$, $[0.45T]$, and $[0.5T]$ respectively. The thin dotted line in the two upper panels is the power envelope – the $LR_T$ test for a change at $[0.1T]$ (left) and $[0.25T]$ (right). Based on 10000 replications for the sample size $T = 200$.

In Figure 3.5.2 we calculate the power functions for changes occurring at particular points in

time. There is not any noteworthy difference between the power of SupQ, MeanQ, and ExpQ, but not surprisingly MidQ is more powerful at $\tau = T/2$, and has low power for $\tau$ far from $T/2$.

Next, consider the alternatives that involve a rotation or rescaling of the cointegration parameter. These take the form $\beta(t) = \beta_o\phi(t)$, where $\varphi(t) = \phi_1$ for $t \leq \tau$, and $\phi(t) = \phi_2$ for $t > \tau$, and where $\phi_1$ and $\phi_2$ are $r \times r$ matrices of full rank. A change of this type need not be associated with a change in $\beta$, but can instead be interpreted as a change in the adjustment coefficients $\alpha$, in a way that leaves the orthogonal compliments of $\alpha$ and $\beta$ unchanged, and hence the stochastic trend, $\beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp \sum_{s=1}^{t} \varepsilon_s$, unchanged, (see Johansen (1996)). This double interpretation is explained by the fact that $\alpha$ and $\beta$ are not identified from a given value of the $p \times p$ matrix $\alpha\beta'$. Thus, $\alpha$ constant and $\beta(t) = \beta\phi(t)$ is equivalent to $\alpha(t) = \alpha\phi(t)'$ and $\beta$ constant. The parameter estimates in these directions are only $T^{1/2}$-consistent, as opposed to the $T$-consistency we had for the other directions. We may therefore expect the power of the test to be lower, and that the tests MeanQ and ExpQ dominate the SupQ test, due to the results by Andrews and Ploberger (1994).

The parameter $\beta$, in equation (3.4.1), was substituted with

$$
\beta(t) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} c \\ -c \end{pmatrix} I(t > \tau)
$$

in the simulations for various values of $c$, where $\tau$ is uniformly distributed on $[\pi_0 T], \ldots, [\pi_1 T]$.

Figure 3.5.3 shows the power function of the test statistics. As can be seen, the MeanQ and ExpQ do dominate the SupQ, but the power of all the tests are quite poor: for a change as large as $\beta = (1, -1)'$ to $\beta = (3, -3)'$ the tests only reject the null hypothesis of no changes in about 50% of the cases, where as a change from $\beta = (1, -1)'$ to $\beta = (1, -2)'$ is rejected in about 85% of the times, see Figure 3.5.1.
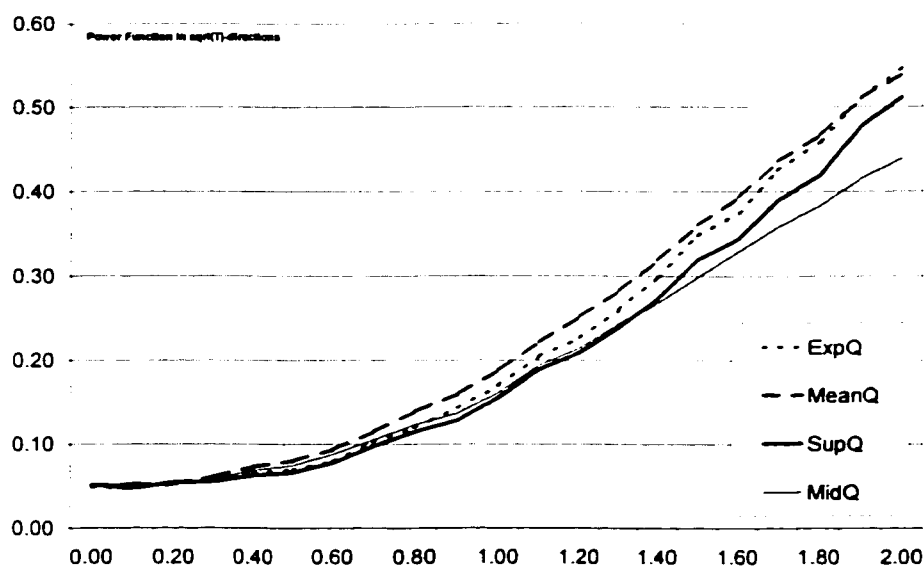
Figure 3.5.3: The power functions in the $T^{1/2}$-convergent directions, based on simulations with 10.000 repetitions

## 3.6. Conclusion

This paper showed how the Sup, Mean. and Exp test can be based on the likelihood ratio test for changes in the cointegrating relations. and that the likelihood ratio test is easy to compute. In a situation where the potential change point is known or chosen independently of the sample. the likelihood ratio test is asymptotically $\chi^2$. When the timing of the change point is unknown. the MeanQ test and the ExpQ test have better power properties than the SupQ in directions in which the parameters are $T^{1/2}$-consistent. In the other directions in which the parameter estimates are $T$-consistent. the MeanQ, the ExpQ, and the SupQ have similar power. The naive MidQ test is dominated by the three other tests. but is easier to compute and evaluate.

The Monte Carlo study indicated some dependence on nuisance parameters in small samples. This bias is not alarming when the stable roots are not too close to unity.

Information on how to correct for the bias in the tests, could be obtained by additional simula-

tions. and derivation of an approximation for the response surface, see Hendry (1984). preferably along with an analytically derived bias correction similar to the results of Johansen (1999a, 1999b). We leave this for future research.

## Appendix C: Asymptotic Analysis and Proofs

In this appendix we give the proofs of the preceding theorems. The asymptotic analysis is a bit involved: what only took half a page to formulate takes several to prove. Part of the analysis is similar to the analysis of the standard model with constant parameters, (see Johansen (1996)), from which much inspiration is taken. In the model with constant parameters the important element of the asymptotics is a $(p - r)$-dimensional Brownian motion. This Brownian motion leads to a stochastic integral that describes the limit distribution of the cointegration parameters. In this model with a structural change the asymptotics involve, in part, a stochastic integral and, in part, standard results from the stationary analysis.

In the following, we derive for simplicity the asymptotic distribution with $\Phi D_t = 0$. More general choices of $\Phi D_t$ would change the rate of convergence for some of the limits we derive. but will not change the main results: That the asymptotic distribution of the cointegration parameter estimates are mixed Gaussian and that the likelihood ratio test. for no change against a change at a known point in time, is asymptotically $\chi^2$.

### C.1. Limits and Rate of Convergence

In the asymptotics we shall keep the proportion of observations in each sub-sample constant as $T$ goes to infinite, and we denote the ratio of observations in the first sub-sample by $\rho = \frac{\tau}{T} \in (0, 1)$.

Under the null hypothesis, the parameters are constant $\beta_1 = \beta_2 = \beta$. and $\beta' X_t$ and $\Delta X_t$ are stationary. So we can adopt many results from Johansen (1996). We denote the covariance matrix

by

$$\begin{pmatrix} \Sigma_{00} & \Sigma_{03} \\ \Sigma_{30} & \Sigma_{33} \end{pmatrix} \equiv \mathrm{var}\left[ \begin{array}{c} \Delta X_t \\ \beta' X_{t-1} \end{array} \middle| \Delta X_{t-1}, \dots, \Delta X_{t-k+1} \right] = \mathrm{var}\left[ \begin{array}{c} Z_{0t} \\ \beta' Z_{1t} \end{array} \middle| Z_{2t} \right].$$

From Johansen (1996) we have the identities

$$\Sigma_{03}\Sigma_{33}^{-1} = \alpha \qquad\qquad\qquad (\text{C.1})$$

$$\Sigma_{00} - \Sigma_{03}\Sigma_{33}^{-1}\Sigma_{30} = \Omega \qquad\qquad\qquad (\text{C.2})$$

$$\left(\Sigma_{33} - \Sigma_{30}\Sigma_{00}^{-1}\Sigma_{03}\right)^{-1} - \Sigma_{33}^{-1} = \alpha'\Omega^{-1}\alpha. \qquad\qquad (\text{C.3})$$

By the law of large numbers we have that $S_{00} \xrightarrow{P} \Sigma_{00}$. $\beta' S_{10} \xrightarrow{P} \Sigma_{30}$. $\beta' S_{11}\beta \xrightarrow{P} \Sigma_{33}$, and that the following limits in probability are well defined:

$$\beta' M_{11}\beta \xrightarrow{P} \Sigma_{33}^u$$

$$\beta' M_{12} \xrightarrow{P} \Sigma_{32}$$

$$M_{22} \xrightarrow{P} \Sigma_{22}$$

$$\beta' M_{12}M_{22}^{-1}M_{21}\beta \xrightarrow{P} \Sigma_{33}^c = \Sigma_{32}\Sigma_{22}^{-1}\Sigma_{23}$$

$$\beta' S_{11}\beta \xrightarrow{P} \Sigma_{33} = \Sigma_{33}^u - \Sigma_{33}^c.$$

where $\Sigma_{23} = \Sigma_{23}'$.

To simplify notation we let $\gamma = 2\rho - 1$. and define

$$\Sigma_{303} = \Sigma_{30}\Sigma_{00}^{-1}\Sigma_{03}$$

$$\Sigma_{33}^c = \Sigma_{32}\Sigma_{22}^{-1}\Sigma_{23}$$

$$\Sigma_{33}^\gamma = \Sigma_{33} + (1 - \gamma^2)\Sigma_{33}^c.$$

where $(1 - \gamma^2)$ is a measure of the change point position in the sample, with $(1 - \gamma^2) = 1$ if the change point is in the middle of the sample and decreases to zero as the change point approaches the beginning or end of the sample.

From Chapter 1 we have the Granger representation which gives the moving average representation of the process

$$X_t = C \sum_{i=1}^{t} \varepsilon_t + C(L)\varepsilon_t + C \left( X_0 - \Gamma_1 X_{-1} - \cdots - \Gamma_{k-1} X_{-k+1} \right).$$

where $C = \beta_\perp \left( \alpha'_\perp \Gamma \beta_\perp \right)^{-1} \alpha'_\perp$ and where $C(L)\varepsilon_t$ is a stationary process. The continuous time limit (of the non-stationary directions) are denoted by

$$T^{-1/2} \beta'_\perp R_{1[Tu]} \xrightarrow{w} \beta'_\perp C W(u) = G(u). \tag{C.4}$$

where $W(u)$ is a Brownian motion with no drift and covariance matrix $\Omega$. and $\xrightarrow{w}$ denotes weak convergence on $D[0,1]$. (see Billingsley (1999)).

Next. we define

$$B_0 = \begin{pmatrix} \beta \\ -\beta \end{pmatrix} \text{ and } B_1 = \begin{pmatrix} \beta_\perp & 0 \\ 0 & \beta_\perp \end{pmatrix}$$

so that $B$. $B_0$ and $B_1$ form a mutually orthogonal basis. and $B$ and $B_0$ define the directions of $\tilde{Z}_{1t} = (I_{(t \leq \tau)} X'_{t-1} \cdot I_{(t > \tau)} X'_{t-1})'$ that are $I(0)$ and $B_1$ defines the directions that are $I(1)$.

The limits of various matrices are given in the following lemma.

**Lemma C.1.** *With the definitions above we have that*

$$B'\tilde{S}_{10} \xrightarrow{p} \Sigma_{30} \tag{C.5}$$

$$B_0'\tilde{S}_{10} \xrightarrow{p} \gamma\Sigma_{30} \tag{C.6}$$

$$(B.B_0)'\tilde{S}_{11}(B.B_0) \xrightarrow{p} \begin{pmatrix} \Sigma_{33} & \gamma\Sigma_{33} \\ \gamma\Sigma_{33} & \Sigma_{33}^\gamma \end{pmatrix} \tag{C.7}$$

$$(B.B_0)'\tilde{S}_{10}S_{00}^{-1}\tilde{S}_{01}(B.B_0) \xrightarrow{p} \begin{pmatrix} \Sigma_{303} & \gamma\Sigma_{303} \\ \gamma\Sigma_{303} & \gamma^2\Sigma_{303} \end{pmatrix} \tag{C.8}$$

$$T^{-1}B_1'\tilde{S}_{11}B_1 \xrightarrow{w} \begin{pmatrix} \int_0^\rho G(u)G'(u)du & 0 \\ 0 & \int_\rho^1 G(u)G'(u)du \end{pmatrix} \tag{C.9}$$

$$B_1'\tilde{S}_{1\varepsilon} = B_1'(\tilde{S}_{10} - \tilde{S}_{11}B\alpha') \xrightarrow{w} \begin{pmatrix} \int_0^\rho G(dW)' \\ \int_\rho^1 G(dW)' \end{pmatrix} \tag{C.10}$$

$$T^{\frac{1}{2}}B_0'\tilde{S}_{1\varepsilon} \xrightarrow{w} Z_\varepsilon \sim N(0.\Sigma_{33}^\gamma \times \Omega) \tag{C.11}$$

$$(B.B_0)'\tilde{S}_{11}B_1 = O_p(1). \tag{C.12}$$

**Proof.** Equation (C.5) follows by the law of large numbers and the identity

$$B'\tilde{S}_{10} = B'\tilde{M}_{10} - B'\tilde{M}_{12}M_{22}^{-1}M_{20} = \beta'M_{10} - \beta'M_{12}M_{22}^{-1}M_{20} = \beta'S_{10}.$$

To prove the other identities. it is convenient to define the sub-sample moment matrices:

$$M_{ij}^{(1)} = T^{-1}\sum_{t=1}^\tau Z_{it}Z_{jt}'. \quad i.j = 0.1.2$$

$$M_{ij}^{(2)} = T^{-1}\sum_{t=\tau+1}^T Z_{it}Z_{jt}' \quad i.j = 0.1.2.$$

and similarly define $\tilde{M}_{ij}^{(k)}$ where $\tilde{Z}_{1t}$ replaces $Z_{1t}$. $k = 1.2$. The stationarity implies that $M_{00}$. $M_{00}^{(1)}/\rho$. $M_{00}^{(2)}/(1-\rho)$ have the same limit, and similarly for $M_{02}$. $M_{22}$. $\beta'M_{10}$. and $\beta'M_{12}$.

We have

$$
\begin{aligned}
B_0' \tilde{S}_{10} &= (\beta' M_{10}^{(1)} - \beta' M_{12}^{(1)} M_{22}^{-1} M_{20}) - (\beta' M_{10}^{(2)} - \beta' M_{12}^{(2)} M_{22}^{-1} M_{20}) \\
&= \rho(\beta' M_{10} - \beta' M_{12} M_{22}^{-1} M_{20}) - (1 - \rho)(\beta' M_{10} - \beta' M_{12} M_{22}^{-1} M_{20}) + o_p(1) \\
&= (2\rho - 1)\beta' S_{10} + o_p(1) \\
&= (2\rho - 1)\Sigma_{30} + o_p(1).
\end{aligned}
$$

which shows (C.6). The upper left element of (C.7) is proven by

$$
\begin{aligned}
B' \tilde{S}_{11} B &= \beta' M_{11}^{(1)} \beta + \beta' M_{11}^{(2)} \beta - (\beta' M_{12}^{(1)} + \beta' M_{12}^{(2)}) M_{22}^{-1} (M_{21}^{(1)} \beta + M_{21}^{(2)} \beta) \\
&= (\rho + (1 - \rho))\Sigma_{33}^u - (\rho + (1 - \rho))\Sigma_{33}^c(\rho + (1 - \rho)) + o_p(1) \\
&= \Sigma_{33}^u - \Sigma_{33}^c + o_p(1) = \Sigma_{33} + o_p(1).
\end{aligned}
$$

and the off diagonal elements by

$$
\begin{aligned}
B' \tilde{S}_{11} B_0 &= \beta' M_{11}^{(1)} \beta - \beta' M_{11}^{(2)} \beta - (\beta' M_{12}^{(1)} + \beta' M_{12}^{(2)}) M_{22}^{-1} (M_{21}^{(1)} \beta - M_{21}^{(2)} \beta) \\
&= (\rho - (1 - \rho))\Sigma_{33}^u - (\rho + (1 - \rho))\Sigma_{33}^c(\rho - (1 - \rho)) + o_p(1) \\
&= (2\rho - 1)\Sigma_{33} + o_p(1) = \gamma \Sigma_{33} + o_p(1),
\end{aligned}
$$

and finally the lower right element by

$$
\begin{aligned}
B_0' \tilde{S}_{11} B_0 &= \beta'(M_{11}^{(1)} + M_{11}^{(2)})\beta - \beta'(M_{12}^{(1)} - M_{12}^{(2)}) M_{22}^{-1} (M_{21}^{(1)} - M_{21}^{(2)})\beta \\
&= (\rho + (1 - \rho))\Sigma_{33}^u - (\rho - (1 - \rho))\Sigma_{33}^c(\rho - (1 - \rho)) + o_p(1) \\
&= \Sigma_{33}^u - \Sigma_{33}^c + (1 - \gamma^2)\Sigma_{33}^c + o_p(1) = \Sigma_{33} + (1 - \gamma^2)\Sigma_{33}^c + o_p(1).
\end{aligned}
$$

Equation (C.8) follows directly from (C.5). (C.6). and the fact that $S_{00} \overset{p}{\to} \Sigma_{00}$. The continuous

mapping theorem and (C.4) prove (C.9) and (C.10). Equation (C.11) is proven by noting that

$$T^{-\frac{1}{2}}\sum_{t=1}^{T} B_0' \tilde{R}_{1t}\varepsilon_t' = T^{-\frac{1}{2}}\sum_{t=1}^{T}\beta'(R_{1t}1_{(t\leq\tau)} - R_{1t}1_{(t\leq\tau)})\varepsilon_t'$$

is a linear combination of two Gaussian variables with mean zero, since $V_t = \beta' R_{1t} = \sum_{i=0}^{\infty} w_i \varepsilon_{t-1-i}$ is a linear process with exponentially decreasing coefficients. So what remains is to derive its asymptotic variance. This is found from

$$\lim_{T\to\infty} \text{var}\left(T^{-\frac{1}{2}}\sum_{1}^{T} B_0'\tilde{R}_{1t}\varepsilon_t'\right) = \lim_{T\to\infty} E\left(T^{-1}\text{vec}\left(\sum_{1}^{T} B_0'\tilde{R}_{1t}\varepsilon_t'\right)\left(\text{vec}\left(\sum_{1}^{T} B_0'\tilde{R}_{1t}\varepsilon_t'\right)\right)'\right)$$

$$= \lim_{T\to\infty} T^{-1}\sum_{1}^{T} E\left(\text{vec}\left(B_0'\tilde{R}_{1t}\varepsilon_t'\right)\left(\text{vec}\left(B_0'\tilde{R}_{1t}\varepsilon_t'\right)\right)'\right)$$

$$= \lim_{T\to\infty} T^{-1}\sum_{1}^{T} E\left(B_0'\tilde{R}_{1t}\tilde{R}_{1t}'B_0\right) \times E\left(\varepsilon_t\varepsilon_t'\right)$$

$$= \lim_{T\to\infty} B_0'\tilde{S}_{11}B_0 \times \Omega = \Sigma_{jj}^? \times \Omega.$$

by the law of iterated expectations.

Finally, (C.12) is a moment matrix of an $I(0)$ variable and an $I(1)$ variable, so the term is $O_p(1)$. This completes the proof. ∎

**Lemma C.2.** $\hat{B} \xrightarrow{p} B$, $\hat{\alpha} \xrightarrow{p} \alpha$, and $\hat{\Omega} \xrightarrow{p} \Omega$.

**Proof.** The estimator of $B$ is found by solving the eigenvalue problem

$$\left|\lambda\tilde{S}_{11} - \tilde{S}_{10}S_{00}^{-1}\tilde{S}_{01}\right| = 0. \tag{C.13}$$

Solving this is equivalent to solving

$$\left|\lambda A_T'\tilde{S}_{11}A_T - A_T'\tilde{S}_{10}S_{00}^{-1}\tilde{S}_{01}A_T\right| = 0 \tag{C.14}$$

where $A_T$ is the full rank matrix

$$A_T = (B, B_0, T^{-1/2}B_1).$$

Then, as $T \to \infty$, the solutions to (C.14), and hence (C.13), converges in probability to the solutions of

$$\left| \lambda \begin{pmatrix} \Sigma_{33} & \gamma\Sigma_{33} & 0 & 0 \\ \gamma\Sigma_{33} & \Sigma_{33}^{\gamma} & 0 & 0 \\ 0 & 0 & \int_0^\rho GG'du & 0 \\ 0 & 0 & 0 & \int_\rho^1 GG'du \end{pmatrix} - \begin{pmatrix} \Sigma_{303} & \gamma\Sigma_{303} & 0 & 0 \\ \gamma\Sigma_{303} & \gamma^2\Sigma_{303} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \right| = 0. \qquad (C.15)$$

Define the full rank matrix $Q = \begin{pmatrix} I & \gamma I \\ 0 & -I \end{pmatrix}$, then

$$Q' \begin{pmatrix} \Sigma_{33} & \gamma\Sigma_{33} \\ \gamma\Sigma_{33} & \Sigma_{33}^{\gamma} \end{pmatrix} Q = \begin{pmatrix} \Sigma_{33} & 0 \\ 0 & (1-\gamma^2)\Sigma_{33}^u \end{pmatrix}$$

$$Q' \begin{pmatrix} \Sigma_{303} & \gamma\Sigma_{303} \\ \gamma\Sigma_{303} & \gamma^2\Sigma_{303} \end{pmatrix} Q = \begin{pmatrix} \Sigma_{303} & 0 \\ 0 & 0 \end{pmatrix}.$$

where we, in the lower right block of the first equality, used that

$$\Sigma_{33}^{\gamma} - \gamma^2\Sigma_{33} = \Sigma_{33} + (1-\gamma^2)\Sigma_{33}^c - \gamma^2\Sigma_{33} = (1-\gamma^2)\Sigma_{33}^u. \qquad (C.16)$$

So the solutions of (C.15) converges to the solutions of

$$\left| \lambda\Sigma_{33} - \Sigma_{303} \right| \left| \lambda(1-\gamma^2)\Sigma_{33}^u \right| \left| \lambda \int_0^\rho GG'du \right| \left| \lambda \int_\rho^1 GG'du \right| = 0.$$

which has $r$ positive real solutions from the first term, since $\Sigma_{\beta\beta}^{-\frac{1}{2}}\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}\Sigma_{\beta\beta}^{-\frac{1}{2}}$ is a real symmetric positive definite matrix, and $r + 2(p - r)$ zero solutions (almost surely) from the last three terms.

So the space spanned by the eigenvectors corresponding to the $r$ largest eigenvalues of (C.14) converges in probability to the space spanned by the first $r$ unit vectors. Hence the space spanned by the eigenvector corresponding to the $r$ largest eigenvalues of (C.13) converges to the space spanned by the first $r$ columns of $A_T^{-1} = (\bar{B}.\bar{B}_0.T^{\frac{1}{2}}\bar{B}_1)'$. which is the space spanned by the true parameter values $B$. since $\bar{B} = \frac{1}{2}(\bar{\beta}.\bar{\beta})$. Let $c$ be a $2p \times r$ matrix with full column rank. and such that $c'B$ has full rank. The result above shows that the normalized estimator, $\hat{B}_c = \hat{B}(c'\hat{B})^{-1}$. is consistent for $B^o = B(c'B)^{-1}$. where the chosen normalization is $c'B = I$.

One can (in theory) normalize with respect to the true parameter. that is

$$\tilde{B} = \hat{B}\left(\bar{B}'\hat{B}\right)^{-1}.$$

We then have that

$$A_T^{-1}\tilde{B} = (\bar{B}.\bar{B}_0.T^{\frac{1}{2}}\bar{B}_1)'\tilde{B} = (I.U_{0T}.T^{\frac{1}{2}}U_{1T}) \xrightarrow{P} (I.0.0)$$

which shows that $U_{0T} = \bar{B}_0'\tilde{B} = o_p(1)$ and $U_{1T} = \bar{B}_1'\tilde{B} = o_p(T^{-\frac{1}{2}})$.

From the identity

$$\tilde{B} = B + B_0\bar{B}_0'\tilde{B} + B_1\bar{B}_1'\tilde{B} = B + B_0U_{0T} + B_1U_{1T}$$

we have $\tilde{B} - B = B_0U_{0T} + B_1U_{1T}$ and

$$
\begin{aligned}
\tilde{B}'\tilde{S}_{11}\tilde{B} &= (B + B_0U_{0T} + B_1U_{1T})'\tilde{S}_{11}(B + B_0U_{0T} + B_1U_{1T}) \\
&= B\tilde{S}_{11}B + o_p(1) = \Sigma_{\beta\beta} + o_p(1)
\end{aligned}
$$

where the last equality follows from Lemma C.1. Similarly $\tilde{B}'\tilde{S}_{10} = \Sigma_{30} + o_p(1)$. so the consistency of $\tilde{\alpha}$ follows from

$$\tilde{\alpha} = \tilde{S}_{01}\tilde{B}(\tilde{B}'\tilde{S}_{11}\tilde{B})^{-1} \xrightarrow{P} \Sigma_{03}\Sigma_{33}^{-1} = \alpha.$$

see (C.1). and the consistency of $\hat{\Omega}$ follows from

$$\hat{\Omega} = S_{00} - \tilde{S}_{01}\tilde{B}(\tilde{B}'\tilde{S}_{11}\tilde{B})^{-1}\tilde{B}'\tilde{S}_{10} \xrightarrow{P} \Sigma_{00} - \Sigma_{03}\Sigma_{33}^{-1}\Sigma_{30} = \Omega.$$

see (C.2). ∎

**Lemma C.3.** $\tilde{B}$ has a mixed Gaussian asymptotic distribution.

**Proof.** The likelihood equations for $\alpha$ and $B$ are given by $\left(\tilde{S}_{01} - \tilde{\alpha}B'\tilde{S}_{11}\right)\tilde{B} = 0$ and $\tilde{\alpha}'\hat{\Omega}(\tilde{S}_{01} - \tilde{\alpha}\tilde{B}'\tilde{S}_{11}) = 0$, so by inserting $\tilde{S}_{01} = \alpha B'\tilde{S}_{11} + \tilde{S}_{\varepsilon1}$ we find

$$0 = \tilde{S}_{\varepsilon1}\tilde{B} - (\tilde{\alpha} - \alpha)\tilde{B}'\tilde{S}_{11}\tilde{B} - \alpha\left(\tilde{B} - B\right)'\tilde{S}_{11}\tilde{B} \tag{C.17}$$

$$0 = \tilde{\alpha}'\hat{\Omega}^{-1}(\tilde{S}_{\varepsilon1} - \tilde{\alpha}\left(\tilde{B} - B\right)'\tilde{S}_{11} - (\tilde{\alpha} - \alpha)B'\tilde{S}_{11}). \tag{C.18}$$

Note that

$$(\tilde{B} - B)'\tilde{S}_{11}B_1 = (U_{0T}'B_0' + TU_{1T}'T^{-1}B_1')\tilde{S}_{11}B_1$$
$$= U_{0T}'O_p(1) + TU_{1T}'[T^{-1}B_1'\tilde{S}_{11}B_1]$$
$$= TU_{1T}'\int_0^1 GG'du + o_p(1).$$

so by multiplying equation (C.18) by $B_1$ from the right we find

$$
\begin{aligned}
0 &= \tilde{\alpha}'\hat{\Omega}^{-1}\left[\tilde{S}_{\epsilon 1} - \tilde{\alpha}\left(U'_{0T}B'_0 + TU'_{1T}T^{-1}B'_1\right)\tilde{S}_{11} - (\tilde{\alpha} - \alpha)B'\tilde{S}_{11}\right]B_1 \\
&= \alpha'\Omega^{-1}\left[\int dWG' - \alpha\,[TU'_{1T}]\left(\int GG'du'\right) + o_p(1)\right].
\end{aligned}
$$

by the consistency of $\tilde{\alpha}$ and $\hat{\Omega}$. Hence the asymptotic distribution of $U_{1T}$ is given by

$$
TU_{1T} \overset{w}{\to} \left(\int GG'du'\right)^{-1}\int GdW'\Omega^{-1}\alpha\left(\alpha'\Omega^{-1}\alpha\right)^{-1}
$$

which is mixed Gaussian, and we have shown that $U_{1T} = O_p(T^{-1})$.

In the last part of the proof, we make use of the following two results

$$
\begin{aligned}
(\tilde{B} - B)'\tilde{S}_{11}\tilde{B} &= (U'_{0T}B'_0 + U'_{1T}B'_1)\tilde{S}_{11}(B + U'_{0T}B'_0 + U'_{1T}B'_1) \\
&= U'_{0T}B'_0\tilde{S}_{11}B + O_p(T^{-1}) \\
&= U'_{0T}\gamma\Sigma_{33} + o_p(T^{-1/2}).
\end{aligned}
$$

and

$$
\begin{aligned}
(\tilde{B} - B)'\tilde{S}_{11}B_0 &= (U'_{0T}B'_0 + U'_{1T}B'_1)\tilde{S}_{11}B_0 \\
&= U'_{0T}B'_0\tilde{S}_{11}B_0 + O_p(T^{-1}) \\
&= U'_{0T}\Sigma^7_{33} + o_p(T^{-1/2}).
\end{aligned}
$$

Next, let $U'_{\alpha T} \equiv \left(\alpha'\Omega^{-1}\alpha\right)^{-1}\alpha'\Omega^{-1}(\tilde{\alpha} - \alpha)$, and multiply (C.17) by $T^{\frac{1}{2}}\left(\alpha'\Omega^{-1}\alpha\right)^{-1}\alpha'\Omega^{-1}$ from the left and (C.18) by $\left(\alpha'\Omega^{-1}\alpha\right)^{-1}$ from the left and by $\left(T^{\frac{1}{2}}B_0\right)$ from the right. This yields the

following two equations

$$\left(\alpha'\Omega^{-1}\alpha\right)^{-1}\alpha'\Omega^{-1}T^{\frac{1}{2}}\tilde{S}_{\epsilon 1}\tilde{B} = T^{\frac{1}{2}}U'_{\alpha T}\Sigma_{\jmath\jmath} + T^{\frac{1}{2}}U'_{0T}\gamma\Sigma_{\jmath\jmath} + o_p(1)$$

$$\left(\alpha'\Omega^{-1}\alpha\right)^{-1}\alpha'\Omega^{-1}T^{\frac{1}{2}}\tilde{S}_{\epsilon 1}B_0 = T^{\frac{1}{2}}U'_{\alpha T}\gamma\Sigma_{\jmath\jmath} + T^{\frac{1}{2}}U'_{0T}\Sigma^{\gamma}_{\jmath\jmath} + o_p(1).$$

which can be expressed in matrix form

$$\begin{pmatrix} \Sigma_{\jmath\jmath} & \gamma\Sigma_{\jmath\jmath} \\ \gamma\Sigma_{\jmath\jmath} & \Sigma^{\gamma}_{\jmath\jmath} \end{pmatrix} \begin{pmatrix} T^{\frac{1}{2}}U_{\alpha T} \\ T^{\frac{1}{2}}U_{0T} \end{pmatrix} = \begin{pmatrix} \tilde{B}' \\ B'_0 \end{pmatrix} \tilde{S}_{1\epsilon}\Omega^{-1}\alpha\left(\alpha'\Omega^{-1}\alpha\right)^{-1} + o_p(1)$$

such that

$$\begin{pmatrix} T^{\frac{1}{2}}U_{\alpha T} \\ T^{\frac{1}{2}}U_{0T} \end{pmatrix} \xrightarrow{w} \begin{pmatrix} \Sigma_{\jmath\jmath} & \gamma\Sigma_{\jmath\jmath} \\ \gamma\Sigma_{\jmath\jmath} & \Sigma^{\gamma}_{\jmath\jmath} \end{pmatrix}^{-1} Z_{bb0}\Omega^{-1}\alpha\left(\alpha'\Omega^{-1}\alpha\right)^{-1}$$

where $Z_{bb0} \sim \mathcal{N}\left(0, \begin{pmatrix} \Sigma_{\jmath\jmath} & \gamma\Sigma_{\jmath\jmath} \\ \gamma\Sigma_{\jmath\jmath} & \Sigma^{\gamma}_{\jmath\jmath} \end{pmatrix} \times \Omega\right)$. The asymptotic normality of $(\tilde{B}, B_0)'\tilde{S}_{1\epsilon}$ can be proven the same way as (C.11) was proven.

Isolating $U_{0T}$ and $U_{1T}$ we find that

$$\left(T^{\frac{1}{2}}U_{0T}, TU_{1T}\right) = \left(T^{\frac{1}{2}}\tilde{B}_0, T\tilde{B}_1\right)'(\tilde{B} - B) \xrightarrow{w} \begin{pmatrix} Z_\epsilon \\ \left[\int_0^\rho GG'du\right]^{-1}\int_0^\rho G\,(dW)' \\ \left[\int_\rho^1 GG'du\right]^{-1}\int_\rho^1 G\,(dW)' \end{pmatrix} \Omega^{-1}\alpha\left(\alpha'\Omega^{-1}\alpha\right)^{-1}$$

which shows that $\tilde{B}$ is mixed Gaussian, with a partly non-stochastic mixing parameter. ∎

## C.2. Expansion and Asymptotic Distribution of the $LR$ Test

**Proof of Theorem 3.2.2.** Recall that $\tilde{\alpha}\tilde{B}' = \hat{\alpha}\hat{B}'$ so that

$$
\begin{aligned}
\hat{\Omega} &= S_{00} - \tilde{\alpha}\tilde{B}'\tilde{S}_{11}\tilde{B}\tilde{\alpha}_0' \\
&= S_{00} - \tilde{S}_{01}\tilde{B}(\tilde{B}'\tilde{S}_{11}\tilde{B})^{-1}\tilde{B}'\tilde{S}_{10}
\end{aligned}
$$

where we substituted in for $\tilde{\alpha} = \tilde{S}_{01}\tilde{B}(\tilde{B}'\tilde{S}_{11}\tilde{B})^{-1}$. and note that

$$
\begin{aligned}
|\hat{\Omega}| &= |S_{00} - \tilde{S}_{01}\tilde{B}(\tilde{B}'\tilde{S}_{11}\tilde{B})^{-1}\tilde{B}'\tilde{S}_{10}| \\
&= |S_{00}|\frac{|\tilde{B}'(\tilde{S}_{11} - \tilde{S}_{10}S_{00}^{-1}\tilde{S}_{01})\tilde{B}|}{|\tilde{B}'\tilde{S}_{11}\tilde{B}|}.
\end{aligned}
$$

For the restricted model, where $B$ is known, we similarly have that

$$
|\hat{\Omega}_0| = |S_{00}|\frac{|B'(\tilde{S}_{11} - \tilde{S}_{10}S_{00}^{-1}\tilde{S}_{01})B|}{|B'\tilde{S}_{11}B|}.
$$

So the quotient test is given by

$$
Q^{-2/T} = \frac{|B'(\tilde{S}_{11} - \tilde{S}_{10}S_{00}^{-1}\tilde{S}_{01})B|}{|B'\tilde{S}_{11}B|} \Bigg/ \frac{|\tilde{B}'(\tilde{S}_{11} - \tilde{S}_{10}S_{00}^{-1}\tilde{S}_{01})\tilde{B}|}{|\tilde{B}'\tilde{S}_{11}\tilde{B}|}.
$$

If we define

$$
\begin{aligned}
C_T &= A_T'(\tilde{S}_{11} - \tilde{S}_{10}S_{00}^{-1}\tilde{S}_{01})A_T \\
D_T &= A_T'\tilde{S}_{11}A_T.
\end{aligned}
$$

where $A_T = (B, B_0, T^{-1/2}B_1)$, we have that $C_T$ and $D_T$ are $O_p(1)$. Now. $(\tilde{B} - B) = B_0 U_{0T} + B_1 U_{1T} = A_T U_T$, where $U_T = (0, U'_{0T}, T^{1/2}U'_{1T})' = O_p(T^{-1/2})$, so we have

$$
\begin{aligned}
(\tilde{B} - B)' \tilde{S}_{11} (\tilde{B} - B) &= U'_T D_T U_T = O_p(T^{-1}) \\
(\tilde{B} - B)' \tilde{S}_{11} B &= U'_{0T} B'_0 \tilde{S}_{11} B = O_p(T^{-1/2}) \\
(\tilde{B} - B)' \tilde{S}_{10} &= U'_{0T} B'_0 \tilde{S}_{10} + O_p(T^{-1}) = O_p(T^{-1/2})
\end{aligned}
$$

and similar identities involving $\tilde{S}_{11} - \tilde{S}_{10}S_{00}^{-1}\tilde{S}_{01}$.

We use the expansion. taken from Johansen (1996), of $f(x) = |x'Mx|/|x'Nx|$:

$$
\begin{aligned}
\log f(x + h) &= \log f(x) \\
&\quad -\mathrm{tr}\{(x'Nx)^{-1}h'(N - Nx(x'Nx)^{-1}x'N)h\} \\
&\quad +\mathrm{tr}\{(x'Mx)^{-1}h'(M - Mx(x'Mx)^{-1}x'M)h\} \\
&\quad +O(\|h\|^3).
\end{aligned}
$$

where $O(\|h\|) = \max_{i,j} |h_{i,j}|$. In our case we have $x = B$. $h = \tilde{B} - B$. $M = \tilde{S}_{11} - \tilde{S}_{10}S_{00}^{-1}\tilde{S}_{01}$. and $N = \tilde{S}_{11}$. The first term is given by

$$
\begin{aligned}
&\mathrm{tr}\{(B'\tilde{S}_{11}B)^{-1}(\tilde{B} - B)'(\tilde{S}_{11} - \tilde{S}_{11}B(B'\tilde{S}_{11}B)^{-1}B'\tilde{S}_{11})(\tilde{B} - B)\} \\
&= \mathrm{tr}\{\Sigma_{33}^{-1}(U'_T D_T U_T - U''_{0T}\gamma\Sigma_{33}\Sigma_{33}^{-1}\Sigma_{33}\gamma U_{0T})\} + o_p(T^{-1}) \\
&= \mathrm{tr}\{\Sigma_{33}^{-1}(U'_T D_T U_T - \gamma^2 U'_{0T}\Sigma_{33}U_{0T})\} + o_p(T^{-1}).
\end{aligned}
$$

and similarly we find the second term to be

$$
\mathrm{tr}\{(\Sigma_{33} - \Sigma_{303})^{-1}(U'_T C_T U_T - \gamma^2 U'_{0T}(\Sigma_{33} - \Sigma_{303})U_{0T})\} + o_p(T^{-1}).
$$

Since $h = (\tilde{B} - B)$ is $O_p(T^{-1/2})$, the remaining term $O(\|h\|^3)$ is $O_p(T^{-3/2})$.

Now $U_T' C_T U_T = U_T' D_T U_T - \gamma^2 U_{0T}' \Sigma_{\vartheta 0 \vartheta} U_{0T} + o_p(T^{-1})$, so the two terms nicely add up to

$$\text{tr}\{[(\Sigma_{\vartheta\vartheta} - \Sigma_{\vartheta 0 \vartheta})^{-1} - \Sigma_{\vartheta\vartheta}^{-1}](U_T' D_T U_T - \gamma^2 U_{0T}' \Sigma_{\vartheta\vartheta} U_{0T})\} + o_p(T^{-1})$$

$$= \text{tr}\{\alpha' \Omega^{-1} \alpha (U_T' D_T U_T - \gamma^2 U_{0T}' \Sigma_{\vartheta\vartheta} U_{0T})\} + o_p(T^{-1}).$$

where we used (C.3).

Finally,

$$U_T' D_T U_T - \gamma^2 U_{0T}' \Sigma_{\vartheta\vartheta} U_{0T} = (U_{0T}' . T^{1/2} U_{1T}') \Xi \begin{pmatrix} U_{0T} \\ T^{1/2} U_{1T} \end{pmatrix} + o_p(T^{-1}).$$

where

$$\Xi = \begin{pmatrix} (1 - \gamma^2)\Sigma_{\vartheta\vartheta}^u & 0 & 0 \\ 0 & \int_0^\rho GG'du & 0 \\ 0 & 0 & \int_\rho^1 GG'du \end{pmatrix}.$$

(see (C.16) for the upper left element). We can conclude that

$$-2 \log Q = \text{tr}\{Z_T Z_T'\} + o_p(1).$$

where

$$Z_T = (\alpha' \Omega^{-1} \alpha)^{\frac{1}{2}} \left( T^{\frac{1}{2}} U_{0T}' . T U_{1T}' \right) \Xi^{\frac{1}{2}}$$

is asymptotically $N(0, I_r \otimes I_{r+2(p-r)})$. which shows that

$$-2 \log Q \xrightarrow{w} \chi^2(2pr - r^2).$$

∎

Consider the test of $\beta_1 = \beta_2 = \beta_0$ against $\beta_1 = \beta_2$. By applying an orthogonalization argument we can give the last proof.

**Proof of Theorem 3.2.3.** The restriction of $\beta_1 = \beta_2$ can be expressed as the linear restriction $B = H\beta$ where $H = (I_p, I_p)'$. Define the projection matrix $P_H = H(H'H)^{-1}H'$. then the likelihood ratio test of the simple hypothesis of $\beta_1 = \beta_2 = \beta_o$ against $\beta_1 = \beta_2$ can be expressed as

$$
\begin{aligned}
&- 2 \log Q(\beta_1 = \beta_2 = \beta_o | \beta_1 = \beta_2) \\
&\quad = T \text{tr} \left\{ \alpha' \Omega^{-1} \alpha (U'_{0T}.T^{1/2}U'_{1T}) P_H \tilde{S}_{11} P_H (U'_{0T}.T^{1/2}U'_{1T})' \right\} + o_p(1) \\
&\quad\quad = tr \left\{ Z_T M (M'M)^{-1} M Z'_T \right\} + O_p(T^{-\frac{1}{2}}).
\end{aligned}
$$

for some $r + 2(p-r) \times p$ matrix $M$ with full rank $p$. This shows that

$$
\begin{aligned}
-2 \log Q(\beta_1 = \beta_2 | B) &= tr \{ Z_T Z'_T \} - tr \left\{ Z_T M (M'M)^{-1} M Z'_T \right\} + O_p(T^{-\frac{1}{2}}) \\
&= tr \left\{ Z_T M_\perp (M'_\perp M_\perp)^{-1} M_\perp Z'_T \right\} + O_p(T^{-\frac{1}{2}}) \\
&\xrightarrow{w} \chi^2(pr).
\end{aligned}
$$

This completes the proof. ∎

# References

ANDREWS. D. W. K. (1993): "Test for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61. 821–856.

ANDREWS. D. W. K., AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

BILLINGSLEY. P. (1999): *Convergence of Probability Measures.* John Wiley and Sons, New York, 2nd edn.

CAMPOS. J.. N. R. ERICSSON, AND D. F. HENDRY (1996): "Cointegration Tests in the Presence of Structural Breaks," *Journal of Econometrics,* 70, 187–220.

GREGORY. A. W.. AND B. E. HANSEN (1996): "Residual-Based Test for Cointegration in Models with Regime Shifts," *Journal of Econometrics,* 70, 99–126.

GREGORY. A. W., J. M. NASON, AND D. G. WATT (1996): "Testing for Structural Breaks in Cointegrated Relationships," *Journal of Econometrics,* 71, 321–341.

HANSEN. B. E. (1992a): "Test for Instability with $I(1)$ Processes," *Journal of Business and Economic Statistics,* 10, 321–335.

——— (1992b): "Testing for Parameter Instability in Linear Models," *Journal of Policy Modeling,* 14, 517–533.

——— (1996): "Inference When a Nuisance Parameter is Not Identified under the Null Hypothesis," *Econometrica,* 64, 413–430.

HANSEN. H.. AND S. JOHANSEN (1999): "Some Tests for Parameter Constancy in Cointegrated VAR-Models," *Econometrics Journal,* 2, 306–333.

HANSEN. H.. AND K. JUSELIUS (1995): *CATS in RATS. Cointegration Analysis of Time Series.* Estima. Evanston. IL.

HENDRY. D. F. (1984): "Monte Carlo Experimentation," in *Handbook of Econometrics,* ed. by Z. Griliches. and M. D. Intriligator. Elsevier. vol. 2, pp. 937–976.

JOHANSEN. S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control,* 12, 231–254.

——— (1996): *Likelihood Based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press. Oxford, 2nd edn.

——— (1999a): "A Bartlett Correction Factor for Tests on the Cointegrating Relations." *EUI Working Paper ECO No. 99/10.*

——— (1999b): "A Small Sample Correction for Tests of Hypotheses on the Cointegrating Vectors." *EUI Working Paper ECO No. 99/9.*

KUO. K.-S. (1998): "Test for Partial Parameter Instability in Regressions with $I(1)$ Processes." *Journal of Econometrics,* 86, 337–368.

LUMSDAINE. R. L.. AND D. H. PAPELL (1997): "Multiple Trend Breaks and the Unit-Root Hypothesis." *Review of Economics and Statistics,* 79, 212–218.

NELSON. C. R.. AND C. I. PLOSSER (1982): "Trends and Random Walks in Macro-Economic Time Series: Some Evidence and Implications," *Journal of Monetary Economics,* 10. 139–162.

NYBLOM. J. (1989): "Testing for the Constancy of Parameters over Time." *Journal of the American Statistical Association,* 84, 223–230.

PERRON. P. (1989): "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis." *Econometrica,* 57, 1361–1401.

———— (1990): "Testing for a Unit Root in a Time Series with a Changing Mean," *Journal of Business and Economic Statistics*, 8, 153–162.

QUINTOS. C. E. (1997): "Stability Tests in Error Correction Models," *Journal of Econometrics*, 82. 289–315.

SEO. B. (1998): "Tests for Structural Change in Cointegrated Systems," *Econometric Theory*. 14. 222–259.

ZIVOT. E., AND D. W. K. ANDREWS (1992): "Further Evidence on the Great Crash. the Oil-Price Shock. and the Unit-Root Hypothesis," *Journal of Business and Economic Statistics*. 10. 251–270.

# Chapter 4

## Determining the Cointegration Rank

## in Processes with Structural Changes[†]

**Abstract**

It is well known that unit root tests are affected by structural changes in the parameters. This chapter provides a general framework for determining the extent of stationarity versus unit roots in a multivariate time series with structural changes. Changes in the mean, trend, slope coefficients, and covariance matrix are all special cases of this framework.

I derive the likelihood ratio test for determination of the cointegration rank in the vector autoregressive model with changing parameters. Its asymptotic distribution is shown to be a convex combination of Dickey-Fuller distributions, when the change points are taken as given. Some tests for the case with unknown changes points are suggested and discussed.

---

[†] I thank Sören Johansen for many valuable suggestions. All errors remain my responsibility.

107

## 4.1. Introduction

It is well known that unit root tests and tests for cointegration can be misleading if the underlying process has structural changes that are not accounted for. A shift in the mean in a univariate process causes Dickey-Fuller type tests to accept the null of a unit root, (see Perron (1990)).

In this paper, I derive the likelihood ratio (LR) test to determine the cointegration rank in a multivariate framework where the number of structural changes may be any integer. The cointegration rank may differ across the regimes. The ranks in the individual regimes are determined simultaneously. Situations with known and unknown change points are both treated, and the framework allows for changes in all parameters, multiple changes, and multiple cointegration relations.

The framework used is the vector autoregressive (VAR) model with parameters that may change their value at the change points. The LR test for the number of cointegration relations is similar to the one of the standard model without changes. When the change points are known, the asymptotic distribution is a weighted average of Dickey-Fuller distributions, known from the standard model. When the change points are unknown, one can use tests that are calculated from the sequence of LR tests. Such tests include the supremum of the tests (Sup-LR) and the average of the tests (Ave-LR).

In models with structural changes, there are two types of problems. One is to determine the number of change points and estimate when the changes occurred, (see Bai (1997, 1999), and Bai and Perron (1998)). Another problem is to analyze the qualitative changes in the parameters, and test for constancy of parameters. The early studies in this field, starting with Chow (1960) took the change points as given, whereas the situation with an unknown change point has been analyzed by Quandt (1960), Nyblom (1989), Andrews (1993), and Andrews and Ploberger (1994). In relation to cointegrated processes, parameter stability has been analyzed by B. E. Hansen (1992a, 1992b), Gregory and B. E. Hansen (1996), Seo (1998), Quintos (1995, 1997), H. Hansen and Johansen (1999), and P. R. Hansen (2000c).

The problem treated in this chapter, is the additional problem that arises from cointegration models with structural changes. That is to determine the extent of cointegration versus stochastic $I(1)$-trends. Naturally, a test for constant parameters is indirectly a test for a constant rank as analyzed by Quintos (1997) and by Inoue (1999) who derived the rank test (test for the number of cointegrating relations) in the situation where the process may have a broken deterministic trend. This chapter generalizes this problem to a situation with multiple structural changes, where the changes may affect any parameter, including the linear trend. I provide a test for determining the cointegration rank that may change at the change points: the test for constant rank is a special case of this test. The method allows for some parameters to be held constant across some (or all) subsamples. This is often desired in practice to avoid an overfit by having too many free parameters. A structural change in all parameters is easy to estimate using parameter estimates based on the different subsamples, whereas the case with a partial structural change (some parameters held constant) can be solved with techniques developed by Boswijk (1995) or the generalized versions thereof, derived in Chapter 5.

## 4.2. The Statistical Model

We consider the $p$-dimensional vector autoregressive model with structural changes in the parameters which may change their values at change points. The case with $m$ structural changes results in $m + 1$ distinct sub-samples with $m + 1$ (possibly) different parameter values. The time of the change points is denoted by $T_1, \ldots, T_m$ where $0 < T_1 < \cdots < T_m < T$, such that subsample $j$, is given by $T_{j-1} + 1, \ldots, T_j$, $j = 1, \ldots, m + 1$, where $T_0 = 0$ and $T_{m+1} = T$.

In the error correction form the model is given by

$$\Delta X_t = \Pi_j X_{t-1} + \sum_{i=1}^{k-1} \Gamma_{j,i} \Delta X_{t-i} + \Phi_j D_t + \varepsilon_t, \quad t = T_{j-1} + 1, \ldots, T_j.$$

where $\varepsilon_t$ is assumed to be independent and Gaussian distributed with mean zero and variance

$\Omega$. The variable $D_t$ contains deterministic terms such as a constant, a linear trend and seasonal dummies.

For each of the subsamples, $j = 1, \ldots, m + 1$, we assume that the characteristic polynomial, $A_j(z) = I(1-z) - \Pi_j z - \sum_{i=1}^{k-1} \Gamma_{j,i}(1-z)z^i$, has its roots outside the unit circle or at one ($z = 1$), and that the number of unit roots equals the reduced rank of $\Pi_j$, $j = 1, \ldots, m + 1$. When there is at least one unit root, these conditions ensure that $X_t$ is integrated of order one within the subsample. (see Johansen (1996)).

If the number of unit roots are less than $p$, then the process is cointegrated. We denote the (cointegration) rank of $\Pi_j$ by $r_j$, and write $\Pi_j = \alpha_j \beta_j'$ where $\alpha_j$ and $\beta_j$ are $p \times r_j$ matrices with full column rank, $j = 1, \ldots, q$. As shown in Johansen (1988), $\beta_j$ defines the $r_j$ cointegrating relations whereas $\alpha_j$ can be interpreted as the adjustment coefficients.

### 4.2.1. Estimation

In the situation where all the parameters $\alpha(t), \beta(t), \Gamma_1(t), \ldots, \Gamma_{k-1}(t)$ and $\Phi(t)$ have structural changes, the model is easily estimated using the reduced rank regression techniques of the standard model, applied to each of the subsamples.

The parameter estimation under partial structural changes is slightly more complicated. Consider first the case where only $\alpha(t)$ and $\beta(t)$ have structural changes.

We define the indicator functions $I_j(t) = I(T_{j-1}+1 \leq t \leq T_j)$ for $j = 1, \ldots, m+1$, and with the conventions $Z_{0t} = \Delta X_t$, $Z_{1t} = (X_{t-1}'I_1(t), \ldots, X_{t-1}'I_{m+1}(t))'$ and $Z_{2t} = (\Delta X_{t-1}', \ldots, \Delta X_{t-k+1}', D_t')'$, the model can be expressed as:

$$Z_{0t} = (\alpha_1, \ldots, \alpha_{m+1})\beta' Z_{1t} + \Psi Z_{2t} + \varepsilon_t.$$

where $\Psi = (\Gamma_1, \ldots, \Gamma_{k-1}, \Phi)$ and where

$$\beta \equiv \begin{pmatrix} \beta_1 & 0 & \cdots & 0 & 0 \\ 0 & \beta_2 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & \beta_m & 0 \\ 0 & 0 & \cdots & 0 & \beta_{m+1} \end{pmatrix}.$$

This structure of $\beta$ leads to a generalized reduced rank regression problem, because the structure can be formulated by the linear restrictions $\text{vec}(\beta) = H\varphi$, where $H$ is a known matrix and $\varphi$ contains the free parameters in $\beta$. To simplify notation we define $\alpha = (\alpha_1, \ldots, \alpha_{m+1})$. The solution to this estimation problem is given by the corollary below, taken from P. R. Hansen (2000c).

Consider the regression problem $Z_{0t} = \alpha\beta' Z_{1t} + \Psi Z_{2t} + \varepsilon_t$ where $\beta$ is restricted by $\text{vec}(\beta) = H\phi$ and where $Z_{1t}$ and $Z_{2t}$ are $\mathcal{F}_{t-1}$-measurable and $\{\varepsilon_t\}$ is a sequence of i.i.d. Gaussian variables with mean zero and variance $\Omega$. The $\sigma$-algebra, $\mathcal{F}_t$, is generated by $Z_{01}, Z_{02}, \ldots, Z_{0t}$ and initial values $(Z_{11}, Z_{21})$, $t = 1, \ldots, T$.

Define the moment matrices $M_{ij} = \frac{1}{T}\sum_{t=1}^T Z_{it}Z_{jt}'$. $i, j = 0, 1, 2$. and the residuals $R_{0t} = Z_{0t} - M_{02}M_{22}^{-1}Z_{2t}$. $R_{1t} = Z_{1t} - M_{12}M_{22}^{-1}Z_{2t}$. and the moment matrices of the residuals $S_{ij} = \frac{1}{T}\sum_{t=1}^T R_{it}R_{jt}'$. $i, j = 0, 1$.

**Corollary 4.2.1.** *With the conventions given above, the maximum likelihood estimates satisfy*

*the equations*

$$\mathrm{vec}(\hat{\beta}(\hat{\alpha}.\hat{\Omega})) = H\left[H'\left(\hat{\alpha}'\hat{\Omega}^{-1}\hat{\alpha} \otimes S_{11}\right)H\right]^{-1}H'\left(\hat{\alpha}' \otimes S_{10}\right)\mathrm{vec}\left(\hat{\Omega}^{-1}\right) \qquad (4.2.1)$$

$$\hat{\alpha}(\hat{\beta}) = S_{01}\hat{\beta}\left(\hat{\beta}'S_{11}\hat{\beta}\right)^{-1}. \qquad (4.2.2)$$

$$\hat{\Omega}(\hat{\beta}) = S_{00} - S_{01}\hat{\beta}\left(\hat{\beta}'S_{11}\hat{\beta}\right)^{-1}\hat{\beta}'S_{10}. \qquad (4.2.3)$$

$$\hat{\Psi} = M_{02}M_{22}^{-1} - \hat{\alpha}\hat{\beta}'M_{12}M_{22}^{-1}. \qquad (4.2.4)$$

The maximum value of the likelihood function is given by

$$L_{\max}^{-2/T}(\hat{\alpha}.\hat{\beta}.\hat{\Psi}.\hat{\Omega}) = (2\pi e)^p |\hat{\Omega}|.$$

where $p$ is the dimension of the process $Z_{0t}$.

Parameter estimates are found by iterating on equations (4.2.1–4.2.4). starting from some initial values for the parameters. and is in nature similar to the *switching algorithm* by Johansen and Juselius (1992). For more on this estimation technique see Boswijk (1995) or P. R. Hansen (2000c).

If the variance is not the same across subsamples. the likelihood equations are slightly more complicated. In fact the likelihood equations are now given by

$$\mathrm{vec}(\hat{\beta}) = H\left[H'\mathbf{Z}_{1A}'\Sigma^{-1}\mathbf{Z}_{1A}H\right]^{-1}H'\mathbf{Z}_{1A}'\Sigma^{-1}\mathrm{vec}(Z_0 - CZ_2)$$

$$\mathrm{vec}(\hat{\alpha}.\hat{\Psi}) = \left[\mathbf{Z}_{1B2}'\Sigma^{-1}\mathbf{Z}_{1B2}\right]^{-1}\mathbf{Z}_{1B2}'\Sigma^{-1}\mathrm{vec}(Z_0).$$

$$\hat{\Omega}_j = (T_j - T_{j-1})^{-1}\sum_{t=T_{j-1}+1}^{T_j}\hat{\varepsilon}_t\hat{\varepsilon}_t'.$$

$$\hat{\varepsilon}_t = Z_{0t} - \hat{\alpha}\hat{\beta}Z_{1t} - \hat{\Psi}Z_{2t}.$$

where $\Sigma$ is the $Tp \times Tp$ block diagonal matrix with $\Omega_1$ in the $T_1$ first blocks. $\Omega_2$ in the next $T_2 - T_1$

and so forth. where

$$\mathbf{Z}_{1B2} = ((Z_1'\hat{\boldsymbol{\beta}}. Z_2') \otimes I_p)).$$

$$\mathbf{Z}_{1A} = (Z_1' \otimes \hat{\alpha})K_{p_1,r}.$$

and where $Z_0 = (Z_{01}, \ldots, Z_{0T})$, $Z_1 = (Z_{11}, \ldots, Z_{1T})$. and $Z_2 = (Z_{21}, \ldots, Z_{2T})$, (see P. R. Hansen (2000a)).

The maximum value of the likelihood function is given by

$$L_{\max}^{-2/T}(\hat{\alpha}, \hat{\beta}, \hat{\Psi}, \hat{\Omega}_1, \ldots, \hat{\Omega}_{m+1}) = (2\pi e)^p |\hat{\Omega}_1|^{\rho_1} \cdots |\hat{\Omega}_{m+1}|^{\rho_{m+1}}.$$

where $\rho_j = (T_j - T_{j-1})/T$. denotes the proportion of observations in the $j$th subsample.

## 4.3. Asymptotic Analysis

We first derive the moving average representation of the process. from which we can derive stochastic properties of estimators and statistics.

For a $p \times r$ matrix $a$ with full column rank, we define its orthogonal complement. denoted by $a_\perp$. as the $p \times (p - r)$ matrix with full column rank that has $a_\perp' a = 0$.

Define $\bar{\Gamma}_j \equiv I - \sum_{i=1}^{k-1} \Gamma_{j,i}$. The standard $I(1)$ assumptions. stated above. implies that $\alpha_{j,\perp}' \bar{\Gamma}_j \beta_{j,\perp}$ has full rank $p - r_j$. and that the moving average representation for subsample $j$ is given by

$$X_t = C_j \sum_{i=1}^{t} \varepsilon_i + D_j(L)\varepsilon_t + C_j(X_{T_{j-1}} - \sum_{i=1}^{k-1} \Gamma_{j,i} X_{T_{j-1}-i}) \quad t = T_{j-1} + 1, \ldots, T_j.$$

where $C_j = \beta_{j,\perp} \left( \alpha_{j,\perp}' \bar{\Gamma}_j \beta_{j,\perp} \right)^{-1} \alpha_{j,\perp}'$ and where $D_j(L)\varepsilon_t$ is a stationary process. (see P. R. Hansen (2000b)). Let $\Upsilon_{j,i} \equiv C_j \bar{\Gamma}_j C_{j-1} \cdots \bar{\Gamma}_{j-i+1} C_{j-i}$ $(\Upsilon_{j,0} = C_j)$.

In order to get the representation in the appropriate form we need to express all representations

with initial values that are functions of $\{X_t,\ t = 0, -1, \ldots\}$, rather than $(X_{T_j}, \ldots, X_{T_j - k + 1})$, $j = 1, \ldots, m$. This representation is given in the following theorem.

**Theorem 4.3.1 (The Granger representation for change point processes).** *The moving average representation is given by*

$$X_t = \Upsilon_{j,0} \sum_{i=T_{j-1}+1}^{t} (\varepsilon_i + \Phi D_i) + \sum_{i=1}^{j-1} \Upsilon_{j,i} \sum_{i=T_{j-i-1}+1}^{T_{j-i}} (\varepsilon_i + \Phi D_i) + \delta_t + V_{j,t}.$$

*where $\delta_t$ is a deterministic variable satisfying $\delta_t / \max_{1 \leq i \leq t} D_i = O(1)$ and $V_{j,t}$ is a stationary process.*

By Donsker's invariance principle we have that $T^{-1/2} \sum_{i=1}^{[uT]} \varepsilon_i \xrightarrow{w} W(u)$, where $W(u)$ is a driftless Brownian motion with piecewise constant covariance matrix, given by $\Omega_j$ for $u \in [u_{j-1}, u_j)$, where $u_j = T_j/T$. In particular, for $u \in [u_{j-1}, u_j)$, we have that

$$T^{-1/2} \sum_{i=T_{j-1}+1}^{[uT]} \varepsilon_i \xrightarrow{w} W(u) - W(u_j).$$

So if we define $X_{1,uT} = X_{1,t}$, $u \in [\frac{t}{T}, \frac{t+1}{T})$, we have, in the case where $\Phi D_t = 0$,

$$T^{-1/2} X_{1,uT} \xrightarrow{w} (\Upsilon_{j,0}(W(u) - W(u_{j-1})) + \sum_{i=1}^{j-1} \Upsilon_{j,i}(W(u_i) - W(u_{i-1})), \quad u \in [u_{j-1}, u_j).$$

and in the case where $\Phi D_t = \mu_j I(T_{j-1} < t \leq T_j)$, $j = 1, \ldots, m+1$, we find

$$T^{-1/2}(X_{1,uT} - \bar{X}_j) \xrightarrow{w} \Upsilon_{j,0}[(W(u) - W(u_{j-1})) - \int_{u_{j-1}}^{u_j} (W(u) - W(u_{j-1})) du]$$

where $\bar{X}_j = T^{-1} \sum_{t=T_{j-1}+1}^{T_j} X_t$. The proportion of observations in the $j$th subsample, is given by

$\rho_j = u_j - u_{j-1}$, and it is convenient to define the rescaled Brownian motion

$$W_j(\lambda) = \frac{1}{\sqrt{\rho_j}}(W(u) - W(u_j)), \quad \lambda = \frac{u - u_{j-1}}{u_j - u_{j-1}}.$$

where $W_j(\lambda)$ is a Brownian motion on $[0,1]$ that has constant covariance matrix $\Omega_j$. $j = 1, \ldots$, $m + 1$. We then have that

$$T^{-1/2} \sum_{t=T_{j-1}+1}^{[uT]} \varepsilon_t \xrightarrow{w} \sqrt{\rho_j} W_j(\lambda), \quad \lambda = \frac{u - u_{j-1}}{u_j - u_{j-1}}.$$

so that

$$T^{-1/2}(X_{1,uT} - \bar{X}_j) \xrightarrow{w} \sqrt{\rho_j} \Upsilon_{j,0}[W_j(\lambda) - \bar{W}_j],$$

where $\bar{W}_j = \int_0^1 W_j(\lambda)du$, and we have that

$$T^{-2} \sum_{t=T_{j-1}+1}^{T_j} (X_{1,t} - \bar{X}_j)(X_{1,t} - \bar{X}_j)' \xrightarrow{w} \rho_j \Upsilon_{j,0} \int_0^1 (W_j - \bar{W}_j)(W_j - \bar{W}_j)'du \Upsilon'_{j,0}$$

$$= \rho_j \Upsilon_{j,0} F_j \Upsilon'_{j,0}.$$

where $F_j = \int_0^1 (W_j - \bar{W}_j)(W_j - \bar{W}_j)'du$. With this result, it is now clear that

$$T^{-1}S_{11} \xrightarrow{w} \begin{pmatrix} \rho_1 \Upsilon_{1,0} F_1 \Upsilon'_{1,0} & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \rho_j \Upsilon_{j,0} F_j \Upsilon'_{j,0} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \rho_{m+1} \Upsilon_{m+1,0} F_{m+1} \Upsilon'_{m+1,0} \end{pmatrix}.$$

If we set

$$S_{ab}^{(j)} = (T_j - T_{j-1})^{-1} \sum_{t=T_{j-1}+1}^{T_j} R_{at}R_{bt}', \quad a,b = 0,1$$

we similarly find that

$$S_{10}^{(j)}(\alpha_j)_\perp \overset{w}{\to} \rho_j \Upsilon_{j,0} \int_0^1 (W_j - \bar{W}_j)dW_j'(\alpha_j)_\perp.$$

The likelihood function can be concentrated to a function that only depends on $\beta$.

$$
\begin{aligned}
L_{conc.}(\beta) \quad &\propto \quad |\Omega(\beta)| = |S_{00} - S_{01}\beta(\beta'S_{11}\beta)^{-1}\beta'S_{10}| \\
&= \quad |S_{00}|\frac{|\beta'(S_{11} - S_{10}S_{00}^{-1}S_{01})\beta|}{|\beta'S_{11}\beta|} \\
&\propto \quad \frac{|\beta'(S_{11} - S_{10}S_{00}^{-1}S_{01})\beta|}{|\beta'S_{11}\beta|}.
\end{aligned}
$$

(see Johansen (1996)).

Define the concentrated likelihood function for subsample $j$, given by

$$L_j(\beta_j) = \frac{|\beta_j'(S_{11}^{(j)} - S_{10}^{(j)}S_{00}^{(j)-1}S_{01}^{(j)})\beta_j|}{|\beta_j'S_{11}^{(j)}\beta_j|}$$

and consider the pseudo likelihood function for the full sample

$$\tilde{L}(\beta_1,\ldots,\beta_{m+1}) = \prod_{j=1}^{m+1} \rho_j \frac{|\beta_j'(S_{11}^{(j)} - S_{10}^{(j)}S_{00}^{(j)-1}S_{01}^{(j)})\beta_j|}{|\beta_j'S_{11}^{(j)}\beta_j|}.$$

that weighs the individual likelihood functions by the proportion of observations in the corresponding subsamples. This pseudo likelihood function corresponds to a change in all parameters. and therefore ignores that some parameters may be constant across the subsamples in the true likelihood function.

It is easily verified that

$$L_{conc.}(\hat{\beta}) = \tilde{L}(\hat{\beta}_1, \dots, \hat{\beta}_{m+1}) + o_p(1),$$

which makes it easy to derive the asymptotic distribution of a likelihood based test, because it simplifies to a situation where the results from the standard model, without structural changes, can be used.

It is well known from the model without structural changes, that the likelihood ratio test for $r = a$ against $r = b$ $(b > a)$ has the asymptotic distribution

$$\chi_{df}(b - a) = \mathrm{tr}\left\{ \int_0^1 dBF' \left( \int_0^1 FF' \right)^{-1} \int_0^1 FdB' \right\}.$$

where $B$ is a $(b - a)$-dimensional standard Brownian motion, and $F$ depends on the deterministic term of the process. If an unrestricted constant is included in the process we have $F(u) = B(u) - \int_0^1 B(u)du$.

We define

$$LR_{T_1/T,\dots,T_m/T}(x_1, \dots, x_{m+1})$$

as the likelihood ratio test of $r_j = a_j$ against $r_j = b_j$, where $b_j - a_j = x_j$, $j = 1, \dots, m+1$. Since $W_i$ is independent of $W_j$, for $i \neq j$, we have that in a situation where $T_1, \dots, T_m$ (or $\rho_1, \dots, \rho_{m+1}$) are known, the asymptotic distribution is given by

$$\sum_{j=1}^{m+1} \rho_j \chi_{df}(x_j). \tag{4.3.1}$$

which is a convex combination of (squared) Dickey-Fuller distributions.

This distribution is not easy to tabulate, because it depends on the values of $\rho_1 < \rho_2 < \cdots < \rho_m$

as well as $x_j = 0. \ldots ,p - 1$. However, it might be well approximated by $\Gamma$-distributions as is the case for the Dickey-Fuller distribution, (see Nielsen (1997) and Doornik (1998)).

### 4.3.1. Testing Scheme for Rank Determination

Let $H(a_1, \ldots ,a_{m+1})$ denote the hypothesis that the cointegration rank in subsample $j$, denoted by $r_j$, equals the integer $a_j$, for $j = 1, \ldots ,m + 1$, and let $H_r$ be the hypothesis that the cointegration rank is constant and equal to $r$ across the subsamples $r_1 = \cdots = r_{m+1} = r$. The rank can then be determined by testing $H_r$ against $H_p$ for $r = 0, 1, \ldots$ until the first acceptance. Let first acceptance define $r^*$. Because the test statistic diverges to $\infty$ if $\max r_i > r$, this procedure picks $r^*$ such that $r^* = \max r_j$, with probability converging to $1 - \tilde{\alpha}$, where $\tilde{\alpha}$ is the size of the test, and $H_{r^*}$ contains the true model, with probability converging to one. Since $r_j$ may be smaller than $r^*$ for some $j$, one can proceed by testing $H(a, r^*, \ldots ,r^*)$ against $H_{r^*}$ for $a = 0, \ldots ,r^* - 1$ until first acceptance, and let this define $r_1^*$. Then test $H(r_1^*, a, r^*, \ldots ,r^*)$ against $H(r_1^*, r^*, \ldots ,r^*)$, thereby defining $r_2^*$, and so forth. The latter part of the testing procedure is arbitrary in the sense that one could have chosen another ordering of the $j$'s rather than starting with $r_1$.

### 4.3.2. Unknown Change Points

In this subsection we give some ideas on how a test for the cointegration rank can be constructed, in the case where the change points are unknown. Unless the change points are estimated it is uninformative to know that the rank was first two, say, and then three. Nevertheless the outcome may be that the rank is constant, and then insight is gained.

The following two statistics might be used to determine the cointegration ranks in the regimes

between the unknown change points:

$$\text{Sup } LR = \sup_{0 < u_1 < \cdots < u_m < 1} LR_{u_1, \ldots, u_m}(x_1, \ldots, x_{m+1})$$

$$\text{Ave } LR = n^{-1} \sum_{0 < T_1 < T_2 < \cdots < T_m < T} LR_{T_1/T, \ldots, T_m/T}(x_1, \ldots, x_{m+1}),$$

where $n$ is the number of elements that the sum is taken over.

Provided that regularity conditions holds (such as the probability measure being tight), then the statistics' asymptotic distributions are given by

$$\sup_{0 < u_1 < \cdots < u_m < 1} \left( \sum_{j=1}^{m+1} \rho_j Z_j^\rho \right) \tag{4.3.2}$$

and

$$\int_0^1 \int_{u_1}^1 \cdots \int_{u_m}^1 \left( \sum_{j=1}^{m+1} \rho_j Z_j^\rho \right) du_1 \cdots du_m. \tag{4.3.3}$$

where $Z_j^\rho \sim \chi_{df}(x_j)$.

The individual subsample statistics $(Z_1^\rho, \ldots, Z_{m+1}^\rho)$ are, for fixed $\rho = (\rho_1, \ldots, \rho_m)$, mutually independent. But for different segmentation of the unit interval, say $\rho = (\rho_1, \ldots, \rho_m)$ and $\tilde{\rho} = (\tilde{\rho}_1, \ldots, \tilde{\rho}_m)$, $(Z_1^\rho, \ldots, Z_{m+1}^\rho)$ and $(Z_1^{\tilde{\rho}}, \ldots, Z_{m+1}^{\tilde{\rho}})$ are dependent, because they are based on the same underlying Brownian motion. For example.

$$Z_1^\rho = \text{tr} \left\{ \int_0^1 dB_{\rho,1} F_{\rho,1}' \left( \int_0^1 F_{\rho,1} F_{\rho,1}' \right)^{-1} \int_0^1 F_{\rho,1} dB_{\rho,1}' \right\}.$$

where $B_{\rho,j}(\lambda) = \rho_1^{-1/2} B(\lambda/\rho_1)$. $F_{\rho,j}(\lambda) = B_{\rho,j}(\lambda) - \int_0^1 B_{\rho,1}(\lambda) d\lambda$. and $Z_1^{\tilde{\rho}}$ is based on $B_{\tilde{\rho},j}(\lambda) = \tilde{\rho}_1^{-1/2} B(\lambda/\tilde{\rho}_1)$.

It might be the case that additional restrictions are needed in this setting. In the $I(0)$ frame-

work, it is well known that the Sup-test diverges if the test statistics are not bounded away from the end points. For example Andrews (1993) showed that the Sup-LM test, for a single change, diverges if the supremum is taken over $[0.1]$ rather than over $[\varepsilon, 1 - \varepsilon]$ for some $\varepsilon > 0$. This is due to the behavior of a Brownian motion near zero. A way to overcome this problem has been to exclude the first and last 15% of the tests, as proposed by Andrews (1993).

In our case with multiple changes one might expect that, in addition to bounding the change points away from the endpoints, we also need to bound the change points apart from each other, which would be equivalent to requiring that $\min \rho_j \geq \varepsilon > 0$.

However, the problem need not exist in this formulation. First of all, this problem does not belong to the framework covered by Andrews (1993); we do not have a tied down Bessel process as our limit distribution. For a given value of $\rho$ we can make the transformation from

$$\mathrm{tr}\left\{\int_{u_{j-1}}^{u_j} dBF' \left(\int_{u_{j-1}}^{u_j} FF'\right)^{-1} \int_{u_{j-1}}^{u_j} FdB'\right\}.$$

to

$$\rho_j \mathrm{tr}\left\{\int_0^1 dB_{\rho,j}F' \left(\int_0^1 F_{\rho,j}F'_{\rho,j}\right)^{-1} \int_0^1 F_{\rho,j}dB'_{\rho,j}\right\} \xrightarrow{p} 0, \text{ for } \rho_j \to 0.$$

However, a deeper analysis is needed to conclude whether Sup $LR$ diverges or not when $\rho_j$ is not bounded away from zero.

The distributions, given in equations (4.3.1), (4.3.2), and (4.3.3), are non-standard and critical values should be simulated in practical applications. In the case where the change point is known, it is impractical to tabulate critical values, because the critical values of $\sum_{j=1}^{m+1} \rho_j \lambda_{df}(x_j)$ depend on the timing of the structural changes, $(\rho_1, \ldots, \rho_{m+1})$. However, it might be possible to obtain simple and practical formulae, from Monte Carlo studies of the response surface, or related approaches to approximate asymptotic $p$-values, see for example B. E. Hansen (1997). We leave this and

tabulation of the distribution given by equations (4.3.2) and (4.3.3) for future work.

## Appendix D: Proofs

**Proof of Theorem 4.3.1.** Let for simplicity $\Phi D_t = 0$. The result is obtained from the expression

$$
C_j(X_{T_j-1} - \sum_{i=1}^{k-1}\Gamma_{j,i}X_{T_j-1-i})
$$

$$
= C_j\left[C_{j-1}\sum_{i=T_{j-2}+1}^{T_{j-1}}\varepsilon_i + D_{j-1}(L)\varepsilon_{T_{j-1}} + C_{j-1}(X_{j-2} - \sum_{i=1}^{k-1}\Gamma_{j-1,i}X_{j-2-i})\right.
$$

$$
- \Gamma_{j,1}\left(C_{j-1}\sum_{i=T_{j-2}+1}^{T_1-1}\varepsilon_i + D_{j-1}(L)\varepsilon_{T_{j-1}-1} + C_{j-1}(X_{j-2} - \sum_{i=1}^{k-1}\Gamma_{j-1,i}X_{j-2-i})\right)
$$

$$
\vdots
$$

$$
\left.-\Gamma_{j,k-1}\left(C_{j-1}\sum_{i=T_{j-2}+1}^{T_1-k+1}\varepsilon_i + D_{j-1}(L)\varepsilon_{T_{j-1}-k+1} + C_{j-1}(X_{j-2} - \sum_{i=1}^{k-1}\Gamma_{j-1,i}X_{j-2-i})\right)\right]
$$

$$
= C_j\left[\bar{\Gamma}_jC_{j-1}\sum_{i=T_{j-2}+1}^{T_{j-1}}\varepsilon_i + D_{j-1}^*(L)\varepsilon_{T_{j-1}} + \bar{\Gamma}_jC_{j-1}(X_{j-2} - \sum_{i=1}^{k-1}\Gamma_{j-1,i}X_{j-2-i})\right]
$$

$$
= C_j\bar{\Gamma}_jC_{j-1}\sum_{i=T_{j-2}+1}^{T_{j-1}}\varepsilon_i + C_jD_{j-1}^*(L)\varepsilon_{T_{j-1}} + C_j\bar{\Gamma}_jC_{j-1}(X_{j-2} - \sum_{i=1}^{k-1}\Gamma_{j-1,i}X_{j-2-i}).
$$

where $D_{j-1}^*(L)\varepsilon_{T_{j-1}}$ is a stationary process. Altogether. we have the Granger representation

$$
X_t = C_1\sum_{i=1}^{t}\varepsilon_i + D_1(L)\varepsilon_t + C_1(X_0 - \sum_{i=1}^{k-1}\Gamma_{1,i}X_{0-i}) \quad t = 1, \ldots . T_1
$$

$$
X_t = C_j\sum_{i=T_{j-1}+1}^{t}\varepsilon_i + C_j\bar{\Gamma}_jC_{j-1}\sum_{i=T_{j-2}+1}^{T_{j-1}}\varepsilon_i + D_j(L)\varepsilon_t + C_jD_{j-1}^*(L)\varepsilon_{T_{j-1}}
$$

$$
+ C_j\bar{\Gamma}_jC_{j-1}(X_0 - \sum_{i=1}^{k-1}\Gamma_iX_{0-i}). \quad t = T_1 + 1, \ldots . T.
$$

and with the convention $\Upsilon_{j,i} = C_j \bar{\Gamma}_j C_{j-1} \cdots \bar{\Gamma}_{j-i+1} C_{j-i}$ $(\Upsilon_{j,0} = C_j)$, we obtain

$$X_t = C_j \sum_{i=T_{j-1}+1}^{t} \varepsilon_i + \sum_{i=1}^{j-1} \Upsilon_{j,i} \sum_{i=T_{j-i-1}+1}^{T_{j-i}} \varepsilon_i + V_j.$$

where $V_j$ is stationary. In the case where $\Phi D_t \neq 0$ for all $t$, we get

$$X_t = C_j \sum_{i=T_{j-1}+1}^{t} (\varepsilon_i + \Phi D_i) + \sum_{i=1}^{j-1} \Upsilon_{j,i} \sum_{i=T_{j-i-1}+1}^{T_{j-i}} (\varepsilon_i + \Phi D_i) + d_t + V_j,$$

where $d_t = B(L)\Phi D_t$. $B(L)$ is a stationary polynomial, so that $d_t / \max_{1 \leq i \leq t} D_i = O(1)$. Note that we have the stationary cointegrating relations in the $j$th sub-sample given by $\beta_j' X_t = \beta_j' D_j(L)\varepsilon_t$, which does not involve elements from sub-samples prior to the $j$th sub-sample. ∎

# References

ANDREWS. D. W. K. (1993): "Test for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

ANDREWS. D. W. K., AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter is Present Only under the Alternative." *Econometrica*, 62, 1383–1414.

BAI. J. (1997): "Estimating Multiple Breaks One at a Time." *Econometric Theory*, 13, 315–352.

———— (1999): "Likelihood Ratio Test for Multiple Structural Changes." *Journal of Econometrics*, 91, 299–323.

BAI. J., AND P. PERRON (1998): "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica*, 66, 47–78.

BOSWIJK. P. H. (1995): "Identifiability of Cointegrated Systems." *Working paper. Tinbergen Institute.*

CHOW. G. C. (1960): "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica*, 28, 591–605.

DOORNIK. J. A. (1998): "Approximations to Teh Asymptotic Distribution of Cointegration Tsts," *Journal of Economic Surveys*, 12, 573–593.

GREGORY, A. W., AND B. E. HANSEN (1996): "Residual-Based Test for Cointegration in Models with Regime Shifts," *Journal of Econometrics*, 70, 99–126.

HANSEN. B. E. (1992a): "Test for Instability with $I(1)$ Processes." *Journal of Business and Economic Statistics*, 10, 321–335.

—————— (1992b): "Testing for Parameter Instability in Linear Models," *Journal of Policy Modeling*, 14, 517–533.

—————— (1997): "Approximate Asymptotic P Values for Structural-Change Tests," *Journal of Business and Economic Statistics*, 15, 60–67.

HANSEN, H., AND S. JOHANSEN (1999): "Some Tests for Parameter Constancy in Cointegrated VAR-Models," *Econometrics Journal*, 2, 306–333.

HANSEN, P. R. (2000a): "Estimation of Cointegration Models with Heteroskedasticity and Autocorrelation under Parameter Restrictions," *Forthcoming UCSD Department of Economics Discussion Papers*.

—————— (2000b): "The Johansen-Granger Representation Theorem: An Explicit Expression for $I(1)$ Processes," *Forthcoming UCSD Department of Economics Discussion Papers*.

—————— (2000c): "Structural Changes in the Cointegrated Vector Autoregressive Model," *Forthcoming UCSD Department of Economics Discussion Papers*.

INOUE, A. (1999): "Test for Cointegrating Rank with a Trend-Break," *Journal of Econometrics*, 90, 215–237.

JOHANSEN, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

—————— (1996): *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford, 2nd edn.

JOHANSEN, S., AND K. JUSELIUS (1992): "Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK," *Journal of Econometrics*, 53, 211–244.

NIELSEN, B. (1997): "Bartlett Correction of the Unit Root Test in Autoregressive Models," *Biometrika*, 84, 500–504.

NYBLOM, J. (1989): "Testing for the Constancy of Parameters over Time," *Journal of the American Statistical Association*, 84, 223–230.

PERRON, P. (1990): "Testing for a Unit Root in a Time Series with a Changing Mean," *Journal of Business and Economic Statistics*, 8, 153–162.

QUANDT, R. E. (1960): "Tests of the Hypothesis That a Linear Regrssion System Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55, 324–330.

QUINTOS, C. E. (1995): "Substainability of the Deficit Process with Structural Shifts," *Journal of Business and Economic Statistics*, 13, 409–417.

—————— (1997): "Stability Tests in Error Correction Models," *Journal of Econometrics*, 82, 289–315.

SEO, B. (1998): "Tests for Structural Change in Cointegrated Systems," *Econometric Theory*, 14, 222–259.

# Chapter 5

# Estimation of Cointegration Models with Heteroskedasticity

# and Autocorrelation under Parameter Restrictions[†]

## Abstract

This chapter derives a general estimation technique that is applicable to the cointegrated vector autoregressive model under parameter restrictions. It allows for a general form of the covariance matrix and is well suited for estimation of models with heteroskedasticity and serial correlation.

Applicability includes: cointegrated VARMA models, panel cointegration models, cointegration models with structural changes, and cointegration models under restrictions imposed by Granger non-causality.

---

[†]I thank James D. Hamilton for many valuable comments. All errors remain my responsibility.

## 5.1. Introduction

Reduced rank regression (RRR) takes the form

$$Z_{0t} = AB'Z_{1t} + CZ_{2t} + \varepsilon_t, \quad t = 1, \dots, T.$$

where $Z_{0t}$, $Z_{1t}$, and $Z_{2t}$ are vectors of dimension $p$, $p_1$, and $p_2$ respectively, and where $A$, $B$, and $C$ are parameters of dimension $p \times r$, $p_1 \times r$, and $p \times p_2$ respectively.

The estimation problem in the cointegrated vector autoregressive model (VAR), for a given cointegration rank, is a reduced rank regression problem, and Johansen (1988) showed how parameter estimates can be obtained in this case by solving an eigenvalue problem. This approach is similar to the methods by Ahn and Reinsel (1990) and the canonical correlations by Anderson (1951).

Johansen's technique is directly applicable to parameter estimation under restrictions that take the form $A = G\psi$ and $B = H\varphi$ for known matrices $G$ and $H$, while problems of the form $B = (H_1\varphi_1, \dots, H_r\varphi_r)$, for known matrices $H_1, \dots, H_r$, can be solved by a switching algorithm of Johansen and Juselius (1992), that reduces the estimation problem to a simple RRR problem in every iteration.

Boswijk (1995) derived a more general estimation technique that solves estimation problems of the form $\mathrm{vec}(A) = G\psi + g$ and $\mathrm{vec}(B) = H\varphi + h$, where $\mathrm{vec}(\cdot)$ is the vectorization operator. $G$ and $H$ are known matrices, and $g$ and $h$ are known vectors.

When $\{\varepsilon_t\}$ is a sequence of i.i.d. Gaussian variables with mean zero and constant variance $\Omega$ (a $p \times p$ matrix), the techniques yield maximum likelihood estimators when $Z_{1t}$ and $Z_{2t}$ are measurable $\mathcal{F}_{t-1}$, where $\mathcal{F}_t = \sigma(Z_{0,1}, Z_{0,2}, \dots, Z_{0,t}, Z_{1,0}, Z_{2,0})$.

This chapter derives a generalized reduced rank regression (GRRR) technique, that contains each of the above techniques as a special case. The technique is applicable to a more general class of parameter restrictions, as well as more complex structures of the covariance matrix, including

heteroskedasticity and autocorrelation. The technique can be extended to non-linear restrictions by localized linear approximation, and thereby include the class of parameter restrictions considered by Elliott (1997, 1998). However, the non-linear aspect is not treated in this paper. The technique by Elliott (1997, 1998) uses minimum distance methods applied to the cointegrated regressions (see Engle and Granger (1987)), whereas the technique in this paper is motivated by likelihood analysis of the cointegrated VAR. In Section 5.3, we show how the GRRR technique is applicable to several estimation problems. This chapter is only concerned with the estimation problem, whereas (asymptotic) probabilistic properties of the estimators are ignored. In most cases, the probabilistic properties will depend on the problem at hand, and most results already exist in the literature.

## 5.2. Generalized Reduced Rank Regression

We define a generalized reduced rank regression, as the following regression problem:

$$Z_{0t} = AB'Z_{1t} + CZ_{2t} + \varepsilon_t. \quad t = 1, \ldots, T \tag{5.2.1}$$

$$\text{s.t. } \text{vec}(A.C) = Gv + g.$$

$$\text{vec}(B) = H\varphi + h.$$

where $G$ and $H$ are known matrices, $g$ and $h$ are known vectors, and where $\varepsilon = \text{vec}(\varepsilon_1, \ldots, \varepsilon_T) \sim N(0, \Sigma)$. The RRR is a special case of the GRRR. This can be seen by setting $G = I$, $H = I$, $g = 0$, $h = 0$, and $\Sigma = I_T \otimes \Omega$.

Notice that $\{\varepsilon_t\}$ need not be i.i.d. In fact the general structure of $\Sigma$ allows for both heteroskedasticity and serial correlation.

## 5.2.1. When the Covariance, $\Sigma$, is Known

It is convenient to define

$$\mathbf{M}_A^{hac} = T^{-1} \sum_{t,\tau=1}^{T} \left[ A' \Sigma_{t\tau}^{-1} A \times Z_{1t} Z_{1\tau}' \right].$$

$$\mathbf{N}_{AC}^{hac} = T^{-1} \sum_{t,\tau=1}^{T} \text{vec} \left( Z_{1t}(Z_{0\tau} - C Z_{2\tau})' \Sigma_{t\tau}^{-1'} A \right).$$

$$\mathbf{M}_B^{hac} = T^{-1} \sum_{t,\tau=1}^{T} \left[ \begin{pmatrix} B' Z_{1t} Z_{1\tau}' B & B' Z_{1t} Z_{2\tau}' \\ Z_{2t} Z_{1\tau}' B & Z_{2t} Z_{2\tau}' \end{pmatrix} \times \Sigma_{t\tau}^{-1} \right].$$

$$\mathbf{N}_B^{hac} = T^{-1} \sum_{t,\tau=1}^{T} \text{vec} \left( \Sigma_{t\tau}^{-1} Z_{0\tau}(Z_{1t}' B, Z_{2t}') \right).$$

where $\Sigma_{t\tau}^{-1}$ is a $p \times p$ block matrix of $\Sigma^{-1}$, such that the $[i,j]$th element of $\Sigma_{t\tau}^{-1}$ is the $[(t-1)p+i, (\tau-1)p+j]$th element of $\Sigma^{-1}$, $i,j = 1,\ldots,p$. (see Lemma E.7). The identities simplify considerably in the situation where autocorrelation of $\{\varepsilon_t\}$ is excluded, a situation we treat separately below.

We can now formulate the most general theorem.

**Theorem 5.2.1.** *Let the parameter $A$, $B$, and $C$ be restricted by $\text{vec}(A.C) = Gv + g$ and $\text{vec}(B) = Ho + h$ and suppose that $\varepsilon = \text{vec}(\varepsilon_1,\ldots,\varepsilon_T) \sim \mathcal{N}(0,\Sigma)$.*

*Then the parameter estimates of $A$, $B$, and $C$ are derived by iterating on the equations*

$$\text{vec}(\hat{A}.\hat{C}) = G \left[ G' \mathbf{M}_B^{hac} G \right]^{-1} G' \left( \mathbf{N}_B^{hac} - \mathbf{M}_B^{hac} g \right) + g. \tag{5.2.2}$$

$$\text{vec}(\hat{B}) = H \left[ H' \mathbf{M}_A^{hac} H \right]^{-1} H' \left( \mathbf{N}_{AC}^{hac} - \mathbf{M}_A^{hac} h \right) + h. \tag{5.2.3}$$

*until convergence, from some initial values $(A^{(0)}, B^{(0)}, C^{(0)})$. The maximum value of the likelihood function is given by*

$$L_{max}(\hat{A}.\hat{B}.\hat{C}) = (2\pi)^{-\frac{Tp}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left( -\frac{1}{2} \hat{\varepsilon}' \Sigma^{-1} \hat{\varepsilon} \right).$$

where $\hat{\varepsilon} = \text{vec}(\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T)$, $\hat{\varepsilon}_t = Z_{0t} - \hat{A}\hat{B}'Z_{1t} - \hat{C}Z_{2t}$.

The equations (5.2.2) and (5.2.3) may seem complicated at first. However, a closer look reveals the structure of a restricted GLS problem, which is indeed what the individual estimation problems reduce to.

Since the likelihood function is bounded by its maximum, then an algorithm based on these equations will eventually converge. It is not obvious that the likelihood function does not have local maxima, where the algorithm can get stuck. One can investigate the presence of local maxima, by starting the algorithm at different initial values, $(A^{(0)}, B^{(0)}, C^{(0)})$, and verifying if the algorithm leads to the same value of the likelihood function. No local maxima have been found in simulations using this approach, so it is possible that there is not a problem with local maxima, and that the global maximum will always be found by iterations on the likelihood equations.

We now turn to some simpler situations. First the case where autocorrelation of $\{\varepsilon_t\}$ is excluded, but heteroskedasticity is not. In this case $\Sigma$ is block diagonal, and we denote the $T$ diagonal matrices by $\Omega(t)$, $t = 1, \ldots, T$. We then have that $\Sigma_{t\tau}^{-1} = \Omega(t)^{-1}$ for $t = \tau$ and $0$ otherwise, and we define

$$\mathbf{M}_A^{he} = T^{-1}\sum_{t=1}^{T}\left[A'\Omega(t)^{-1}A \otimes Z_{1t}Z_{1t}'\right].$$

$$\mathbf{N}_{AC}^{he} = T^{-1}\sum_{t=1}^{T}\text{vec}\left(Z_{1t}(Z_{0\tau} - CZ_{2\tau})'\Omega(t)^{-1}A\right).$$

$$\mathbf{M}_B^{he} = T^{-1}\sum_{t=1}^{T}\left[\begin{pmatrix} B'Z_{1t}Z_{1t}'B & B'Z_{1t}Z_{2t}' \\ Z_{2t}Z_{1t}'B & Z_{2t}Z_{2t}' \end{pmatrix} \otimes \Omega(t)^{-1}\right].$$

$$\mathbf{N}_B^{he} = T^{-1}\sum_{t=1}^{T}\text{vec}\left(\Omega(t)^{-1}Z_{0t}(Z_{1t}'B, Z_{2t}')\right).$$

**Corollary 5.2.2.** *Let the parameter be restricted as before:* $\text{vec}(A, C) = Gv + g$, $\text{vec}(B) = Ho + h$ *and let* $\varepsilon = \text{vec}(\varepsilon_1, \ldots, \varepsilon_T) \sim N(0, \Sigma)$, *where* $\Sigma$ *is block diagonal, with diagonal block given by* $\Omega(t)$.

Then the parameter estimates of $A$, $B$, and $C$ are derived by iterating on the equations

$$\text{vec}(\hat{A}.\hat{C}) = G\left[G'\mathbf{M}_B^{he}G\right]^{-1}G'\left(\mathbf{N}_B^{he} - \mathbf{M}_B^{he}g\right) + g. \tag{5.2.4}$$

$$\text{vec}(\hat{B}) = H\left[H'\mathbf{M}_A^{he}H\right]^{-1}H'\left(\mathbf{N}_{AC}^{he} - \mathbf{M}_A^{he}h\right) + h. \tag{5.2.5}$$

until convergence, from some initial values $(A^{(0)}, B^{(0)}, C^{(0)})$. The maximum value of the likelihood function is given by

$$L_{\max}(\hat{A}.\hat{B}.\hat{C}) = (2\pi)^{-\frac{Tp}{2}}\left(\prod_{t=1}^{T}|\Omega(t)|^{-\frac{1}{2}}\right)\exp\left(-\frac{1}{2}\sum_{t=1}^{T}\hat{\varepsilon}_t'\Omega(t)^{-1}\hat{\varepsilon}_t\right).$$

where $\hat{\varepsilon}_t = Z_{0t} - \hat{A}\hat{B}'Z_{1t} - \hat{C}Z_{2t}$.

The result of Corollary 5.2.2, allows for heteroskedasticity of a known form, whereas the next corollary corresponds to the homoskedastic situation. Define

$$\mathbf{M}_A^{ho} = \left[A'\Omega^{-1}A \times M_{11}\right].$$

$$\mathbf{N}_{AC}^{ho} = \text{vec}\left(M_{10} - M_{02}C'\right)\Omega^{-1}A\right).$$

$$\mathbf{M}_B^{ho} = \left[\begin{pmatrix} B'M_{11}B & B'M_{12} \\ M_{21}B & M_{22} \end{pmatrix} \times \Omega^{-1}\right].$$

$$\mathbf{N}_B^{ho} = \text{vec}\left(\Omega^{-1}(M_{01}B.M_{02})\right).$$

where $M_{ij} = T^{-1}\sum_{t=1}^{T}Z_{it}Z_{jt}'$, $i.j = 0,1,2$.

**Corollary 5.2.3.** Let the parameter be restricted as before: $\text{vec}(A.C) = Gv+g$, $\text{vec}(B) = H\phi+h$ and let $\varepsilon = \text{vec}(\varepsilon_1.....\varepsilon_T) \sim N(0.\Sigma)$, where $\Sigma$ is block diagonal, with diagonal block given by $\Omega$, i.e. $\Sigma = I_T \times \Omega$.

Then the parameter estimates of $A$, $B$, and $C$ are derived by iterating on the equations

$$\text{vec}(\hat{A}.\hat{C}) = G\left[G'\mathbf{M}_B^{ho}G\right]^{-1}G'\left(\mathbf{N}_B^{ho} - \mathbf{M}_B^{ho}g\right) + g.$$

$$\text{vec}(\hat{B}) = H\left[H'\mathbf{M}_A^{ho}H\right]^{-1}H'\left(\mathbf{N}_{AC}^{ho} - \mathbf{M}_A^{ho}h\right) + h.$$

until convergence. from some initial values $(A^{(0)}, B^{(0)}, C^{(0)})$. The maximum value of the likelihood function is given by

$$L_{\max}^{-2/T}(\hat{A}, \hat{B}, \hat{C}) = (2\pi)^{-p}|\Omega|\exp\left(\text{tr}\left(T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t'\right)\Omega^{-1}\right).$$

where $\hat{\varepsilon}_t = Z_{0t} - \hat{A}\hat{B}'Z_{1t} - \hat{C}Z_{2t}$.

## 5.2.2. When the Covariance, $\Sigma$, is Unknown

The estimation technique only requires a minor modification to the situation where the general covariance matrix is unknown and must be estimated. The equations for $\text{vec}(A.C)$ and $\text{vec}(B)$ remain unchanged but a third equation, defining the likelihood equation for $\hat{\Sigma}$. is added. It is not possible to estimate an entirely unrestricted covariance matrix, since only one observation of $\varepsilon =$ $\text{vec}(\varepsilon_1 \ldots \varepsilon_T)$ is available. But semi-parametric heteroskedastic and autocorrelation consistent (HAC) estimators are available. To fit this framework an estimator of $\Sigma$ must be a maximum likelihood estimator. So we need to specify an estimator $\hat{\Sigma}$ that is the solution to the relevant likelihood equation. An obvious candidate is to estimate an $\text{MA}(q)$ process and let $q$ increase with the sample size $T$. similar to the estimators suggested by Newey and West (1987) and Andrews (1991). see also Den Haan and Levin (1997).

**Theorem 5.2.4.** Let the parameters $A$, $B$, and $C$ be restricted by $\text{vec}(A.C) = Gv + g$ and $\text{vec}(B) = Ho + h$ and suppose that $\varepsilon = \text{vec}(\varepsilon_1 \ldots \varepsilon_T) \sim N(0, \Sigma(\theta))$.

Then the parameter estimates of $A$. $B$. $C$. and $\theta$ are derived by iterating on the equations

$$\text{vec}(\hat{A}.\hat{C}) = G\left[G'\mathbf{M}_B^{hac}G\right]^{-1}G'\left(\mathbf{N}_B^{hac} - \mathbf{M}_B^{hac}g\right) + g. \tag{5.2.6}$$

$$\text{vec}(\hat{B}) = H\left[H'\mathbf{M}_A^{hac}H\right]^{-1}H'\left(\mathbf{N}_{AC}^{hac} - \mathbf{M}_A^{hac}h\right) + h. \tag{5.2.7}$$

$$\hat{\theta} = \arg\min_{\theta} L(\hat{A}, \hat{B}, \hat{C}, \theta) \tag{5.2.8}$$

*until convergence, from some initial values* $(A^{(0)}, B^{(0)}, C^{(0)}, \theta^{(0)})$. *Set* $\hat{\varepsilon} = \text{vec}(\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T)$, *the maximum value of the likelihood function is then given by*

$$L_{\max}(\hat{A}, \hat{B}, \hat{C}, \hat{\theta}) = (2\pi)^{-\frac{TP}{2}} |\Sigma(\hat{\theta})|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\hat{\varepsilon}'\Sigma(\hat{\theta})^{-1}\hat{\varepsilon}\right).$$

In the situation without autocorrelation but presence of heteroskedasticity, we also need to assume additional structure on $\Sigma$, because we only have one (estimated) observation for each of the covariance matrices $\Omega(t)$. A general formulation is to express the covariance by $\Omega_\theta(t) = f(\theta; X_{t-1})$ where $\theta$ is a parameter to be estimated, and $X_{t-1}$ is a set of variables that are $\mathcal{F}_{t-1}$-measurable, where $\mathcal{F}_{t-1}$ is such that $Z_{0t}|\mathcal{F}_{t-1} \sim \mathcal{N}(AB'Z_{1t} + CZ_{2t}, \Omega_\theta(t))$.

**Corollary 5.2.5.** *Let the parameters $A$, $B$, and $C$ be restricted by* $\text{vec}(A, C) = Gv + g$ *and* $\text{vec}(B) = H\phi + h$ *and suppose that* $\{\varepsilon_t\}$ *is a sequence of independent variables with* $\varepsilon_t \sim \mathcal{N}(0, \Omega_\theta(t))$.

Then the parameter estimates of $A$, $B$, $C$, and $D$ are derived by iterating on the equations

$$\text{vec}(\hat{A}, \hat{C}) = G\left[G'\mathbf{M}_B^{het}G\right]^{-1} G' \left(\mathbf{N}_B^{het} - \mathbf{M}_B^{het}g\right) + g.$$

$$\text{vec}(\hat{B}) = H\left[H'\mathbf{M}_A^{het}H\right]^{-1} H' \left(\mathbf{N}_{AC}^{het} - \mathbf{M}_A^{het}h\right) + h.$$

$$\hat{\theta} = \arg\min_{\theta} L(\hat{A}, \hat{B}, \hat{C}, \theta).$$

until convergence, from some initial values $(A^{(0)}, B^{(0)}, C^{(0)}, D^{(0)})$, where we set $\hat{\varepsilon}_t = Z_{0t} - \hat{A}\hat{B}'Z_{1t} - \hat{C}Z_{2t}$ and $\hat{\Omega}(t) = \Omega_{\hat{\theta}}(t)$ in the formulae for $\mathbf{M}_B^{het}$, $\mathbf{N}_B^{het}$, $\mathbf{M}_A^{het}$, and $\mathbf{N}_{AC}^{het}$. The maximum value of the likelihood function is then given by

$$L_{\max}(\hat{A}, \hat{B}, \hat{C}, \hat{D}) = (2\pi)^{-\frac{TP}{2}} \left(\prod_{t=1}^{T} |\hat{\Omega}(t)|^{-\frac{1}{2}}\right) \exp\left(-\frac{1}{2}\sum_{t=1}^{T} \hat{\varepsilon}_t'\hat{\Omega}(t)^{-1}\hat{\varepsilon}_t\right).$$

An ARCH-type of $\Omega_\theta(t)$ is presented in Example 5.3.2. The estimators in the homoskedastic case is given by the following theorem.

**Corollary 5.2.6.** *Let the parameters $A$, $B$ and $C$ be restricted by $\text{vec}(A, C) = Gv + g$ and $\text{vec}(B) = H\phi + h$ and suppose that $\{\varepsilon_t\}$ is i.i.d. Gaussian. $\varepsilon_t \sim N(0, \Omega)$.*

**Corollary 5.2.7.** *The parameter estimates of $A$. $B$. $C$ and $\Omega$ are found by iterating on the equations*

$$\text{vec}(\hat{A}, \hat{C}) = G\left[G'\mathbf{M}_B^{ho}G\right]^{-1} G'\left(\mathbf{N}_B^{ho} - \mathbf{M}_B^{ho}g\right) + g.$$

$$\text{vec}(\hat{B}) = H\left[H'\mathbf{M}_A^{ho}H\right]^{-1} H'\left(\mathbf{N}_{AC}^{ho} - \mathbf{M}_A^{ho}h\right) + h.$$

$$\hat{\Omega} = T^{-1}(Z_0 - AB'Z_1 - CZ_2)(Z_0 - AB'Z_1 - CZ_2)'$$

*until convergence, from some initial values $(A^{(0)}, B^{(0)}, C^{(0)}, \Omega^{(0)})$. The maximum value of the likelihood function is given by*

$$L_{\max}(\hat{A}, \hat{B}, \hat{C}, \hat{\Omega}) = (2\pi e)^{-\frac{Tp}{2}} |\hat{\Omega}|^{-\frac{T}{2}}.$$

If $C$ is unrestricted we define

$$\tilde{\mathbf{M}}_A^{ho} = \left[A'\Omega^{-1}A \times S_{11}\right].$$

$$\tilde{\mathbf{N}}_A^{ho} = \text{vec}\left(S_{10}\Omega^{-1}A\right).$$

$$\tilde{\mathbf{M}}_B^{ho} = \left[B'S_{11}B \times I_p\right].$$

$$\tilde{\mathbf{N}}_B^{ho} = \text{vec}\left(S_{01}\Omega^{-1}B\right).$$

where $S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R_{jt}$. $R_{it} = Z_{it} - M_{i2}M_{22}^{-1}Z_{2t}$. $i, j = 0, 1$. So the residuals, $R_{0t}$ and $R_{1t}$ are $Z_{0t}$ and $Z_{1t}$ corrected for $Z_{2t}$. and $S_{ij}$, $i, j = 0, 1$ are the moment matrices of these residuals. We then obtain the following result of Boswijk (1995).

**Corollary 5.2.8.** *Let $A$ and $B$ be restricted by $\text{vec}(A) = G\zeta + g$ and $\text{vec}(B) = H\phi + h$. for known matrices $G$ and $H$. and vectors $g$ and $h$. and let $\Sigma = I_T \times \Omega$ be known. Then the parameter estimates of Model (5.2.1) are derived by iterating on the equations*

$$\text{vec}(\hat{A}) = G\left[G'\bar{\mathbf{M}}_B^{ho}G\right]^{-1}G'\left(\bar{\mathbf{N}}_B^{ho} - \bar{\mathbf{M}}_B^{ho}g\right) + g.$$

$$\text{vec}(\hat{B}) = H\left[H'\bar{\mathbf{M}}_A^{ho}H\right]^{-1}H'\left(\bar{\mathbf{N}}_A^{ho} - \bar{\mathbf{M}}_A^{ho}h\right) + h.$$

$$\hat{\Omega} = S_{00} - S_{01}BA' + AB'S_{11}BA' - AB'S_{10}$$

*until convergence. from some initial values $(A.B.\Omega) = (A^0.B^0.\Omega^0)$. and then calculate*

$$\hat{C} = M_{02}M_{22}^{-1} - \hat{A}\hat{B}'M_{12}M_{22}^{-1}.$$

*The maximum value of the likelihood function is given by*

$$L_{\max}^{-2/T}(\hat{A},\hat{B}.\hat{C}.\hat{\Omega}) = (2\pi e)^P |\hat{\Omega}|.$$

If $g = 0$ and $h = 0$ the equations for $A$ and $B$ simplify to

$$\text{vec}(\hat{A}) = G\left[G'(B'S_{11}B \times I_p)G\right]^{-1}G'\text{vec}\left(S_{01}\Omega^{-1}B\right).$$

$$\text{vec}(\hat{B}) = H\left[H'(A'\Omega^{-1}A \times S_{11})H\right]^{-1}H'\text{vec}\left(S_{10}\Omega^{-1}A\right).$$

If $A$ and $C$ are unrestricted we have the following result.

**Corollary 5.2.9.** *Let $B$ be restricted by $\text{vec}(B) = H\phi + h$. Then the parameter estimates are found by iterating on the equations*

$$\text{vec}(\hat{B}) = H\left[H'\left(A'\Omega^{-1}A \times S_{11}\right)H\right]^{-1}H'\left[\text{vec}\left(S_{10}\Omega^{-1}A\right) - \left(A'\Omega^{-1}A \times S_{11}\right)h\right] + h$$

*and*

$$\hat{A}(B) \;=\; S_{01}B\,(B'S_{11}B)^{-1}$$

$$\hat{\Omega}(B) \;=\; S_{00} - S_{01}B\,(B'S_{11}B)^{-1}\,B'S_{10},$$

*until convergence. from some initial values* $(A.B.\Omega) = (A^0.B^0.\Omega^0)$. *and then calculate*

$$\hat{C} = M_{02}M_{22}^{-1} - \hat{A}\hat{B}'M_{12}M_{22}^{-1}.$$

*The maximum value of the likelihood function is given by*

$$L_{\max}^{-2/T}(\hat{A}.\hat{B}.\hat{C}.\hat{\Omega}) = (2\pi e)^P\,|\hat{\Omega}|.$$

## 5.3. Applicability: Examples

**Example 5.3.1 (Moving average residuals).** *In Theorem 5.2.4. we indicated that the GRRR technique is applicable to cointegrated VARMA models. Since A. B. and C are easily estimated for a fixed value of* $\Sigma$. *all we need to add is an additional equation for the estimator of* $\Sigma$. *Suppose that* $\{\varepsilon_t\}$ *is a Gaussian moving average process of order q. with* $\Omega_i = \mathrm{cov}(\varepsilon_t.\varepsilon_{t-i})$. $i = 1.\ldots.q$. *Then from fixed values of* $\hat{A}$. $\hat{B}$. *and* $\hat{C}$. *that define the residuals.* $\hat{\varepsilon}_1\ldots\ldots\hat{\varepsilon}_T$. *one can estimate the*

*general covariance matrix*

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Omega}_0 & \cdots & \hat{\Omega}'_q & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \ddots & \vdots \\ \hat{\Omega}_q & \cdots & \hat{\Omega}_0 & & & 0 \\ & & & \ddots & & \\ 0 & & & \hat{\Omega}_0 & \cdots & \hat{\Omega}'_q \\ \vdots & \ddots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \hat{\Omega}_q & \cdots & \hat{\Omega}_0 \end{pmatrix}.$$

by maximizing the exact likelihood function. There does not exist a closed form solution to this problem. but numeric methods are available. see Osborn (1977) or Hamilton (1994). See Lütkepohl and Claessen (1997) for a different estimation method of cointegrated VARMA processes.

**Example 5.3.2 (ARCH type heteroskedasticity).** *Let* $\Omega(t)^{-1} = Q'_t D Q_t$. *where* $Q'_t = (I.X_t)$. $X_t$ *is a sequence of* $p \times q$-*dimensional exogenous variables. and* $D$ *is a* $(p+q) \times (p+q)$ *matrix of parameters. The parameter estimate of* $D$ *satisfies*

$$\sum_{t=1}^{T} Q_t (Q'_t \hat{D} Q_t)^{-1} Q'_t = \sum_{t=1}^{T} Q_t \hat{\varepsilon}_t \hat{\varepsilon}'_t Q'_t. \tag{5.3.1}$$

This can be verified from the first order conditions. Let $\hat{A}$. $\hat{B}$. and $\hat{C}$ be given. and define $(\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T)$ accordingly. The log likelihood equation for $D$ is given by

$$l(D) = \sum_{t=1}^{T} \log |Q'_t D Q_t| - \text{tr} \left\{ \sum_{t=1}^{T} \hat{\varepsilon}'_t Q'_t D Q_t \hat{\varepsilon}_t \right\}.$$

and

$$l(D+d) = \sum_{t=1}^{T} \log |Q'_t (D+d) Q_t| - \text{tr} \left\{ \sum_{t=1}^{T} \hat{\varepsilon}'_t Q'_t (D+d) Q_t \hat{\varepsilon}_t \right\}$$

$$= l(D) + \sum_{t=1}^{T} \text{tr}\left\{(Q_t'DQ_t)^{-1}Q_t'dQ_t\right\} - \text{tr}\left\{\sum_{t=1}^{T} \tilde{\varepsilon}_t'Q_t'dQ_t\tilde{\varepsilon}_t\right\} + O(||d||^2).$$

so the first order condition is given by

$$\sum_{t=1}^{T} \text{tr}\left\{\left[Q_t(Q_t'DQ_t)^{-1}Q_t' - Q_t\tilde{\varepsilon}_t\tilde{\varepsilon}_t'Q_t'\right]d\right\} = 0$$

for all $d$. which proves Equation (5.3.1).

**Example 5.3.3 (Structural change).** Consider the cointegrated Gaussian VAR with a structural change in the cointegration relations and covariance matrix. as derived in Chapter 2. This can be expressed as

$$\Delta X_t = (\alpha. \alpha)\begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix}'\begin{pmatrix} X_{t-1}I(t \leq T_1) \\ X_{t-1}I(t > T_1) \end{pmatrix} + \Gamma_1\Delta X_{t-1} + \varepsilon_t. \quad t = 1, \dots, T.$$

where $\varepsilon_t \sim i.i.d.\mathcal{N}(0.\Omega_1)$ for $t \leq T_1$ and $\varepsilon_t \sim i.i.d.\mathcal{N}(0.\Omega_2)$ for $t > T_1$.

Set $A = (\alpha.\alpha)$. $B = \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix}$ and $C = \Gamma_1$. Then this regression problem can be written as a GRRR problem by the definitions $g = 0$. $h = 0$. and

$$G = \begin{pmatrix} G_1 & 0 \\ 0 & G_2 \end{pmatrix}. \quad G_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \times I_{pr}. \quad G_2 = I_{pp}. \quad H = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \times I_{pr}$$

where $p$ denotes the number of rows in $\alpha$ and $\beta$. and $r$ denotes the number of columns in $\alpha$. $\beta_1$.

and $\beta_2$. The covariance matrix is given by

$$\Sigma = \begin{pmatrix} I_{T_1} \otimes \Omega_1 & 0 \\ 0 & I_{(T-T_1)} \otimes \Omega_2 \end{pmatrix},$$

and because the log-likelihood function, for given values of $\hat{A}$, $\hat{B}$, and $\hat{C}$, splits into the sum

$$l(\Omega_1, \Omega_2) \quad \propto \quad T_1 \log |\Omega_1| + \mathrm{tr} \left\{ \sum_{t=1}^{T_1} \hat{\varepsilon}_t \hat{\varepsilon}_t' \Omega_1^{-1} \right\}$$

$$+(T - T_1) \log |\Omega_2| + \mathrm{tr} \left\{ \sum_{t=T_1+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t' \Omega_2^{-1} \right\},$$

the estimators are given by the sum of squares

$$\hat{\Omega}_1 \quad = \quad T_1^{-1} \sum_{t=1}^{T_1} \hat{\varepsilon}_t \hat{\varepsilon}_t'$$

$$\hat{\Omega}_2 \quad = \quad (T - T_1)^{-1} \sum_{t=T_1+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'.$$

**Example 5.3.4 (Granger non-causality).** *Consider a cointegration model where $X_{2t}$ does not Granger cause $X_{1t}$. (see Mosconi and Giannini (1992)). In the VAR(2) model. $\Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \varepsilon_t$. for $t = 1, \ldots, T$. the Granger non-causality is equivalent to the restrictions*

$$\Pi = \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ 0_{p_2 \times p_1} & \Pi_{22} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ 0_{p_2 \times r_1} & \alpha_{22} \end{pmatrix} \begin{pmatrix} \beta_{11}' & \beta_{21}' \\ 0_{r_2 \times p_1} & \beta_{22}' \end{pmatrix}$$

$$\Gamma_1 = \begin{pmatrix} \Gamma_{1,11} & \Gamma_{1,12} \\ 0_{p_2 \times p_1} & \Gamma_{1,22} \end{pmatrix}.$$

*where $p_1$ is the dimension of $X_{1t}$. $p_2 = p - p_1$ is the dimension of $X_{2t}$. $r_2$ is the rank of $\Pi_{22}$. and*

$r_1 = r - r_2$. *These restrictions can be expressed by*

$$
G = \begin{pmatrix} I_{r_1} \otimes \begin{pmatrix} I_{p_1} \\ 0_{p_2} \end{pmatrix} & 0 & 0 & 0 \\ 0 & I_{pr_2} & 0 & 0 \\ 0 & 0 & I_{p_1} \times \begin{pmatrix} I_{p_1} \\ 0_{p_2} \end{pmatrix} & 0 \\ 0 & 0 & 0 & I_{p_2 p} \end{pmatrix}
$$

$$
H = \begin{pmatrix} I_{pr_1} & 0 \\ 0 & I_{r_2} \otimes \begin{pmatrix} 0_{p_1} \\ I_{p_2} \end{pmatrix} \end{pmatrix} .
$$

*When* $\varepsilon_t \sim i.i.d. \mathcal{N}(0, \Omega)$ *the covariance matrix is given by* $\Sigma = I_T \times \Omega$.

**Example 5.3.5 (Panel cointegration).** *The panel cointegration model of Larsson. Lyhagen. and Lothgren (1998). Larsson and Lyhagen (1999). and Groen and Kleibergen (1999). takes the form*

$$
\begin{pmatrix} \Delta X_{1t} \\ \vdots \\ \Delta X_{nt} \end{pmatrix} = \alpha \beta' \begin{pmatrix} X_{1,t-1} \\ \vdots \\ X_{1,t-1} \end{pmatrix} + \Gamma_1 \begin{pmatrix} \Delta X_{1,t-1} \\ \vdots \\ \Delta X_{n,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{pmatrix} .
$$

*where* $\alpha$ *has one of three structures*

$$
\alpha = \begin{pmatrix} \alpha_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \alpha_n \end{pmatrix} . \quad \alpha = I_n \times \alpha.
$$

or $\alpha$ is unrestricted. where $\beta$ has the structure

$$\beta = \begin{pmatrix} \beta_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \beta_n \end{pmatrix}. \quad or\ \beta = I_n \times \beta.$$

where $\Gamma_1$ has the structure

$$\Gamma_1 = \begin{pmatrix} \Gamma_{1,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Gamma_{1,n} \end{pmatrix}. \quad \Gamma_1 = I_n \times \Gamma_1.$$

or $\Gamma_1$ is unrestricted. and where $\Sigma = I_T \times \Omega$. The covariance matrix. $\Omega$. can either be block diagonal

$$\Omega = \begin{pmatrix} \Omega_1 & & 0 \\ & \ddots & \\ 0 & & \Omega_n \end{pmatrix}$$

or be unrestricted.

Similar to the examples above. these (sets of) restrictions can be expressed in terms of $G$ and $H$ matrices.

**Example 5.3.6 (Sector cointegration).** Sector cointegration is similar to panel cointegration. The parameters have a block-diagonal structure except for one set of rows that corresponds to a common set of variables. $X_{0t}$. This can be expressed as

$$\begin{pmatrix} \Delta X_{0t} \\ \Delta X_{1t} \\ \vdots \\ \Delta X_{nt} \end{pmatrix} = \alpha\beta' \begin{pmatrix} X_{0,t-1} \\ X_{1,t-1} \\ \vdots \\ X_{1,t-1} \end{pmatrix} + \Gamma_1 \begin{pmatrix} \Delta X_{0,t-1} \\ \Delta X_{1,t-1} \\ \vdots \\ \Delta X_{n,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{0t} \\ \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{pmatrix}.$$

*where* $\alpha$ *may have the structure:*

$$\alpha = \begin{pmatrix} \alpha_{01} & \cdots & \alpha_{0n} \\ \alpha_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \alpha_n \end{pmatrix}, \quad \alpha = \begin{pmatrix} (\alpha_{01}\ldots\ldots\alpha_{0n}) \\ I_n \times \alpha \end{pmatrix}.$$

*or*

$$\alpha = \begin{pmatrix} (1\ldots\ldots1) \times \alpha_0 \\ I_n \times \alpha \end{pmatrix}.$$

*and* $\beta$ *may have the structure*

$$\beta = \begin{pmatrix} \beta_{01} & \cdots & \beta_{0n} \\ \beta_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \beta_n \end{pmatrix}, \quad \beta = \begin{pmatrix} (\beta_{01}\ldots\ldots\beta_{0n}) \\ I_n \times \beta \end{pmatrix}.$$

*or*

$$\beta = \begin{pmatrix} (1\ldots\ldots1) \times \beta_0 \\ I_n \times \beta \end{pmatrix}.$$

*and a similar structure for* $\Gamma_1$.

## Appendix E: Proofs

For notational convenience, define

$$Z_{1B2} = ((Z_1' B . Z_2') \times I_p)).$$

$$Z_{1A} = (Z_1' \times A) K_{p_1, r}.$$

where $K_{p_1,r}$ is the commutation matrix, uniquely defined by the $K_{p_1,r}\mathrm{vec}(B) \equiv \mathrm{vec}(B')$ for any $p_1 \times r$ matrix $B$.

Let $\Sigma$ be the covariance matrix of $\varepsilon = \mathrm{vec}(\varepsilon_1,\ldots,\varepsilon_T)$. In practice, it can be burdensome to work directly with the $Tp \times Tp$ matrix $\Sigma$ in equations (5.2.2) and (5.2.3). The following lemma is therefore useful.

**Lemma E.7.** Let $\Sigma_{t\tau}^{-1}$ be the $p \times p$ sub-matrix of $\Sigma^{-1}$, such that the $(i,j)$th element of $\Sigma_{t\tau}^{-1}$ is the $(p(t-1)+i,p(\tau-1)+j)$th element of $\Sigma^{-1}$.

Then

$$T^{-1}\mathbf{Z}'_{1A}\Sigma^{-1}\mathbf{Z}_{1A} = \mathbf{M}_A^{hac} = T^{-1}\sum_{t,\tau=1}^{T}\left[A'\Sigma_{t\tau}^{-1}A \times Z_{1t}Z'_{1\tau}\right]. \tag{E.2}$$

$$T^{-1}\mathbf{Z}'_{1A}\Sigma^{-1}\mathrm{vec}(Z_0 - CZ_2) = \mathbf{N}_{AC}^{hac} = T^{-1}\sum_{t,\tau=1}^{T}\mathrm{vec}\left(Z_{1t}(Z_{0\tau} - CZ_{2\tau})'\Sigma_{t\tau}^{-1'}A\right). \tag{E.3}$$

$$T^{-1}\mathbf{Z}'_{1B2}\Sigma^{-1}\mathbf{Z}_{1B2} = \mathbf{M}_B^{hac} = T^{-1}\sum_{t,\tau=1}^{T}\left[\begin{pmatrix} B'Z_{1t}Z'_{1\tau}B & B'Z_{1t}Z'_{2\tau} \\ Z_{2t}Z'_{1\tau}B & Z_{2t}Z'_{2\tau} \end{pmatrix} \times \Sigma_{t\tau}^{-1}\right] \tag{E.4}$$

$$T^{-1}\mathbf{Z}'_{1B2}\Sigma^{-1}\mathrm{vec}(Z_0) = \mathbf{N}_B^{hac} = T^{-1}\sum_{t,\tau=1}^{T}\mathrm{vec}\left(\Sigma_{t\tau}^{-1}Z_{0\tau}(Z'_{1t}B,Z'_{2t})\right). \tag{E.5}$$

If $\Sigma$ is block diagonal, with $T$ blocks of size $p \times p$ given by $\Omega(t), t = 1,\ldots,T$, then

$$T^{-1}\mathbf{Z}'_{1A}\Sigma^{-1}\mathbf{Z}_{1A} = \mathbf{M}_A^{he} = T^{-1}\sum_{t=1}^{T}\left[A'\Omega(t)^{-1}A \times Z_{1t}Z'_{1t}\right].$$

$$T^{-1}\mathbf{Z}'_{1A}\Sigma^{-1}\mathrm{vec}(Z_0 - CZ_2) = \mathbf{N}_{AC}^{he} = T^{-1}\sum_{t=1}^{T}\mathrm{vec}\left(Z_{1t}(Z_{0\tau} - CZ_{2\tau})'\Omega(t)^{-1}A\right).$$

$$T^{-1}\mathbf{Z}'_{1B2}\Sigma^{-1}\mathbf{Z}_{1B2} = \mathbf{M}_B^{he} = T^{-1}\sum_{t=1}^{T}\left[\begin{pmatrix} B'Z_{1t}Z'_{1t}B & B'Z_{1t}Z'_{2t} \\ Z_{2t}Z'_{1t}B & Z_{2t}Z'_{2t} \end{pmatrix} \times \Omega(t)^{-1}\right].$$

$$T^{-1}\mathbf{Z}'_{1B2}\Sigma^{-1}\mathrm{vec}(Z_0) = \mathbf{N}_B^{he} = T^{-1}\sum_{t=1}^{T}\mathrm{vec}\left(\Omega(t)^{-1}Z_{0t}(Z'_{1t}B,Z'_{2t})\right).$$

*If $\{\varepsilon_t\}$ is i.i.d. Gaussian with covariance matrix $\Omega$, the expressions simplify to:*

$$T^{-1}\mathbf{Z}'_{1A}\Sigma^{-1}\mathbf{Z}_{1A} \quad = \quad \mathbf{M}_A^{ho} = \left[ A'\Omega^{-1}A \otimes M_{11} \right],$$

$$T^{-1}\mathbf{Z}'_{1A}\Sigma^{-1}\mathrm{vec}(Z_0 - CZ_2) \quad = \quad \mathbf{N}_{AC}^{ho} = \mathrm{vec}\left( (M_{10} - M_{02}C')\Omega^{-1}A \right),$$

$$T^{-1}\mathbf{Z}'_{1B2}\Sigma^{-1}\mathbf{Z}_{1B2} \quad = \quad \mathbf{M}_B^{ho} = \left[ \begin{pmatrix} B'M_{11}B & B'M_{12} \\ M_{21}B & M_{22} \end{pmatrix} \otimes \Omega^{-1} \right].$$

$$T^{-1}\mathbf{Z}'_{1B2}\Sigma^{-1}\mathrm{vec}(Z_0) \quad = \quad \mathbf{N}_B^{ho} = \mathrm{vec}\left( \Omega^{-1}(M_{01}B, M_{02}) \right).$$

**Proof.** The identity

$$\mathbf{Z}'_{1A}\Sigma^{-1}\mathbf{Z}_{1A} \quad = \quad K'_{p_1,r}(Z_1 \otimes A')\Sigma^{-1}(Z'_1 \otimes A)K_{p_1,r}$$

$$= \quad K_{r,p_1} \sum_{t,\tau=1}^{T} (Z_{1t} \otimes A')\Sigma^{-1}_{t,\tau}(Z'_{1\tau} \otimes A)K_{p_1,r}$$

$$= \quad K_{r,p_1} \sum_{t,\tau=1}^{T} (Z_{1t} \otimes A'\Sigma^{-1}_{t,\tau})(Z'_{1\tau} \otimes A)K_{p_1,r}$$

$$= \quad K_{r,p_1} \sum_{t,\tau=1}^{T} (Z_{1t}Z'_{1\tau} \otimes A'\Sigma^{-1}_{t,\tau}A)K_{p_1,r}$$

$$= \quad \sum_{t,\tau=1}^{T} (A'\Sigma^{-1}_{t,\tau}A \otimes Z_{1t}Z'_{1\tau}).$$

proves (E.2) and

$$\mathbf{Z}'_{1A}\Sigma^{-1}\mathrm{vec}(Z_0 - CZ_2) \quad = \quad K_{r,p_1} \sum_{t,\tau=1}^{T} (Z_{1t} \otimes A')\Sigma^{-1}_{t,\tau}\mathrm{vec}(Z_{0\tau} - CZ_{2\tau})$$

$$= \quad K_{r,p_1} \sum_{t,\tau=1}^{T} (Z_{1t} \otimes A'\Sigma^{-1}_{t,\tau})\mathrm{vec}(Z_{0\tau} - CZ_{2\tau})$$

$$= \quad K_{r,p_1} \sum_{t,\tau=1}^{T} \mathrm{vec}(A'\Sigma^{-1}_{t,\tau}(Z_{0\tau} - CZ_{2\tau})Z'_{1t})$$

$$= \quad \sum_{t,\tau=1}^{T} \mathrm{vec}\left( Z_{1t}(Z_{0\tau} - CZ_{2\tau})'\Sigma^{-1'}_{t\tau}A \right).$$

proves (E.3). Equations (E.4) and (E.5) are proven similarly. The last eight identities follow by setting $\Sigma_{t,\tau}^{-1} = \Omega(t)^{-1}$ or $\Sigma_{t,\tau}^{-1} = \Omega^{-1}$ for $t = \tau$ and zero otherwise. ∎

**Proof of Theorem 5.2.1.** Applying the vec operation to equation (5.2.1) yields the equation

$$
\begin{aligned}
\text{vec}(Z_0) &= (Z_1'B \otimes I_p)\text{vec}(A) + (Z_2' \otimes I_p)\text{vec}(C) + \varepsilon \\
&= [(Z_1'B.Z_2') \otimes I_p)]\,\text{vec}(A.C) + \varepsilon \\
&= \mathbf{Z}_{1B2}(G\upsilon + g) + \varepsilon.
\end{aligned}
$$

which may be rewritten as

$$
\text{vec}(Z_0) - \mathbf{Z}_{1B2}g = \mathbf{Z}_{1B2}G\upsilon + \varepsilon.
$$

Then for fixed values of $B$ and $\Sigma$ this is a restricted GLS problem with the well know solution given in equation (5.2.2), using Equations E.4 and E.5.

Similarly. for fixed $A.$ $C.$ and $\Sigma.$ we have the equation

$$
\begin{aligned}
\text{vec}(Z_0 - CZ_2) &= \text{vec}(AB'Z_1) + \varepsilon \\
&= (Z_1' \otimes A)\text{vec}(B') + \varepsilon \\
&= (Z_1' \otimes A)K_{p_1,r}\text{vec}(B) + \varepsilon \\
&= \mathbf{Z}_{1A}\text{vec}(B) + \varepsilon = \mathbf{Z}_{1A}(H\phi + h) + \varepsilon.
\end{aligned}
$$

which we rewrite as

$$
\text{vec}(Z_0 - CZ_2) - \mathbf{Z}_{1A}h = \mathbf{Z}_{1A}H\phi + \varepsilon.
$$

This is also a restricted GLS problem, with the solution given in equation (5.2.3), using Equations E.2 and E.3. ∎

**Proof of Corollaries 5.2.2 and 5.2.3.** Follows from Theorem 5.2.1 and Lemma E.7. ∎

The remaining corollaries were proven in Appendix B of Chapter 2.

# References

AHN, S. K., AND G. C. REINSEL (1990): "Estimation for Partially Non-Stationary Multivariate Autoregressive Models," *Journal of the American Statistical Association*, 85, 813–823.

ANDERSON, T. W. (1951): "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *Annals of Mathematical Statistics*, 22, 327–351.

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

BOSWIJK, P. H. (1995): "Identifiability of Cointegrated Systems," *Working paper, Tinbergen Institute*.

DEN HAAN, W. J., AND A. T. LEVIN (1997): "A Practitioner's Guide to Robust Covariance Matrix Estimation," in *Handbook of Statistics*, ed. by G. S. Maddala, and C. R. Rao. Elsevier, Amsterdam, vol. 15, pp. 309–327.

ELLIOTT, G. (1997): "Minimum Distance Methods and Cointegrated Models," *UCSD manuscript*.

——— (1998): "Estimating Restricted Cointegrating Vectors," *UCSD manuscript*.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Co-Integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.

GROEN, J. J. J., AND F. R. KLEIBERGEN (1999): "Likelihood-Based Cointegration Analysis in Panels of Vector Error Correction Models," *Tinbergen Institute Discussion Paper TI 99-055/4*.

HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton N.J.

JOHANSEN, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

JOHANSEN, S., AND K. JUSELIUS (1992): "Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK," *Journal of Econometrics*, 53, 211–244.

LARSSON, R., AND J. LYHAGEN (1999): "Likelihood-Based Inference in Multivariate Panel Cointegration Models," *Working Paper Series in Economics and Finance, No. 331, Stockholm School of Economic*.

LARSSON, R., J. LYHAGEN, AND M. LOTHGREN (1998): "Likelihood-Based Cointegration Test in Heterogeneous Panels," *Working Paper Series in Economics and Finance No. 250, Stockholm School of Economics*.

LÜTKEPOHL, H., AND H. CLAESSEN (1997): "Analysis of Cointegrated VARMA Processes," *Journal of Econometrics*, 80, 223–239.

MOSCONI, R., AND C. GIANNINI (1992): "Non-Causality in Cointegrated Systems: Representation Estimation and Testing," *Oxford Bulletin of Economics and Statistics*, 54, 399–417.

NEWEY, W., AND K. WEST (1987): "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

OSBORN, D. R. (1977): "Exact and Approximate Maximum Likelihood Estimators for Vector Moving Average Processes," *Journal of Royal Statistical Society. Series B*, 39, 114–118.