

Wagner, Martin

Working Paper

VAR cointegration in VARMA models

Reihe Ökonomie / Economics Series, No. 65

Provided in Cooperation with:

Institute for Advanced Studies (IHS), Vienna

Suggested Citation: Wagner, Martin (1999) : VAR cointegration in VARMA models, Reihe Ökonomie / Economics Series, No. 65, Institute for Advanced Studies (IHS), Vienna

This Version is available at:

<https://hdl.handle.net/10419/70293>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

VAR Cointegration in VARMA Models

Martin Wagner

VAR Cointegration in VARMA Models

Martin Wagner

Reihe Ökonomie / Economics Series No. 65

May 1999

Martin Wagner
Abteilung Ökonomie
Institut für Höhere Studien
Stumpergasse 56, A -1060 Wien
Fax: ++43/1/599 91-163
Phone: ++43/1/599 91-155
E-mail: mwagner@ihs.ac.at

**Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna**

The Institute for Advanced Studies in Vienna is an independent center of postgraduate training and research in the social sciences. The **Economics Series** presents research done at the Economics Department of the Institute for Advanced Studies. Department members, guests, visitors, and other researchers are invited to contribute and to submit manuscripts to the editors. All papers are subjected to an internal refereeing process.

Editorial

Main Editor:

Robert M. Kunst (Econometrics)

Associate Editors:

Walter Fisher (Macroeconomics)

Klaus Ritzberger (Microeconomics)

Abstract

The method for estimation and testing for cointegration put forward by Johansen assumes that the data are described by a vector autoregressive process. In this article we extend the data generating process to autoregressive moving average models without unit roots in the MA polynomial. We first extend some matrix algebraic relationships for $I(1)$ processes and derive their implications for the structure theory of cointegration. Specifically we show that the cointegrating space is invariant to MA errors which have no unit roots in the MA polynomial. The above results permit to prove the robustness of the Johansen estimates of the cointegrating space in a Gaussian vector autoregressive framework when the true model is vector autoregressive moving average, without unit roots in the MA polynomial. The small sample properties of the theoretical results are examined through a small simulation study.

Keywords

Cointegration, Johansen procedure, misspecification, robustness, simulation, Hausdorff distance

JEL Classifications

C13, C15, C32

Comments

The author thanks S. Johansen, M. Deistler, and W. Polasek for helpful remarks and comments.

Contents

1	Introduction	1
2	Some Matrix Algebra of I(1) Processes	2
3	The Behaviour of the Johansen Estimates under Misspecification	10
4	Results of a Simulation Study	13
5	Conclusions	20
	References	23
A	Appendix to Section 2	25
B	Appendix to Section 3	27
C	Appendix to Section 4	33

1 Introduction

The estimation and testing procedure for cointegration in a vector autoregressive framework put forward by Johansen [11, 12, 13] is probably one of the most important developments in time series econometrics during the last decade. Due to its simplicity, and also because of the elaborate possibilities to test hypotheses on the cointegrating space, this method is the most popular in the literature.

One issue that has – to the author’s knowledge – only been addressed by means of simulation studies (e.g. Bewley and Yang [1], Podivinsky [16] or Toda [19]) is the behaviour of the Johansen estimates under misspecification. This is an important question because the data might not be perfectly described by a low order autoregression.

In principle one could always overcome this problem by adding a sufficient number of lags. Saikkonen [18] has shown that, generalising the work of Said and Dickey [17] to the multivariate case, one can consistently estimate the cointegrating space of a general $I(1)$ process by increasing the lag order of an autoregressive approximation with an appropriate rate for increasing sample size.

For the sample sizes usually provided by macro-econometric applications this result is probably of limited relevance. Due to the short sample sizes one is usually restricted to estimating simple models like low order autoregressions, although on principle one can use the maximum likelihood approach to estimation and testing in cointegrated vector $ARMA$ models developed by Yap and Reinsel [21]. For small samples however, it may be difficult to discriminate between a pure autoregressive model and an autoregressive moving average model. Beside the more empirically oriented questions, the issue of the robustness of the Johansen procedure is also interesting from a purely statistical point of view.

In Section 2 we start with a review of some basic concepts and relationships for integrated processes, like Error Correction Model and the structure theory for $I(1)$ processes related to issues of (co-)integration. In this section we will show that although $I(1)$ - AR processes always have an error correction representation, this does not hold true for $I(1)$ - $ARMA$ processes. We formulate several sets of sufficient conditions on the matrix polynomials of a vector $ARMA$ process to allow for an error correction representation. In Section 2 we also show that the AR process $a(z)y_t = \epsilon_t$ and all left co-prime $ARMA$ systems given by $a(z)y_t = b(z)\epsilon_t$ have the same cointegrating space of order 1, if $\det(b(1)) \neq 0$.

This result forms the basis for the possibility that the Johansen procedure yields consistent estimates of the cointegrating space also under misspecification. In this paper we restrict ourselves to the type of misspecification, following from the above result, where the true model is a left co-prime $ARMA$ system with unit roots only in the AR polynomial, but where a pure AR system is estimated. The above result shows that the cointegrating space is determined by the AR part of the model, which is intuitively clear because the allowed moving average dynamics have a short run.

In Section 3 we prove that the Johansen procedure yields consistent estimates of the cointegrating space under the considered type of misspecification. We also show that the adjustment parameter matrix, α in Johansen notation, and the variance-covariance matrix of the noise process ϵ_t are generally not estimated consistently when the system is misspecified.

In the course of proving the consistency of the estimates for the cointegrating space we will, as a by-product, see that the power for fixed alternatives of the trace and max tests is tending to one also under misspecification.

To assess the empirical relevance of the results above we report the results of a simulation study in Section 4. The aim of this small simulation study is to analyse the effects of the misspecification for different sample sizes on the estimated cointegrating space and on the size of the tests. The sample sizes we use are $T = 50, 100, 150$ and 200 observations. It turns out that the Johansen method is remarkably robust for the larger sample sizes with respect to the discussed misspecifications, with regard to the size of the tests, to the distribution of the estimated cointegrating vectors and also to the quality of the approximation of the true cointegrating space by the estimated cointegrating space. The quality of approximation is measured by the Hausdorff distance between the estimated and the true cointegrating space.

The results of the simulations carry two messages: The first one is already well known, for small sample sizes the size and power of the tests is rather poor, although for small systems the quality of the estimates is good. The second message is that somebody only interested in the cointegrating relationships can restrict himself to pure autoregressive models instead of using *ARMA* models or non-parametric methods, when there are no unit roots in the un-modelled MA polynomial.

In Appendix A we present the proofs of some of the lemmata of Section 2. In Appendix B the proofs of some lemmata of Section 3 are given. In Appendix C we show, by means of a simple example, what may happen to the test statistic when the *MA* polynomial has unit roots, too. In this appendix some tables and figures related to Section 4 are also presented.

2 Some Matrix Algebra of $I(1)$ Processes

In this section we will first review the matrix algebraic properties of integrated processes. We will first discuss the general case of processes integrated of order d and will then go on to look at the specific properties of $I(1)$ systems. The results of the first part are taken from Johansen [10]. The case of variables integrated of order 2 is discussed in Haldrup and Salmon [7], a discussion of general orders of integration is given by Gregoir [5, 6].

Let y_t be an m -dimensional stochastic process integrated of order d ($y_t \in I(d)$), i.e. y_t is stationary after differencing d times but not stationary after differencing $d - 1$ times. We start with the definition of a *Generalised Error Correction Model*

Definition: A time series model of the form

$$D_{-s}\Delta^{-s}y_t + \dots + D_{k-1}\Delta^{k-1}y_t + D_k(z)\Delta^k y_t = f(z)\Delta^p \epsilon_t$$

is called *Generalised Error Correction Model (GECM) of order k*, if the following conditions hold:

1. $D_k(z)$ is holomorphic for $|z| < 1 + \rho$ and $D_k(1) \neq 1$,
2. $D_i\Delta^i y_t$ is stationary for $i = -s, \dots, k - 1$,
3. y_t is integrated of order k , i.e the integration order d has to equal k ,
4. $f(z) \neq 0 \forall |z| < 1 + \rho$.

where Δ denotes the difference operator. The terms $D_i\Delta^i y_t$ for $i = -s, \dots, -1$ are called *integral correction terms*, the terms $D_i\Delta^i y_t$ for $i = 0, \dots, k - 1$ are the *error correction terms* and $D_k\Delta^k y_t$ is the *autoregressive* part of the model. The rows of D_0 give the linear combinations of y_t that are stationary. The rows of D_1 are those linear combinations of Δy_t that are stationary, etc.

For $s = p = 0$ the above reduces to an *Error Correction Model*.

From our assumptions it follows that $\Delta^d y_t$ is stationary and therefore it has a Wold representation $\Delta^d y_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$, where the function $C(z) = \sum_{j=0}^{\infty} c_j z^j$ is existing and finite on the set $|z| < 1 + \epsilon$. We now develop the function $C(z)$ in its power series around the point $z = 1$

$$C(z) = \sum_{j=0}^{\infty} C_j (1 - z)^j \quad |1 - z| < \rho$$

and let $C_n(z)$ denote

$$C_n(z) = \sum_{j=0}^{\infty} (1 - z)^j C_{j+n},$$

then

$$C(z) = \sum_{j=0}^{n-1} (1 - z)^j C_j + (1 - z)^n C_n(z).$$

Also $\bar{C}(z)$, the adjoint matrix function of $C(z)$, is developed as a power series around $z = 1$.

$$\bar{C}(z) := \sum_{j=0}^{\infty} (1 - z)^j \bar{C}_j$$

Next, define the following subspaces of \mathbb{R}^m

$$\begin{aligned} N_j &:= \{x \in \mathbb{R}^m : x' C_j = 0\} & \forall j = 0, 1, \dots \\ M_j &:= N_0 \cap N_1 \cap \dots \cap N_j & \forall j = 0, 1, \dots \end{aligned}$$

and $m_j := \dim(M_j)$. So M_j is the space of all left null vectors of the matrices C_0, \dots, C_j . Since $C(0) = I$, there exists no vector x , for which $x'C_j = 0 \forall j$ holds, i.e. there is a smallest k so that $M_j = \emptyset \forall j \geq k$ holds.

Let n furthermore denote the sum of the dimensions of all the M_j , i.e.

$$n := \sum_{j=0}^{\infty} m_j.$$

For all $x \in M_j$ the following holds:

$$\begin{aligned} \Delta^d y_t &= C(z)\epsilon_t \\ x'\Delta^d y_t &= x'C(z)\epsilon_t \\ x'\Delta^d y_t &= \Delta^{j+1}x'C_{j+1}(z)\epsilon_t \end{aligned}$$

Thus the space M_j contains all cointegration vectors of order greater or equal to $j+1$. The cointegrating vectors of order j are contained in the space V_j which is given by $V_0 = M_0^\perp$ and $V_j = M_{j-1} \cap M_j^\perp$ for $j = 1, \dots, k$. Consequently, $\mathbb{R}^m = V_0 \oplus \dots \oplus V_k$ has to hold.

In Appendix A some of the results given in Johansen [10], which form the basis of the results of Section 2, are briefly stated. (Theorems (A.1) to (A.5))

Against the background of the material summarised in the appendix in the sequel we look at the case of processes integrated of order 1. We will need some results for representations of matrix polynomials and rational functions of matrices, notably the Smith and the Smith-McMillan representation. A discussion of these concepts can be found e.g. in Hannan and Deistler [8]. Let us start with an AR process integrated of order 1

$$a(z)y_t = \epsilon_t, \tag{2.1}$$

where we assume $\det(a(z)) = 0$ implies $z = 1$ or $|z| > 1$. Then we also have

$$\Delta y_t = c(z)\epsilon_t, \tag{2.2}$$

with $c(z) = \sum_{j=0}^{\infty} c_j$ and $\sum_{j=0}^{\infty} \|c_j\|^2 < \infty$.

The Smith representation of $a(z)$ is given by

$$a(z) = u(z) \begin{pmatrix} \lambda_1(z) & & 0 \\ & \ddots & \\ 0 & & \lambda_m(z) \end{pmatrix} v(z)$$

where $u(z)$ and $v(z)$ are unimodular matrices, i.e. they have constant determinants $\neq 0$. Thus we see that $a^{-1}(z)$ is given by

$$a^{-1}(z) = v^{-1}(z) \begin{pmatrix} \lambda_1^{-1}(z) & & 0 \\ & \ddots & \\ 0 & & \lambda_m^{-1}(z) \end{pmatrix} u^{-1}(z)$$

The assumption that $y_t \in I(1)$ implies that $c(z) = (1 - z)a^{-1}(z) = \Delta a^{-1}(z)$ exists on the closed unit circle. Therefore

$$c(z) = (1 - z)a^{-1}(z) = v^{-1}(z) \begin{pmatrix} (1 - z)\lambda_1^{-1}(z) & & 0 \\ & \ddots & \\ 0 & & (1 - z)\lambda_m^{-1}(z) \end{pmatrix} u^{-1}(z)$$

has to be finite at $z = 1$.

Since u and v are unimodular, it suffices to analyse $(1 - z)\Lambda^{-1}(z)$.

The expressions $(1 - z)\lambda_i^{-1}(z)$ exist for $z = 1$ finitely only if the multiplicity of the zero $z = 1$ of $\lambda_i(z)$ is 0 or 1. This means $d = 1$ is equivalent to the existence of a number q , with $1 \leq q \leq m$, so that $\lambda_1, \dots, \lambda_q$ are not equal to 0 at $z = 1$ and $\lambda_{q+1}, \dots, \lambda_m$ have a zero of order 1 at $z = 1$.

For $I(1) - AR$ processes we can prove the following strengthening of Theorem A.5

Lemma 2.1 *If y_t is an $I(1)$ -AR-process, then $r = n$ and $k = 1$ hold.*

Proof: See Appendix A

From Theorem A.5 we only know that $r' = n'$ implies $d = k'$. Lemma 2.1 shows a stronger relationship for the case $d = 1$, where the matrix function is always balanced. For having a representation as an error correction model in Theorem A.4 $d = k$ was a necessary condition. Since we are especially interested in $I(1)$ -systems, i.e. $d = 1$, we want to know when $k = 1$.

A characterisation in terms of the matrix function $c(z)$ is given in the next corollary

Corollary 2.2 *Under the assumption $d = 1$, the following four statements are equivalent*

1. $k = 1$
2. $N_1 \subseteq N_0^\perp$
3. $\text{lker}(\frac{d}{dz}(c(1))) \subseteq (\text{lker}(c(1)))^\perp$
4. $\text{lker}(C_1) \subseteq (\text{lker}(C_0))^\perp$

Proof:

It is only necessary to prove the equivalence between the first two statements, because the last two are only re-formulations of the second.

Since $k = 1$ the space \mathbb{R}^m can be represented as

$$\mathbb{R}^m = V_0 \oplus V_1$$

and all other $V_j, j = 2, 3, \dots$ are equal to the empty set, denoted by \emptyset .

By construction we also have $V_0 = M_0^\perp = N_0^\perp$ and therefore

$$\mathbb{R}^m = V_0 \oplus N_0.$$

This shows that $V_1 = N_0$.

By definition $V_1 = N_0 \cap N_1^\perp$, which directly implies that $N_1^\perp \supseteq N_0$. All implications also hold in the other direction.

qed

This corollary shows that a regular derivative matrix of $c(z)$ at the point $z = 1$ is sufficient for $k = 1$.

We will now see that for $I(1)$ -*ARMA*-systems $k = 1$ and $r = n$ do not have to hold necessarily. We start from a left co-prime *ARMA*(p, q) system

$$a(z)y_t = b(z)\epsilon_t \quad (2.3)$$

integrated of order 1. Again we have an $MA(\infty)$ representation for the first differences of y_t

$$\Delta y_t = c(z)\epsilon_t \quad (2.4)$$

with $c(z)$ given by $\Delta a^{-1}(z)b(z)$. Let the Smith representations of $a(z)$ and $b(z)$ be given by

$$\begin{aligned} a(z) &= u(z)\Lambda(z)v(z) \quad , \Lambda(z) = \text{diag}(\lambda_i(z)) \\ b(z) &= o(z)\Gamma(z)p(z) \quad , \Gamma(z) = \text{diag}(\gamma_i(z)) \end{aligned}$$

then

$$c(z) = (1-z)v^{-1}(z)\Lambda^{-1}(z)u^{-1}(z)o(z)\Gamma(z)p(z)$$

Therefore the determinant of $c(z)$ is given by

$$\det(c(z)) = e(1-z)^m \prod_{i=1}^m \frac{\gamma_i(z)}{\lambda_i(z)},$$

where e is the product of the determinants of the unimodular matrices.

Next we denote by $\delta_{z_0}(f(z))$ the multiplicity of the zero $z = z_0$ of the function $f(z)$.

The index r of the matrix function $c(z)$, which we will now denote with r_c , is, using the above notation, given by

$$r_c = m + \sum_{i=1}^m [\delta_1(\gamma_i(z)) - \delta_1(\lambda_i(z))],$$

because the system is left co-prime.

As for the *AR* case we will again answer the question of the implications of the system to be $I(1)$ on the matrices $a(z)$ and $b(z)$.

We start with

$$\Delta y_t = \bar{u}(z) \begin{pmatrix} \frac{(1-z)\epsilon_1(z)}{\psi_1(z)} & & \\ & \ddots & \\ & & \frac{(1-z)\epsilon_m(z)}{\psi_m(z)} \end{pmatrix} \bar{v}(z)\epsilon_t$$

Here the ϵ_i and ψ_i shall be the polynomials from the Smith-McMillan representation¹. Because the ϵ_i and ψ_i are relative prime, i.e. they have no common zeros, $\delta_1(\psi_i(z))$ has to be 0 or 1, because in the case of multiple zeros at $z = 1$ $c(z)$ would have a pole at $z = 1$ which would imply $y_t \notin I(1)$. Thus there exists a set $T = \{1, \dots, t\} \subseteq \{1, \dots, m\}$ for which we have $\psi_i(1) = 0 \forall i \in T$. The relative primeness implies $\epsilon_i(1) \neq 0 \forall i \in T$. Independently of $\delta_1(\epsilon_i(z))$, $t \leq i \leq m$ the following holds

$$\bar{\Lambda}(1) = \begin{pmatrix} * & & & & & \\ & \ddots & & & & \\ & & * & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

because the factors $(1 - z)$ in the columns $t + 1, \dots, m$ cannot be cancelled by a zero of the corresponding $\psi_i(z)$.

So for $y_t \in I(1)$, independently of $\epsilon_i(z)$,

$$rank(c(1)) = m - \sum_{i=1}^m \delta_1(\psi_i(z))$$

holds. By simple enumeration one also sees that

$$r_c = m - \sum_{i=1}^m \delta_1(\psi_i(z)) + \sum_{i=1}^m \delta_1(\epsilon_i(z)). \tag{2.5}$$

For the indices k_c and n_c the following relationships hold.

Lemma 2.3 Under the assumption that $k_c = d = 1$ $c(z)$ is balanced, i.e. $r_c = n_c$, if and only if $\sum_{i=1}^m \delta_1(\epsilon_i(z)) = 0$.

Proof:

$k_c = 1$ implies $n_c = m_0$, because all $M_i = \emptyset \forall i \geq 1$.

That implies

$$n_c = m_0 = m - \sum_{i=1}^m \delta_1(\psi_i(z))$$

If one now uses (2.12) one sees that equality of r_c and n_c holds exactly when $\sum_{i=1}^m \delta_1(\epsilon_i(z)) = 0$.

qed

¹If $a(z)y_t = b(z)\epsilon_t$ is a left co-prime ARMA system, then for the Smith representations of $a(z) = u(z)\Lambda(z)v(z)$, $\Lambda(z) = \text{diag}(\lambda_i(z))$ and $b(z) = o(z)\Gamma(z)p(z)$, $\Gamma(z) = \text{diag}(\gamma_i(z))$ and the Smith-McMillan representation of $k(z) = a^{-1}(z)b(z) = \bar{u}(z)\bar{\Lambda}(z)\bar{v}(z)$ with $\bar{\Lambda}(z) = \text{diag}(\frac{\epsilon_i(z)}{\psi_i(z)})$ the following relationship holds: $\lambda_i(z) = \psi_{m+1-i}(z)$, $i = 1, \dots, m$ and $\gamma_i(z) = \epsilon_i(z)$, $i = 1, \dots, m$, as has been shown e.g. in Wagner [20]. Therefore, because of this one-to-one relationship, it is not necessary to distinguish between the polynomials from the ARMA representation and the MA(∞) representation.

The above implies that the polynomials $\epsilon_i(z)$, which can be attributed to the MA part, do not under the assumptions stated exert any influence, on the cointegration space of order 1.

Under slightly different assumptions than before one obtains

Lemma 2.4 *Let $a(z)y_t = b(z)\epsilon_t$ be an ARMA-System and let the following assumptions be fulfilled*

1. (a, b) are left co-prime
2. $y_t \in I(1)$
3. $r_c = n_c$
4. $\det(b(1)) \neq 0$

Then it also follows that $k_c = 1$.

Proof:

From Theorem A.1 it is known that r_c is larger or equal than n_c .

From assumption 3 $r_c = n_c = \sum_{i=0}^{\infty} m_i$ holds and relation (2.5) is by assumption 4 now equal to

$$r_c = m - \sum_{i=1}^m \delta_1(\psi_i(z))$$

That means

$$n_c = m - \sum_{i=1}^m \delta_1(\psi_i(z)) = m_o$$

which implies $m_i = 0$ has to hold $\forall i \geq 1$.

This is equivalent to $k_c = 1$.

qed

Lemma 2.3 and Lemma 2.4 show the influence of the MA polynomial on the properties of the matrix function $c(z)$, reflected by the indices k_c, n_c and r_c . We see that under the assumptions of Corollary 2.2 and $\det(b(1)) \neq 0$ the matrix $c(z)$ fulfils the assumptions of Theorem A.1.

The situation is not as simple for $ARMA$ processes as it is for AR processes integrated of order 1, for which Lemma 2.1 holds.

In the $ARMA$ case either $r = n$ or $k = d$ and $\det(b(1)) \neq 0$ have to be assumed to guarantee that both conditions hold.

Now from the point of view of the Johansen AR based approach to cointegration we can interpret $ARMA$ systems as AR systems disturbed by an MA polynomial $b(z)$. This interpretation raises the question of the relationship between the cointegrating spaces of the AR and the $ARMA$ system. Although as we have seen for $I(1)$ - AR systems we always have $k = 1$ and $r = n$, this is not true for all $I(1)$ - $ARMA$ systems. The $ARMA$

system can e.g. be unbalanced. If $a(z)$ and $b(z)$ are left co-prime then the integration order of the AR system $a(z)y_t = \epsilon_t$ is the same as for the $ARMA$ system $a(z)y_t = b(z)\epsilon_t$, because no cancellations of zeros of $a(z)$ can occur. We have the following lemma:

Lemma 2.5 *Let*

$$a(z)y_t = \epsilon_t \quad (2.6)$$

be an $AR(p)$ -System integrated of order 1. Furthermore $a(1)$ shall be singular but $\neq O_m$. Then for all $ARMA(p, q)$ systems

$$a(z)y_t = b(z)\epsilon_t \quad (2.7)$$

with (a, b) relative left prime and $\det(b(1)) \neq 0$ the cointegrating spaces of order 1 of the system (2.6) and all systems (2.7) are identical.

Proof:

For $I(1)$ processes given in the MA representation of the stationary first differences $\Delta y_t = c(z)\epsilon_t$ the cointegrating spaces of order one are given by the left kernel of $c(1)$ denoted by $\text{lker}(c(1))$. This means specifically for the AR systems (2.6)

$$\Delta y_t = c_{AR}(z)\epsilon_t = (1-z)a^{-1}(z)\epsilon_t$$

Representing the $ARMA$ systems (2.7) in the same way one gets

$$\Delta y_t = c_{ARMA}(z)\epsilon_t = (1-z)a^{-1}(z)b(z)\epsilon_t$$

From the regularity assumption on $b(1)$ of course

$$\text{lker}(c_{AR}(1)) = \text{lker}(c_{ARMA}(1))$$

follows, which finishes the proof, because under our assumptions also all the considered $ARMA$ systems are $\in I(1)$.

qed

The above result can also be analysed by looking at the common trends representation of integrated systems of order 1, which is the content of the famous Granger representation theorem, see e.g. Engle and Granger [4]. To make the argument visible we write the $ARMA$ system as follows:

$$\begin{aligned} a(z)y_t &= u_t \\ u_t &= b(z)\epsilon_t \end{aligned}$$

which is reduced to the AR case if $b(z) = I$. Now the Granger representation theorem derives an MA representation of the above “autoregressive” representation, which is given by

$$y_t = \beta_{\perp}(\alpha'_{\perp} a_1(1)\beta_{\perp})^{-1} \alpha'_{\perp} \sum_{t=1}^T u_t + c_1(z)u_t \quad (2.8)$$

Here for a matrix $\gamma \in \mathbb{R}^{m \times r}$ with full rank γ_{\perp} is defined as a matrix of full rank, dimension $m \times (m - r)$ and $\gamma' \gamma_{\perp} = 0$. The matrices α and β stem from $-a(1) = \alpha \beta'$, where by assumption $a(1)$ is rank deficient. Furthermore $a_1(1)$ is given by $a(z) = a(1) + (1 - z)a_1(z)$. The same derivations apply to $c_1(z)$, where $c(z)$ is the inverse of $a(z)/(1 - z)$.

Now replace u_t by $b(z)\epsilon_t$ in (2.8)

$$\begin{aligned} y_t &= \beta_{\perp} (\alpha'_{\perp} a_1(1) \beta_{\perp})^{-1} \alpha'_{\perp} \sum_{t=1}^T \sum_{j=0}^q b_j \epsilon_{t-j} + c_1(z) b(z) \epsilon_t \\ &= \beta_{\perp} (\alpha'_{\perp} a_1(1) \beta_{\perp})^{-1} \alpha'_{\perp} b(1) \sum_{t=1}^T \epsilon_t + c_1(z) b(z) \epsilon_t \end{aligned} \quad (2.9)$$

If we want this to be a common trends representation we need the second term on the right hand side of the above equation to be stationary.

From the assumptions on $a(z)$ we know that $c_1(z)$ has all its roots outside the unit circle. Therefore we have to postulate that also $b(1)$ has no unit roots either, to guarantee the stationarity of that component. Now, if $b(1)$ is regular, we see that the common trends in the first component, in the pure autoregressive case given by $\alpha'_{\perp} \sum_{t=1}^T \epsilon_t$, are subject to a coordinate transformation due to pre-multiplication by $b(1)$ and are now given by $\alpha'_{\perp} b(1) \sum_{t=1}^T \epsilon_t$.

The above lemma shows that the regularity of $b(1)$ is sufficient, together with the assumption of left co-primeness of $a(z)$ and $b(z)$, for the cointegrating space of order 1 to remain unchanged.

These assumptions, however, do not guarantee that the matrix function $c(z) = (1 - z)a^{-1}(z)b(z)$ is balanced, so there could be cointegrating spaces of higher order.

An important remark at this stage is, as can easily be deduced from the above, that, for integrated processes of any order, the cointegrating space that reduces the order of integration by 1 is invariant to MA polynomials $b(z)$ as long as $\det b(1) \neq 0$.

3 The Behaviour of the Johansen Estimates under Misspecification

The results from the previous section, where we have seen that the cointegrating space of order 1 is the same for AR systems $a(z)y_t = \epsilon_t$ and for all left co-prime ARMA systems $a(z)y_t = b(z)\epsilon_t$ without unit roots in the MA polynomial $b(z)$, justify the question whether we can estimate the cointegrating vectors using the Johansen method under this form of "misspecification". We start with a brief description of the Johansen method, a detailed description can be found e.g. in the monograph by Johansen [13]. The starting point is a Gaussian vector autoregressive model of order p :

$$a(z)y_t = \epsilon_t, \quad \epsilon_t \sim NID(0, \Sigma)$$

From Theorem A.5 and Lemma 2.1 we know that the representation

$$a(1)y_t + a_1(z)\Delta y_t = \epsilon_t$$

is an error correction model. We use the equivalent representation

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + \Gamma_p y_{t-p} + \epsilon_t \quad (3.1)$$

where we have put the level term at lag p . The matrices Γ_i are given by $\Gamma_i = -I_m + a_1 + \dots + a_i$ for $i = 1, \dots, p$ and we also have $\Gamma_p = -a(1)$. Under the hypothesis that there are r cointegrating vectors one can rewrite $\Gamma_p = \alpha\beta'$ with α and $\beta \in R^{(m \times r)}$. The space spanned by the columns of β is then the cointegrating space of order 1.

The maximum likelihood estimation of β proceeds as follows: First is terminus technicus the parameter matrices $\Gamma_1, \dots, \Gamma_{p-1}$ out by running two OLS regressions: Regress Δy_t and y_{t-p} on the lagged differences $\Delta y_{t-1}, \dots, \Delta y_{t-p+1}$. The residuals of these two regressions are denoted with R_{0t} and R_{pt} . The product moment matrices of these residuals are then given by

$$S_{ij} = \frac{1}{T} \sum_{t=1}^T R_{it} R'_{jt}, \quad i, j = 0, p$$

Using the above quantities the maximum likelihood estimates of β are given by the eigenvectors corresponding to the r largest eigenvalues $\hat{\lambda}_1, \dots, \hat{\lambda}_r$ of

$$|\lambda S_{pp} - S_{p0} S_{00}^{-1} S_{0p}| = 0$$

The likelihood ratio test statistic of $H_0 : \dim(\beta) \leq r$ against the alternative $\dim(\beta) = m$ is given by

$$-2\ln(Q)_\eta = T \sum_{i=r+1}^m \ln(1 - \hat{\lambda}_i)$$

This test is denoted trace or η test. We can also test the hypothesis $H_0 : \dim(\beta) \leq r$ against the alternative $\dim(\beta) = r+1$. This leads to the max or ξ test with test statistic

$$-2\ln(Q)_\xi = \ln(1 - \hat{\lambda}_{r+1})$$

Let now $a(z)y_t = b(z)\epsilon_t$ be a left co-prime $ARMA(p, q)$ system, where the assumptions concerning $a(z)$ are as in Section 2 and $\det(b(1)) \neq 0$. We can re-parameterise the system as

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + \Gamma_p y_{t-p} + \epsilon_t + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q} \quad (3.2)$$

The model (3.2) does not necessarily form an *ECM* in the sense of the definition given in Section 2, but as we have seen in Lemma 2.4 the cointegrating space of order one is still given by $sp(\beta)$. We denote the variance matrix of the stationary variables $\beta'y_t$ by $\Sigma_{\beta\beta}$. For the model class described we can formulate the following theorem:

Theorem 3.1 *The Johansen estimation procedure for cointegrated vector autoregressive models yields consistent estimates for $\hat{\beta}$ under the form of ARMA misspecification discussed above .*

The estimates for $\hat{\Sigma}$ and $\hat{\Pi}$ are generally not consistent and are given by

$$\hat{\Pi} \xrightarrow{P} \Pi + \xi_p \alpha (\alpha' \alpha)^{-1} \Sigma_{\beta\beta}^{-1} \beta'$$

and

$$\begin{aligned} \hat{\Sigma} \xrightarrow{P} \Sigma - \xi_p \alpha (\alpha' \alpha)^{-1} \Sigma_{\beta\beta}^{-1} \beta' \Sigma_{p0} - \Sigma_{0p} \beta \Sigma_{\beta\beta}^{-1} (\alpha' \alpha)^{-1} \alpha' \xi_p' - \\ - \xi_p \alpha (\alpha' \alpha)^{-1} \Sigma_{\beta\beta}^{-1} (\alpha' \alpha)^{-1} \alpha' \xi_p' + \xi_0 + \xi_p' \end{aligned}$$

With

$$\begin{aligned} \xi_0 &= \sum_{n=1}^q b_n \Sigma c_n' - \tau_1 \mu_{**}^{-1} \mu_{*0} \\ \xi_p &= \tau_2 - \tau_1 \mu_{**}^{-1} \mu_{*0} \end{aligned}$$

and

$$\begin{aligned} \tau_1 &= \left(\sum_{n=1}^q b_n \Sigma c_{n-1}', \sum_{n=2}^q b_n \Sigma c_{n-2}', \dots \right); \\ \tau_2 &= \sum_{m=p}^q b_m \Sigma d_{m-p}' \Gamma_k' \end{aligned}$$

for $q \geq p$ and $\tau_2 = 0$ for $q < p$. The c_n are the coefficients from the Wold representation for Δy_t , i.e. from $\Delta y_t = c(z) \epsilon_t$, the b_m are the coefficients from the MA polynomial $b(z)$ and $d_{m-p} = -\sum_{i=j+1}^{\infty} c_i$.

Proof: See Appendix B.

We have seen in Section 2 that the common trends are given by $\alpha'_{\perp} b(1) \sum_{t=1}^T \epsilon_t$, therefore it is no surprise that the estimate of α is inconsistent since the loading matrix under misspecification is influenced by the MA polynomial. The same holds for the variance matrix of the residuals.

It would be interesting to relate these results to the literature on AR estimation of ARMA processes in the stationary case.

The asymptotic distributions of the likelihood ratio test statistics and the estimated cointegrating vectors remain unclear.² The derivation of the limit distribution of $(T\hat{\lambda}_{r+1}, \dots, T\hat{\lambda}_m)$ uses equation (B.11) and (B.12) from Lemma B.3 and (B.26) from Lemma B.4 which we have seen to depend on $b(z)$.

²When the model is correctly specified the limit of the test statistic of the trace test given by

$$tr \left(\int_0^1 dW W' \left(\int_0^1 W W' du \right)^{-1} \int_0^1 W dW' \right)$$

where W is an m -dimensional Wiener process.

The derivation of the limit distribution of $T[\hat{\beta}\hat{x}^{-1} - \beta]$ uses relation (B.36) from Lemma B.5, a relation that also depends on $b(z)$.

From the discussion after Lemma B.4 we know that, regardless of the misspecification, the first r eigenvalues are converging towards non-zero constants while the latter go to zero as $O_p(\frac{1}{T})$. This directly implies that the asymptotic power of the *trace* test against the alternative that there are $r + s$ cointegrating vectors is tending to 1, because then the test statistic $-T \sum_{i=r+1}^m \ln(1 - \hat{\lambda}_i)$ contains s terms that are diverging.³

On the other hand, since no analytical derivation of the limit distribution is available, it is unclear how much influence on the asymptotic distribution is really exerted by the *MA* polynomial $b(z)$. To deal with these issues a simulation study has been performed.

4 Results of a Simulation Study

The aim of this section is to report some results of simulations that have been performed to analyse the finite sample implications of the above results. Naturally any simulation study can only be interpreted with caution, but it may have some indicative value.⁴

We want to study several aspects. First we want to see whether the actual size of the test statistics really approximates the nominal size under the discussed type of misspecification. Then we want to analyse the behaviour of the estimated cointegrating space, i.e. the distribution of the estimated cointegrating space and the distance of the estimated vectors to the true cointegrating space.

Both of these aspects are investigated in reference to different sample sizes to see whether the established consistency of the estimated cointegrated space under misspecification is of empirical relevance.

As a distance measure between the estimated and the true cointegrating space we use the Hausdorff distance, which is defined thus:

Let ζ and η be two subspaces of \mathbb{R}^m . The intersection of a subspace θ of \mathbb{R}^m with the closed unit circle in \mathbb{R}^m is denoted by $C(\theta)$,

$$C(\theta) = \{z \in \theta \mid \|z\| \leq 1\},$$

where $\|z\|$ is the Euclidean norm of z . Using this notation the distance d of ζ and η is given by the Hausdorff distance d_H of $C(\zeta)$ and $C(\eta)$, i.e.

$$d(\zeta, \eta) = d_H(C(\zeta), C(\eta)) = \max(\rho(C(\zeta), C(\eta)), \rho(C(\eta), C(\zeta)))$$

where $\rho(C_1, C_2)$ is given by

$$\rho(C_1, C_2) = \sup_{x \in C_1} \inf_{y \in C_2} \|x - y\|.$$

³Analogous reasoning also works for the max test.

⁴The simulations have been performed using GAUSS 3.2, the programmes and further results are available from the author upon request.

The simulations have been performed as follows: We have generated 1000 time series of the system(s) of length 250, then we skipped the first 50 observations and constructed four samples of lengths $T = 50, 100, 150$ and 200, by taking the first 50, first 100 etc. observations. This setup is used to see the effects of increasing sample size on effectively growing samples. Then the Johansen procedure was run for an $AR(2)$ model.⁵

The first set of models that has been simulated are 2 dimensional $ARMA(2, 1)$ system with one cointegrating vector adopted from Hargreaves [9]:

$$\begin{aligned} \begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} &= \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = \begin{bmatrix} 1.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} u_{1t-1} \\ u_{2t-1} \end{bmatrix} + \\ &+ \begin{bmatrix} -0.5 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{1t-2} \\ u_{2t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} + \begin{bmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-1} \\ \epsilon_{2t-1} \end{bmatrix} \end{aligned} \quad (4.1)$$

The parameter values of the MA polynomials that we have chosen are $\gamma_1 = -1$ and $\gamma_2 = -0.9$, systems with

$$\gamma_1 = \gamma_2 = -0.8, -0.5, -0.2, 0, 0.2, 0.5, 0.8$$

and $\gamma_1 = 1$ and $\gamma_2 = 0.8$.

The first system has a unit root in the MA polynomial, so it does not fulfill the conditions of Theorem 3.1. Still, it is interesting to see the behaviour of the Johansen estimates in the case of the presence of an MA unit root. The fifth system, the pure $AR(2)$ system, serves as a point of reference.

The true cointegrating vector of the above system(s) is, suitably normed, given by $(1, -3)$, it is in fact the space spanned by the second row of the matrix in the beginning of the first line of (4.1).

Table 1 shows some characteristics of the distribution of the estimated cointegrating vectors. After normalising of the first coordinate to 1 in this example only one element of the estimated cointegrating vector is undetermined. Thus we only have an empirical distribution for the second coordinate which is described by its mean, standard deviation (S.D.), median, skewness (skew.) and kurtosis (kurt.).⁶ Our measure of skewness is

$$\frac{q_{97.5} - q_{50.0}}{q_{50.0} - q_{2.5}} - 1,$$

where q_i is the i -th quartile. As a measure of kurtosis we use

$$\frac{q_{97.5} - q_{2.5}}{q_{99.5} - q_{0.5}} - \frac{1.96}{2.575}.$$

⁵One issue that has to be mentioned here is that on principle one must look at the problem of lag length selection. Especially for small sample sizes and intermediate values of the MA parameters of order 1, information criteria do not reject this choice of the lag length. For large sample sizes the MA misspecification leads to a tendency to choose a higher autoregressive order and this makes the Johansen procedure more robust.

⁶The empirical distribution is calculated by taking the first solution vector of the eigenvalue problem in each repetition of the simulation.

Both measures are equal to zero for a normally distributed random variable.

In Table 1 (and also in Table 4 in Appendix C) HD-mean indicates the Hausdorff distance between the true cointegrating vector(s) and the vector(s) composed of the mean elements, over all replications, of the estimated cointegrating vector(s). The same holds true for HD-medi., of course with the median replacing the mean.

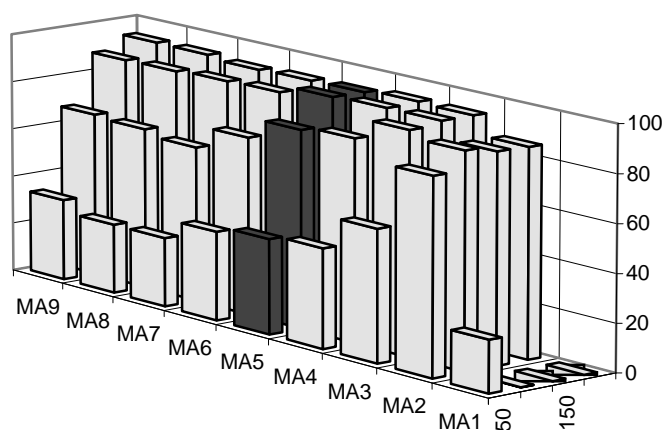
The behaviour of the estimates is very similar for all nine systems, the quality of approximation of the true cointegrating space by the estimated cointegrating space is already very good for only 50 observations. It is merely for sample size 50 that the standard deviation of the estimated cointegrating vector is smaller for the correctly specified AR system than for the misspecified systems.

This means that the (empirical) distribution is almost unchanged under the misspecification discussed in this example for sample sizes that are usually available in macro-econometrics.⁷

In applications we naturally do not only want to estimate a potential cointegrating space usually also determine the dimension of that space.⁸ The probabilities of accepting a specific dimension of the cointegrating space for our examples, using the sequential test procedures with the trace or max test, are given in Tables 2 and 3 in Appendix C. The probabilities of choosing the correct dimension are given in Figure 1 for the trace test and in Figure 2 for the max test.⁹

All test results reported in this paper are at the 95 % level, they are unchanged for other significance levels. Looking at the figures and the tables one sees that the results

Figure 1: Acceptance probability of the correct number of cointegrating vectors for systems (4.1) using the trace test



⁷The differences, although only minor, do not vanish if one looks at sample sizes up to 400 observations.

⁸If one runs the above described sequential test procedures, based on the trace and max test respectively, in the reversed order, one gets the same results as in the original order in more than 99.5 % of the cases, for all systems and also for the 3 dimensional systems described later.

⁹The critical values have been taken from Osterwald-Lenum [14].

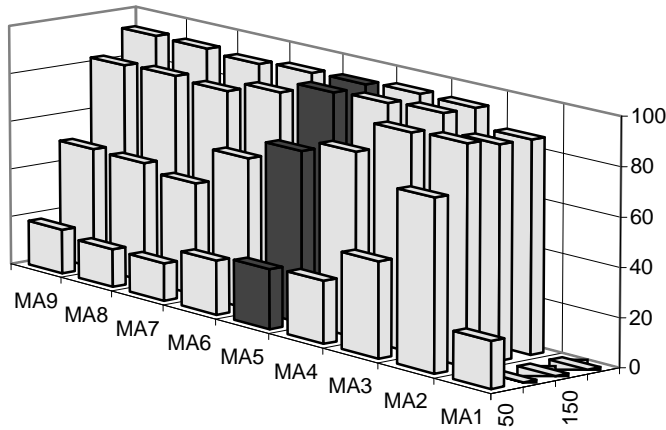
Table 1: Some features of the empirical distribution of the estimated cointegrating vectors of systems (4.1)

T		50	100	150	200
MA1	Mean	-2.9828	-3.009	-3.0006	-2.9998
	S.D.	0.9723	0.1436	0.0755	0.059
	Median	-2.9949	-2.9962	-2.9975	-2.996
	Skew.	-0.4436	-0.3424	-0.2023	-0.2568
	Kurt.	-0.4953	0.7186	0.7903	1.0114
	HD-mean	0.0012	0.0006	0	0
	HD-medi.	0.0004	0.0003	0.0002	0.0003
MA2	Mean	-2.998	-3.0012	-3.0006	-3.0005
	S.D.	0.1861	0.024	0.0152	0.0114
	Median	-3.0012	-3.0004	-3.0003	-3.0007
	Skew.	-0.1419	-0.0483	-0.0662	-0.0625
	Kurt.	0.2717	0.4251	0.6762	0.3711
	HD-mean	0.0001	0.0001	0	0
	HD-medi.	0.0001	0	0	0
MA3	Mean	-3.0031	-3.0009	-3.0007	-3.0002
	S.D.	0.0888	0.0226	0.0135	0.0101
	Median	-3.0004	-3.0004	-2.9999	-3
	Skew.	-0.1662	-0.1096	-0.2304	0.0488
	Kurt.	0.1543	0.4006	0.6506	0.5331
	HD-mean	0.0002	0.0001	0.0001	0
	HD-medi.	0	0	0	0
MA4	Mean	-2.9972	-3	-2.9997	-2.9997
	S.D.	0.2221	0.0267	0.0165	0.0121
	Median	-2.9974	-2.9991	-2.9996	-2.9998
	Skew.	-0.1674	-0.0325	-0.2083	-0.0019
	Kurt.	0.357	0.6263	0.6871	0.5639
	HD-mean	0.0002	0	0	0
	HD-medi.	0.0002	0.0001	0	0
MA5	Mean	-3.0036	-2.9997	-2.9997	-3.0001
	S.D.	0.1024	0.0252	0.0156	0.0111
	Median	-2.9997	-2.9999	-3.0002	-3.0001
	Skew.	-0.2294	-0.0229	0.0518	-0.081
	Kurt.	0.1918	0.6474	0.4198	0.5099
	HD-mean	0.0003	0	0	0
	HD-medi.	0	0	0	0
MA6	Mean	-3.007	-3.001	-2.9995	-3.0001
	S.D.	0.2089	0.0341	0.0172	0.0124
	Median	-2.9988	-2.9993	-2.9995	-3.0001
	Skew.	-0.1599	-0.2015	-0.0451	-0.07
	Kurt.	0.3059	0.4118	0.3297	0.4186
	HD-mean	0.0005	0.0001	0	0
	HD-medi.	0.0001	0.0001	0	0
MA7	Mean	-3.0108	-3.0004	-2.9992	-2.9998
	S.D.	0.4899	0.0285	0.0305	0.0121
	Median	-2.9977	-2.9991	-2.9997	-2.9996
	Skew.	-0.0736	-0.2603	-0.2144	-0.0562
	Kurt.	-0.3618	0.342	0.4373	0.4654
	HD-mean	0.0008	0	0.0001	0
	HD-medi.	0.0002	0.0001	0	0
MA8	Mean	-3.0067	-3.0034	-3.0012	-3.0007
	S.D.	0.4396	0.0769	0.0169	0.0135
	Median	-3.0009	-3.0002	-3.0007	-3.0006
	Skew.	-0.1716	-0.189	-0.2558	-0.0474
	Kurt.	-0.2519	0.5242	0.553	0.4218
	HD-mean	0.0005	0.0002	0.0001	0
	HD-medi.	0.0001	0	0	0
MA9	Mean	-3.0043	-2.9996	-3.0011	-3.0007
	S.D.	0.2944	0.0306	0.0163	0.0105
	Median	-2.999	-3	-3	-3.0005
	Skew.	0.1221	-0.0114	-0.2113	-0.1696
	Kurt.	0.1158	-0.1153	0.4662	0.8067
	HD-mean	0.0003	0	0.0001	0.0001
	HD-medi.	0.0001	0	0	0

are similar in both tests and that for all the systems without a unit root in the MA polynomial (MA2 to MA9) the actual acceptance probabilities of the correct dimension are approaching the nominal value. This happens irrespective of the MA polynomial, only in the case of MA2 the value is only 86 % for 200 observations. In the figures the results for the correctly specified system are drawn in black.

Something completely different happens for the system with the unit root in the MA

Figure 2: Acceptance probability of the correct number of cointegrating vectors for systems (4.1) using the max test



polynomial. Here the probability of choosing a cointegrating space of maximal dimension is going to 1 (see Tables 2 and 3). A simple algebraic example analysed in Appendix C is used to explain why this kind of behaviour occurs. There we show that each element of the trace of the matrix product that forms the basis for the testing procedure is diverging to infinity when the MA polynomial is converging to $b(z) = \text{diag}(1 - z)$. Take a look at equation (2.9),

$$y_t = \beta_{\perp} (\alpha'_{\perp} a_1(1) \beta_{\perp})^{-1} \alpha'_{\perp} b(1) \sum_{t=1}^T \epsilon_t + c_1(z) b(z) \epsilon_t$$

which we have seen to be a common trends representation for regular $b(1)$. The first part of the right hand side of this equation is still composed of the same number of linearly independent random walks as for regular $b(1)$ if the rank of $\alpha'_{\perp} b(1)$ is the same as the rank of α'_{\perp} . For the system at hand $\alpha'_{\perp} = (\frac{1}{2}, -1)$ and

$$b(1) = \begin{bmatrix} -2 & 6 \\ -1 & 3 \end{bmatrix},$$

therefore $\alpha'_{\perp} b(1) = (0, 0)$.

What can also be clearly seen from the tables and figures is the fact that for 50 observations the actual size is very far from the (asymptotic) nominal size. Therefore, and this is a well known fact, using the tables with the asymptotic critical values for just

50 observations is not a good idea.

Since we have also seen before that the quality of the estimated cointegrating vectors in terms of e.g. mean and standard deviation of the estimates is not bad, this implies that the estimation of cointegrating vectors for small systems and few observations is feasible if one uses corresponding tables with critical values for small samples, at least for 2 dimensional systems.

The second set of simulated systems are 3-dimensional *ARMA*(2, 1) systems with a two dimensional cointegrating space.

$$\begin{aligned} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 3 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} &= \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix} = \begin{bmatrix} 0.8 & 0 & 0 \\ 0 & 1.2 & 0 \\ 0 & 0 & 1.5 \end{bmatrix} \begin{bmatrix} u_{1t-1} \\ u_{2t-1} \\ u_{3t-1} \end{bmatrix} + \\ &+ \begin{bmatrix} 0 & 0 & 0 \\ 0 & -0.7 & 0 \\ 0 & 0 & -0.5 \end{bmatrix} \begin{bmatrix} u_{1t-2} \\ u_{2t-2} \\ u_{3t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{bmatrix} + \begin{bmatrix} \gamma_1 & 0 & 0 \\ 0 & \gamma_2 & 0 \\ 0 & 0 & \gamma_3 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-1} \\ \epsilon_{2t-1} \\ \epsilon_{3t-1} \end{bmatrix} \quad (4.2) \end{aligned}$$

The MA polynomials used for simulation are $\gamma_1 = -1, \gamma_2 = \gamma_3 = -0.9$ and the following systems with identical entries γ_s

$$\gamma_1 = \gamma_2 = \gamma_3 = -0.5, 0, 0.6, 0.8.$$

Again the first system, MA1, has a unit root in the MA polynomial.

The true cointegrating space is two-dimensional now, and a basis is given by

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

With a 2-dimensional cointegrating space after normalisation 4 elements are undetermined. We normalise the first element of the first vector to 1 and the second element of the second to 1.

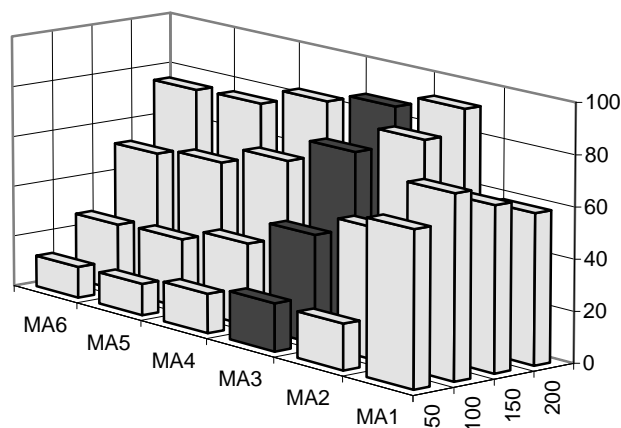
Table 4 in Appendix C describes some features of the empirical distribution of the normed basis of the estimated cointegrating space.

The results are comparable to the results from the simulation of 2-dimensional systems before. The behaviour of the estimates is again similar for the different MA polynomials, but the standard deviations for 50 observations are now much larger than before. Also the Hausdorff distance between the spaces spanned by the mean and median vectors and the true cointegrating space are now much larger for $T = 50$.

The large standard deviations imply that for small sample sizes hypothesis testing on the cointegrating vectors is not useful; at least 100 observations are required for that task.

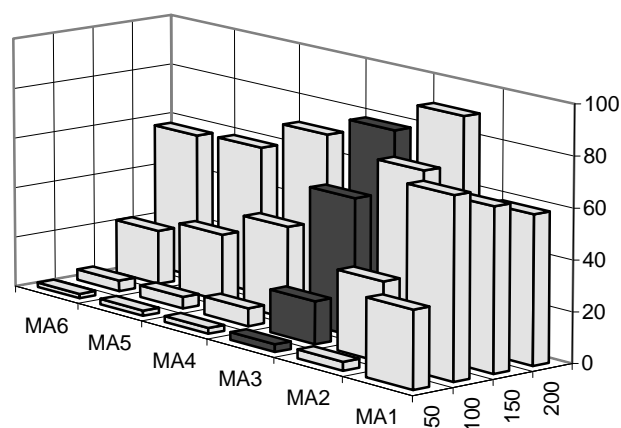
As before the actual probabilities of choosing the correct dimension of the cointegrating space are shown in two figures, again the corresponding tables show the probabilities

Figure 3: Acceptance probability of the correct number of cointegrating vectors for systems (4.2) using the trace test



of all dimensions given in Appendix C. In Figures 3 and 4 we see that the actual size of the test is not as close to the nominal as before, the behaviour also varies more between the different systems. As a rough guideline it seems to be the case that 150 observations are now required for making the use of asymptotical tables a useful exercise. For MA1 (the system with the unit root) the actual size is decreasing from 100 observations onwards, in Tables 5 and 6 one sees that the probability of deciding for a 3-dimensional cointegrating space is increasing with the sample size.¹⁰

Figure 4: Acceptance probability of the correct number of cointegrating vectors for systems (4.2) using the max test



¹⁰For 400 observations and for the trace test the probabilities of choosing a 2- or 3-dimensional space are 50.2 and 49.8 %. For this system, although $b(1)$ is singular, the rank of α'_{\perp} and $\alpha'_{\perp}b(1)$ coincide. This could be a reason for the less rapid divergence of the test statistic compared to the behaviour of the 2-dimensional system with a unit root in the MA polynomial.

In the following figure we try to combine the results from the estimation and the testing step of the simulation. Then we show the Hausdorff distances between the estimated and true cointegrating space. We only show the cases where the trace test indicates the correct number of cointegrating vectors.¹¹

From the front to the back we see the results for the 6 systems, first for 50, then for 100 observations. We only show the interval of Hausdorff distances smaller than 0.05, thereby we exclude those cases with a correct decision about the cointegrating rank but a bad quality of the estimates and all the cases with a wrong decision about the cointegrating rank. The latter would appear as spikes in the “front” of the picture, because the Hausdorff distance between spaces of different dimensions is 1. In other words, the mass, for fixed MA polynomial and sample size, under the curve is equal to the number of cases in which the sequential test procedure based on the trace test indicated the correct dimension of the cointegrating space and the Hausdorff distance of the estimated and true cointegrating space is smaller than 0.05.¹²

The picture shows two things. Only for sample size 50 the (unnormalised) distribution is higher for MA1 than for the others. This stems from the fact that in this case the test statistic diverges for increasing sample size, producing an “upward bias” compared to small sample sizes.

From 100 observations onwards the distribution corresponding to MA1 is very flat and due to the divergence of the test statistic, is having less and less mass. For the other systems we see that the distribution for 100 and 150 observations is higher and the more negative are the coefficients in the MA polynomial.¹³ Finally, for 200 observations the distributions are quite similar, and they are all shifted to 0. The last aspect means that the quality of the approximation is getting better for larger sample sizes, as has to be expected from consistent estimates.

The simulation study indicates that the Johansen procedure is quite robust with respect to MA disturbances for sample sizes that one may have in macro-econometric applications. An important question that remains to be answered is the asymptotic distribution of the estimates and the test statistics in the case of moving average structure in the errors. This question is even more interesting in the light of the results of Yap and Reinsel [21], who have shown that the asymptotic distribution of the test statistic is the same for correctly specified cointegrated *ARMA* models as for *AR* models.

5 Conclusions

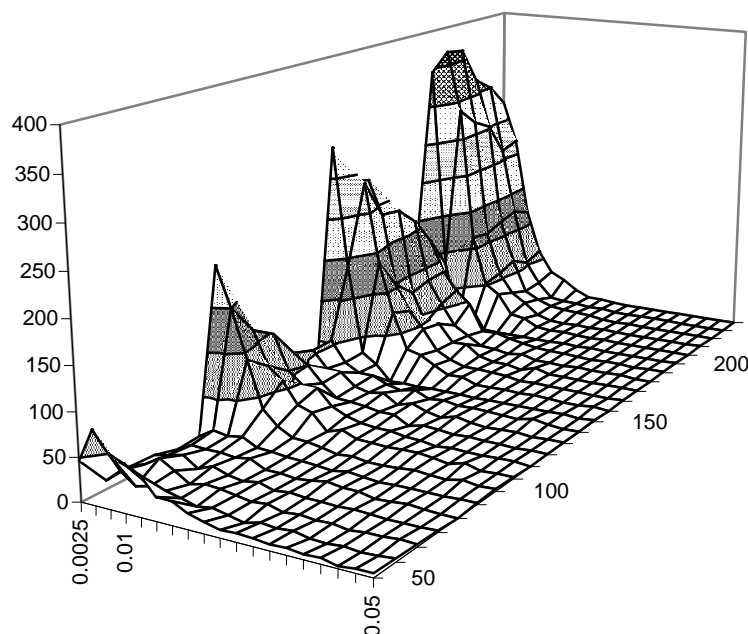
The results of this paper can be summarised as follows. The Johansen procedure for estimation and testing in cointegrated vector autoregressive systems is quite robust with

¹¹The corresponding figure for the max test is given in Appendix C.

¹²Extending the picture to include a larger range of Hausdorff distances just produces an almost flat area in the front of the picture.

¹³This kind of monotony is also seen in Figures 3 and 4.

Figure 5: Hausdorff distances between the true and estimated cointegrating space of systems (4.2) using the trace test



respect to misspecification in the form of un-modelled moving average error dynamics. This robustness has been established theoretically and it has been shown to be of practical relevance by means of a small simulation study carried out with sample sizes that are usually available in macro-econometric applications.

In the case of proving the consistency of the estimates of the co-integrating space under the discussed form of misspecification we have derived some matrix algebraic relationships for $I(1) - ARMA$ systems that may be of some interest in themselves.

Once again it has to be noted that the misspecification leads to biased estimates of the short-run adjustment parameters and the variance-covariance matrix. This is as expected, because we have seen that the MA polynomial leads to a coordinate transformation of the common trends, and the variance matrix estimated by assuming a VAR structure for the likelihood is generally biased.

As long as there are no unit roots in the errors, the procedure appears very robust. The result is especially intended to cover cases where the misspecification is “small”, i.e. there are only minor deviations from white noise behaviour of the errors. This is a situation that especially in small samples may be hard to be distinguished from a pure VAR by model selection procedures, like AIC, BIC or the like. For cases like this the results of the estimation and testing procedure are not affected to a big extent.

The results also show that when one is interested only in the long-run relationships, one can estimate them with a high degree of precision by a VAR model. This has the advantage that one does not have to use VARMA models or non-parametric methods.

The relative performance of several different methods is currently investigated.¹⁴ Still unresolved issues are the asymptotic distributions of the estimators and test statistics under misspecification as in Theorem 3.1. The simulations, and also the super-consistency of $\hat{\beta}$, suggest that, if there is some influence of the MA polynomial at all, the effects seem to be rather small.

Although already known in (part of) the literature it should be stressed once more that the estimation of cointegrating spaces requires some minimum sample sizes to allow for the sensible use of the tabulated critical values and also for a good quality of approximation of the true cointegrating space by the estimated cointegrating space.

¹⁴The author is currently comparing Johansen's method and Bierens' [2] nonparametric cointegration method. The latter uses Chebycheff polynomials to approximate VARMA models. For a very brief comment comparing the two approaches see Deistler and Wagner [3].

References

- [1] BEWLEY R & M. YANG, *On the size and power of system tests for cointegration*, The Review of Economics and Statistics, **80**(1998), 675 – 879.
- [2] BIERENS H.J., *Nonparametric cointegration analysis*, Journal of Econometrics, **77**(1997), 379 – 404.
- [3] DEISTLER M. & M. WAGNER, *Comment on Cointegration Analysis by Bierens H. J.*, (1997), 241 – 243, in CHRISTIAAN HEIJ ET AL. (EDS.), *System dynamics in economic and financial models*, Chichester: Wiley, Series in Financial Economics and Quantitative Analysis.
- [4] ENGLE R.F. & C.W.J. GRANGER, *Co-Integration and Error Correction: Representation, Estimation and Testing*. Econometrica, **55**(1987), 251 – 276.
- [5] GREGOIR S., *Multivariate Time Series With Various Hidden Unit Roots: Part I: Integral Operator Algebra and Representation Theorem*, Mimeo, CREST - I.N.S.E.E.
- [6] GREGOIR S., *Multivariate Time Series With Various Hidden Unit Roots: Part II: Estimation and Test*, Mimeo, CREST - I.N.S.E.E.
- [7] HALDRUP N. & M. SALMON, *Representations of $I(2)$ cointegrated systems using the Smith-McMillan form*, Journal of Econometrics, **84**(1998), 303 – 325.
- [8] HANNAN E.J. & M. DEISTLER, *The Statistical Theory of Linear Systems*, New York: Wiley, (1988).
- [9] HARGREAVES C., *A Review of Methods of Estimating Cointegrating Relationships*, (1994), 87 – 131, in HARGREAVES C. (ED.), *Nonstationary Time Series Analysis and Cointegration*, Oxford: Oxford University Press.
- [10] JOHANSEN S., *The Mathematical Structure of Error Correction Models*, Contemporary Mathematics, **80**(1988), 359 – 386.
- [11] JOHANSEN S., *Statistical Analysis of Cointegration Vectors*, Journal of Economic Dynamics and Control, **12**(1988), 231 – 254.
- [12] JOHANSEN S., *Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models*, Econometrica, **59**(1991), 1551 – 1580.
- [13] JOHANSEN S., *Likelihood-Based Inference on Cointegration in the Vector Autoregressive Model*, Oxford: Oxford University Press, (1995).
- [14] OSTERWALD-LENUM M., *A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics*, Oxford Bulletin of Economics and Statistics, **54**(1992), 461 – 472.

-
- [15] PHILLIPS P.C.B. & V. SOLO, *Asymptotics for Linear Processes*, The Annals of Statistics, **20**(1992), 971 – 1001.
- [16] PODIVINSKY J.M., *Testing misspecified cointegrating relationships*, Economics Letters, **60**(1998), 1 – 9.
- [17] SAID S.E. & D.A. DICKEY, *Testing for Unit Roots in Autoregressive- Moving Average Models of Unknown Order*, Biometrika, **71**(1984), 599 – 607.
- [18] SAIKKONEN P., *Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation*, Econometric Theory, **8**(1992), 1 – 27.
- [19] TODA H.Y., *Finite Sample Performance of Likelihood Ratio Tests for Cointegrating Ranks in Vector Autoregressions*, Econometric Theory, **11**(1995), 1015 – 1032.
- [20] WAGNER M., *Robustness Analysis of the Johansen cointegration estimation and testing procedure (in German)*, Vienna: Master Thesis at the Vienna University of Technology, (1995).
- [21] YAP S.F. & REINSEL G.C., *Estimating and Testing for Unit Roots in a Partially Nonstationary Vector Autoregressive Moving Average Model*, Journal of the American Statistical Association, **90**(1995), 253 – 267.

A Appendix to Section 2

The following Theorems A.1 to A.5 are derived by Johansen [10].

Theorem A.1 *The multiplicity r of the zero $z = 1$ of $\det(C(z))$ is larger than or equal to n . Thus there exists a function $f(z) \neq 0$ so that*

$$\det(C(z)) = (1 - z)^r f(z), \quad r \geq n$$

holds.

Theorem A.2 *For the coefficients of the adjoint matrix function*

$$\bar{C}(z) = \sum_{j=0}^{\infty} (1 - z)^j \bar{C}_j$$

$$\bar{C}_j = 0 \quad j = 0, \dots, n - k - 1 \quad (\text{A.1})$$

$$\text{therefore: } \bar{C}(z) = (1 - z)^{n-k} \bar{C}_{n-k}(z) \quad (\text{A.2})$$

$$\bar{C}_{n-j} C_i = 0 \quad 0 \leq i < j \leq k \quad (\text{A.3})$$

$$\bar{C}_{n-j} C(z) = (1 - z)^j \bar{C}_{n-j} C_j(z) \quad j = 1, \dots, k \quad (\text{A.4})$$

$$\text{and } \bar{C}(z) C_i = (1 - z)^{n-i} \bar{C}_{n-i}(z) C_i \quad i = 0, \dots, k - 1 \quad (\text{A.5})$$

holds.

Relation (A.3) shows that the rows of \bar{C}_{n-j} are contained in the left null spaces of the matrices C_i , for $i < j$. This means that the space spanned by the rows of $\bar{C}_{n-k}, \dots, \bar{C}_{n-i-1}$ is contained in $M_i = N_0 \cap \dots \cap N_i$.

A sufficient condition for these two spaces to coincide is formulated in the next theorem.

Theorem A.3 *If $r = n$, then:*

1. $\sum_{j=0}^k \bar{C}_{n-j} C_j$ is proportional to the identity matrix.
2. $\text{rg}(\bar{C}_{n-j} C_i) = m_{j-1} - m_j = \dim V_j \quad j = 0, \dots, k$.
3. The rows of $\bar{C}_{n-k}, \dots, \bar{C}_{n-i-1}$ span M_i .

The following theorem shows when the process $\Delta^d y_t = C(z) \epsilon(t)$ has an error correction representation.

Theorem A.4 *The process y_t , satisfies an error correction model of the form*

$$\bar{C}_{n-k} y_t + \dots + \bar{C}_{n-1} \Delta^{k-1} y_t + \bar{C}_n(z) \Delta^k y_t = f(z) \epsilon_t \quad t = 0, 1, \dots \quad (\text{A.6})$$

if:

1. $d = k$
2. $C(z)$ is balanced, i.e. $r = n$.

Where the indices k, r, n are defined from $C(z)$.

If the assumptions of the preceding theorem are not fulfilled y_t satisfies a *GE*CM of order d , which may for $r > n$ contain a non-invertible stationary process on the right hand side and integral correction terms on the left hand side for $d < k$.

For the process given by its autoregressive representation $A(z)y_t = \epsilon_t$ the corresponding theorem is

Theorem A.5 *The equation $A(z)y_t = \epsilon_t$, satisfies an error correction model of order k'*

$$A_0 y_t + \cdots + A_{k'-1} \Delta^{k'-1} y_t + A_{k'}(z) \Delta^{k'} y_t = \epsilon_t. \quad (\text{A.7})$$

if $A'(z)$ is balanced, i.e. $r' = n'$.

In the above theorem $A'(z)$ denotes the transposed matrix function of $A(z)$ and k' , r' and n' its indices.

As with error correction models in *MA* representation, for the *AR* representation one sees that the cointegrating vectors of order i are contained in the space formed by the rows of the A_i .

Proof of Lemma 2.1:

We can represent the determinant of $c(z)$ from the $MA(\infty)$ -representation of Δy_t by $\det(c(z)) = (1-z)^r f(z)$

So we can write $c(z)$, where $\Lambda(z)$ shall be given in Smith form, as

$$c(z) = (1-z)v^{-1}(z) \begin{pmatrix} \lambda_1^{-1}(z) & & & & & \\ & \ddots & & & & \\ & & \lambda_{m-r}^{-1}(z) & & & 0 \\ & & & \frac{\tilde{\lambda}_{m-r+1}^{-1}(z)}{(1-z)} & & \\ 0 & & & & \ddots & \\ & & & & & \frac{\tilde{\lambda}_m^{-1}(z)}{(1-z)} \end{pmatrix} u^{-1}(z)$$

therefore

$$c(z) = v^{-1}(z) \begin{pmatrix} \lambda_1^{-1}(z)(1-z) & & & & & \\ & \ddots & & & & \\ & & \lambda_{m-r}^{-1}(z)(1-z) & & & 0 \\ & & & \tilde{\lambda}_{m-r+1}^{-1}(z) & & \\ 0 & & & & \ddots & \\ & & & & & \tilde{\lambda}_m^{-1}(z) \end{pmatrix} u^{-1}(z)$$

Now inserting $z = 1$ results in

$$c(1) = v^{-1}(1) \begin{pmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & 0 \\ & & & \tilde{\lambda}_{m-r+1}^{-1}(1) & & \\ 0 & & & & \ddots & \\ & & & & & \tilde{\lambda}_m^{-1}(1) \end{pmatrix} u^{-1}(1)$$

This means that the rank of $c(1)$ is $m-r$, because $\tilde{\lambda}_i(1) \neq 0$ has to hold for $i = m-r+1, \dots, m$. This directly implies that the dimension of the space M_0 denoted by $m_0 = r$.

From theorem 2.1 we know $r \geq n = \sum_{i=0}^{\infty} m_i$.

Combining this and $m_0 = r$ we obtain $r = n = m_0$.

From this one also sees that all m_i for $i \geq 1$ have to equal 0, which shows $k = 1$.

qed

B Appendix to Section 3

This appendix is devoted to proving Theorem 3.1. The structure is thereby taken to parallel the proof sequence in Johansen [11], with the advantage that one clearly sees where the misspecification exerts influence on the results.

We write y_t as the sum of lagged differences

$$y_t = \sum_{j=1}^t \Delta y_j + y_0 \quad t = 1, 2, \dots$$

and, to simplify notation, we will assume $y_0 = 0$.

We denote, following the notation in Johansen [10], by

$$\psi(i) := E \Delta y_t \Delta y_{t+i}'$$

and

$$\begin{aligned} \mu_{ij} &= \psi(i-j) & i, j = 0, 1, \dots, p-1 \\ \mu_{pi} &= \sum_{j=k-i}^{\infty} \psi(j) & i = 0, 1, \dots, p-1 \\ \mu_{pp} &= - \sum_{j=-\infty}^{\infty} |j| \psi(j) \\ \Psi &= \sum_{j=-\infty}^{\infty} \psi(j) \end{aligned}$$

From $\Delta y_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$ one immediately obtains

$$\psi(i) = \sum_{j=0}^{\infty} c_j \Sigma c_{j+i}'$$

and using the notation $C = c(1) = \sum_{j=0}^{\infty} c_j$ yields

$$\Psi = \sum_{j=0}^{\infty} c_j \Sigma \sum_{j=0}^{\infty} c_j' = C \Sigma C'$$

Also

$$\text{var}(y_{t-p}) = \sum_{j=-t+p}^{t-p} (t-p-|j|) \psi(j) \quad (B.1)$$

and

$$\text{cov}(y_{t-p}, \Delta y_{t-i}) = \sum_{j=p-i}^{t-i-1} \psi(j). \quad (B.2)$$

Asymptotically (3.3) and (3.4) have limits

$$\text{var}(T^{-1/2}y_t) \xrightarrow{T \rightarrow \infty} \sum_{i=-\infty}^{\infty} \psi(i) = \Psi \quad (B.3)$$

$$\text{cov}(y_{T-p}, \Delta y_{T-i}) \xrightarrow{T \rightarrow \infty} \sum_{p-i}^{\infty} \psi(i) = \mu_{pi} \quad i = 0, 1, \dots, p-1 \quad (B.4)$$

For the variance of $\beta' y_{T-p}$ one gets from

$$\text{var}(\beta' y_{T-p}) = (T-p) \sum_{j=-T+p}^{T-p} \beta' \psi(j) \beta - \sum_{j=-T+p}^{T-p} |j| \beta' \psi(j) \beta$$

for the limit

$$\lim_{T \rightarrow \infty} \text{var}(\beta' y_{T-p}) = \beta' \mu_{pp} \beta = \mu_{\beta\beta}, \quad (B.5)$$

because $\beta' C = 0$ implies that $\beta' \Psi = 0$, for that reason the first term vanishes asymptotically.

All the above relations hold for general $y_t \in I(1)$.

We, furthermore, define the following sample moment matrices:

$$M_{ij} := \frac{1}{T} \sum_{t=1}^T \Delta y_{t-i} \Delta y'_{t-j} \quad i, j = 0, 1, \dots, p-1$$

$$M_{pi} := \frac{1}{T} \sum_{t=1}^T y_{t-p} \Delta y'_{t-i} \quad i = 0, 1, \dots, p-1$$

$$M_{pp} := \frac{1}{T} \sum_{t=1}^T y_{t-p} y'_{t-p}$$

The asymptotic behaviour of these is given by

Lemma B.2 For $T \rightarrow \infty$ the following holds

$$\frac{1}{\sqrt{T}} y_{[Tt]} \xrightarrow{w} CW(t) \quad (B.6)$$

$$M_{ij} \xrightarrow{a.s.} \mu_{ij} \quad i, j = 0, 1, \dots, p-1 \quad (B.7)$$

$$M_{pi} \xrightarrow{p} C \int_0^1 W dW' C' + \mu_{pi} \quad i = 0, 1, \dots, p-1 \quad (B.8)$$

$$\beta' M_{pp} \beta \xrightarrow{a.s.} \mu_{\beta\beta} \quad (B.9)$$

$$\frac{1}{T} M_{pp} \xrightarrow{w} C \int_0^1 W(u) W'(u) du C' \quad (B.10)$$

where W is an m -dimensional Brownian motion with covariance matrix Σ .

For the matrices S_{ij} defined in the preceding description of the Johansen procedure the following relationships hold

$$S_{ij} = M_{ij} - M_{i*} M_{**}^{-1} M_{*j} \quad i, j = 0, p$$

where

$$\begin{aligned} M_{0*} &:= (M_{01}, \dots, M_{0p-1}) \\ M_{p*} &:= (M_{p1}, \dots, M_{pp-1}) \\ M_{**} &:= \begin{pmatrix} M_{11} & \dots & M_{1p-1} \\ \vdots & & \vdots \\ M_{p-11} & \dots & M_{p-1p-1} \end{pmatrix} \end{aligned}$$

and $M_{*j} = (M_{j*})'$.

In exactly the same way we introduce for the matrices μ_{ij}

$$\Sigma_{ij} := \mu_{ij} - \mu_{i*} \mu_{**}^{-1} \mu_{*j}$$

Using this notation one obtains the following lemma, which already shows effects of the misspecification

Lemma B.3 *The following identities hold*

$$\Sigma_{00} = \Gamma_p \Sigma_{p0} + \Sigma + \xi_0 \quad (B.11)$$

$$\Sigma_{0p} \Gamma_p' = \alpha \Sigma_{\beta\beta} \alpha' + \xi_p \quad (B.12)$$

and using $\Gamma_p = -\alpha \beta'$,

$$\Sigma_{00} = \alpha \Sigma_{\beta\beta} \alpha' + \Sigma + \xi_0 + \xi_p \quad (B.13)$$

where ξ_0 and ξ_p depend on $b(z)$ and are defined in the proof.

Proof:

The stated relations are reduced to the relations given in *Lemma 2* in Johansen in the case that the true model is *AR*. The derivation there exploits the *AR* structure, therefore here one has to expect changes. From the defining equation for y_t

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + \Gamma_p y_{t-p} + \epsilon_t + b_1 \epsilon_{t-1} + \dots + b_l \epsilon_{t-l}$$

one obtains by multiplication with $\Delta y'_{t-i}$ for $i = 0, 1, \dots, p-1$. division by $\frac{1}{T}$ and summation over $t = 1, \dots, T$ the following equations

$$M_{0i} = \Gamma_1 M_{1i} + \dots + \Gamma_{p-1} M_{p-1i} + \Gamma_p M_{pi} + \frac{1}{T} \sum_{t=1}^T b(z) \epsilon_t \Delta y'_{t-i} \quad (B.14)$$

for $i = 0, 1, \dots, p-1$

and by multiplication with y'_{t-p}

$$M_{0p} = \Gamma_1 M_{1p} + \dots + \Gamma_{p-1} M_{p-1p} + \Gamma_p M_{pp} + \frac{1}{T} \sum_{t=1}^T b(z) \epsilon_t y'_{t-p} \quad (B.15)$$

If one now solves equations (B.14) for $i = 1, \dots, p-1$ for $(\Gamma_1, \dots, \Gamma_{p-1})$ one gets

$$(\Gamma_1, \dots, \Gamma_{p-1}) = M_{0*} M_{**}^{-1} - \Gamma_p M_{p*} M_{**}^{-1} - Z_1 M_{**}^{-1} \quad (B.16)$$

with $Z_1 = \left(\frac{1}{T} \sum_{t=1}^T b(z) \epsilon_t \Delta y'_{t-1}, \dots, \frac{1}{T} \sum_{t=1}^T b(z) \epsilon_t \Delta y'_{t-p+1} \right)$.

Inserting this solution in (B.1) for $i = 0$ yields

$$S_{00} = \Gamma_p S_{p0} + \frac{1}{T} \sum_{t=1}^T b(z) \epsilon_t \Delta y'_t - Z_1 M_{**}^{-1} M_{*0} \quad (B.17)$$

By inserting in (B.15) und multiplication from the right with Γ_p' one gets

$$S_{0p} \Gamma_p' = \Gamma_p S_{pp} \Gamma_p' + \frac{1}{T} \sum_{t=1}^T b(z) \epsilon_t y'_{t-p} \Gamma_p' - Z_1 M_{**}^{-1} M_{*p} \Gamma_p' \quad (B.18)$$

Let

$$\tau_1 := \lim_{T \rightarrow \infty} Z_1$$

i.e.

$$\tau_1 = \left(\sum_{n=1}^q b_n \Sigma c'_{n-1}, \sum_{n=2}^q b_n \Sigma c'_{n-2}, \dots \right);$$

and

$$\tau_2 := \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T b(z) \epsilon_t (\Gamma_p y_{t-p})'$$

For $q < p$ $\tau_2 = 0$ holds and for $q \geq p$ the result will be derived now by

$$\begin{aligned} \Delta y_{t-p} &= c(z) \epsilon_{t-p} \\ &= [C + \Delta c^*(z)] \epsilon_{t-p} \\ \Gamma_p \Delta y_{t-p} &= [\Gamma_p \Delta c^*(z)] \epsilon_{t-p} \end{aligned}$$

Since $\Gamma_p = \alpha \beta'$, we have $\Gamma_p C = 0$ and after cancellation of Δ

$$\Gamma_p y_{t-p} = \Gamma_p c^*(z) \epsilon_{t-p}$$

remains. By straightforward calculation one sees that

$$c^*(z) = \sum_{j=0}^{\infty} \left(- \sum_{i=j+1}^{\infty} c_i \right) z^j.$$

Using this representation for $c^*(z)$ for τ_2 in the case $q \geq p$ one derives

$$\tau_2 = \sum_{m=p}^q b_m \Sigma d'_{m-p} \Gamma'_k$$

with $d_{m-p} = - \sum_{i=j+1}^{\infty} c_i$. Here b_m denote of course the coefficients from the MA polynomial. The limits for $T \rightarrow \infty$ of eqs. (B.17) and (B.18). using $b_0 = c_0 = I_m$, are given by:

$$\Sigma_{00} = \Gamma_p \Sigma_{p0} + \Lambda + \sum_{n=1}^q b_n \Sigma c'_n - \tau_1 \mu_{**}^{-1} \mu_{*0} \quad (B.19)$$

and

$$\Sigma_{0p} \Gamma'_p = \alpha \Sigma_{\beta\beta} \alpha' + \tau_2 - \tau_1 \mu_{**}^{-1} \mu_{*0} \quad (B.20)$$

Now using the following notation

$$\begin{aligned} \xi_0 &= \sum_{n=1}^q b_n \Sigma c'_n - \tau_1 \mu_{**}^{-1} \mu_{*0} \\ \xi_p &= \tau_2 - \tau_1 \mu_{**}^{-1} \mu_{*0} \end{aligned}$$

one obtains

$$\Sigma_{00} = \Gamma_p \Sigma_{p0} + \Sigma + \xi_0 \quad (B.21)$$

$$\Sigma_{0p} \Gamma'_p = \alpha \Sigma_{\beta\beta} \alpha' + \xi_p \quad (B.22)$$

Combining the above two equations one also obtains

$$\Sigma_{00} = \alpha \Sigma_{\beta\beta} \alpha' + \Sigma + \xi_0 + \xi'_p \quad (B.23)$$

If the true model coincides with the estimated model, one additionally gets the following identity $\alpha = -\Sigma_{0p}\beta\Sigma_{\beta\beta}^{-1}$.

For this relation we find no analogon under *ARMA* misspecification, we obtain

$$\alpha = \Sigma_{0p}\beta\Sigma_{\beta\beta}^{-1} - \xi_p\alpha(\alpha'\alpha)^{-1}\Sigma_{\beta\beta}^{-1} \quad (B.24)$$

where α is only defined implicitly.

qed

The next lemma discusses the asymptotic behaviour of the matrices S_{ij} , only the second relationship changes under the presence of an unmodeled MA part.

Lemma B.4 For $T \rightarrow \infty$ and δ chosen so that $\delta'\alpha = 0$

$$S_{00} \xrightarrow{a.s.} \Sigma_{00} \quad (B.25)$$

$$\delta' S_{0p} \xrightarrow{w} \delta' b(1) \int_0^1 dW W' C' + \delta' \left(\sum_{s=p}^q b_s \Sigma \left(\sum_{h=0}^{s-p} c'_h \right) - \tau_1 \mu_{**}^{-1} \mu_{*p} \right) \quad (B.26)$$

$$\beta' S_{p0} \xrightarrow{a.s.} \beta' \Sigma_{k0} \quad (B.27)$$

$$\frac{1}{T} S_{pp} \xrightarrow{w} C \int_0^1 W(u) W'(u) du C' \quad (B.28)$$

$$\beta' S_{pp} \beta \xrightarrow{a.s.} \beta' \Sigma_{pp} \beta \quad (B.29)$$

hold. τ_1 is the limit of $Z_1 = \left(\frac{1}{T} \sum_{t=1}^T b(z) \epsilon_t \Delta y'_{t-1}, \dots, \frac{1}{T} \sum_{t=1}^T b(z) \epsilon_t \Delta y'_{t-p+1} \right)$

Proof: Only relationship (B.26) has to be shown, because it is the only one that has changed. Use the solution for $(\Gamma_1 \dots \Gamma_{p-1})$ given in eq. (B.17) and insert in eq. (B.16) to get

$$M_{0p} = M_{0*} M_{**}^{-1} M_{*p} - \Gamma_p M_{p*} M_{**}^{-1} M_{*p} - Z_1 M_{**}^{-1} M_{*p} + \Gamma_p M_{pp} + \frac{1}{T} \sum_{t=1}^T b(z) \epsilon_t y'_{t-k}.$$

which we rewrite as

$$S_{0p} = \Gamma_p S_{pp} - Z_1 M_{**}^{-1} M_{*p} + \frac{1}{T} \sum_{t=1}^T b(z) \epsilon_t y'_{t-k}. \quad (B.30)$$

Multiply with δ' to get

$$\delta' S_{0p} = \delta' \frac{1}{T} \sum_{t=1}^T b(z) \epsilon_t y'_{t-p} - \delta' Z_1 M_{**}^{-1} M_{*p}$$

The difference for the asymptotic behaviour between the *AR* case where $b(z) = I$ and the *ARMA* misspecified case is that in the first term on the right hand side for $q \geq p$ correlated random variables appear. To derive the limit of this expression we use the results obtained in Phillips and Solo [15] and find

$$\delta' S_{0p} \xrightarrow{w} \delta' b(1) \int_0^1 dW W' C' + \delta' \left(\sum_{s=p}^q b_s \Sigma \left(\sum_{h=0}^{s-p} c'_h \right) - \tau_1 \mu_{**}^{-1} \mu_{*p} \right).$$

qed

We see that as soon as the MA order is ≥ 1 τ_1 is not equal to 0.

Regardless of the nature of the MA part the ordered eigenvalues of

$$|\lambda S_{pp} - S_{p0} S_{00}^{-1} S_{0p}| = 0 \quad (B.31)$$

still converge in probability to $(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$,

where $\lambda_1, \dots, \lambda_r$ are the ordered eigenvalues of the equation

$$|\lambda \Sigma_{\beta\beta} - \beta' \Sigma_{p0} \Sigma_{00}^{-1} \Sigma_{0p} \beta| = 0 \quad (B.32)$$

This is true because the matrices S_{ij} and Σ_{ij} are defined "model-free", as they are moment matrices resp. limits of moments matrices. This implies that there are still r eigenvalues positive in the limit and $m - r$ are converging to zero. The above means that also under misspecification $T\hat{\lambda}_i$ diverges for $1 \leq i \leq r$.

As in Johansen [11] the eigenvectors $\hat{\beta}_i$, $i = 1, \dots, r$ are decomposed into two orthogonal components.

Let $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_r)$ denote the matrix of the eigenvectors. Then using

$$\begin{aligned} \hat{x}_i &:= (\beta' \beta)^{-1} \beta' \hat{\beta}_i \\ \hat{y}_i &:= (\gamma' \gamma)^{-1} \gamma' \hat{\beta}_i \end{aligned}$$

where γ is the matrix spanning the orthogonal complement of the space spanned by β , the vectors $\hat{\beta}_i$ are decomposed as:

$$\hat{\beta}_i = \beta \hat{x}_i + \gamma \hat{y}_i$$

With $\hat{x} := (\hat{x}_1, \dots, \hat{x}_r)$ we have $\hat{x} = (\beta' \beta)^{-1} \beta' \hat{\beta}$, so $\beta \hat{x}$ is the projection of $\hat{\beta}$ on the column space of β . For the vectors \hat{y}_i the matrix \hat{y} is defined accordingly. Finally $S(\lambda) := \lambda S_{pp} - S_{p0} S_{00}^{-1} S_{0p}$, then

Lemma B.5 For $i = 1, 2, \dots, r$

$$\gamma' S(\hat{\lambda}_i) \gamma / T \xrightarrow{w} \lambda_i \gamma' C \int_0^1 W W' du C' \gamma \quad (B.33)$$

$$\hat{y}_i \in O_p\left(\frac{1}{T}\right), \quad \hat{x} \in O_p(1), \quad \hat{x}^{-1} \in O_p(1) \quad (B.34)$$

$$\beta' S(\hat{\lambda}_i) \beta \hat{x}_i \in O_p\left(\frac{1}{T}\right) \quad (B.35)$$

$$\begin{aligned} \gamma' S(\hat{\lambda}_i) \beta \hat{x}_i &= -\gamma' \left[\frac{1}{T} \sum_{t=1}^T y_{t-p}(b(z) \epsilon_t)' \right] \Sigma_{00}^{-1} \Sigma_{0p} \beta \hat{x}_i + Z_1 M_{**}^{-1} M_{*p} S_{00}^{-1} S_{0p} \beta \hat{x}_i - \\ &\quad - \gamma' S_{pp} \beta (\beta' S_{pp} \beta)^{-1} (\alpha' \alpha)^{-1} \alpha' \xi_p' S_{00}^{-1} S_{0p} \beta \hat{x}_i + o_p(1) \end{aligned} \quad (B.36)$$

hold

Proof: Up to the last equation all the relationships are invariant under $ARMA$ misspecification. This directly follows from the essentially unchanged behaviour of the moment matrices M_{ij} and S_{ij} and the fact that the cointegrating space is invariant under the assumed class of misspecification. The last relationship follows from straightforward but tedious calculations.

qed

The essential point of the last lemma is the fact that relation (B.34) holds regardless of $b(z)$, this implies

$$\hat{\beta}\hat{x}^{-1} - \beta = \gamma\hat{y}\hat{x}^{-1} \in O_p\left(\frac{1}{T}\right)$$

which shows that $\hat{\beta}$ is still a consistent estimate of β .

Using relations (B.13) and (B.24) the asymptotic behaviour of $\hat{\Pi}$ and $\hat{\Sigma}$ is found to be

$$\hat{\Pi} \xrightarrow{P} \Pi + \xi_p \alpha (\alpha' \alpha)^{-1} \Sigma_{\beta\beta}^{-1} \beta' \quad (B.37)$$

and

$$\begin{aligned} \hat{\Sigma} \xrightarrow{P} \Sigma - \xi_p \alpha (\alpha' \alpha)^{-1} \Sigma_{\beta\beta}^{-1} \beta' \Sigma_{p0} - \Sigma_{0p} \beta \Sigma_{\beta\beta}^{-1} (\alpha' \alpha)^{-1} \alpha' \xi_p' - \\ - \xi_p \alpha (\alpha' \alpha)^{-1} \Sigma_{\beta\beta}^{-1} (\alpha' \alpha)^{-1} \alpha' \xi_p' + \xi_0 + \xi_p' \end{aligned} \quad (B.38)$$

This completes the proof of Theorem 3.1.

C Appendix to Section 4

We analyse the behaviour of the test statistic in the conceivably simplest case of misspecification. We start with the following model

$$\Delta y_t = \Pi y_{t-1} + \epsilon_t + \Theta \epsilon_{t-1}$$

where we furthermore assume that $Var(\epsilon) = I$ and the matrix Θ is a diagonal matrix with identical entries θ . We want to test the hypothesis that $\Pi = 0$, i.e. the case that y_t is a vector random walk of dimension $dim(y_t) = n$. The test statistic for this hypothesis is given by

$$tr\{S_{00}^{-1} S_{01} S_{11}^{-1} S_{10}\}$$

using the notation of the previous appendix and $p = 1$.

The limits of the above matrices are given by

$$\begin{aligned} S_{00} &\xrightarrow{P} I + \Theta\Theta \\ S_{01} &\xrightarrow{w} (I + \Theta) \int_0^1 dW W' (I + \Theta) - \Theta \\ \frac{1}{T} S_{11} &\xrightarrow{w} (I + \Theta) \int_0^1 W W' du (I + \Theta) \end{aligned}$$

The asymptotic test statistic is therefore given by

$$\begin{aligned} T_0 = &tr\{(I + \Theta\Theta)^{-1} \left((I + \Theta) \int_0^1 dW W' (I + \Theta) - \Theta \right) * \\ &\left((I + \Theta) \int_0^1 W W' du (I + \Theta) \right)^{-1} \left((I + \Theta) \int_0^1 W dW' (I + \Theta) - \Theta \right)\} \end{aligned}$$

Multiplying yields

$$\begin{aligned} T_0 = &tr \left((I + \Theta\Theta)^{-1} (I + \Theta)^2 \left(\int_0^1 dW W' \left[\int_0^1 W W' du \right]^{-1} \int_0^1 W dW' \right) \right) + \\ &+ tr \left((I + \Theta\Theta)^{-1} \Theta^2 (I + \Theta)^{-2} \left[\int_0^1 W W' du \right]^{-1} \right) - \\ &- 2 * tr \left((I + \Theta\Theta)^{-1} \Theta \left(\int_0^1 W W' du \right)^{-1} \int_0^1 W dW' \right) \end{aligned}$$

Table 2: Acceptance probabilities of the different dimensions of the cointegrating space for systems (4.1) using the trace test

$dim(\beta)$	0				1				2			
T	50	100	150	200	50	100	150	200	50	100	150	200
MA1	0.3	0	0	0	21.6	0.1	0	0	78.1	99.9	100	100
MA2	10.4	0	0	0	81.3	89.1	86.5	86.2	8.3	10.9	13.5	13.8
MA3	38.6	1.7	0	0	54.6	92.1	93.7	93.7	6.8	6.2	6.3	6.3
MA4	53.4	10.1	0.4	0	41.1	83.5	93.7	94.2	5.5	6.4	5.9	5.8
MA5	54.9	13.1	0.9	0	39.3	81.8	93.1	92.9	5.8	5.1	6.0	7.1
MA6	57.3	20.1	1.8	0	36.7	73.4	90.0	92.7	6.0	6.5	8.2	7.3
MA7	65.6	29.7	4.3	0.2	28.6	64.1	89.5	92.2	5.8	6.2	6.2	7.6
MA8	64.2	27.3	4.1	0.4	28.6	66.5	88.7	93.6	7.2	6.2	7.2	6.0
MA9	61.2	25.1	4.7	0.6	33.5	67.5	88.4	93.8	5.3	7.4	6.9	5.6

Table 3: Acceptance probabilities of the different dimensions of the cointegrating space for systems (4.1) using the max test

$dim(\beta)$	0				1				2			
T	50	100	150	200	50	100	150	200	50	100	150	200
MA1	10.2	0	0	0	19.0	0.1	0	0	70.8	99.9	100	100
MA2	23.1	0	0	0	69.9	89.1	86.5	86.2	7.0	10.9	13.5	13.8
MA3	58.1	5.9	0	0	39.0	88.2	93.7	93.7	2.9	5.9	6.3	6.3
MA4	72.1	19.4	1.5	0	25.5	75.2	92.6	94.2	2.4	5.4	5.9	5.8
MA5	73.1	25.8	2.4	0	24.5	70.2	91.8	92.9	2.4	4.0	5.8	7.1
MA6	75.7	33.0	5.3	0.1	22.3	61.9	86.7	92.6	2.0	5.1	8.0	7.3
MA7	82.7	49.9	12.4	1.2	15.5	46.2	82.2	91.3	1.8	3.9	5.4	7.5
MA8	82.0	47.3	10.1	1.3	15.7	49.3	83.4	92.7	2.3	3.4	6.5	6.0
MA9	80.1	46.3	10.6	1.4	18.3	49.8	82.8	93.1	1.6	3.9	6.6	5.5

In the case of an AR(1) model, where $\Theta = 0$, the expressions in the second and third line above vanish and the expression in the first line is reduced to the known asymptotic test distribution. If $\Theta \rightarrow -I$ one can easily show that the expression in the first line above goes to 0 while the expression in the second line explodes and the limit of the expression in the third line is $-tr \left([\int_0^1 WW' du]^{-1} \int_0^1 W dW' \right)$. So for $\Theta \rightarrow -I$ we have $T_0 \rightarrow \infty$. Applying the trace or max test in that situation always yields a decision for a cointegrating space of maximal dimension, regardless of the true dimension of the cointegrating space.

In the case of the true Π being 0 the limiting system is

$$\Delta y_t = \Delta \epsilon_t,$$

which is naturally not left co-prime and observationally equivalent to $y_t = \epsilon_t$. This shows once more that left co-primeness is an essential condition for all the results.

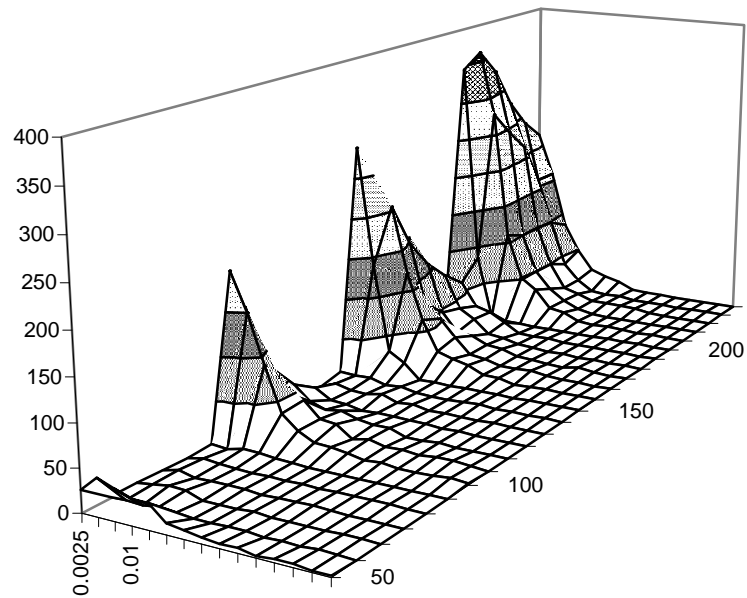
Table 5: Acceptance probabilities of the different dimensions of the cointegrating space for systems (4.2) using the trace test

$dim(\beta)$	0				1			
T	50	100	150	200	50	100	150	200
MA1	0	0	0	0	23.4	0	0	0
MA2	3.9	0	0	0	74.0	42.5	10.4	1.4
MA3	10.8	0.1	0	0	66.5	52.4	21.6	4.8
MA4	13.9	0.6	0	0	65.7	61.2	32.7	11.5
MA5	17.8	1.5	0.1	0	65.2	66.3	39.7	16.4
MA6	18.0	1.6	0	0	66.0	67.5	41.6	20.2
$dim(\beta)$	2				3			
T	50	100	150	200	50	100	150	200
MA1	60.8	71.6	64.3	58.5	15.8	28.4	35.7	41.5
MA2	18	51	82.9	92.2	4.1	6.5	6.7	6.4
MA3	18.8	42.4	71.9	86.9	3.9	5.1	6.5	8.3
MA4	15.5	32.3	62.1	82.9	4.9	5.9	5.2	5.6
MA5	12.5	27.3	55.1	75.9	4.5	4.9	5.1	7.7
MA6	12.3	26.5	52.3	75.4	3.7	4.4	6.1	4.4

Table 6: Acceptance probabilities of the different dimensions of the cointegrating space for systems (4.2) using the max test

$dim(\beta)$	0				1			
T	50	100	150	200	50	100	150	200
MA1	35.2	0	0	0	25.9	0.3	0	0
MA2	72.2	7.0	0	0	23.5	58.3	22.9	3.8
MA3	82.2	41.8	3.0	0.1	14.7	39.6	38.5	14.6
MA4	84.8	67.1	17.6	0.5	12.1	25.1	43.4	24.5
MA5	86.8	76.2	32.4	3.9	10.9	18.8	38.7	32.1
MA6	87.6	77.1	34.8	5.8	10.3	18.2	41.8	33.6
$dim(\beta)$	2				3			
T	50	100	150	200	50	100	150	200
MA1	30.7	71.4	64.3	58.5	8.2	28.3	35.7	41.5
MA2	3.5	31.6	71.6	90.0	0.8	3.1	5.5	6.2
MA3	3.0	16.9	54.4	78.3	0.1	1.7	4.1	7.0
MA4	2.3	7.1	36.5	70.4	0.8	0.7	2.5	4.6
MA5	2.1	4.6	26.7	58.9	0.2	0.4	2.2	5.1
MA6	1.8	4.3	21.6	57.6	0.3	0.4	1.8	3.0

Figure 6: Hausdorff distances between the true and the estimated cointegrating space of systems (4.2) using the max test



Institut für Höhere Studien
Institute for Advanced Studies

Stumpergasse 56

A-1060 Vienna

Austria

Phone: +43-1-599 91-145

Fax: +43-1-599 91-163