Norges Bank

ARBEIDS NOTAT

Research Department

Oslo, December 22, 1991
1991/9

Inference in small cointegrated systems

Some Monte Carlo results

by

Øyvind Eitrheim

ISSN 0801-2504
ISBN 82-7553-031-8

Les

Notatet fås ved henvendelse
til Norges Bank, Biblioteket,
Boks 1179 Sentrum, 0107 Oslo 1

Norges Banks arbeidsnotater inneholder
forskningsarbeider og utredninger som
vanligvis ikke har fått sin endelige form.
Hensikten er blant annet at forfatteren
kan motta kommentarer fra kolleger og
andre interesserte.

Synspunkter og konklusjoner
står for forfatterens regning.

Copies of this Working Paper are obtainable
from Norges Bank, Library,
P.B. 1179, Sentrum, 0107 Oslo 1, Norway.

Norges Bank's Working papers present
research projects and reports
(not usually in their final form),
and are intended inter alia to enable
the author to benefit from the comments
of colleagues and other interested parties.

Views and conclusions expressed are the responsibility
of the author alone.

Inference in small cointegrated systems
Some Monte Carlo results

by
Øyvind Eitrheim[1]

22.December 1991

[1]The research for this paper was carried out while the author was visiting the Department of Economics, University of California, San Diego in 1990/1991. Thanks to Tony Hall, Eugen Nowak and Timo Teräsvirta for many stimulating discussions and to Clive W.J. Granger and Rob Engle for providing such a stimulating research environment. The paper was presented at the Nordic workshop on cointegration analysis in Helsingør, oct.28.-nov.1 1991 and I am indebted to Søren Johansen for his stimulating comments and suggestions. Finally I thank my colleages Kjell G. Bleivik, Eilev S. Jansen and Ragnar Nymoen for their comments to a previous version. Authors adress: Research Department, C51, Bank of Norway, P.B. 1179 Sentrum, 0107 Oslo 1, Norway. Phone: +47 2 316161, Fax: +47 2 424062, Bitnet: EXT90004 at NOBIVM.

Abstract

The Johansen procedure for testing and estimating cointegration models is analysed from a practitioner's perspective. We adress the robustness of the cointegration tests in small samples and with respect to particular types of misspecification of the model.

A small cointegrated system is parameterized and forms the basis for the Monte Carlo simulations. Non-parametric estimates of the distribution of the Trace and λ-Max tests are reported, as well as for some of the estimators for long- and short-run parameters in the model respectively. Power properties and finite sample performance for the cointegration test and estimators are discussed and the results are interpreted in the light of available asymptotics.

The types of model misspecification considered include the case with wrong dynamic specification (i.e. wrong order k in the VAR model) and the case when we ignore non-normality in the DGP residuals (i.e. when the DGP residuals are subject to ARCH (Auto Regressive Conditional Heteroscedasticity) or are serially correlated). We also discuss how data properties like temporal aggregation or systematic sampling may affect the inference on cointegration, and how the Johansen procedure performs under those conditions. Finally, we consider the case with cointegration between non-stationary latent variables which are observed with measurement errors.

Contents

1 Introduction

1.1 Background and motivation

Empirical analysis of small systems of cointegrated time series has become a popular approach to modelling simultaneous relationships between $I(1)$ non-stationary economic variables. The flourishing literature on cointegration has provided insights about postulated long run relationships between economic variables and has stimulated to the research on how these should be modelled in a dynamic context. The pioneering work on cointegration was presented in Granger(1981,1983) and Granger and Weiss(1983) and came as a natural extension of the univariate analysis of time series with a unit root in Fuller(1976), Dickey and Fuller(1979), Phillips(1987) and others. Engle and Granger(1987) proposed a simple two step approach to cointegration based on OLS estimation which immideately caught widespread attention. Stock(1987) compared OLS and NLS estimators for the parameters in a single equation error correction model (ECM) and analysed the asymptotic distribution for long run and short run parameter estimates. The single equation ECM has later been further analysed in Phillips and Hansen(1990), Stock and Watson(1990) and Phillips and Loretan(1991).

An alternative approach to cointegration has been proposed in Johansen(1988). He suggested to use the theory of reduced rank regression, inspired by the seminal work by Anderson(1951), cf. also e.g. Velu, Reinsel and Wichern(1986). This approach has been further developed in Johansen(1989,1992) and in Johansen and Juselius(1990). Gonzalo(1989) has compared five different approaches to model bivariate cointegration, and his Monte Carlo results came up with very favourable properties for Johansen's approach, which stimulated us to further analysis of this approach. See also Phillips and Loretan(1991) for additional Monte Carlo evidence on the relative merits of different approaches to cointegration.

The Johansen method is reasonably simple to implement on a PC, and it is also fairly general in the sense that it can be used in a multivariate context to estimate a small cointegrated system as well as in the usual single equation error correction model (ECM) framework. A VAR(k) model with Gaussian errors forms the backbone in Johansen's approach while e.g. Phillips(1991) analyses cointegration under

more general assumptions for the error distribution (i.e. assuming weakly dependent and heterogeneously distributed errors).

1.2 Non-stationarity and persistence in economic data

There is a controversy in the literature between people who believe in $I(1)$ non-stationary time series representations of economic time series in contrast to those who favor trend stationary models (where the time series are $I(0)$ but non-stationary and the non-stationarity stems from a deterministic polynomial trend (possibly including break points for shift in the trend)). At the centre of this debate is how literally one should interpret the unit root assumption. While most econometricians probably agree that economic variables seem to be characterized by substantial persistence in memory (that shock die out very slowly), and frequently also by some degree of time heterogeneity like heteroscedasticity, (e.g. with ARCH characteristics), the claim that the variables contain a unit root seem to be more controversial.

Phillips(1991) argues that prior knowledge about unit roots and cointegration represent important information about the data generating mechanism, DGP, and that the restrictions which follows should be imposed during estimation. Phillips and Loretan(1989), Johansen(1988) and Johansen and Juselius(1990) follow this view while e.g. Sims, Stock and Watson(1990) are in favour of using unrestricted models. Campbell and Perron(1991) take a more pragmatic standpoint in a recent survey paper and suggest that the strategy in each case should be motivated by the particular purpose of the analysis.

1.3 Some remarks about the methodology

A small cointegrated system is parameterized and used as basis for the Monte Carlo simulations in this paper. We have specified a four dimensional stochastic process and a system approach is preferred in order to determine the rank of the cointegrating space and estimate the parameters in the model. The approach to cointegration in Johansen(1988) satisfies this requirement and has been used throughout the paper. A reduced rank (canonical regression) procedure, similar to the one developed in Anderson(1951), forms the basis for the tests and the asymptotics

has been worked out for $I(1)$ and cointegrated variables in Johansen(1988)[1]. Two test statistics denoted *Trace* and λ_{Max} respectively were derived in order to determine the rank of the cointegration space. These methods have later been further developed by Johansen and Juselius(1990). Their approach have attracted particular interest among proponents of structural econometric modelling (SEM), since it allows for quite general classes of (linear) parameter restrictions to be tested within this framework[2]. The relevant test statistics turn out to be surprisingly simple, with a limiting χ^2 distribution, cf. e.g. Johansen(1992) for details. Phillips(1991) proves that this result holds more generally and in particular for the class of LAMN (Locally Asymptotic Multivariate Normal) models.

The distribution of the two rank tests for cointegration and the estimators for the long run parameters in the model turn out to be more complicated and have non-standard asymptotic distributions. It is the performance of these tests and estimators we will focus on in the following chapters.

A prototype DGP is constructed such that it contains a small cointegrated system with three relationships (part 2). Some properties of the cointegration tests are discussed in part 3 along with a discussion of the small sample performance. Part 4 presents some results from experiments where the model is misspecified in some direction. Part 5 concludes the paper.

[1] Anderson(1991) give interesting insights in the early development of econometric techniques in the Cowles Commision in the 1940-1950s and claim that many of the recent contributions in econometrics have aspects of "rediscovery" to them.

[2] Cf. e.g. Hendry and Mizon(1991) and Clements and Mizon(1991) who have applied the VAR cointegration model in this context.

2 The model

A popular definition of cointegration is the following:

Definition 2.1 *A vector of time series, x_t, with elements x_i, $i = 1,..,p$, each of which are assumed to be $I(d)$ are said to be cointegrated $I(d,k)$ if a linear combination $z_t = \beta' x_t$ is integrated $I(d-k)$ with $0 < k \leq d$.*

2.1 A small cointegrated system

Let the p-dimensional vector x_t be generated from a linear VAR(k) model with Gaussian errors[3]. All variables are assumed to be $I(1)$, and we apply the compact notation $x_t = (x_{1t}, x_{2t}, \ldots, x_{pt})'$. The VAR($k$) model is given by

$$x_t = \sum_{i=1}^{k} \Pi_i x_{t-i} + \mu_0 + \epsilon_t \tag{2.1}$$

and we assume $\epsilon_t \sim \text{Niid}(0, \Omega)$. The long-run multipliers in this model may be expressed by the matrix $\Pi(1) = I_p - \sum_{i=1}^{k} \Pi_i$, where I_p is the identity matrix. It is convenient to rewrite the model for x_t using the *interim multiplier* reparameterization[4].

$$\Delta x_t = \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \alpha\beta' x_{t-k} + \epsilon_t \tag{2.2}$$

The parameters in $\Gamma_j = -I_p + \sum_{i=1}^{j} \Pi_i \quad j = 1, \ldots, (k-1)$ and α may be thought of as *short run* parameters while β contain the *long run* parameters in the model. The prototype model is given by:

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \\ \Delta x_{3t} \\ \Delta x_{4t} \end{bmatrix} = + \begin{bmatrix} 0.1 & 0.4 & 0 & 0 \\ 0.3 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{1,t-1} \\ \Delta x_{2,t-1} \\ \Delta x_{3,t-1} \\ \Delta x_{4,t-1} \end{bmatrix} \tag{2.3}$$

$$\begin{bmatrix} -0.4 & 0.2 & 0 \\ 0.2 & -0.2 & 0 \\ 0 & 0 & 0.25 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{1,t-2} - x_{2,t-2} \\ x_{2,t-2} - x_{3,t-2} \\ x_{3,t-2} - x_{4,t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \epsilon_{4t} \end{bmatrix}$$

where $\epsilon_t | \mathcal{I}_{t-1} \sim \text{Niid}\ (0,\ \sigma_\epsilon\ I_4)$.

[3]This model has been extensively analysed by Johansen(1988) and Johansen and Juselius(1990).
[4]Cf. e.g. Johansen(1988), Johansen and Juselius(1990) or Hylleberg and Mizon(1989).

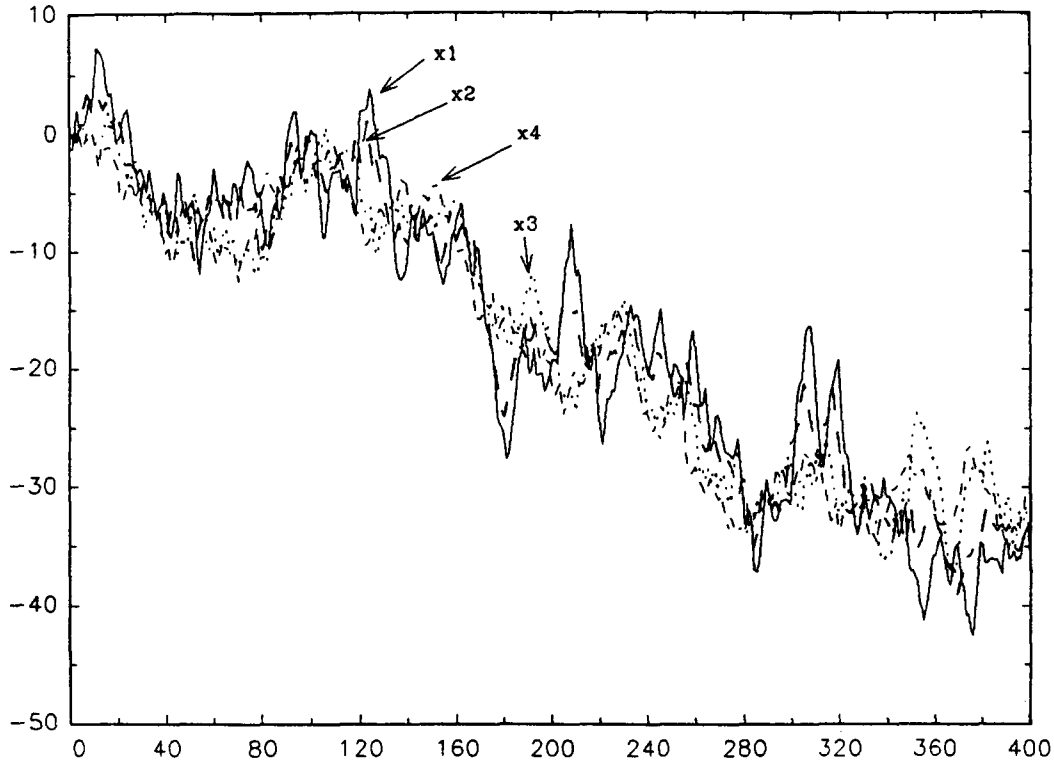

Figure 2.1: A simulated trajectory for the model (2.3)

2.2 A simulated example

The intuition behind cointegration can be expressed as follows. Consider a group of variables, for which the development over time is determined by (at least) one unit root, and hence creates a non-stationary typically $I(1)$ type of time trajectory. Cointegration means that the non-stationarity is a common characteristic (or feature) among the variables, and we expect the variables to stay together and not drift too far apart from each other over time. A trajectory for cointegrated variables is illustrated in figure 2.1 and shows a particular realization of the four-dimensional stochastic process (2.3).

2.3 The dynamic properties

Some further remarks about the DGP are useful. (2.3) is parameterized with $p = 4$, $r = 3$ $k = 2$ and with a particular normalization of the long run system imposed such that the differences between the series are stationary. In addition, the DGP contain some short run dynamics but such that only one root has modulus one. The unit root corresponds to the variable x_{4t} which is specified as a random walk. The

other roots have moduli less than one and there is also one pair of complex conjugate roots which contributes to stable cycles in the trajectory. The error correction part of the model will prevent the four variables from drifting too far apart and we will typically see a long swing (unit root) behaviour for the cluster of non-stationary variables, cf. figure 2.1 above.

How do the dynamic properties change when we change one of the columns in α? In order to describe the dynamics in the DGP it is useful to look at the characteristic roots of the following determinant[5].

$$\det \begin{bmatrix} \Pi_1 & \Pi_2 \\ I & 0 \end{bmatrix}$$

We have considered multiplicative changes in either the 2. or 3. column in α using a scalar $s \in [4, 2, 1, 1/5, 1/100]$. Table 2.1 decompose the roots of the system determinant.

The roots for the prototype model (2.3) are shown in the middle part of table 2.1. The two upper sets belong to models where we have increased the absolute values of either α_{j2} or α_{j3} and the two bottom sets belong to models where we let either α_{j2} or α_{j3} approach 0. Figure 2.2 show simulated trajectories for the five cases in the left part of table 2.1 (different α_{j2}s) while figure 2.3 show the five right hand cases (for different α_{j3}s).

Since x_{4t} is a simple random walk process and enters into the system only through the parameter α_{33}, we expect to see at least one unit root in table 2.1. In the limit when $\alpha_{j2} \to 0$ or $\alpha_{j3} \to 0$, we see that there will be two unit roots and the rank of the cointegration matrix will be reduced from 3 to 2. The model has one or two pairs of complex conjugate roots which give rise to cyclical movements in the series. When the absolute values of either α_{j2} or α_{j3} are sufficiently increased the modulus becomes larger than 1 for one of the pairs, and we obtain the explosive cyclical pattern which is evident in the upper left part of figure 2.2 and 2.3. The block diagonal structure in α will generate some interesting differences in the two limiting cases when either $\alpha_{j2} \to 0$ or $\alpha_{j3} \to 0$. When $\alpha_{j3} \to 0$, the system will be driven by two stochastic trends, one which drives the three series x_{1t}, x_{2t} and x_{3t} and another which drives x_{4t}. When $\alpha_{j2} \to 0$, the pairs (x_{1t}, x_{2t}) and (x_{3t}, x_{4t}) will be

[5]Cf. Hendry and Mizon(1991) where they suggests a similar procedure based on the companion form representation of the VAR(2) model.

Table 2.1 Dynamic characteristics of the DGP (2.3). Multiplicative changes in the columns α_{j2} or α_{j3}.

Changes in α_{j2}			Changes in α_{j3}		
Real root	Imaginary root	Modulus	Real root	Imaginary root	Modulus
	$4 \times \alpha_{j2}$			$4 \times \alpha_{j3}$	
0.75	0.24	0.79	0.35	0.61	0.70
0.75	-0.24	0.79	0.35	-0.61	0.70
0.35	0.95	1.01	0.85	-	0.85
0.35	-0.95	1.01	0.64	-	0.64
0.50	-	0.50	0.50	0.87	1.00
0.50	-	0.50	0.50	-0.87	1.00
0.00	-	0.00	0.00	-	0.00
1.00	-	1.00	1.00	-	1.00
	$2 \times \alpha_{j2}$			$2 \times \alpha_{j3}$	
0.35	0.73	0.81	0.35	0.61	0.70
0.35	-0.73	0.81	0.35	-0.61	0.70
0.76	0.15	0.77	0.85	-	0.85
0.76	-0.15	0.77	0.64	-	0.64
0.50	-	0.50	0.50	0.50	0.71
0.50	-	0.50	0.50	-0.50	0.71
0.00	-	0.00	0.00	-	0.00
1.00	-	1.00	1.00	-	1.00
	$1 \times \alpha_{j2}$			$1 \times \alpha_{j3}$	
0.35	0.61	0.70	0.35	0.61	0.70
0.35	-0.61	0.70	0.35	-0.61	0.70
0.86	0	0.86	0.85	0	0.85
0.64	0	0.64	0.64	0	0.64
0.50	0	0.50	0.50	0	0.50
0.50	0	0.50	0.50	0	0.50
0.00	0	0.00	0.00	0	0.00
1.00	0	1.00	1.00	0	1.00
	$1/5 \times \alpha_{j2}$			$1/5 \times \alpha_{j3}$	
0.38	0.48	0.61	0.35	0.61	0.70
0.38	-0.48	0.61	0.35	-0.61	0.70
0.98	0	0.98	0.85	0	0.85
0.45	0	0.45	0.64	0	0.64
0.50	0	0.50	0.01	0	0.01
0.50	0	0.50	0.99	0	0.99
0.00	0	0.00	0.00	0	0.00
1.00	0	1.00	1.00	0	1.00
	$1/100 \times \alpha_{j2}$			$1/100 \times \alpha_{j3}$	
0.39	0.46	0.61	0.35	0.61	0.70
0.39	-0.46	0.61	0.35	-0.61	0.70
1.00	0	1.00	0.85	0	0.85
0.42	0	0.42	0.64	0	0.64
0.50	0	0.50	0.00	0	0.00
0.50	0	0.50	1.00	0	1.00
0.00	0	0.00	0.00	0	0.00
1.00	0	1.00	1.00	0	1.00

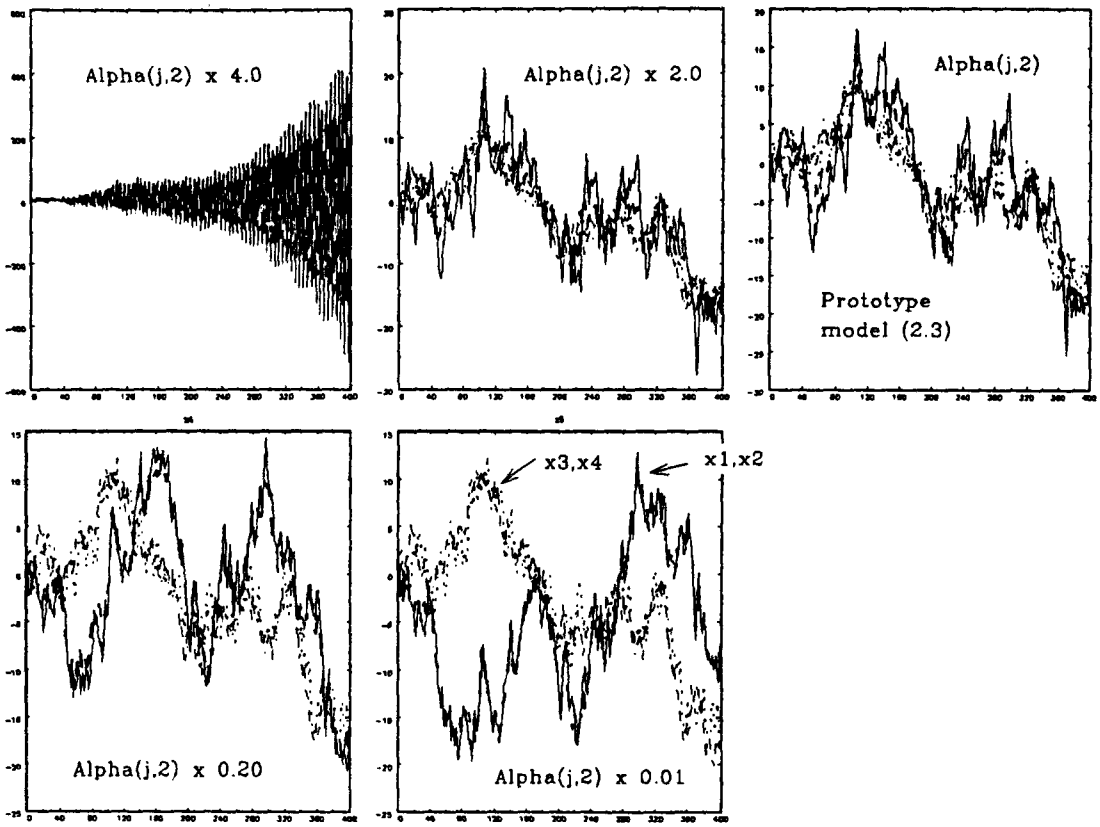

Figure 2.2: Simulated trajectories for (2.3) when we change α_{j2} to $\alpha_{j2} \times c \in [4, 2, 1, 1/5, 1/100]$ $j = 1, \ldots, 4$.

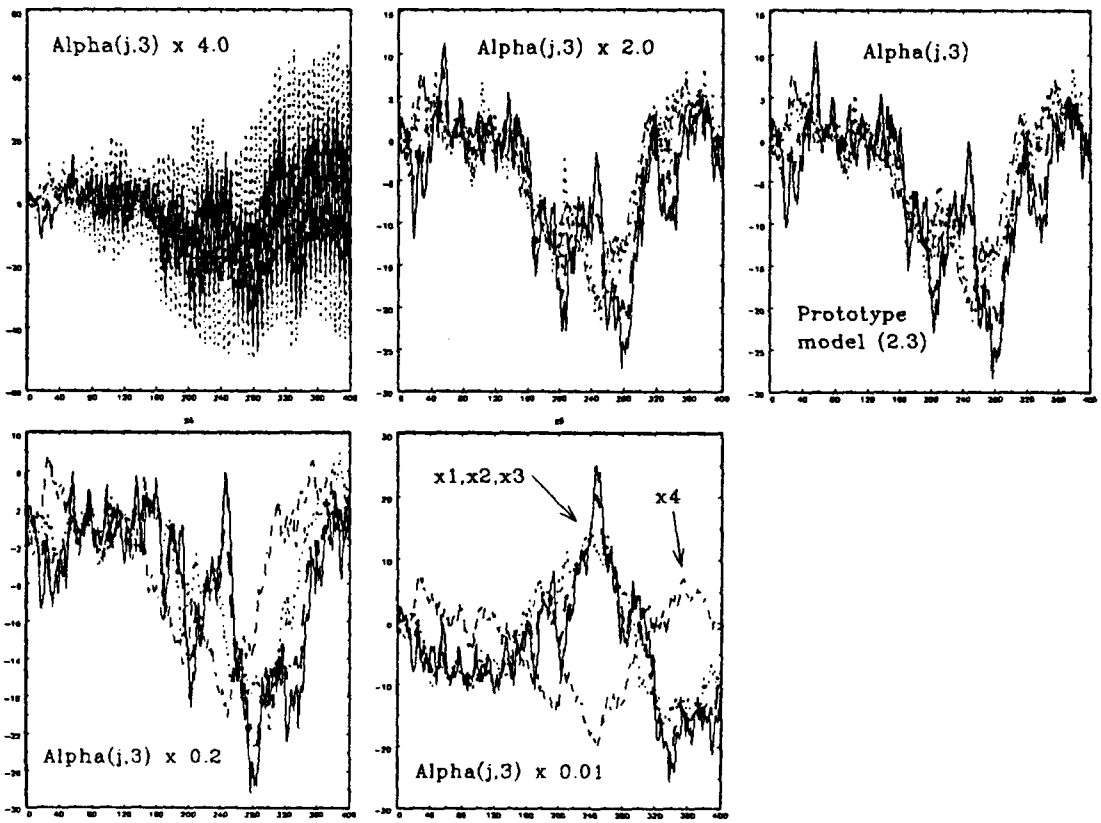

Figure 2.3: Simulated trajectories for (2.3) when we change α_{j3} to $\alpha_{j3} \times c \in [4, 2, 1, 1/5, 1/100]$ $j = 1, \ldots, 4$.

driven by separate stochastic trends. In both cases a typical long swing behaviour is reckognized for each of the two sub-clusters.

2.4 A note on renormalization

The model (2.3) is deliberately overparameterized (cf. Johansen(1989)), and the $2pr$ $(= 24)$ parameters in α and β are identified only up to some arbitrary linear transformation. Any non-singular $r \times r$ matrix D can be used to obtain 'new' (but observationally equivalent) matrices $\alpha^\dagger$ and $\beta^\dagger$ such that $\beta^\dagger = \beta D$, i.e.

$$-\Pi(1) = \alpha\beta' = \underbrace{\alpha D'^{-1}}_{\alpha^\dagger} \underbrace{D'\beta'}_{\beta^\dagger},$$

In order to facilitate the interpretation of the estimated results, it is often necessary to renormalize the original parameters $\hat{\beta}$ and $\hat{\alpha}$. In applied work we often impose a normalization of $\hat{\beta}$ by dividing each column with a column specific element k_j, say $\hat{\beta}_{k_j j}$. More generally we can renormalize $\hat{\beta}$ by defining the matrix $D = (\hat{\beta}'R)^{-1}$ for an appropriate choice of the $p \times r$ matrix R. For instance, to obtain the normalization rule above we set $R = [\iota_1, \ldots, \iota_r]$ where ι_j are $p \times 1$ vectors with $(p-1)$ 0's and a 1 in the k_jth element.

For given rank r, we suggest the following procedure to renormalize the estimated parameters $\hat{\beta}$. We let $\hat{\bar{\beta}}$ denote the renormalized estimates when we use a particular transformation matrix $\bar{D}$, i.e. $\hat{\bar{\beta}} = \hat{\beta}\bar{D}$. $\bar{D}$ has to be designed carefully, and we use whatever available information we can think of about the long run system and its natural representation in order to construct it. First we assign prior values r parameters in each column of a tentative parameter matrix $\bar{\beta}$ on the basis of this information. The number of restricted parameters is r^2 (=9) which leaves $(p-r)r$ (=3) free parameters to be calculated to obtain $\hat{\bar{\beta}}$. The r^2 elements in the transformation matrix $\bar{D}$ can be obtained from the expression

$$S\text{vec}(\bar{\beta}) = S(I \otimes \hat{\beta}')\text{vec}(\bar{D})$$

We have premultiplied with a $r^2 \times pr$ $(= 9 \times 12)$ matrix S which picks out the r^2 restricted parameters in $\bar{\beta}$. This expression is solved for $\text{vec}(\bar{D})$.

$$\text{(2.4)} \qquad \text{vec}(\bar{D}) = [S(I \otimes \hat{\beta}')]^{-1} S\text{vec}(\bar{\beta})$$

Finally, we calculate the free parameters and obtain $\hat{\tilde{\beta}} = \hat{\beta}\bar{D}$.

In our case it is natural to select the a priori values in $\bar{\beta}$ on the basis of the DGP given by (2.3). The preassigned values are represented by the 0's and 1's in the matrix below, leaving $\beta_{21}, \beta_{32}, \beta_{43}$ as free parameters. Note that this transformation is performed in order to facilitate the interpretation and that the procedure does not restrict the estimated long run matrix $-\widehat{\Pi(1)}$ in any way.

Tentative long run parameters $\bar{\beta}$ and renormalized estimator $\hat{\tilde{\beta}}$

$$\bar{\beta} = \begin{bmatrix} 1 & 0 & 0 \\ \bar{\beta}_{21} & 1 & 0 \\ 0 & \bar{\beta}_{32} & 1 \\ 0 & 0 & \bar{\beta}_{43} \end{bmatrix} \qquad \hat{\tilde{\beta}} = \begin{bmatrix} 1 & 0 & 0 \\ \hat{\tilde{\beta}}_{21} & 1 & 0 \\ 0 & \hat{\tilde{\beta}}_{32} & 1 \\ 0 & 0 & \hat{\tilde{\beta}}_{43} \end{bmatrix}$$

Appropriate selection matrix S

$$S = \left[\begin{array}{cccc:cccc:cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right]$$

2.5 The Trace test for the cointegrating rank

Let $z_0 = \Delta x_t, z_1 = \Delta x_{t-1}$ and $z_2 = x_{t-2}$. The sample moments between the z's are denoted M_{ij} for $j = 0, 1, 2$. Now, define by R_{0t}, R_{kt} the residuals from two sets of auxilliary regressions, i.e. those of z_0 and z_2 on z_1 respectively, and let $S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it} R'_{jt}$, for $i, j = 0, k$ denote their sample covariances. To determine the rank of the long run parameter matrix $\Pi(1)$, we solve for the eigenvalues in the following equation.

$$(2.5) \qquad |\lambda S_{kk} - S_{k0} S_{00}^{-1} S_{0k}| = 0$$

We test the null hypothesis $H_0 : r \leq \bar{r}$ using two different test statistics which are constructed in the following way. We first sort the elements in $\hat{\lambda}$ in descending order. If there are $\bar{r}$ cointegrating equations, we would expect the $\bar{r}$ largest eigenvalues to be greater than zero and represent the $\bar{r}$ linearly independent columns in $\Pi(1)$ which determines its rank. The last $(p-\bar{r})$ eigenvalues are expected to be zero. The *Trace*-test for cointegration is based on the sum of these $(p - \bar{r})$ smallest $\hat{\lambda}_i$'s which are zero under H_0, hence we test $H_0 : r \leq \bar{r}$ against $H_1 : \bar{r} + 1 \leq r \leq p$.

$$(2.6) \qquad \text{Trace} = -2 \ln Q = -T \sum_{i=\bar{r}+1}^{n} \ln(1 - \lambda_i)$$

The alternative test statistic is called λ_{Max} and test $H_0 : r \leq \bar{r}$, but now against the alternative $H_1 : r = \bar{r} + 1$. In this case the test is based on the largest of the $(p - \bar{r})$ $\hat{\lambda}_i$'s which are zero under H_0.

$$(2.7) \qquad \lambda_{\text{Max}} = -2 \ln Q = -T \ln(1 - \lambda_{\bar{r}+1})$$

2.6 The asymptotic distribution of the Trace test

Johansen(1988) derived the asymptotic distribution of the Trace-test statistic under the null hypothesis above, and it can be shown that the limiting distribution is a function of a m dimensional Brownian motion B $(m = p - \bar{r})$. Apart from m it can be shown that the limiting distribution of *Trace* is independent of other "nuisance parameters", cf. theorem 4.1 in Johansen(1989). More precisely, it can be shown that

$$(2.8) \quad \text{Trace} \simeq T \sum_{i=\bar{r}+1}^{p} \hat{\lambda}_i \rightarrow T \sum_{i=\bar{r}+1}^{p} \hat{\rho}_i = tr\{\int dBB'[\int BB'du]^{-1} \int BdB'\}$$

where B is the m-dimensional Brownian motion, ρ_i for $i = 1, \ldots, m$ are eigenvalues associated with the stochastic matrix above and all integrals are defined on the unit interval. Critical values have been simulated for the case with no linear trend in Johansen(1988). Johansen and Juselius(1989) have extended the simulations to situations when there may be a deterministic trend in the DGP and they shown how the asymptotic distribution of *Trace* change in this case[6].

2.7 The estimators for α , Γ_1 and β

When the variables are cointegrated, we can write the residuals ϵ_t in (2.2) as $\epsilon_t = R_{0t} + \alpha[\beta' R_{kt}]$. In the hypothetical case when β is known, α can be estimated by OLS since the model is linear in α. In practice, β is of course unknown and has to be estimated. If we know a ML estimator for β, say $\hat{\beta}$, we can still estimate $\hat{\alpha}$ by regressing R_{0t} on $\hat{\beta}' R_{kt}$ and obtain

$$(2.9) \quad \widehat{\alpha(\hat{\beta})} = -S_{0k}\hat{\beta}(\hat{\beta}' S_{kk}\hat{\beta})^{-1}$$

A similar estimator for $\hat{\Gamma}_1$ is obtained from the expression

$$(2.10) \quad \widehat{\Gamma_1(\hat{\beta})} = M_{01}M_{11}^{-1} - \widehat{\Pi(1)}M_{k1}M_{11}^{-1}$$

where $-\widehat{\Pi(1)} = \hat{\alpha}\hat{\beta}'$. The estimators are consistent and converge to their true values at the normal rate $T^{1/2}$, cf. Johansen(1989,1992) for details.

Johansen(1988) showed that maximum likelihood estimators for the long run parameters, $\hat{\beta}$, can be found in terms of the estimated eigenvectors associated with the $\bar{r}$ greatest eigenvalues in $\hat{\lambda}$. We set $\hat{\beta} = (v_1, \ldots, v_{\bar{r}})$ where v is the corresponding matrix of eigenvectors. It can be shown that these estimators are superconsistent (in the sense of Stock(1987)) and converge at the rate T, cf. Johansen(1989,1992) for details.

[6] Extended tables which cover higher dimensions of the VAR have been simulated by Osterwald-Lenum(1990).

2.8 A case study: Estimation results based on simulated data

In this part we will briefly discuss some experiences from the estimation of (2.2) based on a simulated sample with $T = 400$ observations. A correctly specified model is estimated "recursively" using the abovementioned procedures and we report the tests for cointegration and parameter estimates. The initial sample size is 25 observations and subsequent observations are added until the entire sample is used, which yields a time sequence of 376 estimates of test statistics, parameters and so on.

Recursive plots of the four eigenvalues are shown in figure 2.4. As we would expect, one of the eigenvalues turn out to be close to zero, while the other three approach non-zero limiting values. The convergence towards stable values is reasonably fast, but we note that all eigenvalues seem to be overstated at very small sample sizes (in particular for less than 40 observations). The scaled *Trace*- and λ_{Max} tests in figure 2.5 (scaled by the appropriate 5% fractiles reported in Osterwald-Lenum(1990)), indicate that $r = 3$ (which is correct since the prototype model (2.3) is constructed such that the true rank $\bar{r} = 3$.). Interestingly, there is a striking similarity between the two sequences of rank tests reported in figure 2.5. In order to understand this similarity, and consequently the advocated use of the cointegration tests $H_0 : r \leq= i$ for $i = 0, 1, \ldots, p-1$, it is useful to apply some results from Johansen(1991) which are inspired by the work by Pantula(1989).

Let $C_\eta(\epsilon)$ denote the (ϵ) critical region of the test sequence $\{T_0 \geq c_0(\epsilon), \ldots, T_\eta \geq c_\eta(\epsilon)\}$, which can be constructed for $\eta = 0, 1, \ldots, p-1$ where T_i denotes the test statistic used to test $H_0 : r \leq i$ and $c_i(\epsilon)$ the ϵ-quantile in the corresponding limiting distribution. It can be shown that $\lim \text{Prob}(r \in C_\eta) = 1$ if the true rank $\bar{r}$ is not contained in the range defined by η, (i.e. $\eta < \bar{r}$), such that the test will reject a false null hypothesis with certainty (at least asymptotically). Similarly, we will have that $\lim \text{Prob}(r \in C_\eta) = \epsilon$ when $\eta = \bar{r}$ and $\lim \text{Prob}(r \in C_\eta) \leq \epsilon$ when $\eta \geq \bar{r}$. For further details about the advocated sequence of tests, cf. Johansen(1991). In simple terms this procedure may be formulated as follows. Start out with $\eta = 0$, $\eta = 1, \ldots$ and so on and continue until the first non-rejection. In figure 2.5 this corresponds to simply count the number of curves above 1. The first non-rejection occurred for

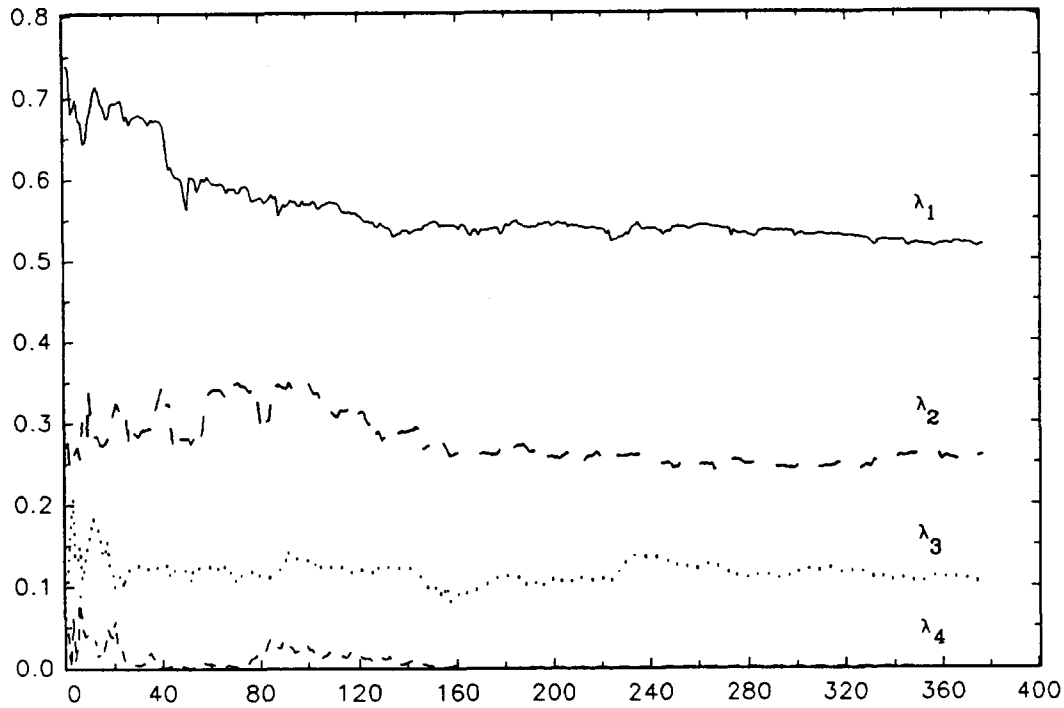

Figure 2.4: Estimated eigenvalues from simulated data. Recursive plots for observations 25 to 400. Correct VAR(2) specification.

$\eta = 3$ in our case so we conclude that $r = 3$.

The estimated values $\hat{\beta}$ and $\hat{\alpha}$ show a more problematic pattern. Figure 2.6 shows recursive plots of the estimated parameters for the four elements in the third column (i.e. $\hat{\beta}_{j3}$ and $\hat{\alpha}_{j3}$ $j = 1, \ldots, 4$ respectively). The estimates are unstable and seem to jump around as we increase the sample size. This picture is however turned around when we renormalize the parameters by the procedure suggested above, and we obtain new and remarkably stable estimates given by $\hat{\tilde{\beta}}_{j3}$ and $\hat{\tilde{\alpha}}_{j3}$, cf. figure 2.7. It is easily verified that the renormalized parameters lie close to their true values in the DGP (2.3) and the former instability is a pure artifact which (in this case) is easily removed. More evidence about these properties will be discussed later in this paper, in connection with the Monte Carlo experiments.

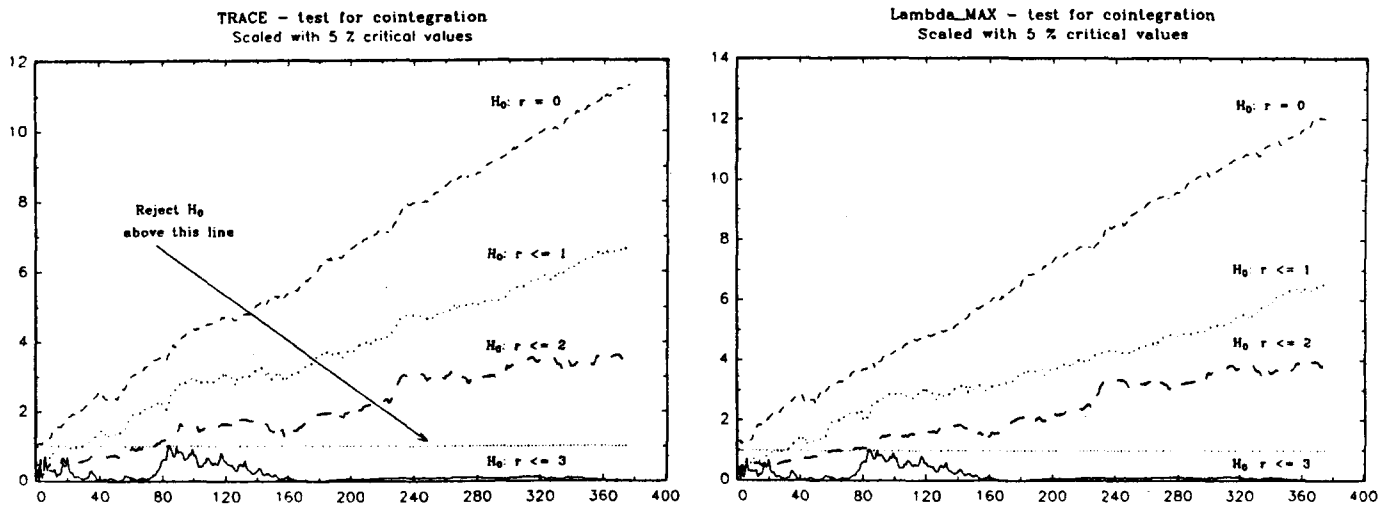

Figure 2.5: Scaled *Trace* tests (left) and λ_{Max} tests (right) for cointegration. Recursive plots for observations 25 to 400. Correct VAR(2) specification.

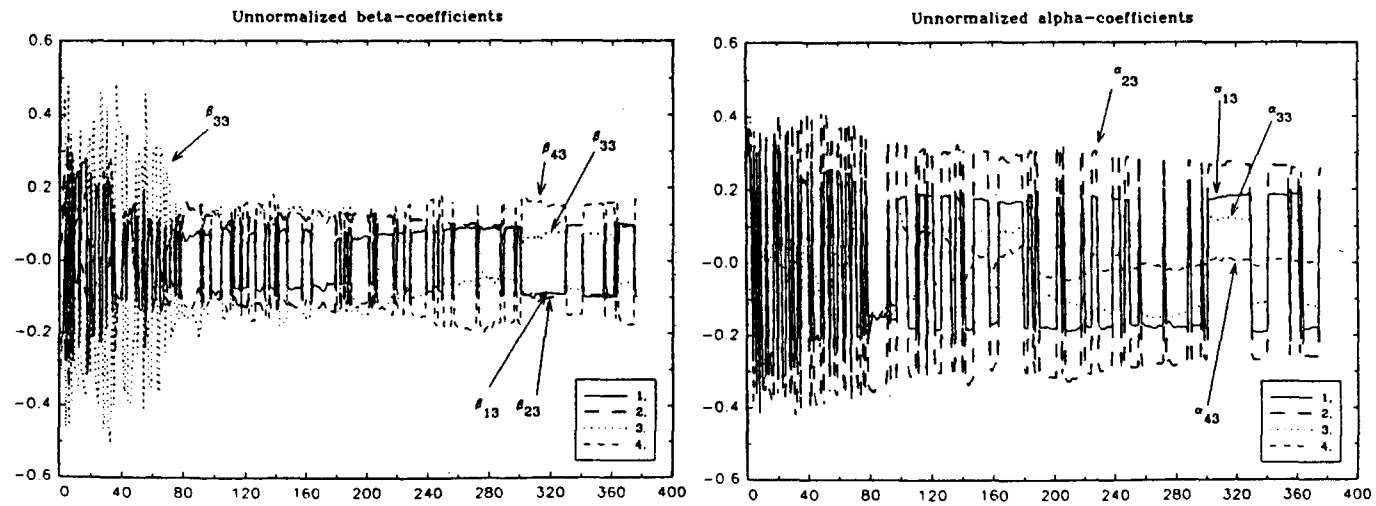

Figure 2.6: Estimates of the parameters $\hat{\beta}_{j3}$ and $\hat{\alpha}_{j3}$,$\forall j$ (eigenvectors). Recursive plots for observations 25 to 400. Correct VAR(2) specification

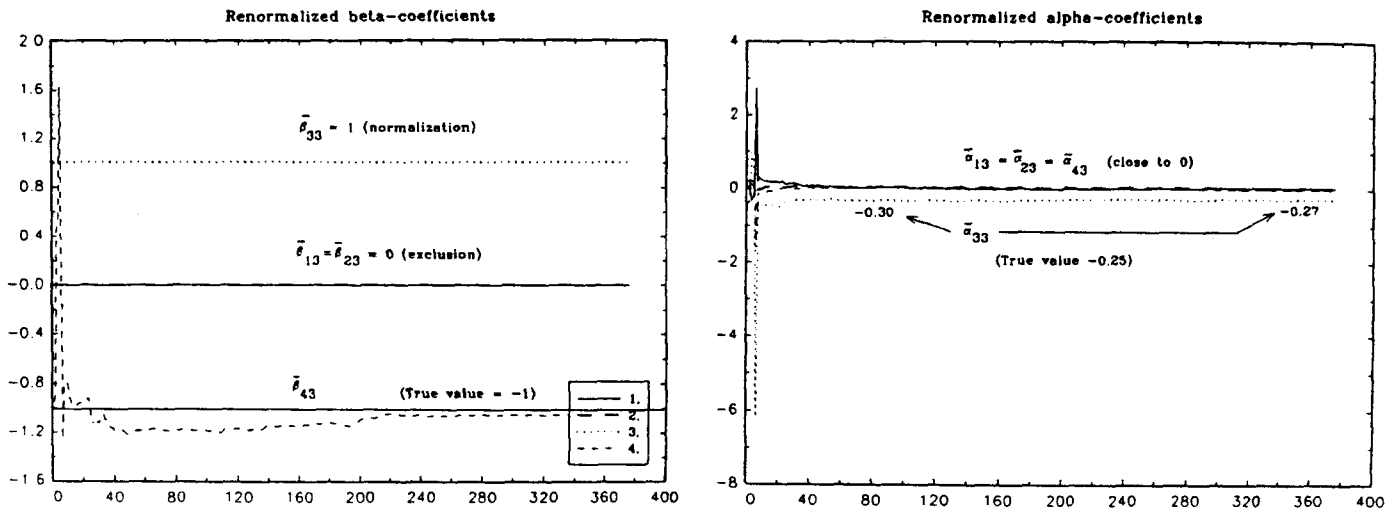

Figure 2.7: Renormalized estimates $\hat{\tilde{\beta}}_{j3}$ and $\hat{\tilde{\alpha}}_{j3}$, $\forall j$. Recursive plots for observations 25 to 400. Correct VAR(2) specification

3 Monte Carlo results for a correctly specified model

The small cointegrated system (2.3) is simulated n times and we have analysed the distribution of the *Trace-* and λ-max statistics. The objectives have been to learn more about the small sample behaviour for these tests, and to see how the different estimators perform in finite samples. To simplify the presentation, we have focused on the distribution of three particular estimators (for one element in each of the three matrices β, α and Γ_1), namely $\hat{\beta}_{21}, \hat{\alpha}_{11}$ and $\hat{\gamma}_{11}$. The corresponding "true" values in the DGP (2.3) are $\beta_{21} = -1$, $\alpha_{11} = -0.4$ and $\gamma_{11} = 0.1$.

The number of replications n vary between the experiments, between ca. 1500 and 5000 and we have presented the results by simple non-parametric estimates of the Monte Carlo distributions[7].

For the Trace- and λ-max test statistics we have also reported the rejection frequencies, based on the 5 % fractiles in the limiting distribution reported in Osterwald-Lenum(1990).

[7]A standard kernel estimator is used with a Gaussian kernel, and the densities are estimated from the expression $f_n(x) = 1/(nh_n)\sum_{i=1}^{n} K((x - x_i)/h_n)$ where $K(y) = 1/\sqrt{2\pi}\exp(-y^2/2)$. Cf. Hendry(1989) or Silverman(1986) for details.

Table 3.1 Size and power for the Trace and λ_{Max} tests for cointegration. Correct VAR(2) specification, $T = 400, n = 5000$, (per cent)

	Trace	λ_{Max}
$Pr[RejectH_0 : r \leq 3]$	5.02	5.02
$Pr[RejectH_0 : r \leq 2]$	100	100
$Pr[RejectH_0 : r \leq 1]$	100	100
$Pr[RejectH_0 : r = 0]$	100	100

3.1 Baseline results for the prototype model

The first Monte Carlo results for the Trace and λ-max tests and the estimators are based on a correct model specification (i.e. a VAR(2) model with Niid residuals) and we approximate the large sample properties using $T = 400$ observations. The relevant asymptotics for appropriately normalized estimators of the long run parameters, $\hat{\tilde{\beta}}_{ij}$, has been derived in Johansen(1989). The limiting distribution is non-standard and can be expressed as a function of Brownian Motions. The distribution of the short run parameters $\hat{\alpha}_{ij}$ and $\hat{\gamma}_{ij}$ can be shown to be asymptotically normal. For details on the limiting distributions, see Johansen(1989).

The Trace and λ_{Max} tests

The Trace test rejected the hypothesis $H_0 : r \leq 3 | \bar{r} = 3$ (i.e. that r is less than or equal to the true rank 3) in 5.02 % of the replications for a sample size $T = 400$, which indicates a correctly sized test[8]. The *Trace* test has substantial power according to these results (at least when evaluated at this particular parameter point and we note e.g. that the hypothesis $H_0 : r \leq 2 | \bar{r} = 3$ was rejected in 100 % of the cases. Rejection frequencies are reported in table 3.1. The estimated distribution of *Trace* is shown in figure 3.1. The 5% critical value is 12.54, cf. Osterwald-Lenum(1990), and is marked with a vertical line in the figure.

In figure 3.2 the estimated distribution of the λ_{Max} test observator shows similar results as for *Trace*. The 5% critical value is 11.44.

[8]The size evaluated at $T = 400$ observations corresponds to the number of observations at which the critical values were originally simulated. In simulations of the asymptotical distribution of *Trace* for $T = 1000$ using $n = 10000$ replications, we have obtained ca. 1 percent point larger critical values at the 5 % level.

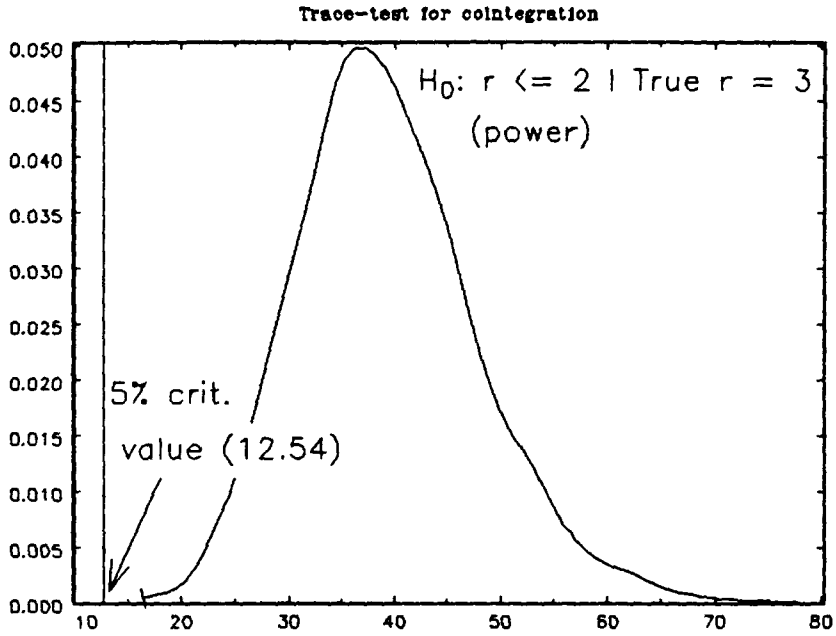

Figure 3.1: Estimated density function for the Trace test statistic. Correct VAR(2) specification, $T = 400, n = 5000$

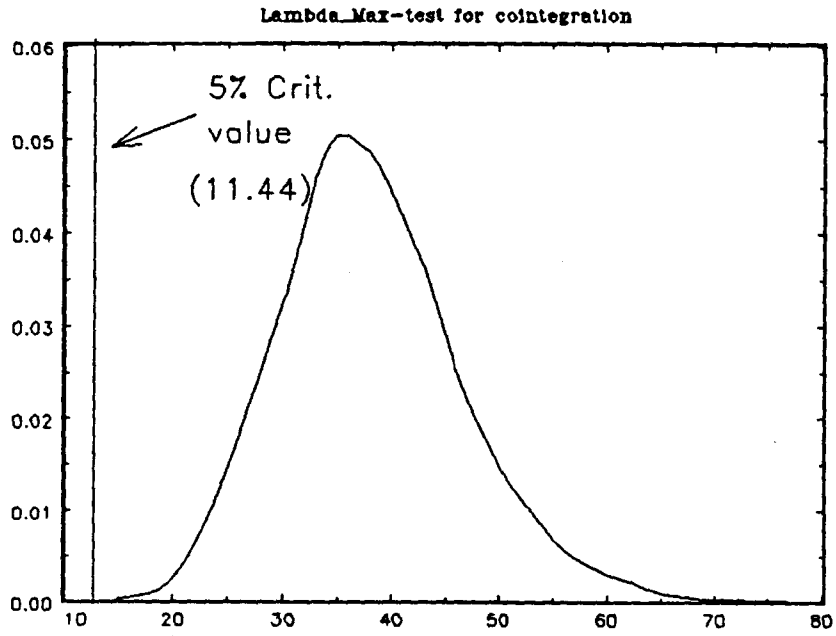

Figure 3.2: Estimated density function for the λ_{Max} test statistic. Correct VAR(2) specification, $T = 400, n = 5000$

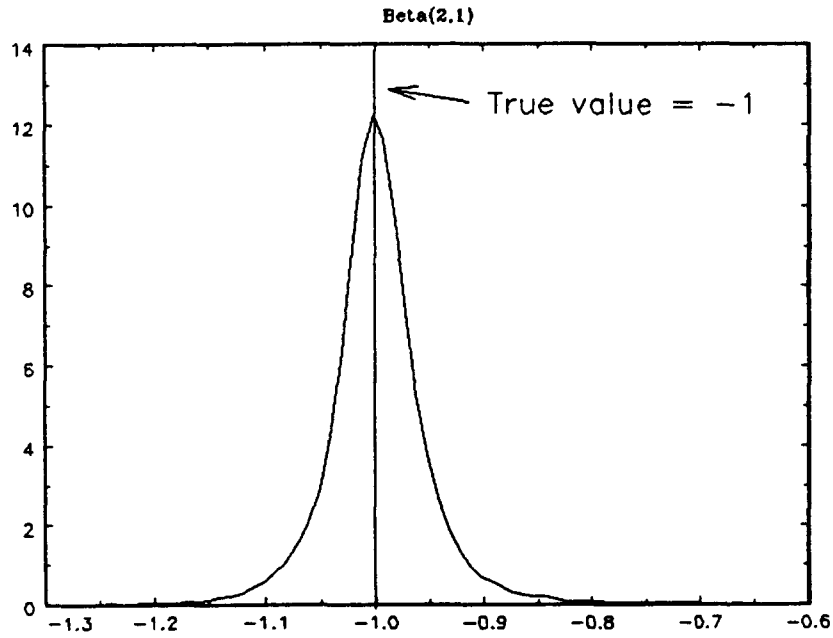

Figure 3.3: Estimated density functions for $\hat{\tilde{\beta}}_{21}$. Correct VAR(2) specification, $T = 400, n = 5000$.

The estimators

The Monte Carlo results for the estimators seem to support the asymptotic results in Johansen(1989,1992). Figure 3.3 shows the estimated distribution of the free long run coefficient in the first cointegrating equation, $\hat{\tilde{\beta}}_{21}$.

The estimator is median unbiased, symmetrically distributed around its true value -1 and seem to be highly concentrated around this value. The asymptotical results in Johansen(1989,1992) (to which we refer the readers for technical details and proofs) tells us to expect the limiting distributions for the estimators of the long run and the short run parameters respectively to be very different. The essential difference can be interpreted in terms of the *super consistency* property which hold for the estimators of the long run parameters, $\hat{\tilde{\beta}}$. The Monte Carlo distributions for the short run parameters $\hat{\tilde{\alpha}}_{11}$ and $\hat{\tilde{\gamma}}_{11}$ are shown in figure 3.4 and 3.5. Both seem to be median unbiased and symmetrically distributed around the true values but both distributions seem to be less concentrated than we observed for the long run estimators. The predicted difference between the two sets of parameters in this model, in terms of conducting inference, seem to be confirmed by the Monte Carlo simulations reported here.

The parameter restriction $\mu = 0$ has been imposed during the estimation of this

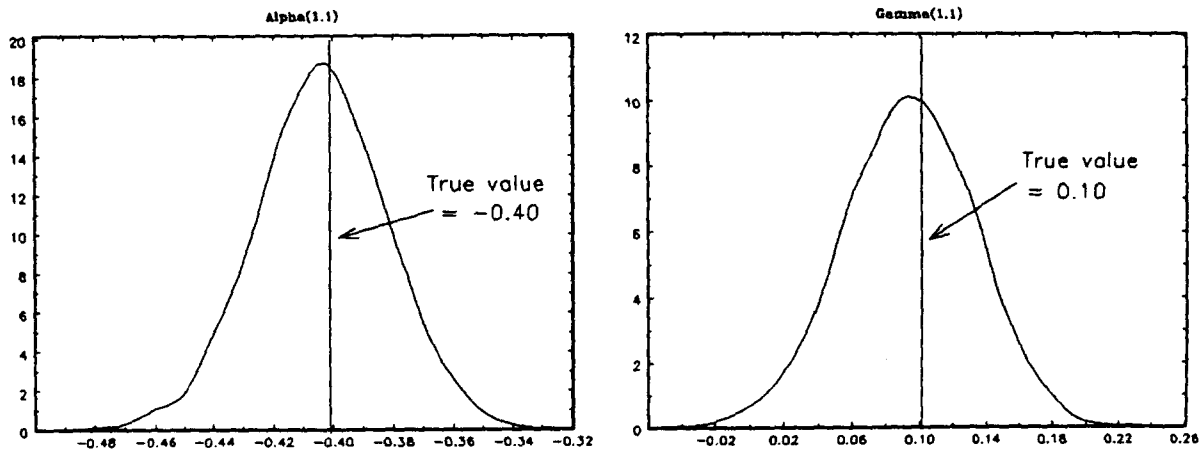

Figure 3.4: Estimated density functions for $\hat{\tilde{\alpha}}_{11}$ and $\hat{\tilde{\gamma}}_{11}$. Correct VAR(2) specification, $T = 400, n = 5000$.

model. We have also estimated models including a constant term, in which case the Monte Carlo distribution for $\hat{\tilde{\beta}}_{21}$ came out with heavier tails, indicating a loss of efficiency.

3.2 Power properties in the case of "near cointegration"

The parameters in (2.3), for simplicity abbreviated θ_0, belong to a finite dimensional parameter space Θ ($\theta_0 \in \Theta$). We will now demonstrate how the distributions of the cointegration tests will be affected when we change θ_0 in certain directions[9].

The particular changes we consider here have already been introduced in part 2. We compare five DGPs where the values in α_{j2} differ and gradually approach zero. Some of the consequences from this were briefly discussed above and we saw that in the limit (when $\alpha_{j2} = 0$), the cointegrating rank was reduced from three to two. For very small values in a particular column in α, the feed-back from the corresponding cointegrating relationship to the rest of the system will be weak, and we denote this as *"near cointegration"*. In (2.3) there are originally three cointegrating vectors,

[9]Ideally, we would of course prefer to analyse a more complete power surface for the cointegration tests but this would have blown up the computational costs considerably. We have therefore only considered a few points in the parameter space. All calculations are based on GAUSS-386 version 2.1.

but when $\alpha_{j2} \to 0$, only two will be left and the difference $x_{2t} - x_{3t}$ will become $I(1)$ non-stationary in the limit. This property is evident from the bottom right trajectory in figure 2.2 where x_{2t} and x_{3t} seem to be driven by separate stochastic trends.

The Trace test

The power of the cointegration test can be defined by the probability $Pr[Reject\ H_0 : r \leq 2|\bar{r} = 3])$. When $\alpha_{j2} \to 0$, this probability will approach the nominal *size* of a different test, namely $Pr[Reject\ H_0 :\ r \leq 2|\bar{r} = 2])$ since the limiting true rank will be reduced by 1.

It follows that the numerical value α_{j2}, or conversely the particular parameter point $\theta_0 \in \Theta$ in which we simulate the model, may have considerable effects on the power of the Trace test. Hence, the ability to determine r (and conduct correct inference about the long run parameters in the model) will depend on the true parameters in the DGP. Johansen(1989) has shown that in the case with "near cointegration", (e.g. with asymptotically vanishing values of say α_{j2}), the limiting distributions of the *trace* and λ-*Max* tests for cointegration have to be modified such that terms involving Brownian motions (cf. (2.8) are replaced by similar terms from an Ornstein-Uhlenbeck type stochastic process. Some simulation results are reported in Johansen(1989) for the *Trace* test in the case when $\bar{r}$ is reduced with 1 asymptotically.

A different approach is used here. Instead of simulating the limiting distribution for *Trace* under the Ornstein-Uhlenbeck assumptions, we have investigated the power properties on the basis of simulations of the small cointegrated system (2.3). Five experiments are compared where we gradually reduce the weights in α_{j3} towards zero. The results turn out to yield strong support to the power results in Johansen(1989), cf. figure 3.6. and table 3.2. It becomes increasingly difficult to detect cointegrating vectors when they have a sufficiently small weight in the error correction representation. Similar results hold for both the *Trace* and the λ-*Max* test for cointegration.

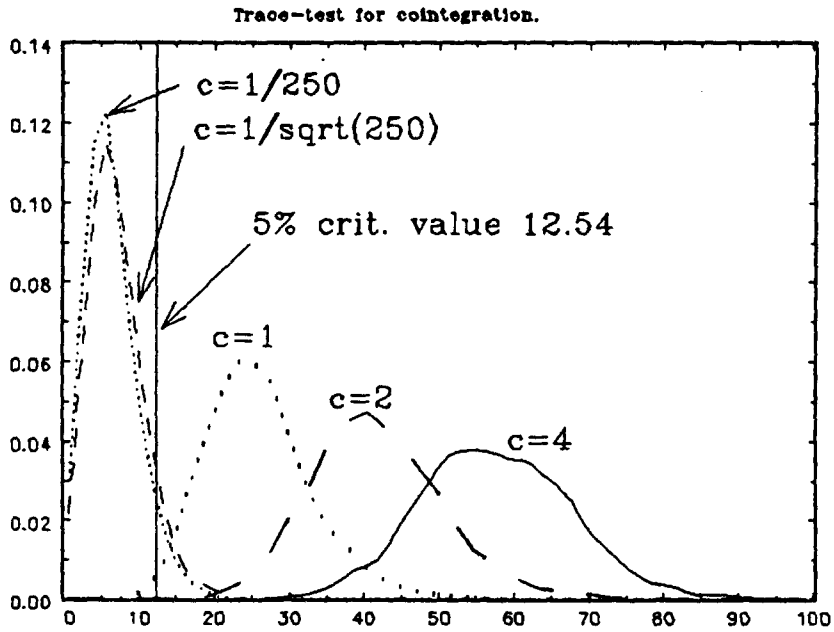

Figure 3.5: Estimated density function for the Trace test statistic for $\alpha_{j2} \times c \in [4, 2, 1, 1/5, 1/100]$ $j = 1, \ldots, 4$. Correct VAR(2) specification, $T = 250, n = 2000$

Table 3.2 Size and power of the Trace test for cointegration.
Different parameterizations $\alpha_{j2} \times c \in [4, 2, 1, 1/\sqrt{250}, 1/250]$
$j = 1, \ldots, 4$ $T = 250, n = 2000$, (per cent).

	Size	Power	Power
	$Pr[Reject\ H_0 : r \leq 3]$	$Pr[Reject\ H_0 : r \leq 2]$	$Pr[Reject\ H_0 : r \leq 1]$
$c = 4$	5.90	100.00	100.00
$c = 2$	6.00	100.00	100.00
$c = 1$	5.90	99.15	100.00
$c = 1/\sqrt{250}$	1.60	7.05	100.00
$c = 1/250$	1.05	5.05	100.00

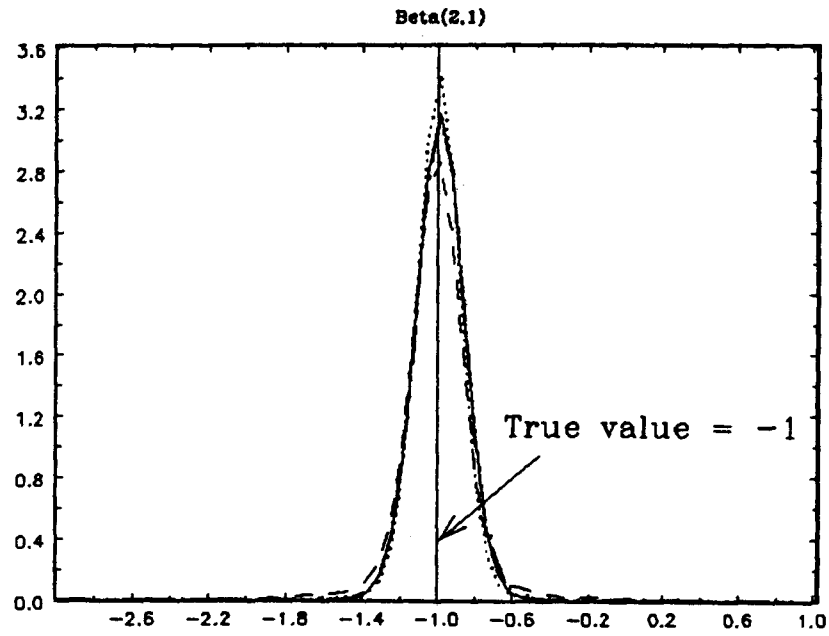

Figure 3.6: Estimated density functions for $\hat{\tilde{\beta}}_{21}$ for $\alpha_{j2} \times c \in [4, 2, 1, 1/\sqrt{250}, 1/250]$, $j = 1, \ldots, 4$. Correct VAR(2) specification, $T = 400, n = 2000$.

The Estimators

We saw above that a reduction in α_{j2} caused gradual shifts to the left in the distribution of *Trace*. In contrast, there seems to be almost no effect on the distribution of the estimator $\hat{\tilde{\beta}}_{21}$ from changing α_{j2} in this way, cf. figure 3.6. Hence, the inference about long run parameters in the model appears to be robust with respect to the size of α_{j2}, provided that we conduct correct inference about r and use the relevant information to renormalize the system.

On the other hand, we see from figure 3.7 that the effects on the distribution of $\hat{\tilde{\alpha}}_{11}$ and $\hat{\tilde{\gamma}}_{11}$ are significant. When we decrease α_{j2} the shape of the distributions of the short run estimators change, and the variance seem to increase substantially (although the estimators seem to remain median unbiased).

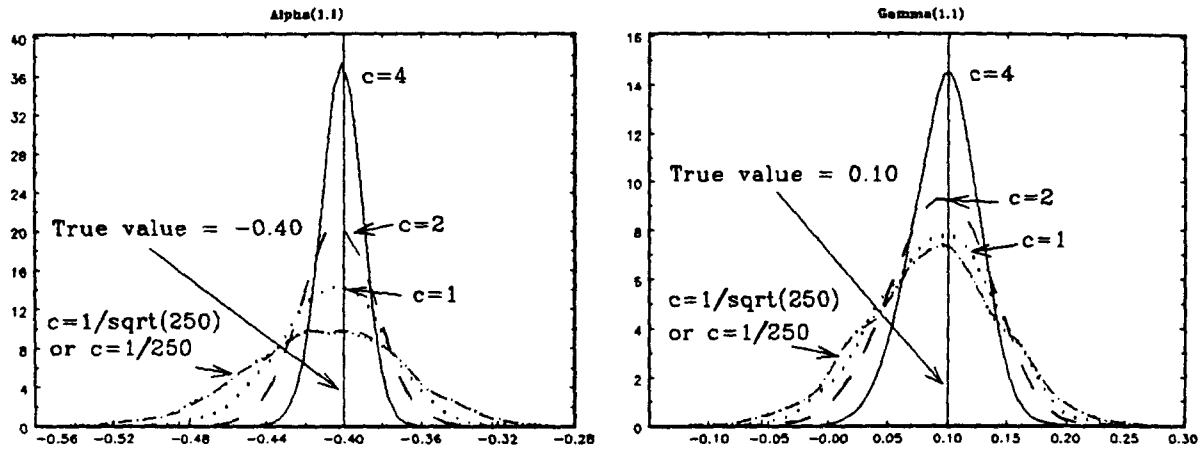

Figure 3.7: Estimated density functions for $\hat{\tilde{\alpha}}_{11}$ and $\hat{\tilde{\gamma}}_{11}$ for $\alpha_{j2} \times c \in [4, 2, 1, 1/\sqrt{250}, 1/250]$ $j = 1, \ldots, 4$. Correct VAR(2) specification, $T = 400, n = 2000$.

3.3 Finite sample performance

In applied work, the available data consist of a finite (and often relatively small) number of observations. In many cases we will try to determine the cointegrating rank on the basis of e.g. 50 or 100 observations using annual or quarterly time series. It would therefore be of great advantage to know the small sample properties of the different tests and estimators. We have conducted a number of experiments where we gradually increase the sample size by 50 or 100 observations and compare results for $T \in [50, 100, 150, 200, 250]$ and $T \in [100, 200, 300, 400, 500]$. Figures from the latter set of experiments are shown below.

The Trace and λ_{Max} tests

Figure 3.8 shows the Monte Carlo distributions for *Trace* and *λ-Max* for the sequence of hypotheses described in part 2.8. It is evident that the distributions of *Trace* and *λ-Max* go off to infinity with T in the cases when the true rank ($\bar{r} = 3$) is not contained in the range η of r-values satisfying the null hypothesis. We see that the distributions shifts to the right when we increase the sample size. This reflects the consistency property of the test sequence in Johansen(1991), i.e. that $\lim \text{Prob}(r \in C_\eta) = 1$ for $\eta < \bar{r}$. Similarly, note that for $\eta = \bar{r} = 3$, the distribution

corroborate reasonably well the conjecture that the nominal size of the test ($\epsilon = 5\%$) is correct, although the results in column 1 of table 3.3 show some variation. This is shown in the bottom of figure 3.8 where we see that the location of the distributions is independent of the sample size.

The results indicate furthermore that we loose power when we reduce the sample size and we note that all three cointegrating relationships are detected only for a sufficiently large sample, cf. the reported rejection frequencies in table 3.3. In small samples, $T = 50$, we find the correct rank in only 15 % of the simulations. The power improves when we increase the sample size but we need almost 200 observations before the power exceeds 90 %. Asymptotically the power is 100% as we have demonstrated above.

In part 3.2 we saw how the power of the cointegration test depend on the true parameters in the DGP (in particular on the values in α) and that it becomes increasingly difficult to detect cointegration in a direction where the feedback back onto the system is weak. It is interesting to note that even when the feedback is weak, as long as it remains different from zero the asymptotic result above will apply and the cointegration test has 100% power. When $\alpha_{jk} \to 0$ for some k, the cointegration test has only trivial power since the test degenerates. Since the location of the distribution of *Trace* depend on T, the combined effect is that the power of the cointegration test may be very low in finite samples. Given the formulation of the hypothesis, we will tend to underestimate r.

How important are these problems in practice? If the model is built primarily for forecasting, it may not matter too much if one erroneously excludes a cointegrating vector, which only has a weak link back to the rest of the system. If, on the other hand the purpose is to detect and learn about deep structural parameters to which there may be attached a particular theoretical interpretation, low power of the cointegration test may complicate the interpretation of the results considerably (it may e.g. be difficult to impose the relevant prior restrictions on $\bar{\beta}$ if r is underestimated).

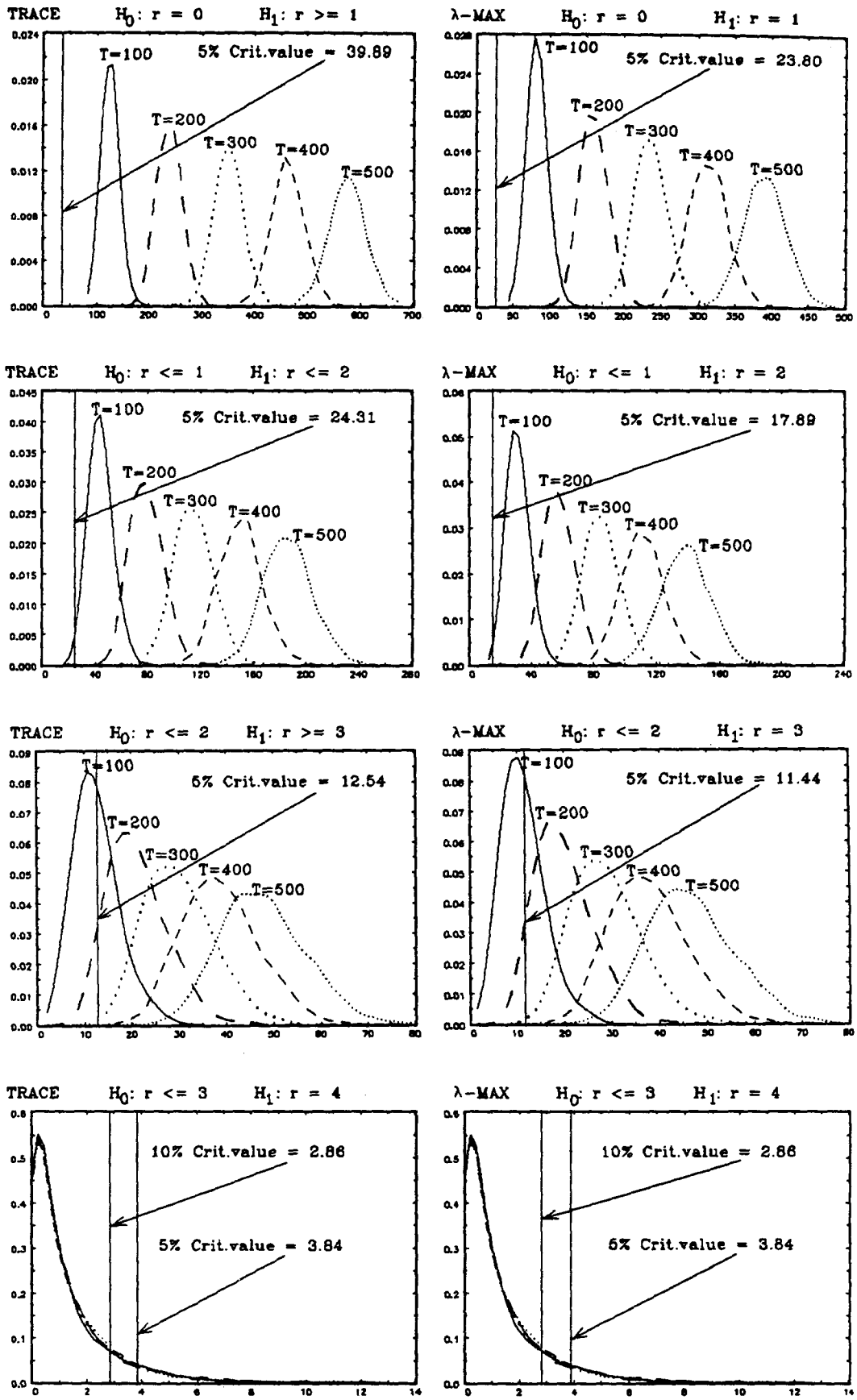

Figure 3.8: Estimated density functions for *Trace* and *λ-Max*. Correct VAR(2) specification, $T \in [100, 200, 300, 400, 500], n = 1850$

Table 3.3 Size and power for the Trace test for cointegration. Results in small samples. Correctly specified VAR(2) model with true rank $\bar{r} = 3$, $n = 2000$

	Size	Power	Power
T	$Pr[Reject\ H_0 :\ r \leq 3]$	$Pr[Reject\ H_0 :\ r \leq 2]$	$Pr[Reject\ H_0 :\ r \leq 1]$
50	4.40	14.85	65.95
100	5.80	47.65	99.40
150	5.85	77.00	100.00
200	6.20	95.15	100.00
250	5.90	99.55	100.00
100	6.61	44.43	99.41
200	6.83	92.96	100.00
300	5.76	99.95	100.00
400	5.97	100.00	100.00
500	5.65	100.00	100.00

The estimators

Figure 3.9 shows the distribution of $\hat{\tilde{\beta}}_{21}$ for different sample sizes. The estimator seem to be median unbiased and also symmetrically distributed around its true value -1. The variance decreases when we increase the sample size. When we use more than 300 observations, the shape of the distribution seem to become invariant with respect to further increases.

The distributions of $\hat{\tilde{\alpha}}_{11}$ and $\hat{\tilde{\gamma}}_{11}$ are shown in figure 3.10. Both seem to have a negative bias in small samples, but as we increase the sample size, they converge towards their true values. The distribution also becomes more smooth, symmetric and concentrated with a smaller variance. This is in line with the theoretical asymptotics which in this case predict normality.

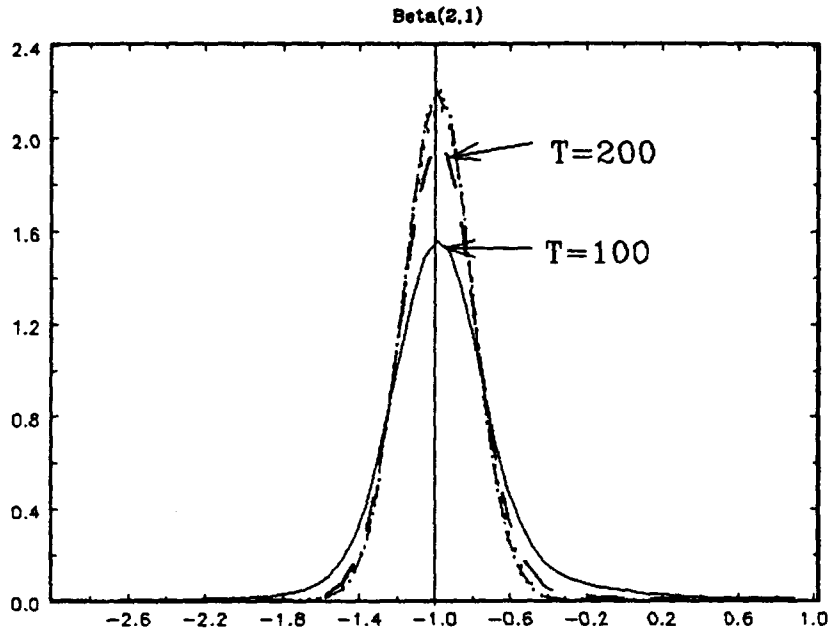

Figure 3.9: Estimated density functions for $\hat{\tilde{\beta}}_{21}$. Correct VAR(2) specification, $T \in [100, 200, 300, 400, 500], n = 1850$

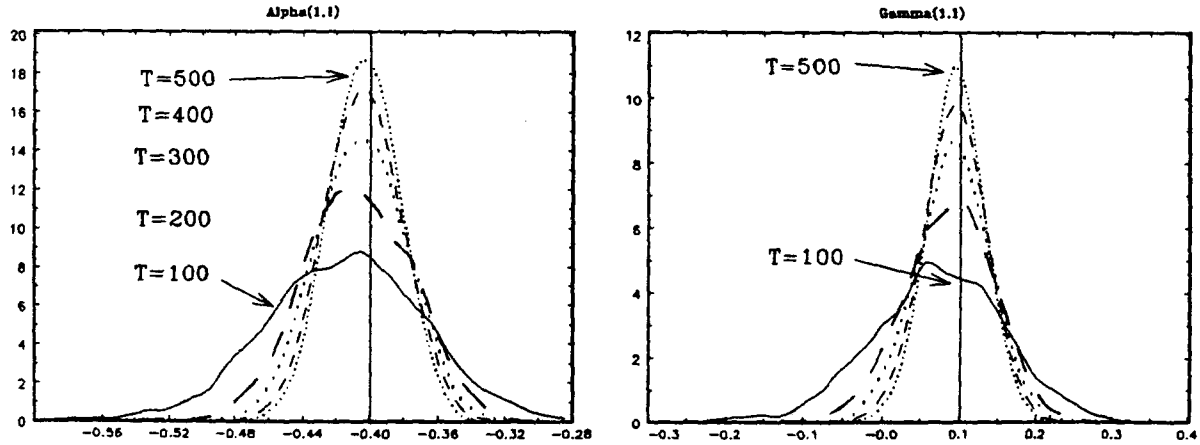

Figure 3.10: Estimated density functions for $\hat{\tilde{\alpha}}_{11}$ and $\hat{\tilde{\gamma}}_{11}$. Correct VAR(2) specification, $T \in [100, 200, 300, 400, 500], n = 1850$.

4 Monte Carlo results for some misspecified models

4.1 Misspecified dynamics, the VAR(k) has wrong order

In applied work the order k of the VAR will generally be unknown to the researcher. A number of model selection criteria (to determine k) have been discussed in the literature, usually based on tests for serial correlation or (multivariate) normality. Recent results in Jacobsson(1991) suggests that a multivariate portmanteau statistic like the multivariate Box-Pierce test, will work well in a number of situations. Intuitively we would expect it to be more harmful to chose k too small rather than too large. Too many lags would violate the principle of parsimony in the model specification and give rise to superfluous parameters and loss of efficiency in estimation. Too few lags however, will in most cases lead to an insufficient representation of the model dynamics and may also cause problems for the inference on cointegration and estimation of the long run parameters in the model.

We can illustrate this by simulation results for different model specifications where we let the order k vary. Five cases are considered with $k \in [1,2,3]$. For $k \in [2,3]$ the model has been estimated twice, with the level part of (2.2) placed on the kth lag as well as on the 1st lag. The interpretation of the long run matrix $-\Pi(1) = \alpha\beta'$ remain unchanged but the the short run parameters in Γ_1 must be redefined, i.e. such that $\Gamma^{(1)} = \Gamma^{(k)} - \alpha\beta'$. Hence, the "true value" $\gamma_{11}^{(k)} = 0.1$ change to $\gamma_{11}^{(1)} = 0.5$ when the level part is put on lag 1.

The Trace test

Figure 4.1 shows only minor differences between the estimated distributions of the Trace statistic in the five experiments. The dotted curve represents the VAR(1) case when the order k is underspecified. The other curves represents the pairs of VAR(2)s and VAR(3)s respectively. Some power is lost in the VAR(1) case and the most favourable results are obtained for a correct choice $k = 2$. As we might expect, it has absolutely no effect on the distribution of *Trace* whether we put the level part on the 1. or the k. lag. Interestingly, the cointegration test seem to be able to determine the correct rank r reasonably well even when k is specified too small.

Table 4.1 Size and power for the Trace test for cointegration.
Incorrect order k in the VAR, true value is $k = 2$.
Long run parameters in (2.2) on 1. or k. lag. $n = 2000$

	Size	Power	Power
	$Pr[Reject\ H_0\ r \leq 3]$	$Pr[Reject\ H_0\ r \leq 2]$	$Pr[Reject\ H_0\ r \leq 1]$
k=2, LR=2	5.50	100	100
k=2, LR=1	5.50	100	100
k=1, LR=1	3.65	97.95	100
k=3, LR=3	5.00	99.90	100
k=3, LR=1	5.00	99.90	100

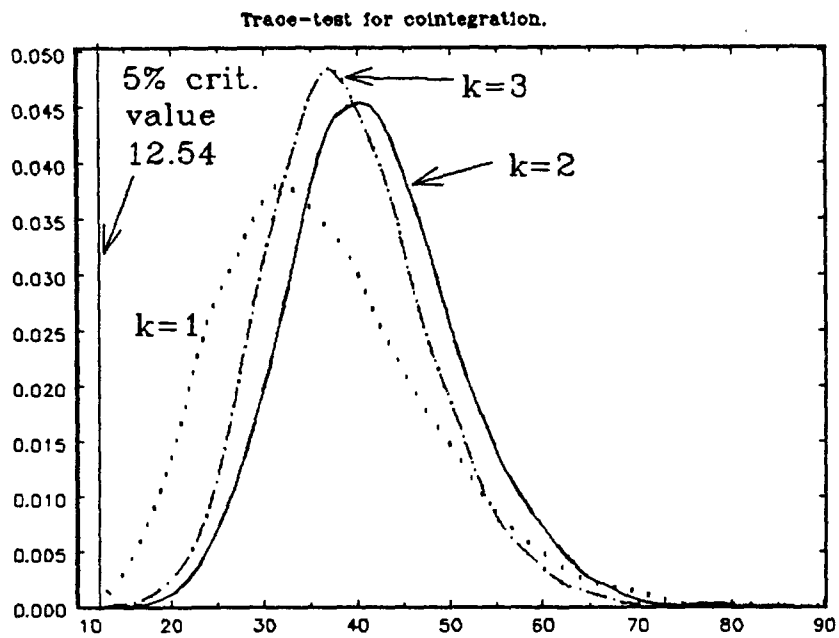

Figure 4.1: Estimated density function for the Trace test statistic. Incorrectly specified VAR(k) models, true value is $k = 2$. Long run parameters in (2.2) on 1. or k. lag. $n = 2000$.

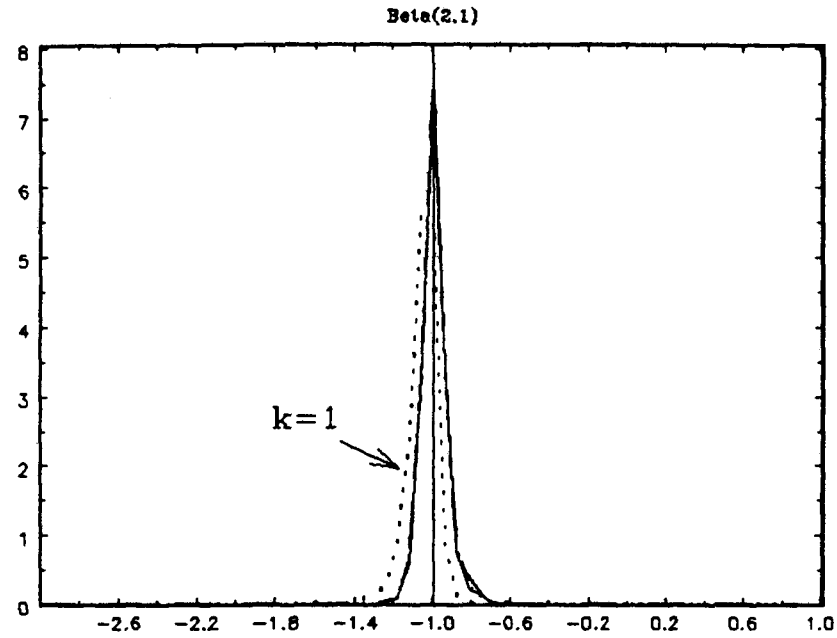

Figure 4.2: Estimated density functions for $\hat{\tilde{\beta}}_{21}$. Incorrectly specified VAR(k) models when the true value is $k = 2$. Level part of (2.2) on 1. or k. lag. $n = 2000$.

The estimators

For a correct or overspecified dynamic representation of the VAR, i.e. when $k \geq 2$, we find that the distribution of $\hat{\tilde{\beta}}_{21}$ is median unbiased and symmetric, while the results for $k = 1$ indicate a serious bias problem when we attempt to renormalize the estimated results in the usual way. In this case we are not able to recover the "true" value, hence the estimator $\hat{\tilde{\beta}}_{21}$ seems to be inconsistent when the dynamic specification is too restricted. There seem to be no visible loss of efficiency from overspecification of the VAR, in the sence that more of the probability mass ends up in the tails of the distribution of $\hat{\tilde{\beta}}_{21}$.

We find similar results for the estimators for the short run parameters. When k is too small, $k = 1$, the distribution of $\hat{\tilde{\alpha}}_{11}$ is seriously dislocated, indicating inconsistency, while for $k \geq 2$ we obtain consistent estimates of $\tilde{\alpha}_{11}$. For $k = 3$ there is a loss of efficiency compared to the results when $k = 2$. The results for $\hat{\tilde{\gamma}}_{11}$ (for $k \in [2, 3]$) yield similar results. Note again that when the level part of the model is placed on the 1st lag rather than the kth, the "true value" changes from 0.1 to 0.5. This explains the shift in the distribution in figure 4.3. The estimators are clearly consistent and we see again somewhat larger variance in the case when $k = 3$[10].

[10]For technical reasons the models are estimated with a constant term and three seasonal dummies included. In the case when $k = 1$ it is the distribution of the estimated constant term

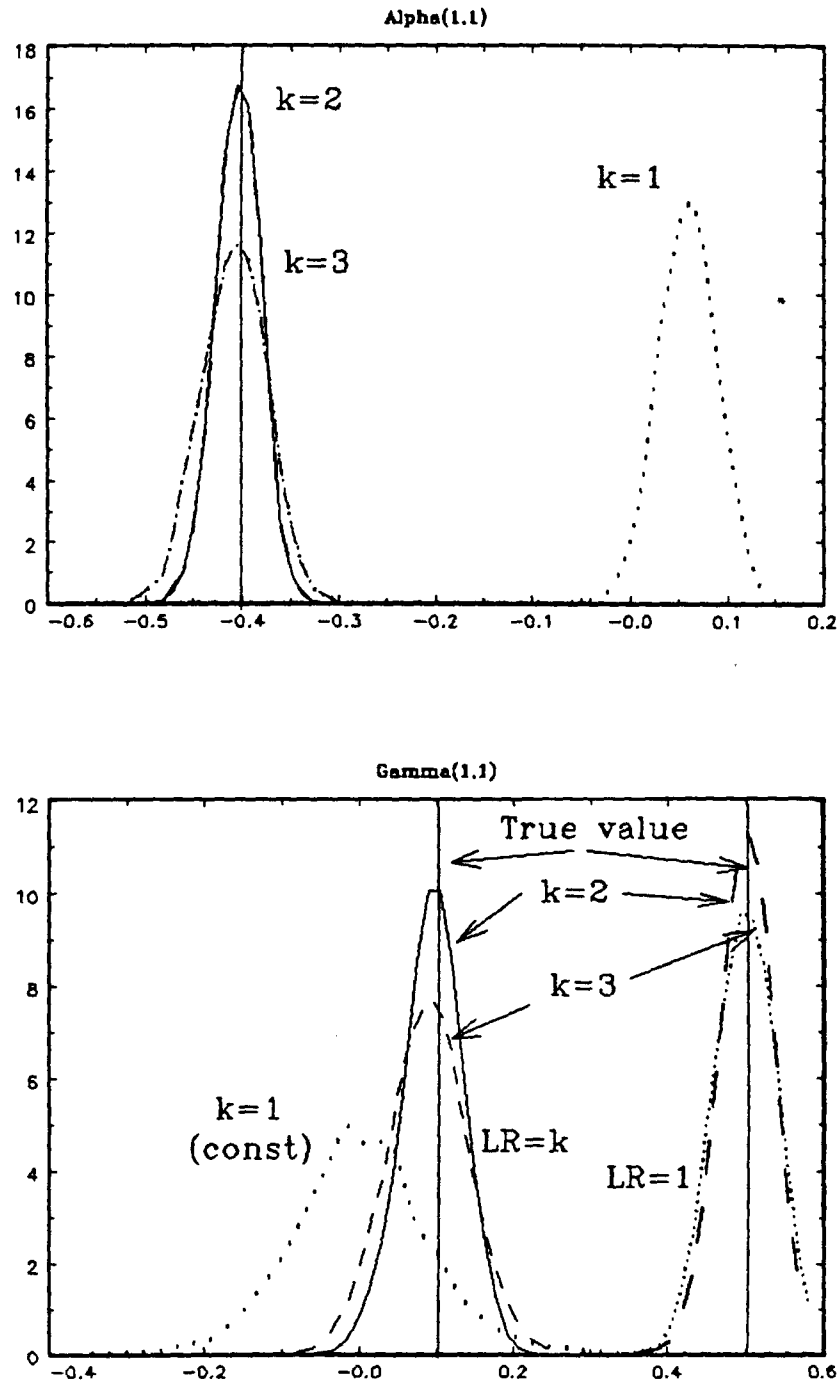

Figure 4.3: Estimated density functions for $\hat{\tilde{\alpha}}_{11}$ and $\hat{\tilde{\gamma}}_{11}$. Incorrectly specified VAR(k) models when the true value is $k = 2$. Level part of (2.2) on 1. or k. lag. $n = 2000$.

4.2 Data are filtered through systematic sampling or temporal aggregation

The DGP is given by (2.3) above, and give rise to stochastic sequences $\{x_t\}_{t=1}^T$. In this case we assume that we can only observe data from this process after the sequence have been filtered. Two cases are considered in the following: 1) We can observe every kth observation of the process and obtain a sequence of N observations given by $\{x_\tau\}_{\tau=1}^N = [x_{1k}, \ldots, x_{Nk}]$. 2) We can only observe sums (or averages) across k-period subintervals, i.e. $\{x_\tau\}_{\tau=1}^N = [\sum_{j=0}^{k-1} x_{1k-j}, \ldots, \sum_{j=0}^{k-1} x_{Nk-j}]$ or $\{x_\tau\}_{\tau=1}^N = [\bar{x_1}, \ldots, \bar{x_N}]$ where $\bar{x_\tau} = k^{-1} \sum_{j=0}^{k-1} x_{\tau k-j}$.

In the first case we say that the data are observed subject to *systematic sampling* and in the latter case they are *temporally aggregated.* Data which are systematically sampled are e.g. different types of *stock data*, state variables like interest rates, prices, asset holdings etc. Data which are temporally aggregated are typically *flow data* like e.g. production, consumption, export, import, which are accumulated per unit of time during a fixed time interval.

Similar situations have been analysed in the context of univariate time-series models, cf. e.g. Brewer(1973), Wei(1978) and Weiss(1984). They have shown that it is necessary, in order to analyse the properties of the observed process $\{x_\tau\}_{\tau=1}^N$, to transform the DGP by an appropriate filter.

We consider first the case with *systematic sampling* of the data. Granger's representation theorem for cointegrated variables is useful to obtain the following expression for $\{x_t\}_{t=1}^T$.

$$(4.1) \qquad \Delta x_t = C(L)\epsilon_t$$

for $t = 1, \ldots, T$, cf. e.g. Johansen(1989) for details. $C(L)$ is a matrix lagpolynomial and $C(1)$ has reduced rank. A problem with (4.1) is that the LHS variable Δx_t is unobservable. If we premultiply with the sum operator $S_k(L) = (1 - L^k)/(1 - L)$ and rearrange the expression, we can write

$$(4.2) \quad \begin{aligned} \Delta_k x_t &= C(L)S_k(L)\epsilon_t \\ &= C(1) \underbrace{S_k(L)\epsilon_t}_{\epsilon_t+\epsilon_{t-1}+\ldots+\epsilon_{t-k+1}} + C_1(L)\Delta_k\epsilon_t \end{aligned}$$

If we multiply in (2.2) (the interim multiplier representation) by $S_k(L)$ and rearrange, we obtain the expression

$$(4.3) \quad \Delta_k x_t = \Gamma(L)\Delta_k x_{t-1} + \alpha\beta' S_k(L)x_{t-k} + S_k(L)\epsilon_t$$

Note that while the LHS variable in (2.2) is unobservable in this case, (4.3) has a LHS variable for which each kth value is available. We see from (4.2) that $\Delta_k x_t$ can be expressed as a distributed lag in sums and differences of the residuals. The lagged values of $\Delta_k x_t$ on the right hand side of (4.3) will therefore be explained by previous values of the residuals and will in general be correlated with the composite residual $S_k(L)\epsilon_t$. We will also have *missing observations* since only each kth observation of x_t (and $\Delta_k x_t$) can be observed.

To illustrate the consequences for the estimation, consider the hypothetical case when β is known. The least squares estimators for $\Gamma(L)$ in (4.3) will then be inconsistent since $\lim_{t\to\infty} \sum \Delta_k x_{t-1}(S(L)\epsilon_t)' \neq 0$. Similar problems apply for the estimator for α since $\lim_{t\to\infty} \sum \hat{\beta}' S_k(L)x_{t-k}(S(L)\epsilon_t)' \neq 0$, and we expect asymptotic biases to show up in the Monte Carlo distributions.

Next, if the data are *temporally aggregated*, and we only observe sums across k-period subintervals, the Granger representation and summation filter can be applied to obtain the following expression:

$$(4.4) \quad \begin{aligned} \Delta_k S_k(L)x_t &= S_k(L)C(L)S_k(L)\epsilon_t \\ &= C(1)C(L) \cdot \underbrace{S_k(L)\epsilon_t}_{\epsilon_t+\epsilon_{t-1}+\ldots+\epsilon_{t-k+1}} + C_1(L)\Delta_k\epsilon_t \end{aligned}$$

If we manipulate the interim multiplier representation in (2.2) we obtain

$$(4.5) \quad \Delta S_k(L)x_t = \Gamma(L)S_k(L)x_{t-1} + \alpha\beta' \frac{S_k(L)}{1-L}x_{t-1} + \frac{S_k(L)}{1-L}\epsilon_t$$

In (4.5), we observe that the summation filter is insufficient to obtain an observable expression on the left hand side of the equation. By applying the filter twice however, (4.5) is transformed into an expression where each kth value of the LHS variable is observable.

$$(4.6) \qquad \Delta_k S_k(L)x_t = \Gamma(L)\Delta_k S_k(L)x_{t-1} + \alpha\beta' S_k(L)[S_k(L)x_{t-1}] + [S_k(L)]^2\epsilon_t$$

Again there will be missing observations on the right hand side of the equation, arising in the distributed lagpolynomials in kth differences of sums $\Delta_k S_k(L)x_{t-1}$. Taking probability limits, $\lim_{t\to\infty} \sum \Delta_k S(L)x_{t-1}([S(L)]^2\epsilon_t)' \neq 0$ and OLS will yield inconsistent estimates of the short run parameters $\Gamma(L)$ (and α).

It is less clear what the properties of $\hat{\beta}$ will be. Some evidence is provided below using Monte Carlo simulations.

Two sets of experiments are reported, one for systematic sampling (where we sample each kth observation of the DGP) and one for temporal aggregation (where we observe sums across k-period subintervals). In both cases we compare the results for $k = 1, 2, 3, 4, 5$ and we use a fixed sample size $N = 100$ in order to account for the previously documented effects from changes in the sample size.

The Trace test

Given the sample size $N = 100$ it is not unexpected to find relatively low power in the case without any aggregation. The results for this case are in line with the previously reported finite sample results, cf. table 3.3. It is perhaps more surprising that the power of the *Trace* test is so dramatically improved (for given sample size $N = 100$) when the width of the aggregation window k is increased. This result holds for both temporal filters used in the experiments. A number of interesting implications arises from this result. In particular we see that to guarantee proper inference on cointegration (in terms of the power of the test), we should not only be concerned with the number of observations but also that the observations cover a sufficiently wide range of time (and variability in the series). Or put it in a different way, we would prefer to have 100 annual observations rather than e.g. 100 weekly (or other highfrequent) observations, at least for the purpose of conducting inference about the cointegrating rank.

Table 4.2 The size and power for the Trace test for cointegration. Correctly specified VAR(2) model. Data are systematically sampled or temporally aggregated with $k \in [1, 2, 3, 4, 5]$, $N = 100$, $n = 2000$

	Size	Power	Power
T	$Pr[Reject\ H_0 : r \leq 3]$	$Pr[Reject\ H_0 : r \leq 2]$	$Pr[Reject\ H_0 : r \leq 1]$
Systematic sampling			
k=1	6.00	47.30	99.45
k=2	6.35	92.35	100
k=3	5.85	99.15	100
k=4	6.45	99.70	100
k=5	5.75	99.95	100
Temporal aggregation			
k=1	5.70	46.30	99.40
k=2	6.15	92.35	100
k=3	5.75	98.85	100
k=4	6.65	99.85	100
k=5	5.80	100	100

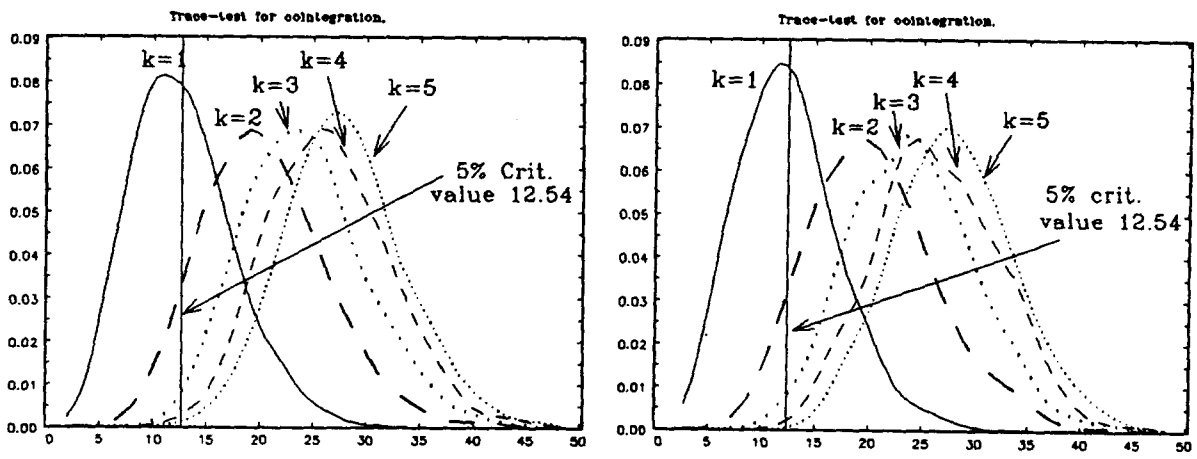

Figure 4.4: Estimated density function for the Trace test statistic. Correctly specified VAR(2) model. Systematic sampling (left), temporal aggregation (right). $k \in [1, 2, 3, 4, 5]$, $N = 100$, $n = 2000$

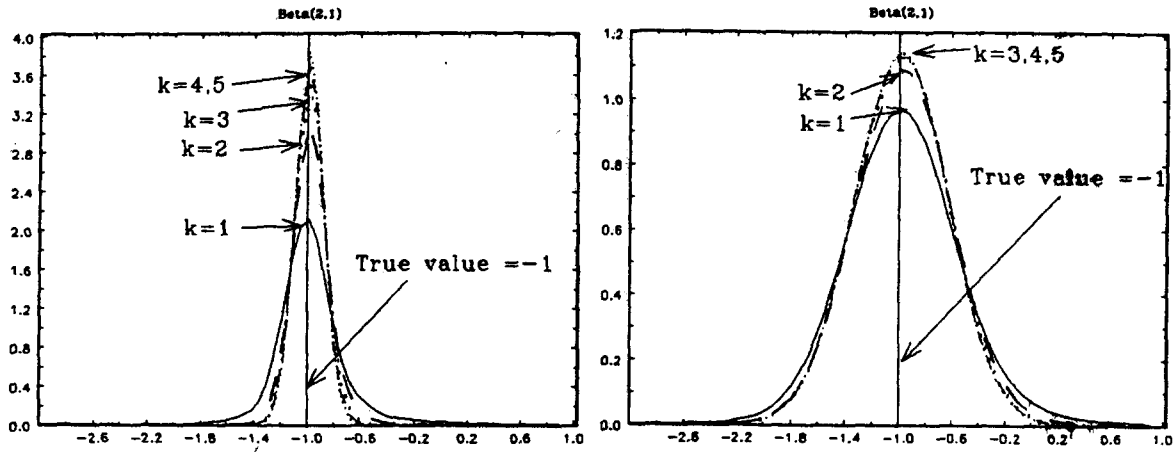

Figure 4.5: Estimated density functions for $\hat{\hat{\beta}}_{21}$. Correctly specified VAR(2) model. Systematic sampling (left), temporal aggregation (right). $k \in [1,2,3,4,5]$, $N = 100$, $n = 2000$

The estimators

The long run parameter estimator $\hat{\hat{\beta}}_{21}$ seem to be consistent throughout the experiments. The distributions are median unbiased and symmetrical around the correct parameter value, cf. figure 4.5. We also note that the efficiency, illustrated by the shape of the distributions, seem to be improved (although at a decreasing rate) as we increase the window width. We saw above how the two types of aggregation could contribute to improve the power of the *Trace* test and also improve the efficiency of the long run parameter estimators. The situation is very different for the short run parameter estimators, cf. figure 4.6. As shown above, the orthogonality between the regressors and (composite) residuals will be violated when $k > 1$. When $k = 1$ the estimators $\hat{\hat{\alpha}}_{11}$ and $\hat{\hat{\gamma}}_{11}$ are median unbiased, but the distributions shifts to the left when we increase k. We also note that the variance increases substantially so that the short run estimators become not only asymptotically biased but also less efficient under temporal aggregation.

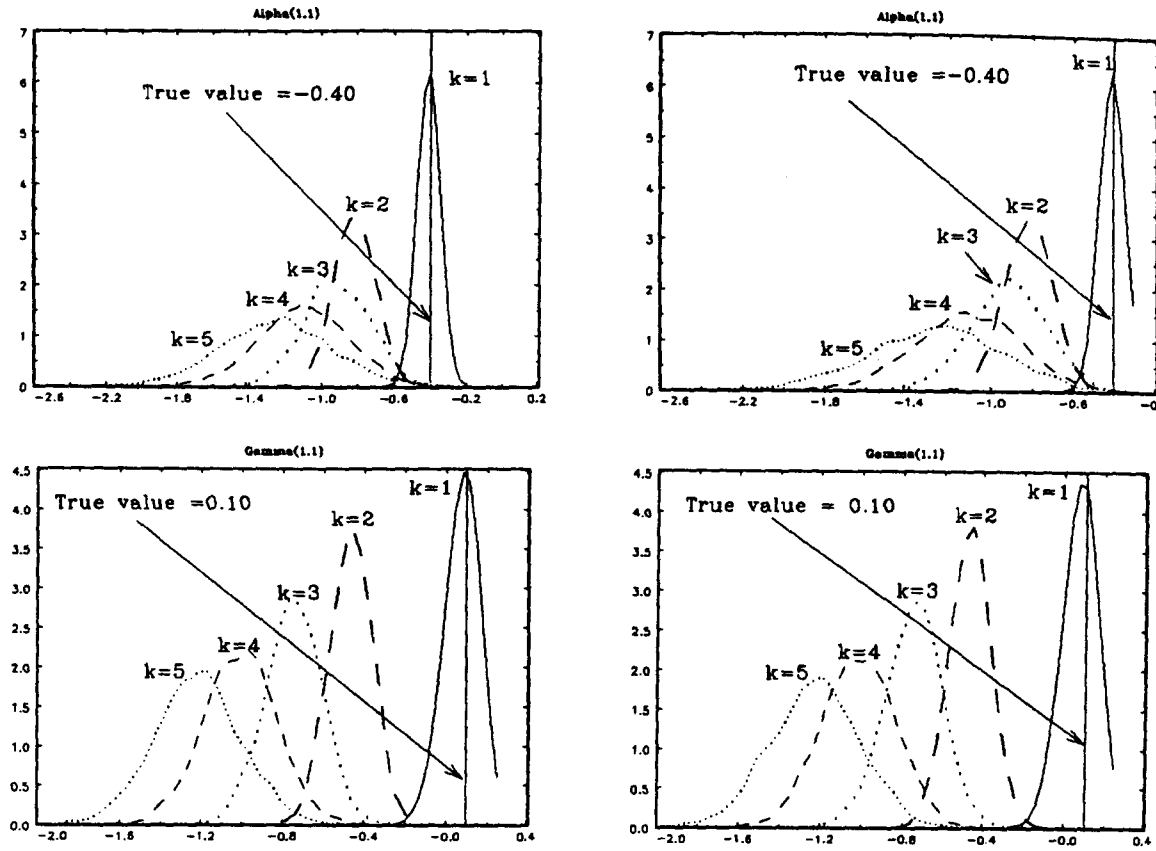

Figure 4.6: Estimated density functions for $\hat{\tilde{\alpha}}_{11}$ and $\hat{\tilde{\gamma}}_{11}$. Correct VAR(2) specification. Systematic sampling (left), temporal aggregation (right). $k \in [1,2,3,4,5]$, $N = 100$, $n = 2000$

4.3 DGP residuals are heterogeneous with weak to medium strong ARCH

In this case the model (2.2) is misspecified in a different sense. The residuals ϵ_t are assumed to be Niid in the model specification while their DGP happen to be a heterogeneous process with ARCH. The previous DGP-assumption $E[\epsilon_t\epsilon_t'|\mathcal{I}_{t-1}] = \Sigma$ is replaced by

$$(4.7) \quad \begin{aligned} E[\epsilon_t\epsilon_t'|\mathcal{I}_{t-1}] &= H_t \\ &= \Sigma + (I_p \otimes \epsilon_{t-1}')'(I_p \otimes cI_p)(\epsilon_{t-1} \otimes I_p) \end{aligned}$$

The first order ARCH residuals are basically driven by the scalars c which determines the contribution from lagged own squares of residuals on the the diagonal elements in H_t. If we replace cI_p in (4.7) by a matrix C_1, a more general (but still diagonal) ARCH process can be obtained which has effects from lagged cross squares and cross products of residuals as well. Here we report results from experiments with $c \in [0, 0.1, 0.2, 0.3, 0.5]$, which covers the range from no ARCH to medium strong ARCH in the residuals.

Table 4.3 The size and power for the Trace test for cointegration. Incorrectly specified VAR(2) model, true model has moderate to medium strong ARCH in the residuals. $c \in [0, 0.1, 0.2, 0.3, 0.5]$, $T = 400$, $n = 1350$

	Size	Power	Power
	$Pr[Reject\ H_0 :\ r \leq 3]$	$Pr[Reject\ H_0 :\ r \leq 2]$	$Pr[Reject\ H_0 :\ r \leq 1]$
c=0	5.14	100	100
c=0.1	4.87	100	100
c=0.2	5.14	99.91	100
c=0.3	5.40	99.91	100
c=0.5	5.58	99.91	100

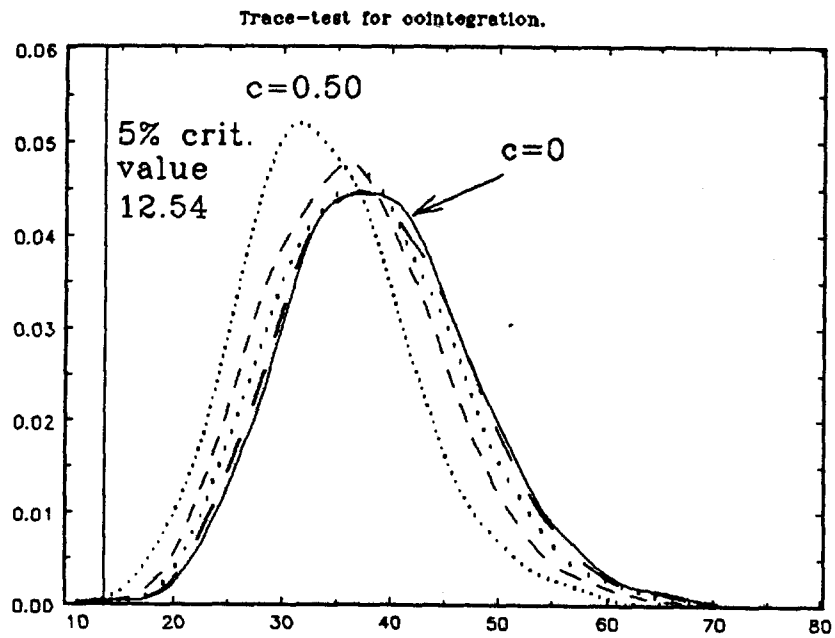

Figure 4.7: Estimated density function for the Trace test statistic. Incorrectly specified VAR(2) model, true model has moderate to medium strong ARCH. $T = 400$, $n = 1350$

The Trace test

The distributin of *Trace* is only slightly shifted to the left when we introduce ARCH into the residuals. The power of the test is good throughout the experiments, cf. figure 4.7 and table 4.3.

The estimators

The distribution of $\hat{\tilde{\beta}}_{21}$ is shown in figure 4.8. The shape is relatively unchanged when we gradually increase the ARCH parameter c in (4.7). A loss of efficiency from applying QMLE should be expected when c is increased and we see that slightly

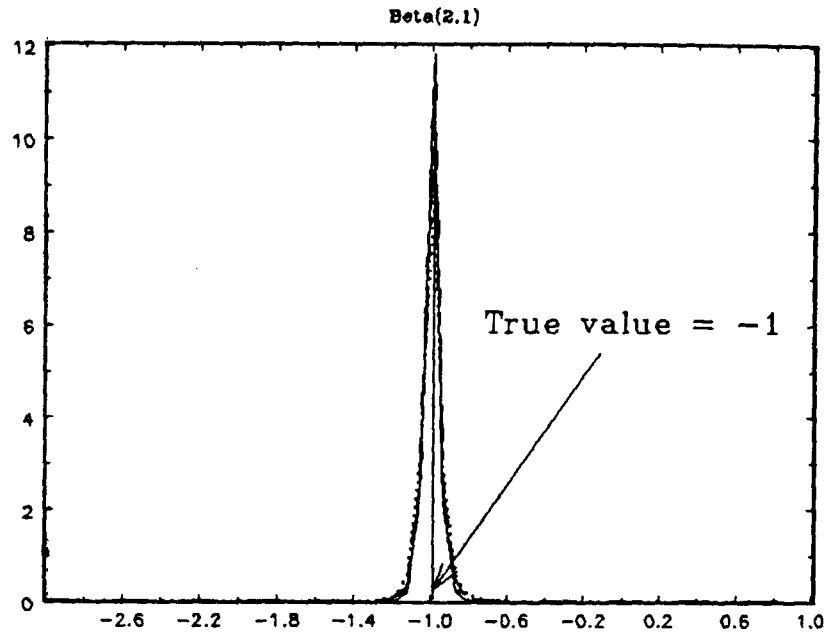

Figure 4.8: Estimated density functions for $\hat{\tilde{\beta}}_{21}$. Incorrectly specified VAR(2) model, true model has moderate to medium strong ARCH. $T = 400$, $n = 1350$

more of the probability mass ends up in the tails of the distribution in this case. Since $\{\epsilon_t, \mathcal{I}_t\}$ is a martingale difference sequence also when ϵ_t have ARCH such that $E[\epsilon_t|\mathcal{I}_{t-1}] = 0$, the OLS estimators for the short run parameters will still be consistent. We note however that the estimated distributions in figure 4.9, which are all estimated from a sample size $T = 400$, look very different. The solid line denote the distribution without ARCH and the asymmetry and bias in the other distributions seem to be monotonously increasing in c. Without attempting to dig very deep into small sample distribution theory for OLS estimators under different degree of ARCH, we ascribe this difference to the *effective sample size*. While $T = 400$ is sufficiently large to yield a good approximation to the asymptotics in the absence of ARCH it is clearly insufficient when the residuals are very heterogenously distributed. Hence, a larger sample seem to be necessary to obtain median unbiasedness in these cases.

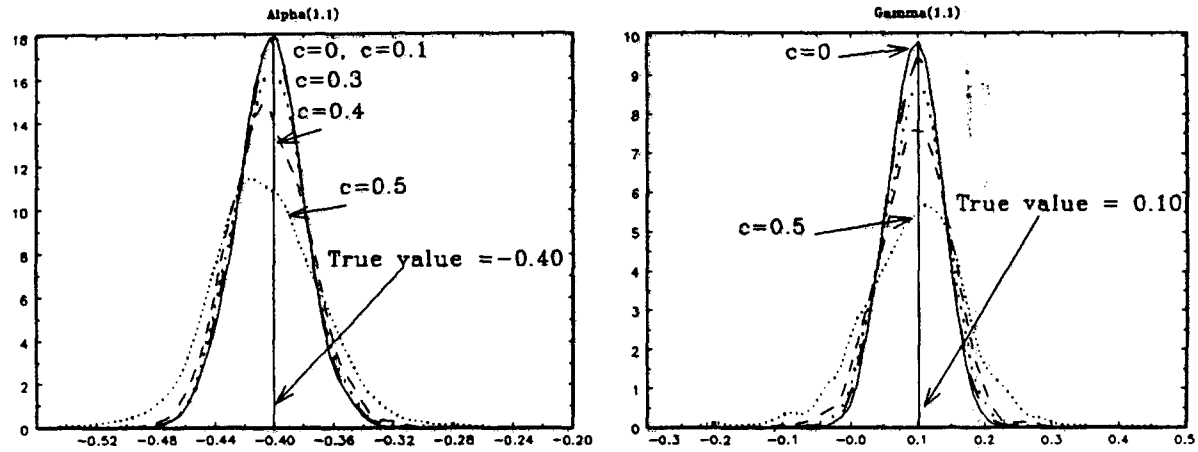

Figure 4.9: Estimated density functions for $\hat{\tilde{\alpha}}_{11}$ and $\hat{\tilde{\gamma}}_{11}$. Incorrectly specified VAR(2) model, true model has moderate to medium strong ARCH. $T = 400$, $n = 1350$

4.4 DGP residuals are serially correlated

In this case the DGP-residuals in (2.3) are serially correlated and generated by the following 1.order process:

$$\epsilon_t = c\epsilon_{t-1} + \omega_t \tag{4.8}$$

ω_t is Niid and the scalar c determines the strength in the autocorrelation. A more general autocorrelated process could have been obtained by replacing c with a $p \times p$ matrix R_1. Here we report results for $c \in [0, 0.1, 0.3, 0.6, 0.75, 0.90, 0.98, 0.99, 1.0]$ when the model specification is correct except that we pay no attention to the residual structure and proceed using the Q(uasi)ML approach.

The Trace test

The case with no serial correlation $c = 0$ is used as benchmark (the solid line in all figures). Figure 4.10 shows that the Trace test is robust with respects to weak to medium strong serial correlation in the DGP residuals. We are able to detect the correct cointegrating rank from the misspecified model (2.2). As shown in table 4.4 the power of the test deteriorates rapidly when c exceed 0.9. The abrupt fall in power when $c \to 1$ indicate a dramatic change in the model properties when ϵ_t

Table 4.4 The size and power for the Trace test for cointegration.
Misspecified VAR(2) model, true model has serially correlated residuals
$c \in [0, 0.1, 0.3, 0.6, 1.0]$, $T = 400$, $n = 1500$

	Size	Power	Power
T	$Pr[Reject\ H_0\ r \leq 3]$	$Pr[Reject\ H_0\ r \leq 2]$	$Pr[Reject\ H_0\ r \leq 1]$
c=0	6.25	100	100
c=0.1	6.25	100	100
c=0.3	6.05	100	100
c=0.6	5.80	100	100
c=0.75	5.80	100	100
c=0.90	5.41	99.21	100
c=0.98	3.69	29.66	80.75
c=0.99	3.36	15.36	61.70
c=1.0	2.37	9.82	52.34

become "near" $I(1)$. Table 4.4 shows that the cointegration test also has wrong size in this case.

The estimators

The distribution of the long run parameter $\hat{\tilde{\beta}}_{21}$ seem on the other hand to be robust with respect to changes in c. There is no significant dislocation, even in the limiting case when $c = 1$, and the distribution seem to be median unbiased. We note that slightly more probabilty mass fill the tails of the distribution for large c-values, indicating a loss of efficiency.

Given that all variables are "endogenous" in a VAR(k) model, we expect well known textbook problems to arise when we use OLS in models with serially correlated residuals and lagged endogenous regressors. The Monte Carlo results confirm this intuition. All short run parameters in experiments with non-zero c-values are clearly inconsistent, cf. figure 4.12. When we increase c, the dislocation of the distributions becomes more significant. Since the experiments are based on fairly large samples ($T = 400$) we should have good approximations to the asymptotic bias for the estimators.

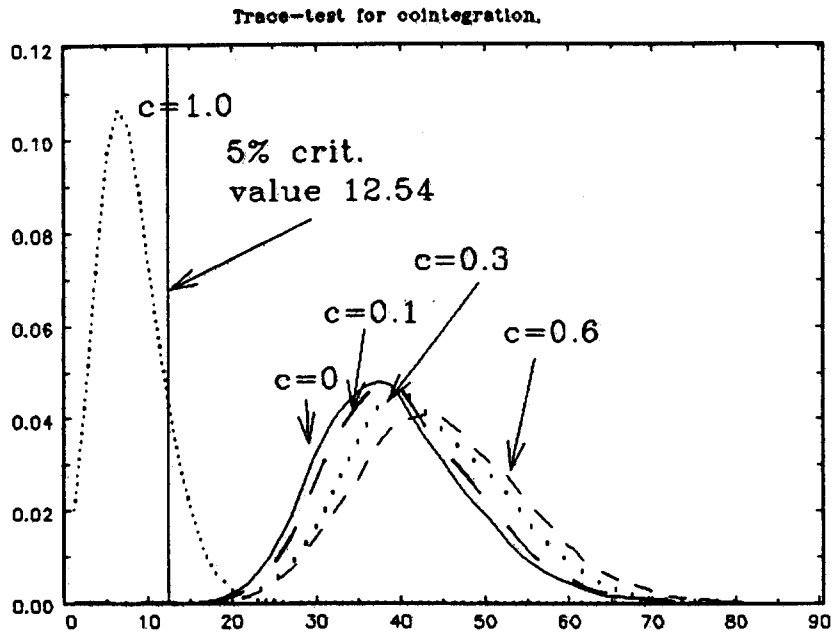

Figure 4.10: Estimated density function for the Trace test statistic. Misspecified VAR(2) model. True model has serially correlated residuals. $c \in [0, 0.1, 0.3, 0.6, 1.0]$, $T = 400$, $N = 1500$

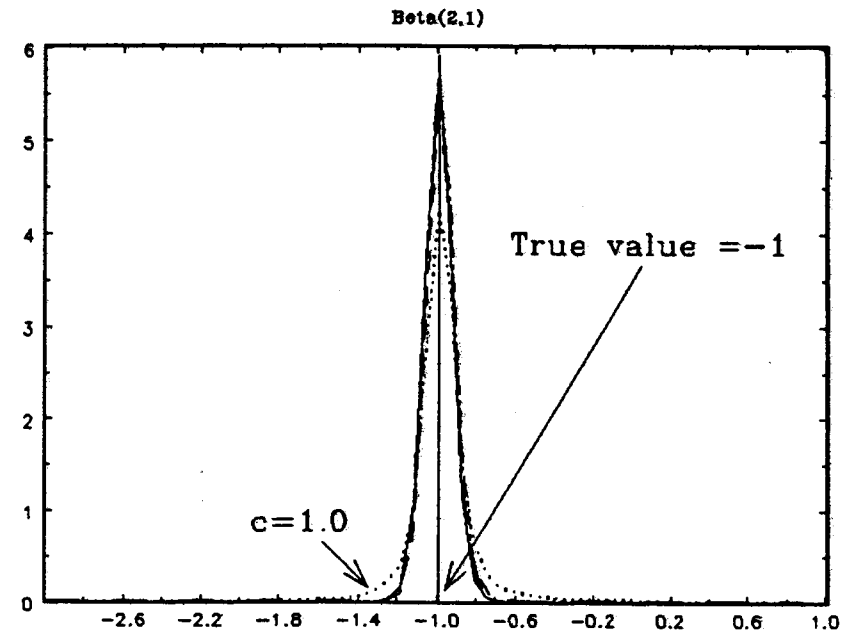

Figure 4.11: Estimated density functions for $\hat{\tilde{\beta}}_{21}$. Incorrectly specified VAR(2) model. True model has serially correlated residuals. $c \in [0, 0.1, 0.3, 0.6, 1.0]$, $T = 400$, $N = 2000$

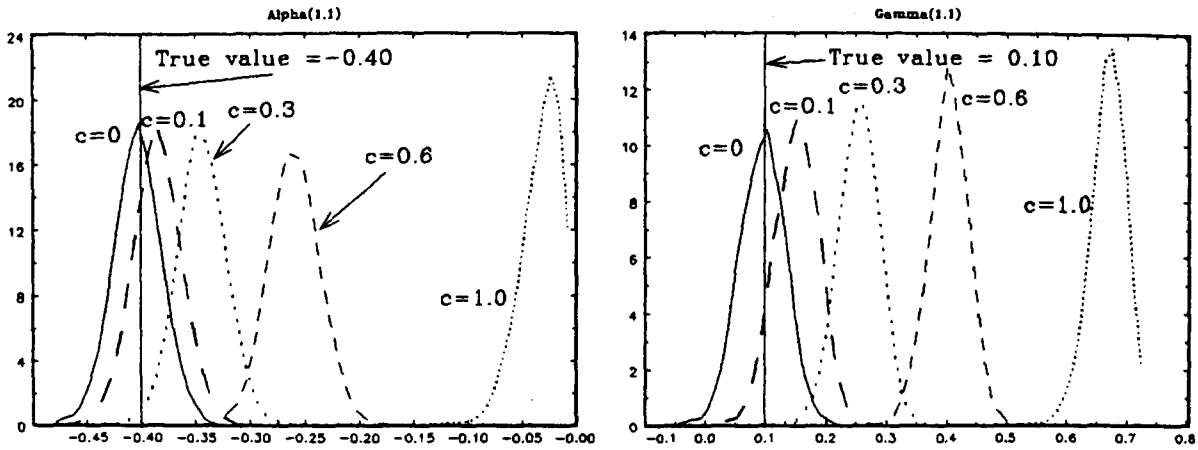

Figure 4.12: Estimated density functions for $\hat{\tilde{\alpha}}_{11}$ and $\hat{\tilde{\gamma}}_{11}$. Incorrectly specified VAR(2) model. True model has serially correlated residuals. $c \in [0, 0.1, 0.3, 0.6, 1.0]$, $T = 400$, $N = 2000$

4.5 Data are observed with measurement errors

We assume that x_t is generated by the small cointegrated system (2.3) but now we regard x_t as a latent vector process which is unobservable for the researcher. The data are instead given by a sequence $\{y_t\}_{t=1}^{T}$ which consist of two unobservable components, x_t (the signal process) and stochastic measurement errors η_t (noise). We have that

$$y_t = x_t + \eta_t \tag{4.9}$$

The measurement errors η_t are generated from

$$\eta_t = c\eta_{t-1} + \omega_t \tag{4.10}$$

The stochastic variables ω_t and ϵ_t (in (2.3)) are independent and Gaussian.

The two questions we have tried to answer with the Monte Carlo simulations are the following. 1) If the observed data consist of two unobservable components, one of which is a measurement error process like (4.10), how much persistence in memory (i.e. 1. order autocorrelation) can we allow for in the measurement error process η_t and still be able to detect the correct number of cointegrating relationships

among the latent x_t's? And 2) to what extent do the measurement errors affect the distributions of the estimated parameters?

A similar problem was studied by Fischer(1990) in the context of iid stationary measurement errors, using the DF- or CRDW-tests for cointegration. In our case, we have applied the *Trace* test for cointegration since the problem with a latent, cointegrated process observed with measurement errors can easily be analysed within the multivariate cointegration framework presented above (cf. e.g. Eitrheim(1991) for further details and Nowak(1991) for theoretical results).

The persistence in the measurement errors is determined by the parameter c in (4.12). When we increase c from 0 to 1, the measurement error process will gradually change from being iid white noise (with no memory) to a $I(1)$ non-stationary process with infinite memory. A sample size of 400 has been used throughout the Monte Carlo analysis, which allows for "large sample" comparisons.

The Trace-test

Results for five different values of $c \in [0, 0.5, 0.75, 0.90, 1.00]$ are compared in figure 4.13 (left). The distribution of *Trace* shifts to the left and jeopardize the power of the test when we increase c. This is also evident from the rejection frequencies in table 4.5. Interestingly, a substantial degree of persistence (high c-values) is necessary before the power of the cointegration test deteriorates. To obtain a more detailed picture of what happens at large c-values we have repeated the experiment with a finer grid for $c \in [0.75, 0.95, 0.98, 0.99, 1.00]$ (right).
We see that the third (and marginal) cointegrating vector becomes "hidden" in the data when $c \geq 0.98$. Nor are we able to detect the second relationship and there is evidence that even $H_0 : r = 0$ may not be rejected in the case when $c = 1$ (although this is less likely). On the other hand, the large shift in the distribution when c increases from 0.98 to 0.99 indicates that correct inference on the cointegrating rank r can be drawn even when there is a substantial degree of persistence in the measurement error process.

The estimators

The estimated distributions of $\hat{\beta}_{21}$ are shown in figure 4.14. The distribution of

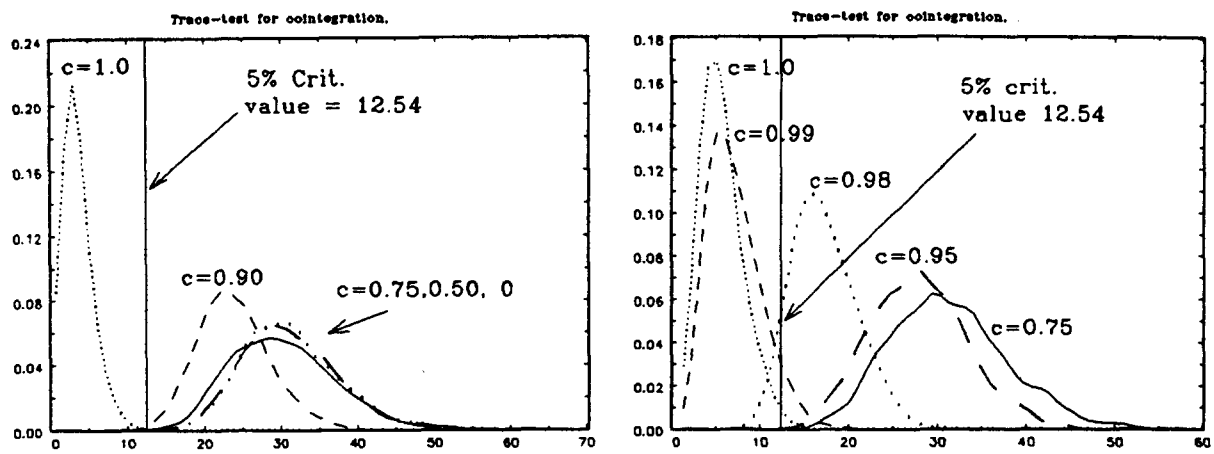

Figure 4.13: Estimated density functions for the Trace test statistic. $c \in [0.75, 0.95, 0.98, 0.99, 1.00]$ (left) and $c \in [0.75, 0.95, 0.98, 0.99, 1.00]$ (right) $T = 400$, $n = 1650$

Table 4.5 Empirical rejection frequencies for the Trace test for different values of the persistence parameter c in (4.12), T = 400 , n= 1650

c =	0	0.50	0.75	0.90	0.95	0.98	0.99	1.00
$Pr[Reject\ H_0 : r \leq 3]$	7.40	7.46	7.89	7.75	8.51	7.22	0.72	0.00
$Pr[Reject\ H_0 : r \leq 2]$	100	100	100	100	100	92.12	5.12	0.85
$Pr[Reject\ H_0 : r \leq 1]$	100	100	100	100	100	100	39.00	15.36
$Pr[Reject\ H_0 : r = 0]$	100	100	100	100	100	100	91.40	76.17

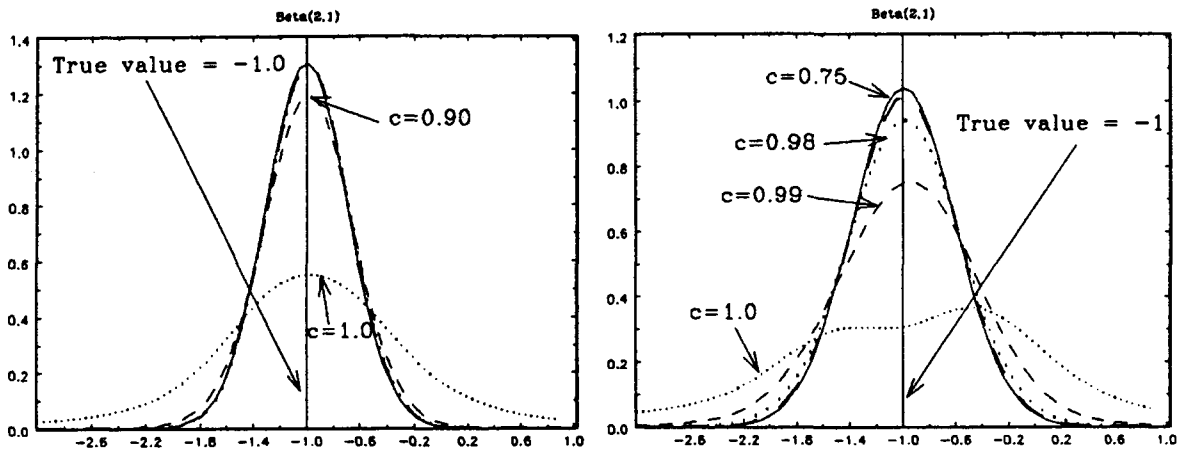

Figure 4.14: Estimated density functions for $\hat{\tilde{\beta}}_{21}$. $c \in [0, 0.5, 0.75, 0.90, 1.00]$ (left) and $c \in [0.75, 0.95, 0.98, 0.99, 1.00]$ (right) $T = 400$, $n = 1650$

$\hat{\tilde{\beta}}_{21}$ is median unbiased and symmetrical for and we note in particular that there seem to be no dislocation of the distribution even when there is a substantial degree of persistence in the measurement error process. In the limiting case $c = 1$, we have seen that the cointegration becomes hidden, and in this case we can also see a significant change in the shape of the distribution of $\hat{\tilde{\beta}}_{21}$.

The estimated density functions for the short run parameters are shown in figure 4.15. Both estimators are clearly inconsistent, and the degree of dislocation depend on the autoregressive parameter c. When we increase c, both distributions shift to the right. Again we can interpret this in light of standard textbook results. Consider the case with measurement errors in the explanatory variables in a standard regression model. As is well known, OLS will be inconsistent since the RHS variables (which includes the measurement errors) will be correlated with the composite residuals. The estimation bias will depend on nuisance parameters in the model like the signal to noise ratio. In our model, the ratio $var(x_t)/var(\eta_t)$ will depend on c and we would expect the estimation bias to occur for all parameters which converge at the ordinary rate $T^{-1/2}$ when these are estimated by OLS.

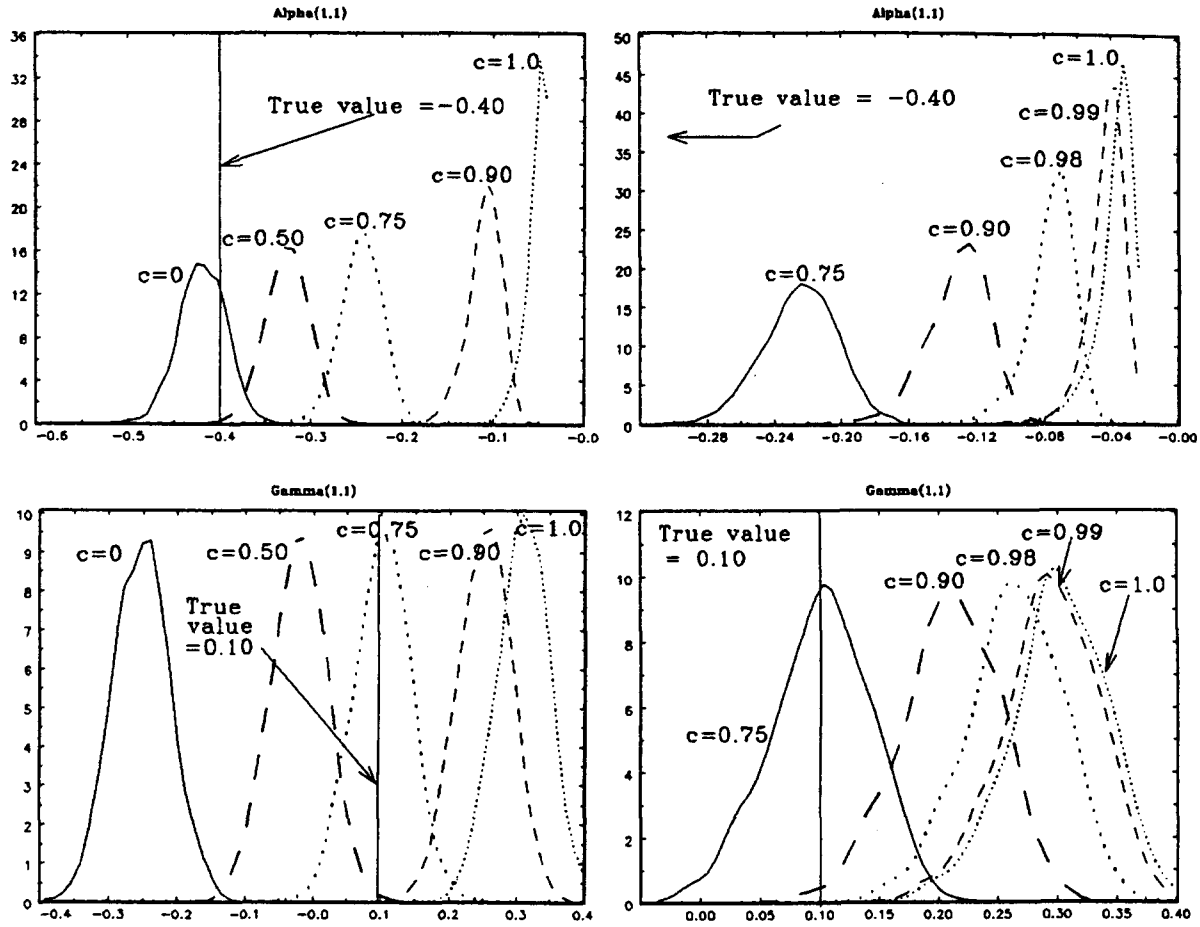

Figure 4.15: Estimated density functions for $\hat{\tilde{\alpha}}_{11}$ and $\hat{\tilde{\gamma}}_{11}$. $c \in [0, 0.5, 0.75, 0.90, 1.00]$ (left) and $c \in [0.75, 0.95, 0.98, 0.99, 1.00]$ (right) $T = 400$, $n = 1650$

5 Concluding remarks

The *Trace* test for cointegration has many interesting properties which are important to understand in order to interpret and evaluate empirical results. We have illustrated the test in different situations in order to learn more about its properties. As previous theoretical studies has shown, the test lacks power against "near cointegrated" alternatives and we demonstrate how the distribution of *Trace* depend on some key parameters in the DGP and also how the location of the test statistic depend on the sample size. The Monte Carlo results corroborate the theoretical analysis in Johansen(1991) with respect to the advocated testing sequence, starting out with $H_0 : r = 0$, $H_0 : r \leq 1$ and so on and stopping at the first non-rejection. Interestingly, both the "near cointegration" case and the small sample evidence indi-

cate that r may be underestimated in finite samples.

On the other hand, provided that we conduct correct inference about r, it turns out that the Monte Carlo distribution of the estimators of parameters in the long run cointegrating vectors $\hat{\tilde{\beta}}$ are surprisingly robust with respect to changes in sample size, changes in the key parameters in the DGP and the different types of model misspecification considered. The suggested normalization procedure simplifies the interpretation of the results, given relevant available prior information. Only in the case when we underspecified the order of the VAR (and imposed insufficient flexibility in the dynamic structure of the model) the renormalization failed to yield a Monte Carlo distribution which was median unbiased around the true value given by the DGP.

The short run parameters are consistently estimated by OLS in the prototype model and we note that the short run estimators seem to be more biased in small samples than the long run parameter estimators. This is also in accordance with theory since the two sets of estimators are expected to have different limiting distributions and only the long run parameter estimators are *super-consistent.* When the model is misspecified, many well known textbook problems apply for the distribution of the short run OLS-estimators, and we have provided examples where inconsistency arise from breakdown in the orthogonality condition because of simultaneity (temporal aggregation), contemporaneous presence of lagged endogenous variables and serially correlated residuals and measurement errors in the data.

Although the distribution of the long run parameter estimators seem to be robust against different types of model misspecification, alternative estimation methods may be required to estimate the short run parameters (including $\hat{\tilde{\alpha}}$). A large number of methods are available in the econometric literature, such as e.g. GIVE estimators, which could be considered in order to correct for these biases.

References

[1] T.W. Anderson(1951). Estimation linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, 22:327–351, 1951.

[2] T.W. Anderson(1991). Trygve Haavelmo and simultaneous equation models. *Scandinavian Journal of Statistics*, 18:1–19, 1991.

[3] J.Y. Campbell and P.Perron(1991). Pitfalls and opportunities: What macroeconomists should know about unit roots. 1991.

[4] M.P. Clements and G.E. Mizon(1991). Empirical analysis of macroeconomic time series. VAR and structural models. *European Economic Review*, 35:887–932, 1991.

[5] D.A. Dickey and W.A.Fuller(1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431, 1981.

[6] Ø. Eitrheim(1991). *Hidden cointegration. Some Monte Carlo evidence.* Discussion Paper 1991/10, Oslo: Bank of Norway.

[7] R.F. Engle and C.W.J.Granger(1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, 55:251–276, 1987.

[8] W.A. Fuller(1976). *Introduction to Statistical Time Series.* New York: John Wiley, 1976.

[9] J. Gonzalo(1989). *Comparison of five alternative methods of estimating long run equilibrium relationships.* Discussion paper 55, University of California, San Diego, 1989.

[10] C.W.J. Granger and A.A.Weiss(1983). Time series analysis of error correction models. In T.Amemiya Karlin, S. and L.A.Goodman, editors, *Studies in Econometric Time Series and Multivariate Statistics*, 1983.

[11] C.W.J. Granger(1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16:121–130, 1981.

[12] C.W.J. Granger(1983). *Cointegrated Variables and Error Correcting Models.* Discussion paper 13, University of California, San Diego, 1983.

[13] D.F. Hendry and G.E. Mizon(1991). Evaluating dynamic econometric models by encompassing the VAR. 1991. Unpublished manuscript.

[14] D.F. Hendry(1989). *PC-GIVE An Interactive Econometric Modelling System.* Institute of Economics and Statistics, University of Oxford, 1989.

[15] S. Hylleberg and G.E.Mizon(1988). Cointegration and error correction mechanisms. *Economic Journal*, 99:113–125, 1989.

[16] S. Johansen and K.Juselius(1990). Some structural hypotheses in a multivariate cointegration analysis of the purchasing power parity and the uncovered interest parity for UK. 1990. Institute of Mathematical Statistics, University of Copenhagen.

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

[18] S. Johansen(1989). *Likelihood based Inference on Cointegration. Theory and Applications.* Technical Report, 1989. Lecture notes for a course on cointegration held at the Seminario Estivo di Econometria, Centro Studi Sorelle Clarke, Bagni di Lucca, Italy in June 1989.

[19] S. Johansen(1990). Determination of cointegration rank in the presence of a linear trend. 1990. Institute of Mathematical Statistics, University of Copenhagen.

[20] S. Johansen(1992). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580, 1992.

[21] E. Nowak(1991). *Discovering hidden cointegration.* UCSD Discussion Paper 20, University of California, San Diego, 1991.

[22] M. Osterwald-Lenum(1990). Recalculated and extended tables of the asymptotic distribution of some important maximum likelihood cointegration test statistics. 1990. Unpublished paper from Institute of Economics, University of Copenhagen.

[23] S.G. Pantula(1989). Testing for unit roots in time series data. *Econometric Theory*, 5:256–271, 1989.

[24] P.C.B. Phillips and B.E.Hansen(1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies*, 57:99–125, 1990.

[25] P.C.B. Phillips and M. Loretan(1991). Estimating long run equilibria. *Review of Economic Studies*, 58:407–436, 1991.

[26] P.C.B. Phillips(1987). Time series regression with a unit root. *Econometrica*, 55:277–301, 1987.

[27] P.C.B. Phillips(1991). Optimal inference in cointegrated systems. *Econometrica*, 59(2):283–306, 1991.

[28] B.W. Silverman(1986). *Density Estimation for Statistical and Data Analysis.* London: Chapman and Hall, 1986.

[29] C.A. Sims, J.H. Stock and M.W.Watson. Inference in linear time series models with some unit roots. *Econometrica*, 58(1):113–144, 1990.

[30] J.H. Stock and M.W.Watson(1990). *A simple MLE of cointegrating vectors in higher order integrated systems.* Technical Working Paper 83, 1990.

[31] J.H. Stock(1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica*, 55(5):1035–1056, 1987.

[32] R.P. Velu, G.C.Reinsel, and D.W.Wichern(1986). Reduced rank models for multiple time series. *Biometrika*, 73:105–118, 1986.

Issued in the series ARBEIDSNOTATER from Bank of Norway 1989 - 1991

1989/1 Brodin, P. Anders

Makrokonsumfunksjonen i RIKMOD.
Utredningsavdelingen, 1989, 86 s.

1989/2 Gulbrandsen, Bjarne

Prising og risikodekning ved utstedelse av valutaopsjoner.
Utredningsavdelingen, 1989, 66 s.

1989/3 Grønvik, Gunnvald

Bankene som samfunnsinstitusjoner.
Utredningsavdelingen, 1989, 26 s.

1989/4 Naug, Bjørn og Ragnar Nymoen

Dynamisk modellering av norsk import.
Utredningsavdelingen, 1989, 54 s.

1989/5 Haraldsen, Jan Fr.

En rettslig analyse av norske valutareguleringsbestemmelser.
Finansmarkedsavdelingen, 1989, 79 s.

1989/6 Berg, Sigbjørn Atle, Finn R. Førsund og Eilev S. Jansen

Bank output measurement and the contruction of best practice frontiers.
Research Department, 1989, 28 p.

1989/7 Brodin, P. Anders and Ragnar Nymoen

The consumption function in Norway. Breakdown and reconstruction.
Research Department, 1989, 54 s.

1989/8 Vale, Bent

Impact of Central Bank Lending under Asymmetric Information in Credit Markets.
Research Department, 1989, 35 p.

1989/9 Mundaca, B. Gabriela

The Volatility of the Norwegian Currency Basket.
Research Department, 1989, 25 p.

1989/10 Mundaca, B. Gabriela

A GARCH-Switching Equations Model: The Effect of Official Interventions on the Exchange Rate Volatility.
Research Department, 1989, 32 p.

1989/11 Vikøren, Birger

En Empirisk Studie av foretakssektorens Netto Valutafordringer.
Utredningsavdelingen, 1989, 73 s.

1990/1 Grønvik, Gunnvald

Scale-effects and unintended dynamics arising from banking regulations.
Research Department, 1990, 34 s.

1990/2 Brodin, P. Anders, Eilev S. Jansen og Erik Nesset (red)

RIMINI. Teknisk dokumentasjon av en aggregert makro-økonometrisk modell.
Utredningsavdelingen, 1990, 115 s.

1990/3 Grønvik, Gunnvald, Bjarne Gulbrandsen, Jan Tore Larsen og Rønnaug Teige

Dataprosjektet FINDATR. Kvartalsvise finansielle sektorbalanser 1975 1 - 1989 3.
Statistikkavdelingen og Utredningsavdelingen, 1990, 51 s.

1990/4 Eitrheim, Øyvind

The demand for money in Norway. Econometric tests of the long run relationship.
Research Department, 1990, 48 s.

1990/5 Eitrheim, Øyvind

Testing long run relationships between economic time series using likelihood-based inference on cointegration.
Research Department, 1990, 39 s.

1990/6 Vale, Bent

Reserve requirements, asymmetric information, credit rationing and screening costs.
Research Department, 1990, 51 p.

1990/7 Ness, Torunn

Reguleringen av finansmarkedene i Norge. En studie av private kredittforetak og finansieringsselskaper.
Utredningsavdelingen, 1990, 67 s.

1990/8 Naug, Bjørn

Importvolum og importpriser.
Utredningsavdelingen, 1990, 109 s.

1990/9 Biblioteket

Indeks til Penger og Kreditt Nr.1/1973 - Nr.2/1990. Index to articles in Economic Bulletin No.1973/1 - 1990/2.
Dokumentasjonsseksjonen, 1990, 72 s.

1991/1 Brodin, P. Anders and Ragnar Nymoen

Wealth Effects and Exogeneity: The Norwegian Consumption Function 1966(1) - 1989(4).
Research Department, 1991, 49 p.

1991/2 Berg, Sigbjørn Atle, Finn R. Førsund and Eilev S. Jansen

Malmquist indices of productivity growth during the deregulation of Norwegian banking 1980-89.
Research Department, 1991, 30 p.

1991/3 Moene, Karl O. and Ragnar Nymoen

The dependency of labour demand on unemployment: Norwegian industry 1966(1)-1989(4).
Research Department, 1991, 24 p.

1991/4 Vale, Bent

An example of economies of scope in banking under asymmetric information.
Research Department, 1991, 32 p.

1991/5 Berg, Sigbjørn Atle and Moshe Kim

Oligopolistic Interdependence and the Structure of Production in Banking: an Empirical Evaluation.
Research Department, 1991, 29 p.

1991/6 Vikøren, Birger

Exchange Rate Expectations and Risk Premium: An Empirical Investigation.
Research Department, 1991, 30 p.

1991/7 Vikøren, Birger

The Saving-Investment correlation in the short and in the long run.
Research Department, 1991, 33 p.

1991/8 Grønvik, Gunnvald

The transition from regulated to free credit markets: The Norwegian case
Research Department, 1991, 48 p.

1991/9 Eitrheim, Øyvind

Inference in small cointegrated systems. Some Monte Carlo results. Research Department, 1991, 59 p.

KEYWORDS:

Multivariate analysis

Cointegration

Monte Carlo simulations