



**EUROPEAN UNIVERSITY INSTITUTE**  
**DEPARTMENT OF ECONOMICS**

EUI Working Paper **ECO** No. 2004 /20

**Practical Problems with Reduced Rank ML Estimators  
for Cointegration Parameters and a Simple Alternative**

RALF BRÜGGEMANN

and

HELMUT LÜTKEPOHL

BADIA FIESOLANA, SAN DOMENICO (FI)

All rights reserved.  
No part of this paper may be reproduced in any form  
Without permission of the author(s).

©2004 Ralf Brüggemann and Helmut Lütkepohl  
Published in Italy in May 2004  
European University Institute  
Badia Fiesolana  
I-50016 San Domenico (FI)  
Italy

# Practical Problems with Reduced Rank ML Estimators for Cointegration Parameters and a Simple Alternative

by

Ralf Brüggemann and Helmut Lütkepohl<sup>1</sup>

European University Institute, Florence and Humboldt University, Berlin

## Abstract

Johansen's reduced rank maximum likelihood (ML) estimator for cointegration parameters in vector error correction models is known to produce occasional extreme outliers. Using a small monetary system and German data we illustrate the practical importance of this problem. We also consider an alternative generalized least squares (GLS) system estimator which has better properties in this respect. The two estimators are compared in a small simulation study. It is found that the GLS estimator can indeed be an attractive alternative to ML estimation of cointegration parameters.

*Keywords:* Vector autoregressive process, vector error correction model, cointegration, reduced rank estimation, maximum likelihood estimation

*JEL classification:* C32

*Word count:* 4191

---

<sup>1</sup>Corresponding author: Helmut Lütkepohl, Department of Economics, European University Institute, Via della Piazzuola 43, I-50133 Firenze, ITALY, email: [helmut.luetkepohl@iue.it](mailto:helmut.luetkepohl@iue.it)

# 1 Introduction

In vector error correction models (VECMs), the Johansen (1995) reduced rank (RR) maximum likelihood (ML) approach is by now the dominant method for estimating the cointegration parameters. Reasons for its popularity are its sound theoretical basis, its computational simplicity and its superior performance relative to some other estimators. Stock & Watson (1993), Hargreaves (1994), Gonzalo (1994) and Caporale & Pittis (2004) for example, found a better small sample performance of the ML estimator (MLE) than of some other estimators. For instance, in Gonzalo's study a Monte Carlo comparison was conducted and the MLE was found to be the best choice at least when the median and the interquartile range are used as evaluation criteria. Gonzalo (1994) also mentions that the mean squared error may be a problematic criterion for estimator comparison in this case because finite sample moments of the estimator do not exist (see Phillips (1994)). There is, in fact, also some recent evidence, that the small sample properties of the ML estimator are not well approximated by its asymptotic distribution and, in particular, that the ML estimator produces occasional outliers which are far away from the true parameter values (see Phillips (1994, p. 74) for further references and Hansen, Kim & Mittnik (1998) for related results). Moreover, Gredenhoff & Jacobson (2001) investigated the small sample properties of the likelihood ratio (LR) test for restrictions on the cointegrating parameters and found that its asymptotic  $\chi^2$  distribution is a poor guide for small sample inference.

In this study we will consider a small German monetary system and demonstrate that the problem is not only an academic one but can arise in a standard empirical setting. In other words, if the MLE is used in applied work and implausible cointegration parameters are found, this can well be a reflection of the potentially poor small sample performance of this estimator. It is not necessarily a consequence of a deficient model setup.

We will compare the MLE to a simple feasible two-step generalized least squares (GLS) estimator which was considered by Ahn & Reinsel (1990) and Saikkonen (1992) (see also Reinsel (1993, Chapter 6) and Phillips (1994) for a closely related estimator) and we show that it does not produce the kind of outlying estimates observed for the MLE. It has not attracted much attention from applied researchers in the past possibly because it is not implemented in some standard econometric software packages. We will show, however, that in some respects it has substantially better small sample properties than the MLE.

The paper is structured as follows. In the next section the model setup is presented and the estimators are briefly introduced in Section 3. A small quarterly German monetary

system is considered in Section 4 which illustrates the undesirable properties of the MLE. A Monte Carlo comparison of the MLE and the GLS estimator is presented in Section 5 and conclusions are drawn in Section 6.

## 2 The Model

It is assumed that the  $K$ -dimensional process  $y_t$  has the VECM representation

$$\begin{aligned}\Delta y_t &= \alpha(\beta' y_{t-1} + \delta^{co'} d_{t-1}^{co}) + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + C d_t^s + u_t \\ &= \alpha \beta^{*'} y_{t-1}^* + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + C d_t^s + u_t,\end{aligned}\tag{2.1}$$

where  $\alpha$  and  $\beta$  are  $(K \times r)$  matrices of rank  $r$  associated with the long-run part of the model,  $\Gamma_i$  ( $i = 1, \dots, p-1$ ) are  $(K \times K)$  coefficient matrices associated with the short-run part. Moreover,  $d_t^{co}$  is a vector of deterministic variables which are included in the cointegration relations. The corresponding coefficient matrix is denoted by  $\delta^{co'}$ . The vector  $d_t^s$  includes the remaining deterministic variables with coefficient matrix  $C$ . The matrix  $\beta^{*'} = [\beta' : \delta^{co'}]$  is  $(r \times K^*)$  and  $y_{t-1}^* = [y'_{t-1}, d_{t-1}^{co}]$  is  $(K^* \times 1)$  with  $K^* = K + \text{dimension}(d_t^{co})$ . The error term  $u_t$  consists of a white noise process with zero mean and nonsingular covariance matrix  $\Sigma_u$ . All variables are assumed to be integrated of at most order 1.

We are interested in estimating the parameters  $\beta^*$ . Because the matrix is not unique we will impose just-identifying restrictions such that  $\beta^{*'} = [I_r : \beta_{(K^*-r)}^{*'}]$ , that is, the first  $r$  rows of  $\beta^*$  constitute an  $(r \times r)$  identity matrix. This form of the cointegration matrix can always be obtained by a suitable ordering of the variables. In practice, if the cointegration properties of all variables and subsets of variables are known, it is not difficult to set up the system properly. In case of doubt, there is also a possibility to perform a statistical test of the right ordering of the variables (Luukkonen, Ripatti & Saikkonen (1999)).

## 3 The Estimators

### 3.1 Reduced Rank ML Estimation

For a given cointegrating rank  $r$  and lag order  $p$ , a sample with  $T$  observations and  $p$  presample values, ML estimation of the VECM (2.1) can be done by RR regression, as shown in Johansen (1995). The estimator may be determined by denoting the residuals from regressing  $\Delta y_t$  and  $y_{t-1}^*$  on  $\Delta Y'_{t-1} = [\Delta y'_{t-1}, \dots, \Delta y'_{t-p+1}, d_t^{s'}]$  by  $R_{0t}$  and  $R_{1t}$ , respectively,

defining  $S_{ij} = T^{-1} \sum_{t=1}^T R_{it} R'_{jt}$  ( $i, j = 0, 1$ ), and solving the generalized eigenvalue problem

$$\det(\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}) = 0. \quad (3.1)$$

Let the ordered eigenvalues be  $\lambda_1 \geq \dots \geq \lambda_{K^*}$  with corresponding matrix of eigenvectors  $B = [b_1, \dots, b_{K^*}]$  satisfying  $\lambda_i S_{11} b_i = S'_{01} S_{00}^{-1} S_{01} b_i$  and normalized such that  $B' S_{11} B = I_{K^*}$ . An estimator of  $\beta^*$  is given by  $\tilde{\beta}^* = [b_1, \dots, b_r]$ . Post-multiplying by the inverse of the first  $r$  rows of  $\tilde{\beta}^*$  gives an MLE

$$\tilde{\beta}'_{ML} = [I_r : \tilde{\beta}'_{(K^*-r)}]. \quad (3.2)$$

Its asymptotic properties can be recovered from the convergence result (e.g., Johansen (1995), Reinsel (1993, Chapter 6))

$$\begin{aligned} & \text{vec} \left\{ (\tilde{\beta}'_{(K^*-r)} - \beta'_{(K^*-r)}) \left( R_1^{(2)} R_1^{(2)'} \right)^{1/2} \right\} \\ &= \left[ \left( R_1^{(2)} R_1^{(2)'} \right)^{1/2} \otimes I_{K^*-r} \right] \text{vec}(\tilde{\beta}'_{(K^*-r)} - \beta'_{(K^*-r)}) \xrightarrow{d} N(0, I_{K^*-r} \otimes (\alpha' \Sigma_u^{-1} \alpha)^{-1}), \end{aligned} \quad (3.3)$$

where  $R_1^{(2)} = [R_{11}^{(2)}, \dots, R_{1T}^{(2)}]$  and  $R_1^{(2)}$  contains the last  $K^* - r$  components of  $R_{1t}$ . This result implies that LR or Wald tests for restrictions on the parameters  $\beta^*_{(K^*-r)}$  have standard asymptotic  $\chi^2$  distributions under the null hypothesis.

### 3.2 A Feasible GLS Estimator

Ahn & Reinsel (1990) and Saikkonen (1992) proposed another estimator for the cointegration parameters which can be viewed as a feasible GLS estimator. To focus on the cointegration parameters, we consider the concentrated model corresponding to the VECM (2.1),

$$R_{0t} = \alpha \beta^{*'} R_{1t} + \tilde{u}_t. \quad (3.4)$$

Using the normalization  $\beta^{*'} = [I_r : \beta^{*'}_{(K^*-r)}]$ , this model can be written in the form

$$R_{0t} - \alpha R_{1t}^{(1)} = \alpha \beta^{*'}_{(K^*-r)} R_{1t}^{(2)} + \tilde{u}_t = (R_{1t}^{(2)'} \otimes \alpha) \text{vec}(\beta^{*'}_{(K^*-r)}) + \tilde{u}_t, \quad (3.5)$$

where  $R_{1t}^{(1)}$  and  $R_{1t}^{(2)}$  consist of the first  $r$  and last  $K^* - r$  components of  $R_{1t}$ , respectively. For a given  $\alpha$ , the GLS estimator of  $\text{vec}(\beta^{*'}_{(K^*-r)})$  is

$$\text{vec}(\check{\beta}'_{(K^*-r)}) = \left[ \left( \sum_t R_{1t}^{(2)} R_{1t}^{(2)'} \right)^{-1} \otimes (\alpha' \Sigma_u^{-1} \alpha)^{-1} \right] \sum_t (R_{1t}^{(2)'} \otimes \alpha' \Sigma_u^{-1}) (R_{0t} - \alpha R_{1t}^{(1)})$$

and, thus,

$$\check{\beta}_{(K^*-r)}^* = (\alpha' \Sigma_u^{-1} \alpha)^{-1} \alpha' \Sigma_u^{-1} \left( \sum_t (R_{0t} - \alpha R_{1t}^{(1)}) R_{1t}^{(2)'} \right) \left( \sum_t R_{1t}^{(2)} R_{1t}^{(2)'} \right)^{-1}. \quad (3.6)$$

Hence,  $\check{\beta}_{(K^*-r)}^*$  can be estimated by a two-step procedure. In the first step, the parameters in the model  $R_{0t} = \Pi R_{1t} + \tilde{u}_t$  are estimated by ordinary least squares (OLS). We denote the OLS estimator of  $\Pi$  by  $\hat{\Pi}$ . The first  $r$  columns of  $\Pi$  are equal to  $\alpha$  and hence these columns from the estimator  $\hat{\Pi}$  are used as an estimator  $\hat{\alpha}$  of  $\alpha$ . This estimator and the usual residual covariance estimator are then used in (3.6) to obtain a feasible GLS estimator which will be denoted by  $\check{\beta}_{(K^*-r)GLS}^*$  and

$$\check{\beta}_{GLS}^* = [I_r : \check{\beta}_{(K^*-r)GLS}^*].$$

This two-step estimator has the same asymptotic distribution as the MLE (see Ahn & Reinsel (1990), Reinsel (1993, Chapter 6), Saikkonen (1992)). The small sample properties of the two estimators are quite different, however, as we will see in the next two sections.

There are also other system methods for estimating the cointegrating parameters. For example, Phillips (1991) considered nonparametric estimation of the short-run parameters, Stock & Watson (1988) discussed an estimator based on principal components, Bossaerts (1988) used canonical correlations and Stock & Watson (1993) proposed dynamic OLS and GLS estimators of the cointegration parameters. The latter one of these estimators uses leads and lags of differenced variables to account for short-run dynamics. Stock and Watson's principal components and Bossaerts' canonical correlations estimators were shown to be inferior to Johansen's ML method in the comparison by Gonzalo (1994). Moreover, Stock & Watson (1993) found in their simulation study that the dynamic OLS estimator tends to be inferior to the dynamic GLS estimator which in turn was inferior in some respects to Johansen's ML estimator. The dynamic GLS estimator also performed quite poorly in the simple bivariate setup of Caporale & Pittis (2004).<sup>2</sup> In the latter study it is also found that the MLE outperforms many of the standard single equation estimators. Therefore we do not consider other estimators here but focus on a comparison between the MLE and the feasible GLS estimators proposed in the foregoing. Nonparametric estimation of the short-run parameters is not treated because we are interested in the properties of the MLE under ideal conditions, which includes knowledge of the autoregressive order and, hence, using a

---

<sup>2</sup>Note that this estimator corresponds to the dynamic OLS estimator of Caporale & Pittis (2004) in their single equation context.

parametric model is more plausible here. In the next section we consider an example model which illustrates some important properties of the ML and GLS estimators.

## 4 A German Monetary System

We have estimated a small textbook-style monetary system for quarterly German data over the period 1975Q1-1998Q4 using both the ML and GLS estimators. Our model includes the log of real money M3,  $m$ , the log of real gross national product,  $gnp$ , and an average bond rate as a long-term interest rate,  $R$ .<sup>3</sup> A detailed description of the data and a preliminary analysis including time series plots as well as unit root and cointegration tests are given in Lütkepohl (2004). All series in  $y_t = (m_t, gnp_t, R_t)'$  appear to be I(1). Cointegration tests point to one cointegration relation ( $r = 1$ ). Using a VECM with  $r = 1$ , four lags of  $\Delta y_t$ , seasonal dummies, an impulse dummy for modelling the effects of German reunification, a constant and a deterministic trend, which is restricted to lie in the cointegration space, the MLE of  $\beta^{*}$  is given by

$$\tilde{\beta}_{ML}^{*'} = (1, -0.863, 3.781, -0.002),$$

(0.228)      (0.801)      (0.002)

whereas using the GLS estimator yields

$$\check{\beta}_{GLS}^{*'} = (1, -1.003, 1.819, -0.002).$$

(0.148)      (0.519)      (0.001)

Here estimated standard errors are given in parentheses. Both estimates allow to interpret the cointegration vector as a money demand function. For instance, using the ML results, the estimated money demand equation can be written as

$$m_t = 0.863 gnp_t - 3.781 R_t + 0.002 t + ec_t.$$

(0.228)                      (0.801)                      (0.002)

Thus, the income elasticity is not far from the theoretically plausible value of 1 and the interest rate coefficient is negative. Hence,  $R_t$  can be interpreted as an opportunity cost variable.

Although there are some differences for the point estimates, we get a more informative picture by looking at the implied confidence intervals. For this purpose, we have computed

---

<sup>3</sup>The data are available at <http://www.jmulti.de/download/datasets/GermanMonetarySystem.dat>.



asymptotic 95% confidence intervals around the estimates of  $\beta_{(K^*-r)}^*$ . Moreover, we provide results for different sample periods. To be more precise, we truncated  $t^* = 1, \dots, 50$  observations at the beginning of the sample and for each  $t^*$  we computed the confidence intervals for the ML and GLS estimators. A graphical representation of these intervals is given in Figure 1. For most sample periods the two estimators lead to overlapping, similar confidence intervals. Note, however, that the reduced rank ML estimator produces rather extreme estimates and intervals for some of the samples considered. For example, using data from 1979Q2-1998Q4 ( $T = 74$ ) the ML and GLS estimators lead to completely different results:

$$\tilde{\beta}_{ML}^* = (1, -14.04, 64.91, 0.125),$$

(3.54)      (13.16)      (0.032)

and

$$\check{\beta}_{GLS}^* = (1, -0.563, 0.213, -0.006).$$

(0.248)      (0.922)      (0.002)

Clearly, the ML income elasticity estimate of 14.04 does not make sense anymore. Although the GLS estimate of the income elasticity also has changed markedly relative to the estimate for the full sample period, the 95% GLS interval for the negative income elasticity still includes the theoretically plausible value of  $-1$ . The same is obviously not true for the ML interval. In Figure 1 this is reflected by a large ‘spike’. The behavior of the MLE seems to be governed by a few observations in the sample. Adding or truncating a few observations shows that ML and GLS lead again to similar point and interval estimates.

Our example illustrates the theoretical property of the MLE to produce occasional outliers in finite samples (see Phillips (1994)). Clearly, it is somewhat worrying that estimating the model by the MLE gives a completely misleading picture of the cointegration relation and may, in fact, lead to the conclusion that the model is useless. Notice, however, that a set of standard diagnostic tests for the VECM residuals does not point to misspecification of this model. Thus, a statistically sound model is obtained with estimated parameter values which are totally implausible from an economic point of view. Hence, using the ML estimator may be quite misleading in practice. In our example, the simple GLS estimator proves to be a more robust alternative as it does not produce the outliers of the MLE. Of course, this result for our example may be a very special case for a specific data set. To show that it has more general significance we present an illustrative Monte Carlo comparison of the two estimators in the next section.

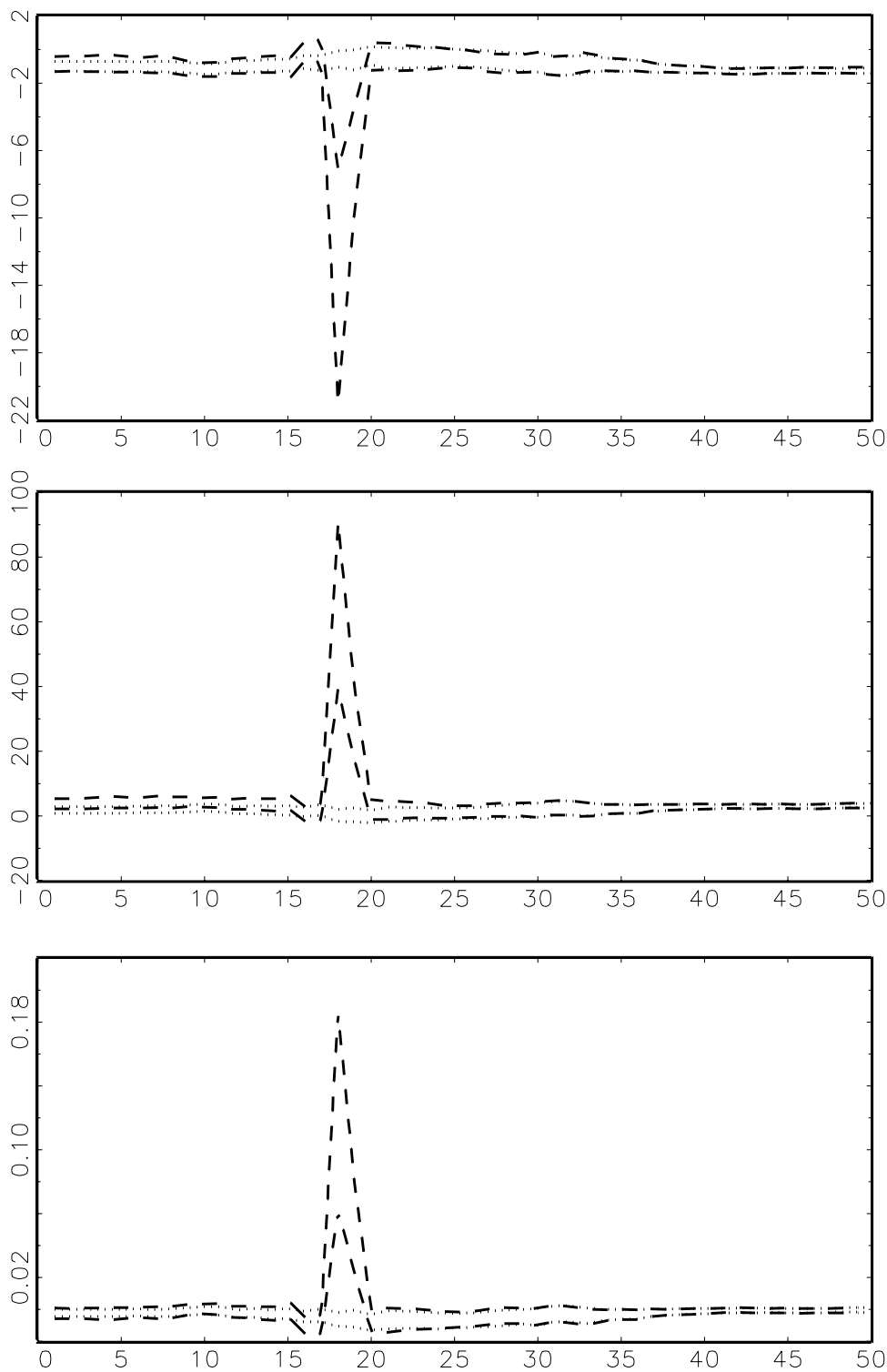


Figure 1: Confidence intervals based on reduced rank ML (- - -) and GLS ( $\cdot\cdot\cdot$ ) estimator for German monetary system. The horizontal axis shows number of truncated observations at the beginning of the sample (relative to 1975Q1). Panels are for  $\beta_{21}^*$  (top),  $\beta_{31}^*$  (middle), and  $\beta_{41}^*$  (bottom).

## 5 Monte Carlo Comparison

We present the Monte Carlo design in Section 5.1 and the evaluation criteria that form the basis for our comparison are discussed in Section 5.2. The main results are summarized in Section 5.3.

### 5.1 Monte Carlo Design

Our Monte Carlo experiments are designed to compare different aspects of estimation precision of the two systems estimators. We base the simulations on data generating processes (DGPs) obtained from empirical models on the one hand and on DGPs that have been previously used in comparative studies on the other hand. Using empirical model specifications leads to DGP characteristics similar to those encountered in applied time series analysis with a typical number of variables, number of lags and number of cointegration vectors. The results for these DGPs are supplemented by simulations for a simple bivariate DGP used by Gonzalo (1994) which, in principle, allows to vary some key characteristics. The main features of the DGPs are summarized in Table 1.

DGP (A) is based on our empirical example from Section 4. For our simulation the cointegration vector  $\beta^*$  is chosen to closely resemble the ML estimates for  $\beta^*$  based on the full sample period 1975Q1-1998Q4, when the coefficient for  $m$  is normalized to one and a unit income elasticity is imposed. The remaining DGP parameters (including a constant and seasonal dummies) are then obtained by estimating a VECM with four lags of  $\Delta y_t$  for given  $\beta$ . In DGP (A) the coefficient of  $m$  is normalized to 1. Consequently, we base the comparison on the remaining parameters, which have the true values  $\beta_{21}^* = -1$ ,  $\beta_{31}^* = 4$  and  $\beta_{41}^* = -0.0015$ . Using these parameters and the first 5 data observations as presample values, sets of time series are generated such that  $T = 91$  observations are available for estimation (as in the original study).

DGP (B) in Table 1 is based on the empirical model for the log of consumption  $c$ , the log of investment  $i$  and the log of private output  $y$  specified by King et al. (1991). The balanced-growth conditions imply two cointegration relations in the three-dimensional system. The cointegration analysis in King et al. (1991) shows that estimates for  $\beta_{31}$  and  $\beta_{32}$  are very close to  $-1$ . Therefore, we have imposed  $\beta_{31} = \beta_{32} = -1$  for the simulations. The remaining parameters are obtained by estimating a VECM with five lags of  $\Delta y_t$  and an unrestricted constant. Sets of Monte Carlo time series are generated using the observed presample values as starting values. In addition to the original number of observations ( $T = 160$ ), we also use

Table 1: DGP Properties

	$K$	$r$	$p$		Reference
DGP (A)	3	1	5	$y_t^{*'} = (m_t, gnpt, R_t, t)$ $\beta^{*'} = (1, -1, 4, -.0015)$	Lütkepohl (2004)
DGP (B)	3	2	6	$y_t' = (c_t, i_t, y_t), \quad \beta' = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$	King et al. (1991)
DGP (C)	2	1	1	$y_t' = (y_{1t}, y_{2t}), \quad \beta' = \begin{pmatrix} 1 & -1 \end{pmatrix}$	Gonzalo (1994)

*Note:* The remaining VECM parameters for DGP (A) and (B) are obtained by estimating a VECM with  $p - 1$  lags of  $\Delta y_t$  for given  $\beta^*$  and  $\beta$  using the data provided by Lütkepohl (2004) and King, Plosser, Stock & Watson (1991). The remaining parameters for DGP (C) are given in Table 4 in Gonzalo (1994).

time series of length  $T = 80$  to investigate the small sample properties of the estimators.

The results for the empirical DGPs are supplemented by using the simple bivariate DGP (C) that has previously been used by Gonzalo (1994) to compare different estimators for cointegration relations. The comparison will be based on the parameter  $\beta_{21}$ , which has a true value of  $-1$ . We generate sets of time series, such that  $T = 30, 50, 100, 300$  observations are available for estimation.

For each design point in our Monte Carlo experiment we have generated  $M = 1000$  sets of time series and applied the ML and GLS estimators to obtain estimates of the cointegration vectors. Due to the normalization of the cointegration vectors the comparison is based on  $\beta_{(K^*-r)}^*$  or the corresponding matrix without parameters of deterministic terms,  $\beta_{(K-r)}$ .

## 5.2 Evaluation Criteria

We compare the precision of the estimators on the basis of various criteria including standard measures such as the mean bias and the mean squared error (MSE). As pointed out already, the finite sample moments of the MLE do not exist and therefore, basing the comparison on these moments may be ‘very unfair’ (see Gonzalo (1994)). Therefore, our criteria include other characteristics of the empirical distribution of estimators. To account for the effects of possible outliers we include the median bias and the sample dispersion is measured by the interquartile range  $IQR_{50} = |q_{75} - q_{25}|$  ( $q_i$  is the  $i$ -th quantile of the empirical distribution). The small and finite sample properties of  $t$ - and Wald tests and of the implied confidence intervals for the parameters in  $\beta_{(K^*-r)}^*$  or  $\beta_{(K-r)}$  are also analyzed. We evaluate the empirical relative rejection frequencies (sizes) of  $t$ -tests for the hypotheses that the parameters in  $\beta_{(K^*-r)}^*$  or  $\beta_{(K-r)}$  are equal to the true values in Table 1. The null hypotheses of our Wald tests state that all parameters in  $\beta_{(K^*-r)}^*$  or  $\beta_{(K-r)}$  equal their true parameters. For DGP

(B), for instance, the Wald test null hypothesis is  $H_0 : \beta_{31} = \beta_{32} = -1$ .

### 5.3 Monte Carlo Results

Some of our Monte Carlo results are summarized in Table 2. The results shown are obtained by applying the two multivariate estimation methods to the generated Monte Carlo time series using a correctly specified model. In other words, the results have been obtained using the correct ('true') specification for lag length, cointegration rank and deterministic terms.

The results for DGP (A) can be summarized as follows. For two of the three parameters, the mean bias of the MLE is larger than that of the GLS estimator. In terms of the MSE the MLE is clearly dominated by the GLS estimator with impressively large MSEs of the Johansen estimator. These results are due to some extreme outliers produced by the MLE (see also the illustration in Section 4). In fact, computing a 1% trimmed mean bias (not shown) the biases of both estimators turn out to be quite similar. Comparing the median biases reveals that the GLS estimator is dominated by the MLE under this criterion. In contrast, the  $\text{IQR}_{50}$  is smaller for the GLS estimator, indicating a smaller sample dispersion. Thus, there appears to be a trade-off between bias and variance. The properties of  $t$ - and Wald tests are clearly not satisfactory for either estimation method. For instance, the true hypothesis of a unit income elasticity,  $H_0 : \beta_{21} = -1$ , is rejected by a  $t$ -test in 36.5% (MLE) and 33.3% (GLS) of the cases, when the nominal significance level is 5%. This outcome also implies that the empirical coverage of asymptotic confidence intervals is substantially lower than the desired nominal level and hence, such confidence intervals are not very reliable. The ranking of the estimators in terms of the  $t$ -test properties varies depending on the tested coefficient, e.g., GLS is better than MLE for  $\beta_{21}$  and vice versa for  $\beta_{31}$ . The mean confidence interval (CI) length of the MLE for this DGP is about four times that of the GLS estimator. The mean interval length is also governed by the extreme outlying observations of the MLE. However, even the 10% trimmed mean CI length of the MLE is about two times larger than the corresponding GLS quantity (results not shown). The empirical relative rejection frequency of the Wald test exceeds the nominal size of 5% substantially for both estimators. In fact, the large size distortion makes this test very unreliable with either of the two estimators. For the MLE this result is in line with findings by Gredenhoff & Jacobson (2001) and Fachin (2000) who propose bootstrap corrections to alleviate the problem.

Moving to the results for DGP (B) with two cointegration relations we find for  $T = 80$  again substantially larger values of the mean bias and the MSE for the MLE. This can be

Table 2: Properties ML and GLS estimates

DGP (A): Lütkepohl (2004)				true: $\beta_{21} = -1, \beta_{31} = 4, \beta_{41} = -0.0015$			
		ML	GLS			ML	GLS
Bias in mean	$\beta_{21}$	-0.0767	0.0521	MSE	$\beta_{21}$	36.53	0.1052
	$\beta_{31}$	-5.0923	-1.7768		$\beta_{31}$	5867.6	5.0996
	$\beta_{41}$	0.0000	-0.0013		$\beta_{41}$	0.0028	0.0000
Bias in median	$\beta_{21}$	0.0583	0.0698	IQR <sub>50</sub>	$\beta_{21}$	0.5396	0.3461
	$\beta_{31}$	-0.6484	-1.8262		$\beta_{31}$	2.8482	1.6172
	$\beta_{41}$	-0.0009	-0.0014		$\beta_{41}$	0.0042	0.0027
Size <i>t</i> -test	$\beta_{21}$	0.365	0.333	Mean CI length	$\beta_{21}$	2.416	0.606
	$\beta_{31}$	0.639	0.771		$\beta_{31}$	8.696	1.883
	$\beta_{41}$	0.455	0.473		$\beta_{41}$	0.018	0.004
Size Wald test		0.762	0.862				

DGP (B): King et al. (1991)				true: $\beta_{31} = \beta_{32} = -1$					
	<i>T</i>	ML	GLS		<i>T</i>	ML	GLS		
Bias in mean	80	$\beta_{31}$	-0.3024	-0.0294	MSE	80	$\beta_{31}$	123.39	0.0128
		$\beta_{32}$	-0.5402	-0.0007			$\beta_{32}$	447.69	0.0241
	160	$\beta_{31}$	-0.0038	-0.0064		160	$\beta_{31}$	0.0010	0.0009
		$\beta_{32}$	0.0007	-0.0005			$\beta_{32}$	0.0020	0.0017
Bias in median	80	$\beta_{31}$	-0.0210	-0.0308	IQR <sub>50</sub>	80	$\beta_{31}$	0.1238	0.1163
		$\beta_{32}$	0.0005	-0.0004			$\beta_{32}$	0.1727	0.1613
	160	$\beta_{31}$	-0.0040	-0.0063		160	$\beta_{31}$	0.0397	0.0383
		$\beta_{32}$	-0.0008	-0.0018			$\beta_{32}$	0.0523	0.0518
Size <i>t</i> -test	80	$\beta_{31}$	0.353	0.329	Mean CI length	80	$\beta_{31}$	0.731	0.205
		$\beta_{32}$	0.302	0.258			$\beta_{32}$	1.281	0.305
	160	$\beta_{31}$	0.172	0.172		160	$\beta_{31}$	0.085	0.084
		$\beta_{32}$	0.155	0.146			$\beta_{32}$	0.123	0.123
Size Wald test	80		0.534	0.509					
	160		0.264	0.255					

DGP (C): Gonzalo (1994)				true: $\beta_{21} = -1$					
	<i>T</i>	ML	GLS		<i>T</i>	ML	GLS		
Bias in mean	30	$\beta_{21}$	2.8627	0.2349	MSE	30	$\beta_{21}$	2365	1.2937
	50	$\beta_{21}$	7.3849	0.1327		50	$\beta_{21}$	59644	0.4612
	100	$\beta_{21}$	-0.0438	0.0430		100	$\beta_{21}$	0.4771	0.1203
	300	$\beta_{21}$	-0.0054	0.0043		300	$\beta_{21}$	0.0111	0.0102
Bias in median	30	$\beta_{21}$	0.0838	0.2504	IQR <sub>50</sub>	30	$\beta_{21}$	1.5622	1.0580
	50	$\beta_{21}$	0.0221	0.1236		50	$\beta_{21}$	0.8583	0.7070
	100	$\beta_{21}$	0.0024	0.0446		100	$\beta_{21}$	0.4055	0.3793
	300	$\beta_{21}$	-0.0003	0.0038		300	$\beta_{21}$	0.1127	0.1130
Size <i>t</i> -test	30	$\beta_{21}$	0.485	0.344	Mean CI length	30	$\beta_{21}$	7.048	2.075
	50	$\beta_{21}$	0.348	0.257		50	$\beta_{21}$	10.476	1.491
	100	$\beta_{21}$	0.193	0.176		100	$\beta_{21}$	0.926	0.885
	300	$\beta_{21}$	0.089	0.085		300	$\beta_{21}$	0.337	0.336

Note: IQR<sub>50</sub> denotes the interquartile range. Nominal size for all tests is 5%. Mean CI length is the mean confidence interval length of an asymptotic 95% interval.

attributed to a few extreme estimates. In fact, both estimators perform similarly if judged by the median bias criterion and the  $\text{IQR}_{50}$ . In this case the median bias is actually quite small for both estimators. As discussed for DGP (A), the empirical relative rejection frequencies of the  $t$ - and Wald tests exceed the nominal level of 5% substantially. Although the rejection frequencies based on the GLS estimates are somewhat smaller than those based on MLE, the differences are small enough to conclude that both estimation methods perform similarly. The mean CI length for the MLE is again governed by the few extreme outliers. Trimming by 10% yields similar interval lengths for both methods. Increasing the number of observations to  $T = 160$  is sufficient to eliminate the extreme ML estimates and consequently the large differences in mean bias and MSEs. Differences between MLE and GLS in the other statistics are now small enough to be neglected.

Finally, results for the bivariate DGP (C) show again that the extreme ML estimates are a small sample phenomenon that can arise even for very simple DGPs. For small sample sizes,  $T = 30$  and  $T = 50$ , that have not been considered by Gonzalo (1994), we find large values for the mean bias and the MSE. Increasing the sample size to  $T = 100$  and  $T = 300$  leads to similar absolute mean bias for MLE and GLS, however, with opposite signs. Note that GLS is dominated by MLE in terms of the median bias for all considered sample sizes. Interestingly, despite the slightly larger median bias, the empirical relative rejection frequencies of the  $t$ -tests based on GLS are somewhat closer to the nominal level than those associated with the MLE.

In addition to the results shown in Table 2 we have conducted variations of the experiments described above. For example, the effects of over- and underfitting the true lag dynamics have been investigated. Not surprisingly, model misspecification in the form of using fewer lags than necessary adversely affects the performance of both estimation methods. Overfitting the true lag length leads to a loss in efficiency as reflected by even larger rejection frequencies of true hypotheses than for the true lag length. The corresponding results are not shown as there is no sign of one method being more severely affected by over- or underfitting than the other.

In summary, our results indicate that the reduced rank ML estimator may produce rather extreme estimates of the cointegration parameters even for DGPs of practical relevance. We find that the asymptotically equivalent GLS estimator performs much better in this respect as it does not produce extreme outliers. Moreover, we conclude from our experiments that the GLS estimator typically has a smaller sample variation but a larger median bias than the MLE. Gonzalo (1994, p. 219) suggests to choose ‘*as the best estimator the one that has*

*smaller bias in median and smaller IQR*'. According to this criterion none of the considered methods dominates the other: In finite samples the MLE has typically a smaller median bias, but a larger interquartile range. Moreover, the properties of  $t$ -tests, Wald tests and confidence intervals based on both estimation methods are quite similar.

## 6 Conclusions

We have shown that the Johansen ML estimator of cointegration parameters has to be used cautiously in applied work because it can produce extremely distorted and unreliable estimates in small samples. This feature of the estimator is a reflection of its lack of moments in small samples. Using a quarterly German monetary system we have illustrated the magnitude of the problem for a particular example. A problematic feature here is that the estimated model passes the usual diagnostic tests and hence these checks do not help detecting the distorted estimates. We have also considered a simple GLS estimator which does not produce similarly outlying estimates and performs in other respects not very differently from the MLE. In applied work it may be a useful strategy to use both estimators and do not trust the ML estimator if it differs drastically from the GLS estimator.

Overall we conclude that the simple GLS estimator is an attractive alternative to the MLE especially in situations where the latter produces extreme estimates. It must be emphasized, however, that for both estimation methods the performance of  $t$ - and Wald tests is far from satisfactory in small samples. Solutions to this problem have been suggested, e.g., by Gredenhoff & Jacobson (2001) and Fachin (2000) who use bootstrap methods to test restrictions based on the MLE. These methods may also be applied to the simple GLS methods to improve finite sample inference.

## References

- Ahn, S. K. & Reinsel, G. C. (1990). Estimation of partially nonstationary multivariate autoregressive models, *Journal of the American Statistical Association* **85**: 813–823.
- Bossaerts, P. (1988). Common nonstationary components of asset prices, *Journal of Economic Dynamics and Control* **12**: 347–364.
- Caporale, G. M. & Pittis, N. (2004). Estimator choice and Fisher's paradox: A Monte Carlo study, *Econometric Reviews* **23**: 25–52.
- Fachin, S. (2000). Bootstrap and asymptotic tests of long-run relationships in cointegrated systems, *Oxford Bulletin of Economics and Statistics* **62**: 543–551.



- Gonzalo, J. (1994). Five alternative methods of estimating long-run equilibrium relationships, *Journal of Econometrics* **60**: 203–233.
- Gredenhoff, M. & Jacobson, T. (2001). Bootstrap testing linear restrictions on cointegrating vectors, *Journal of Business & Economic Statistics* **19**: 63–72.
- Hansen, G., Kim, J. R. & Mittnik, S. (1998). Testing cointegrating coefficients in vector autoregressive error correction models, *Economics Letters* **58**: 1–5.
- Hargreaves, C. P. (1994). A review of methods of estimating cointegrating relationships, in C. P. Hargreaves (ed.), *Nonstationary Time Series Analysis and Cointegration*, Oxford University Press, pp. 87–131.
- Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.
- King, R. G., Plosser, C. I., Stock, J. H. & Watson, M. W. (1991). Stochastic trends and economic fluctuations, *American Economic Review* **81**: 819–840.
- Lütkepohl, H. (2004). Vector autoregressive and vector error correction models, in H. Lütkepohl & M. Kräzig (eds), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge.
- Luukkonen, R., Ripatti, A. & Saikkonen, P. (1999). Testing for a valid normalization of cointegrating vectors in vector autoregressive processes, *Journal of Business & Economic Statistics* **17**: 195–204.
- Phillips, P. C. B. (1991). Optimal inference in cointegrated systems, *Econometrica* **59**: 283–306.
- Phillips, P. C. B. (1994). Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models, *Econometrica* **62**: 73–93.
- Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*, Springer-Verlag, New York.
- Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation, *Econometric Theory* **8**: 1–27.
- Stock, J. H. & Watson, M. W. (1988). Testing for common trends, *Journal of the American Statistical Association* **83**: 1097–1107.
- Stock, J. H. & Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica* **61**: 783–820.