On Determination of Cointegration Ranks*

Qiaoling Li[2] Jiazhu Pan[1,2] Qiwei Yao[3]

[1]Department of Statistics & Modelling Science, University of Strathclyde, Glasgow, G1 1XH, UK

[2]*School of Mathematical Sciences, Peking University, Beijing 100871, China*

[3]*Department of Statistics, London School of Economics, London, WC2A 2AE, UK*

Abstract

We propose a new method to determine the cointegration rank in the error correction model (ECM). The cointegration rank, together with the lag order, is determined by a penalized goodness-of-fit measure. We show that the estimated cointegration vectors are consistent with a convergence rate T, and our estimation for the cointegration rank is consistent. Our approach is more robust than the conventional likelihood based methods, as we do not impose any assumption on the form of the error distribution in the model. Furthermore we allow the serial dependence in the error sequence. The proposed methodology is illustrated with both simulated and real data examples. The advantage of the new method is particularly pronounced in the simulation with non-Gaussian and/or serially dependent errors.

Keywords: Cointegration; error correction models; penalized goodness-of-fit criteria; model selection

*Partially supported by the National Basic Research Program of China (grant No. 2007CB814902) and an EPSRC research grant of UK.

1 Introduction

The concept of cointegration dates back to Granger (1981), Granger and Weiss (1983), Engle and Granger (1987). It was introduced to reflect the long-run equilibrium among several economic variables while each of them might exhibit a distinct nonstationary trend. The cointegration research has made enormous progress since the seminal Granger representation theorem was presented in Engle and Granger (1987). It has a significant impact in economic and financial applications. While the large body of literature on cointegration contains splendid and also divergent ideas, the most frequently used representations for cointegrated systems include, among others, the error correction model (ECM) of Engle and Granger (1987), the common trends form of Stock and Watson (1988), and the triangular model of Phillips (1991).

From the view point of the economic equilibrium, the term "error correction" reflects the correction on the long-run relationship by short-run dynamics. The ECM has been successfully applied to solve various practical problems including the determination of exchange rates, capturing the relationship between consumer's expenditure and income, modelling and forecasting of inflation to establish monetary policy, etc. One of the critical questions in applying ECM is to determine the cointegration rank, which is often done by using some test-based procedures such as the likelihood ratio test (LRT) advocated by Johansen (1988, 1991). The key assumption for Johansen's approach is that the errors in the model are independent and normally distributed. It has been documented that the LRT may lead to either under- or over-estimates for cointegration ranks; see Gonzalo and Lee (1998), Gonzalo and Pitarakis (1998). Moreover, for the models with dependent and/or non-Gaussian errors, the LRT tends to reject the null hypothesis of no cointegration even when it actually presents; see Huang and Yang (1996). Other methods based on tests to determine the rank include Lagrange multiplier and Wald type tests, lag augmentation tests, tests based on canonical correlations, the Stock-Watson tests and Bierens' nonparametric tests; see Hubrich Lutkepohl and Saikkonen (2001) for a survey on the relevant methods. More recently, Aznar and Salvador (2002) proposed to determine the cointegration ranks by minimizing appropriate information criteria for the models with i.i.d. Gaussion errors, and Kapetanios (2004) established the asymptotic distribution of the estimate for the cointegration rank obtained by AIC.

In this paper we propose a new method for determining the cointegration ranks in the ECM with uncorrelated errors. We do not impose any further assumptions on the error distribution. In fact the errors may be serially dependent with each other. This makes our setting more general than those in the papers cited above. We first estimate the cointegration vectors using a method which may be viewed as a version of the reduced rank regression technique introduced by Anderson (1951); see also Johansen (1988, 1991), Ahn and Reinsel (1988), Ahn and Reinsel (1990), Bai (2003). We then determine the cointegration rank by minimizing an appropriate penalized goodness-of-fit measure which is a trade-off between goodness of fit and parsimony. We consider both the cases when the lag order is known or unknown. For the latter, we determine the cointegration rank and the lag order simultaneously. The simulation results reported in Wang and Bessler (2005) support such a simultaneous approach. The numerical results in section 4 indicate that the new method performs better than the conventional LRT-based procedures when the errors in the models are serially dependent and/or non-Gaussian.

At the theoretical front, we have shown that the estimated cointegration vectors are consistent with a convergence rate T which is the same as that of the ML estimator proposed by Johansen Johansen (1988, 1991). Furthermore, our estimation for the cointegration rank is consistent regardless if the lag order is known or not.

The rest of the paper is organized as follows. The estimation for cointegrating vectors and its asymptotic properties are presented in Section 2. Section 3 presents a criterion for determining cointegration ranks and its consistency. Section 4 contains a numerical comparison of the proposed method with the likelihood-based procedures for two simulated examples. An illustration with a real data set is also reported.

2 Estimation of Cointegrating Vectors

2.1 Vector error correction models

Suppose that $\{Y_t\}$ is a $p \times 1$ time series. The error correction model is of the form

$$\Delta Y_t = \mu + \Gamma_1 \Delta Y_{t-1} + \Gamma_2 \Delta Y_{t-2} + \cdots + \Gamma_{k-1} \Delta Y_{t-k+1} + \Gamma_0 Y_{t-1} + e_t, \tag{2.1}$$

where $\Delta Y_t = Y_t - Y_{t-1}$, μ is a $p \times 1$ vector of parameters, Γ_i is a $p \times p$ matrix of parameters, and e_t is covariance stationary with mean 0 and

$$E(e_t e_{t-\tau}) = \begin{cases} \Omega, & \tau = 0, \\ 0, & \text{otherwise.} \end{cases}$$

In the above expression, Ω is a positively definite matrix. The rank of Γ_0, denoted by r, is called the cointegration rank. Note that we assume e_t to be merely weakly stationary and uncorrelated. In fact, e_t, for different t, may be dependent with each other.

Let $\|A\| = [\text{tr}(A'A)]^{1/2}$ denote the norm of matrix A. Some regularity conditions are in order.

Assumption A. The process Y_t satisfies the basic assumptions of the Granger representation theorem (Engle and Granger (1987)):

1. For the characteristic polynomial of (2.1) given by $\Pi(z) = (1-z)I - (1-z)\sum_{i=1}^{k-1}\Gamma_i z^i - \Gamma_0 z$, it holds that $|\Pi(z)| = 0$ implies that either $|z| > 1$ or $z = 1$.
2. It holds that $\Gamma_0 = \gamma\alpha'$, where γ and α are $p \times r$ matrices with rank $r(< p)$.
3. $\gamma_\perp'(I - \sum_{i=1}^{k-1}\Gamma_i)\alpha_\perp$ has full rank, where $\gamma_\perp$ and $\alpha_\perp$ are the orthogonal complements of γ and α respectively.

Assumption B. The covariance stationary sequence $\{e_t\}$ with mean 0 is strongly mixing and the mixing coefficients β_m satisfy $\sum_1^\infty \beta_m^{1/2} < \infty$. Furthermore there exists a finite positive constant $0 < M < \infty$ such that $E\|e_t\|^4 \le M$ and $E\|\alpha' Y_{t-1}\|^4 \le M$ for all t.

By the Granger representation theorem, if there are exactly r cointegrating relations among the components of Y_t, and Γ_0 admits the decomposition $\Gamma_0 = \gamma\alpha'$, then α is an $p \times r$ matrix with linearly independent columns and $\alpha' Y_t$ is stationary. In this sense, α consists of r cointegrating vectors. Note that α and γ are not separately identifiable. The goal is to determine the rank of α and the space spanned by the columns of α.

2.2 Estimating cointegrating vectors

We assume that the cointegration rank r is known in this section. The determination of r will be discussed in section 3 below.

Model (2.1) can be rewritten as

$$\Delta Y_t = \Theta X_t + \gamma\alpha' Y_{t-1} + e_t, \tag{2.2}$$

where $\Theta = (\mu, \Gamma_1, \ldots, \Gamma_{k-1})$, $X_t = (1, \Delta Y_{t-1}', \ldots, \Delta Y_{t-k+1}')'$. Denote

$$\Theta_0 = \sum_{t=1}^T \Delta Y_t X_t' (\sum_{t=1}^T X_t X_t')^{-1}, \quad \Theta_1 = \sum_{t=1}^T Y_{t-1} X_t' (\sum_{t=1}^T X_t X_t')^{-1}, \quad \Theta_2 = \sum_{t=1}^T e_t X_t' (\sum_{t=1}^T X_t X_t')^{-1}.$$

Then it is easy to see from (2.2) that

$$\Theta \equiv \Theta_0 - \gamma\alpha'\Theta_1 - \Theta_2, \tag{2.3}$$

Now replacing Θ in (2.2) by (2.3), (2.2) reduces to

$$R_{0t} = \gamma\alpha' R_{1t} + e_t^*, \tag{2.4}$$

where $R_{0t} = \Delta Y_t - \Theta_0 X_t$, $R_{1t} = Y_{t-1} - \Theta_1 X_t$ and $e_t^* = e_t - \Theta_2 X_t$.

We may estimate the cointegration parameters γ and α by solving the optimization problem

$$\min_{\gamma,\alpha} \frac{1}{T} \sum_{t=1}^{T} (R_{0t} - \gamma\alpha' R_{1t})'(R_{0t} - \gamma\alpha' R_{1t}), \tag{2.5}$$

Although this can be considered as a standard least squares problem, we are unable to derive an explicit solution for α even with the regularity condition to make it identifiable.

Note that for any given α, the sum in (2.5) is minimized at $\gamma = \gamma(\alpha) \equiv S_{01}\alpha(\alpha' S_{11}\alpha)^{-1}$, where $S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R_{jt}'$. Replacing γ with this $\gamma(\alpha)$, (2.5) leads to

$$\min_{\alpha} \mathrm{tr}(S_{00} - S_{01}\alpha(\alpha' S_{11}\alpha)^{-1}\alpha' S_{10}). \tag{2.6}$$

It can be found that if α is a solution of (2.6), so is αA for any invertible matrix A. To choose one solution, we may apply the normalization $\alpha' S_{11}\alpha = I_r$. Now (2.6) is further reduced to

$$\max_{\alpha' S_{11}\alpha = I_r} \mathrm{tr}(\alpha' S_{10} S_{01}\alpha). \tag{2.7}$$

Obviously, the solution of (2.7) is $\hat{\alpha} \equiv (\hat{\alpha}_1, \cdots, \hat{\alpha}_r)$, where $\hat{\alpha}_1, \cdots, \hat{\alpha}_r$ are the r generalized eigenvectors of $S_{10}S_{01}$ with respect to S_{11} corresponding to the r largest generalized eigenvalues.[1] Note that $\hat{\gamma} = S_{01}\hat{\alpha}$ is the cointegration loading matrix.

Gonzalo (1994) compared numerically five different methods for estimating the cointegrating vectors: ordinary least squares (Engle and Granger (1987)), nonlinear least squares (Stock (1987)), maximum likelihood in an error correction model (Johansen (1988)), principal components (Stock and Watson (1988)), and canonical correlations (Bossaerts (1988)). The numerical results indicate that the maximum likelihood method outperformed the other methods for fully and correctly specified models as far as the estimation for cointegration vectors was concerned. However, the likelihood based methods are sensitive to the assumption that the errors are independent and normally distributed. The estimator proposed in this paper tends to overcome these shortcomings.

2.3 Asymptotic properties

By the Granger representation theorem, the ECM (2.1) may be equivalently represented as

$$\Delta Y_t = \delta + \Psi(L)e_t \tag{2.8}$$

where $\delta = \Psi(1)\mu$ and

$$\Psi(1) \equiv \Psi_0 + \Psi_1 + \Psi_2 + \cdots = \alpha_\perp(\gamma_\perp'(I - \sum_{i=1}^{k-1}\Gamma_i)\alpha_\perp)^{-1}\gamma_\perp'. \tag{2.9}$$

[1] If $Ax = \lambda Bx$, λ is called a generalized eigenvalue of A w.r.t. B, and x is the corresponding generalized eigenvector.

Consequently, $\alpha'\Psi(1) = 0$ and (2.8) implies that

$$Y_t = Y_0 + \delta t + \Psi(L)(e_t + e_{t-1} + \cdots + e_1). \tag{2.10}$$

We introduce two technical lemmas first. The proof of Lemma 2.1 may be found in Phillips and Durlauf (1986), Park and Phillips (1988). The notation " $\Rightarrow$ " denotes weak convergence, and " $\xrightarrow{P}$ " denotes convergence in probability. We always assume that Assumptions A & B hold and Y_t satisfies (2.10) in the sequel of this subsection. Note that (2.10) is implied by Assumption A.

Lemma 2.1. *Let* $\mathbf{u}_t \equiv \Psi(L)e_t$ *and* $\mathbf{v}_t \equiv \Psi(L)(e_t + e_{t-1} + \cdots + e_1)$. *As* $T \to \infty$, *it holds that*

$$(a) \qquad T^{-3/2}\sum\nolimits_{t=1}^{T}\mathbf{v}_{t-1} \Rightarrow \Psi(1)\int_0^1 W(s)ds, \quad T^{-1/2}\sum\nolimits_{t=1}^{T}\mathbf{u}_t \Rightarrow \Psi(1)W(1),$$

$$(b) \qquad T^{-2}\sum\nolimits_{t=1}^{T}\mathbf{v}_{t-1}\mathbf{v}_{t-1}' \Rightarrow \Psi(1)\int_0^1 W(s)W(s)'ds\Psi(1)',$$

$$(c) \begin{cases} T^{-1}\sum\nolimits_{t=1}^{T}\mathbf{v}_{t-1}e_t' \Rightarrow \Psi(1)\int_0^1 W(s)dW(s)', \\ T^{-1}\sum\nolimits_{t=1}^{T}\mathbf{v}_{t-1}e_{t-j}' \Rightarrow \Psi(1)\int_0^1 W(s)dW(s)' + (\Psi_{j-1} + \cdots + \Psi_1 + \Psi_0)\Omega, \end{cases}$$

$$(d) \qquad T^{-5/2}\sum\nolimits_{t=1}^{T}t\mathbf{v}_{t-1}' \Rightarrow \int_0^1 sW(s)'ds\Psi(1)',$$

$$(e) \qquad T^{-3/2}\sum\nolimits_{t=1}^{T}te_{t-j}' \Rightarrow \int_0^1 sdW(s)',$$

where $W(s)$ *is a vector Wiener process on* $C[0,1]^p$ *with covariance matrix* $\Omega = E(e_t e_t')$.

Under Assumption A & B, the lemma below can be derived by the results listed in Lemma 2.1. The details of the proof are omitted since there are too much repetitive algebra operations to display. [2]

Lemma 2.2. *Let* τ *be a* $p \times (p - r - 1)$ *matrix which is orthogonal to* α *and* δ *such that* (α, δ, τ) *spans* R^p. *As* $T \to \infty$,

(a)

$$\begin{aligned}
\frac{\delta' S_{11}\delta}{T^2} &\xrightarrow{P} \frac{1}{12}(\delta'\delta)^2, \\
\frac{\tau' S_{11}\tau}{T} &\Rightarrow \tau'\Psi(1)[\int_0^1 W(s)W(s)'ds - \int_0^1 W(s)ds\int_0^1 W(s)'ds]\Psi(1)'\tau, \\
\frac{\delta' S_{11}\tau}{\sqrt{T^3}} &\Rightarrow \delta'\delta(\int_0^1 sW(s)'ds - \frac{1}{2}\int_0^1 W(s)'ds)\Psi(1)'\tau;
\end{aligned}$$

[2]Readers can ask the authors for the whole proof.

(b)

$$
\begin{aligned}
\alpha' S_{10} &\xrightarrow{P} \alpha' \Sigma_{10}, \\
\tau' S_{10} &\Rightarrow \tau'[\Sigma_{10} + \Psi(1)(\int_0^1 W(s)dW(s)' - \int_0^1 W(s)dsW(1)')\Psi(1)'C], \\
\frac{\delta' S_{10}}{\sqrt{T}} &\Rightarrow \delta'\delta(\int_0^1 sdW(s)' - \frac{1}{2}W(1)')\Psi(1)'C,
\end{aligned}
$$

where $C = I_p - (\iota_{k-1}' \otimes I_p)\mathbf{X}^{-1}\mathbf{B_0}$, *and* $\Sigma_{10} = \sum_{l=1}^{\infty}\sum_{i=0}^{l-1}\Psi_i\Omega\Psi_l' - \mathbf{B_1}'\mathbf{X}^{-1}\mathbf{B_0}$. *Here,* ι_{k-1} *is a* $k-1$ *dimensional vector of ones,* $\mathbf{B_0} = (D'_1, \cdots, D'_{k-1})'$, $\mathbf{B_1} = (F_1', \cdots, F_{k-1}')'$, *where*

$$
D_i = \sum_{l=0}^{\infty}\Psi_{l+i}\Omega\Psi_l', \quad F_i = \sum_{l=0}^{\infty}\sum_{j=0}^{i+l-1}\Psi_j\Omega\Psi_l',
$$

and $\mathbf{X}$ *is a* $p(k-1)\times p(k-1)$ *symmetric block matrix with the* ij*-th* ($j \geq i = 1, \ldots, k-1$) *block* D_{j-i}*;*

(c)

$$
\begin{aligned}
\frac{\delta' S_{11}\alpha}{\sqrt{T}} &\Rightarrow \delta'\delta[(\int_0^1 sdW(s)' - \frac{1}{2}W(1)')(\Psi(1)'C - I_p)]\gamma(\gamma'\gamma)^{-1}, \\
\tau' S_{11}\alpha &\Rightarrow \tau'[\Sigma_{10} + \Psi(1)(\int_0^1 W(s)dW(s)' - \int_0^1 W(s)dsW(1)')(\Psi(1)'C - I_p)]\gamma(\gamma'\gamma)^{-1}, \\
\alpha' S_{11}\alpha &\xrightarrow{P} \alpha'\Sigma_{11}\alpha,
\end{aligned}
$$

where $\Sigma_{11} = -\sum_{i=-\infty}^{\infty}|i|\sum_{l=0}^{\infty}\Psi_{l+i}\Omega\Psi_l' - \mathbf{B_1}'\mathbf{X}^{-1}\mathbf{B_1}$*;*

(d)

$$
\begin{aligned}
\frac{1}{\sqrt{T^3}}\sum_{t=1}^{T}\delta' R_{1t}e_t' &\Rightarrow \delta'\delta(\int_0^1 sdW(s)' - \frac{1}{2}W(1)'), \\
\frac{1}{T}\sum_{t=1}^{T}\tau' R_{1t}e_t' &\Rightarrow \tau'\Psi(1)(\int_0^1 W(s)dW(s)' - \int_0^1 W(s)dsW(1)'), \\
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\alpha' R_{1t}e_t' &= O_p(1);
\end{aligned}
$$

(e) $V_T \xrightarrow{P} V$, *where* $V_T = diag(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_r)$ *and* $V = diag(\lambda_1, \lambda_2, \ldots, \lambda_r)$, $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_r$ *be the* r *largest generalized eigenvalues of* $S_{10}S_{01}$ *with respect to* S_{11}*, and* $\lambda_1 \geq \cdots \geq \lambda_r > 0$ *are constants.*

Now we present the asymptotic distribution of $\hat{\alpha}$ in the theorem below.

Theorem 2.1. *Let* τ *be a* $p \times (p-r-1)$ *matrix which is orthogonal to* α *and* δ *such that* (α, δ, τ) *spans* R^p*. There exists a* $r \times r$ *invertible matrix* H_T *for which*

$$
\begin{aligned}
& T(\hat{\alpha}H_T^{-1} - \alpha) \\
= & \begin{bmatrix} \frac{\delta}{\sqrt{T}} & \tau \end{bmatrix} \begin{bmatrix} \frac{\delta' S_{11}\delta}{T^2} & \frac{\delta' S_{11}\tau}{\sqrt{T^3}} \\ \frac{\tau' S_{11}\delta}{\sqrt{T^3}} & \frac{\tau' S_{11}\tau}{T} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{\sqrt{T^3}}\sum_{t=1}^{T}\delta' R_{1t}e_t' \\ \frac{1}{T}\sum_{t=1}^{T}\tau' R_{1t}e_t' \end{bmatrix} S_{01}\alpha(\alpha' S_{10}S_{01}\alpha)^{-1}\alpha' S_{11}\alpha + o_p(1)
\end{aligned}
$$

as $T \to \infty$.

Remark 1. The asymptotic distribution of each cointegrating vector $\hat{\alpha}_i$ is determined by the first term on the right hand side of the equality above. The limit of each component in the matrix and vectors can be found in Lemma 2.2 respectively.

Proof. According to the definition of eigenvectors $\hat{\alpha} \equiv (\hat{\alpha}_1, \ldots, \hat{\alpha}_r)$, we decompose them as follows

$$\hat{\alpha}_i = \alpha H_{iT} + \alpha_\perp L_{iT}, \qquad i = 1, \ldots, r \tag{2.11}$$

where $H_{iT} = (\alpha'\alpha)^{-1}\alpha'\hat{\alpha}_i$, $L_{iT} = (\alpha'_\perp\alpha_\perp)^{-1}\alpha'_\perp\hat{\alpha}_i$ and $\alpha_\perp$ is the orthogonal complement of α. Thus, $H_T \equiv (H_{1T}, \ldots, H_{rT}) = (\alpha'\alpha)^{-1}\alpha'\hat{\alpha}$ is an invertible matrix with rank r.

Let $S(\lambda) = \lambda S_{11} - S_{10}S_{01}$. The eigenvectors $\hat{\alpha}_i$ and eigenvalues $\hat{\lambda}_i$ satisfy $\alpha'_\perp S(\hat{\lambda}_i)\hat{\alpha}_i = 0$, or equivalently

$$\alpha'_\perp S(\hat{\lambda}_i)\alpha H_{iT} + \alpha'_\perp S(\hat{\lambda}_i)\alpha_\perp L_{iT} = 0 \tag{2.12}$$

We have from the decomposition (2.11) and the equality above

$$\hat{\alpha}_i - \alpha H_{iT} = \alpha_\perp L_{iT} = -\alpha_\perp(\alpha'_\perp S(\hat{\lambda}_i)\alpha_\perp)^{-1}(\alpha'_\perp S(\hat{\lambda}_i)\alpha)H_{iT}.$$

Since $\alpha'\delta = 0$, if an $p \times (p-r-1)$ matrix τ is chosen orthogonal to α and δ, then (α, δ, τ) spans the whole R^p. For $\alpha_\perp = (T^{-1/2}\delta, \tau)$, we get

$$\begin{aligned}
\frac{\alpha'_\perp S(\hat{\lambda}_i)\alpha_\perp}{T} &= \begin{bmatrix} \frac{\delta' S(\hat{\lambda}_i)\delta}{T^2} & \frac{\delta' S(\hat{\lambda}_i)\tau}{\sqrt{T^3}} \\ \frac{\tau' S(\hat{\lambda}_i)\delta}{\sqrt{T^3}} & \frac{\tau' S(\hat{\lambda}_i)\tau}{T} \end{bmatrix} = \begin{bmatrix} \frac{\delta' S_{11}\delta}{T^2} & \frac{\delta' S_{11}\tau}{\sqrt{T^3}} \\ \frac{\tau' S_{11}\delta}{\sqrt{T^3}} & \frac{\tau' S_{11}\tau}{T} \end{bmatrix} \hat{\lambda}_i + o_p(1) = O_p(1), \\
\alpha'_\perp S(\hat{\lambda}_i)\alpha &= \begin{bmatrix} \hat{\lambda}_i \frac{\delta' S_{11}\alpha}{\sqrt{T}} - \frac{\delta' S_{10}}{\sqrt{T}} S_{01}\alpha \\ \hat{\lambda}_i \tau' S_{11}\alpha - \tau' S_{10}S_{01}\alpha \end{bmatrix} = O_p(1)
\end{aligned}$$

because it follows from Lemma 2.2(b) that $\delta' S_{10} = O_p(\sqrt{T})$, $\tau' S_{10} = O_p(1)$ and $\alpha' S_{10} = O_p(1)$.

Moreover, the eigenvectors $\hat{\alpha}_i$ and eigenvalues $\hat{\lambda}_i$ satisfy $\alpha' S(\hat{\lambda}_i)\hat{\alpha}_i = 0$, or

$$\alpha' S(\hat{\lambda}_i)\alpha H_{iT} + \alpha' S(\hat{\lambda}_i)\alpha_\perp L_{iT} = 0. \tag{2.13}$$

Therefore, it follow from (2.12), (2.13) and $H_{iT} = O_p(1)$ that

$$\alpha' S(\hat{\lambda}_i)\alpha H_{iT} = \alpha' S(\hat{\lambda}_i)\alpha_\perp(\alpha'_\perp S(\hat{\lambda}_i)\alpha_\perp)^{-1}(\alpha'_\perp S(\hat{\lambda}_i)\alpha)H_{iT} = O_p(\frac{1}{T}). \tag{2.14}$$

Recall that $S_{10} = S_{11}\alpha\gamma' + \frac{1}{T}\sum_{t=1}^{T} R_{1t}e_t'$ by (2.4)[3] and $\frac{1}{T}\sum_{t=1}^{T} \alpha' R_{1t}e_t' = o_p(1)$ by Lemma 2.2(d), we can obtain that

$$\begin{aligned}
\alpha_\perp' S(\hat{\lambda}_i)\alpha H_{iT} &= \alpha_\perp' S_{11}\alpha(\alpha' S_{11}\alpha)^{-1}\alpha' S(\hat{\lambda}_i)\alpha H_{iT} - \alpha_\perp' \frac{1}{T}\sum_{t=1}^{T} R_{1t}e_t' S_{01}\alpha H_{iT} + o_p(1) \\
&= -\alpha_\perp' \frac{1}{T}\sum_{t=1}^{T} R_{1t}e_t' S_{01}\alpha H_{iT} + o_p(1).
\end{aligned}$$

The second equality holds from (2.14). Thus,

$$T(\hat{\alpha}_i - \alpha H_{iT}) = \begin{bmatrix} \frac{\delta}{\sqrt{T}} & \tau \end{bmatrix} \begin{bmatrix} \frac{\delta' S_{11}\delta}{T^2} & \frac{\delta' S_{11}\tau}{\sqrt{T^3}} \\ \frac{\tau' S_{11}\delta}{\sqrt{T^3}} & \frac{\tau' S_{11}\tau}{T} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{\sqrt{T^3}}\sum_{t=1}^{T} \delta' R_{1t}e_t' S_{01}\alpha \\ \frac{1}{T}\sum_{t=1}^{T} \tau' R_{1t}e_t' S_{01}\alpha \end{bmatrix} H_{iT}\hat{\lambda}_i^{-1} + o_p(1).$$

Furthermore, (2.14) implies that

$$\alpha' S_{11}\alpha H_T V_T = \alpha' S_{10} S_{01}\alpha H_T + o_p(1),$$

or equivalently

$$H_T V_T^{-1} H_T^{-1} = (\alpha' S_{10} S_{01}\alpha)^{-1}\alpha' S_{11}\alpha + o_p(1).$$

Thus,

$$\begin{aligned}
&T(\hat{\alpha} H_T^{-1} - \alpha) \\
&= \begin{bmatrix} \frac{\delta}{\sqrt{T}} & \tau \end{bmatrix} \begin{bmatrix} \frac{\delta' S_{11}\delta}{T^2} & \frac{\delta' S_{11}\tau}{\sqrt{T^3}} \\ \frac{\tau' S_{11}\delta}{\sqrt{T^3}} & \frac{\tau' S_{11}\tau}{T} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{\sqrt{T^3}}\sum_{t=1}^{T} \delta' R_{1t}e_t' \\ \frac{1}{T}\sum_{t=1}^{T} \tau' R_{1t}e_t' \end{bmatrix} S_{01}\alpha(\alpha' S_{10} S_{01}\alpha)^{-1}\alpha' S_{11}\alpha + o_p(1).
\end{aligned}$$

□

Theorem 2.1 implies that $\hat{\alpha}$ is a $T-$consistent estimator of αH_T for an invertible matrix H_T. In the theorem below, we show that $\hat{\gamma} \equiv S_{01}\hat{\alpha}$ is a $\sqrt{T}-$consistent estimator of $\gamma(H_T')^{-1}$ and $\hat{\Gamma}_0 \equiv \hat{\gamma}\hat{\alpha}'$ is a $\sqrt{T}-$consistent estimator of $\Gamma_0 = \gamma\alpha'$.

Theorem 2.2. $\sqrt{T}(\hat{\gamma} - \gamma(H_T')^{-1}) = O_p(1)$, $\sqrt{T}(\hat{\Gamma}_0 - \Gamma_0) = O_p(1)$

Proof. Note $\hat{\gamma} = S_{01}\hat{\alpha} = (\gamma\alpha' S_{11} + \frac{1}{T}\sum_{t=1}^{T} e_t R_{1t}')\hat{\alpha}$ and $\hat{\alpha}' S_{11}\hat{\alpha} = I_r$. It holds that

$$\hat{\gamma} - \gamma(H_T')^{-1} = \gamma(\alpha - \hat{\alpha} H_T^{-1})' S_{11}\hat{\alpha} + \frac{1}{T}\sum_{t=1}^{T} e_t R_{1t}'\hat{\alpha}. \tag{2.15}$$

[3]Here and in the sequel of this paper we use the fact that $\sum_{t=1}^{T} R_{1t}e_t^{*'} = \sum_{t=1}^{T} R_{1t}e_t'$.

Let $M_T = \frac{1}{T}\sum_{t=1}^{T}\alpha_\perp' R_{1t}e_t' S_{01}\alpha(\alpha' S_{10}S_{01}\alpha)^{-1}\alpha' S_{11}\alpha$. Then $M_T = O_p(1)$ by Lemma 2.2 $(b) \sim (d)$, and

$$\hat{\alpha}H_T^{-1} - \alpha = \alpha_\perp(\alpha_\perp' S_{11}\alpha_\perp)^{-1}M_T. \tag{2.16}$$

The first term on the right side of (2.15) can be rewritten as $\gamma(\alpha - \hat{\alpha}H_T^{-1})' S_{11}(\hat{\alpha} - \alpha H_T) + \gamma(\alpha - \hat{\alpha}H_T^{-1})' S_{11}\alpha H_T$. From (2.16), we have

$$(\alpha - \hat{\alpha}H_T^{-1})' S_{11}(\hat{\alpha} - \alpha H_T) = -M_T'(\alpha_\perp' S_{11}\alpha_\perp)^{-1}M_T H_T = O_p(\frac{1}{T}), \tag{2.17}$$

$$(\alpha - \hat{\alpha}H_T^{-1})' S_{11}\alpha H_T = M_T'(\alpha_\perp' S_{11}\alpha_\perp)^{-1}\alpha_\perp' S_{11}\alpha H_T = O_p(\frac{1}{T}). \tag{2.18}$$

But, for the second term on the right side of (2.15),

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} e_t R_{1t}'\hat{\alpha} &= \frac{1}{T}\sum_{t=1}^{T} e_t R_{1t}'(\hat{\alpha} - \alpha H_T) + \frac{1}{T}\sum_{t=1}^{T} e_t R_{1t}'\alpha H_T \\
&= \frac{1}{T}\sum_{t=1}^{T} e_t R_{1t}'\alpha_\perp(\alpha_\perp' S_{11}\alpha_\perp)^{-1}M_T H_T + \frac{1}{T}\sum_{t=1}^{T} e_t R_{1t}'\alpha H_T \\
&= O_p(\frac{1}{T}) + \frac{1}{T}\sum_{t=1}^{T} e_t R_{1t}'\alpha H_T. \qquad (2.19)
\end{aligned}$$

Therefore,

$$\sqrt{T}(\hat{\gamma} - \gamma(H_T')^{-1}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_t R_{1t}'\alpha H_T + o_p(1) = O_p(1).$$

The second equality holds by Lemma 2.2(d).

Consider the second relation now. It holds that

$$\begin{aligned}
\hat{\Gamma}_0 - \Gamma_0 &= \hat{\gamma}\hat{\alpha}' - \gamma\alpha' \\
&= (\hat{\gamma} - \gamma(H_T')^{-1})(\hat{\alpha} - \alpha H_T)' + (\hat{\gamma} - \gamma(H_T')^{-1})H_T'\alpha' + \gamma(H_T')^{-1}(\hat{\alpha} - \alpha H_T)' \\
&= O_p(\frac{1}{\sqrt{T^3}}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{T}) = O_p(\frac{1}{\sqrt{T}}).
\end{aligned}$$

This completes the proof of Theorem 2.2. □

3 Estimation of the Cointegration Rank

Let r_0 be the true value of the cointegration rank of model (2.1). In this section, we discuss how to estimate r_0 based on the estimated cointegration vector $\hat{\alpha}$ derived in section 2. The basic idea is to treat the rank as part of the "order" of model (2.1) and to determine the order in terms of an appropriate information criterion. In this section we always assume that Assumptions A and B hold. First we deal with the case when the lag order k is known.

3.1 Determining the cointegration rank r with the lag order k given

Consider the sum of squared residuals

$$\begin{aligned}\mathbb{R}(r,\hat{\alpha}) &= \min_{\gamma} \frac{1}{T}\sum_{t=1}^{T}(R_{0t}-\gamma\hat{\alpha}'R_{1t})'(R_{0t}-\gamma\hat{\alpha}'R_{1t}) \\ &= \operatorname{tr}(S_{00}-S_{01}\hat{\alpha}(\hat{\alpha}'S_{11}\hat{\alpha})^{-1}\hat{\alpha}'S_{10}).\end{aligned} \tag{3.1}$$

To avoid possible overfitting, we add a penalty term. Our penalized goodness-of-fit criterion is defined as

$$\mathbb{M}(r) = \mathbb{R}(r,\hat{\alpha}) + n_r g(T), \tag{3.2}$$

where $g(T)$ is the penalty for "overfitting" and n_r is the number of freely estimated parameters. Note that $n_r = p + p^2(k-1) + 2pr - r^2$ for model (2.1). We may estimate r_0 by minimizing

$$\hat{r} = \arg\min_{0\le r\le p} \mathbb{M}(r).$$

The following theorem shows that $\hat{r}$ is a consistent estimator of r_0 provided that the penalty function $g(T)$ satisfies some mild conditions.

Theorem 3.1. *As $T\to\infty$, $\hat{r}\xrightarrow{P} r_0$ provided that $g(T)\to 0$ and $Tg(T)\to\infty$.*

Remark 2. Note that both the BIC criterion with $g(T)=\ln(T)/T$ (Schwarz (1978)) and the HQ criterion with $g(T)=2\ln(\ln(T))/T$ (Hannan and Quinn (1979)) lead to consistent estimators for the cointegration order. To prove Theorem 3.1, we need a slightly generalized form of Theorem 2.1.

Lemma 3.1. *For any $1\le r\le p$, there exists a $r_0\times r$ matrix H_T^r with full rank such that, as $T\to\infty$, $T(\hat{\alpha}-\alpha H_T^r)=O_p(1)$.*

Proof. The proof is the same as that of Theorem 2.1 without any modification, except that $H_T=(\alpha'\alpha)^{-1}\alpha'\hat{\alpha}$ is not necessarily invertible matrix anymore if $r\neq r_0$. The reason is that r_0 denotes the true rank of γ and α now. □

Let A^l denote a matrix with rank l. In particular, α^{r_0} and $\hat{\alpha}^r$ $(1\le r\le p)$ denote the matrices α and $\hat{\alpha}$ with ranks r_0 and r respectively.

Lemma 3.2. *For any $r_0 \le r \le p$, $\mathbb{R}(r,\hat{\alpha}^r) - \mathbb{R}(r_0,\hat{\alpha}^{r_0}) = O_p(\frac{1}{T})$.*

Proof. Since

$$|\mathbb{R}(r,\hat{\alpha}^r) - \mathbb{R}(r_0,\hat{\alpha}^{r_0})| \le |\mathbb{R}(r,\hat{\alpha}^r) - \mathbb{R}(r_0,\alpha^{r_0})| + |\mathbb{R}(r_0,\alpha^{r_0}) - \mathbb{R}(r_0,\hat{\alpha}^{r_0})|$$
$$\le 2\max_{r_0\le r\le p}|\mathbb{R}(r,\hat{\alpha}^r) - \mathbb{R}(r_0,\alpha^{r_0})|,$$

then, it is sufficient to prove for any $r_0 \le r \le p$,

$$\mathbb{R}(r,\hat{\alpha}^r) - \mathbb{R}(r_0,\alpha^{r_0}) = O_p(T^{-1}).$$

Notice that $S_{01} = \gamma\alpha^{r_0'}S_{11} + \frac{1}{T}\sum_{t=1}^T e_t R_{1t}'$ and

$$\begin{aligned}\mathbb{R}(r,\hat{\alpha}^r) &= tr(S_{00} - S_{01}\hat{\alpha}^r(\hat{\alpha}^{r'}S_{11}\hat{\alpha}^r)^{-1}\hat{\alpha}^{r'}S_{10}),\\ \mathbb{R}(r_0,\alpha^{r_0}) &= tr(S_{00} - S_{01}\alpha^{r_0}(\alpha^{r_0'}S_{11}\alpha^{r_0})^{-1}\alpha^{r_0'}S_{10}).\end{aligned}$$

We have

$$\begin{aligned}&\mathbb{R}(r,\hat{\alpha}^r) - \mathbb{R}(r_0,\alpha^{r_0})\\ =&tr[\gamma\alpha^{r_0'}S_{11}\alpha^{r_0}\gamma' - \gamma\alpha^{r_0'}S_{11}\hat{\alpha}^r\hat{\alpha}^{r'}S_{11}\alpha^{r_0}\gamma']\\ &+2tr[\frac{1}{T}\sum_{t=1}^T\gamma\alpha^{r_0'}R_{1t}e_t' - \gamma\alpha^{r_0'}S_{11}\hat{\alpha}^r\frac{1}{T}\sum_{t=1}^T\hat{\alpha}^{r'}R_{1t}e_t']\\ &+tr[\frac{1}{T}\sum_{t=1}^T e_tR_{1t}'\alpha^{r_0}(\alpha^{r_0'}S_{11}\alpha^{r_0})^{-1}\frac{1}{T}\sum_{t=1}^T\alpha^{r_0'}R_{1t}e_t' - \frac{1}{T^2}\sum_{t=1}^T e_tR_{1t}'\hat{\alpha}^r\sum_{t=1}^T\hat{\alpha}^{r'}R_{1t}e_t']\\ \equiv&I + II + III.\end{aligned}$$

It follows straightly from Lemma 2.2(b) and (c) that $III = O_p(\frac{1}{T})$.

Now, for $r \ge r_0$, $H_T^r = (\alpha^{r_0'}\alpha^{r_0})^{-1}\alpha^{r_0'}\hat{\alpha}^r$ has rank r_0. Let H_T^{r+} denote the generalized inverse of H_T^r such that $H_T^rH_T^{r+} = I_{r_0}$, then it can be written as $H_T^{r+} = (\hat{\alpha}^{r'}\alpha^{r_0})^{r+}\alpha^{r_0'}\alpha^{r_0}$. It follows that,

$$I = tr[\gamma H_T^{r+'}(\hat{\alpha}^r - \alpha^{r_0}H_T^r)'S_{11}^{1/2}(I_p - S_{11}^{1/2}\hat{\alpha}^r\hat{\alpha}^{r'}S_{11}^{1/2})S_{11}^{1/2}(\hat{\alpha}^r - \alpha^{r_0}H_T^r)H_T^{r+}\gamma']$$

where I_p is an identity matrix with rank p. Furthermore, it is easy to see that $I_p - S_{11}^{1/2}\hat{\alpha}^r\hat{\alpha}^{r'}S_{11}^{1/2}$ is an idempotent matrix with eigenvalues 0 or 1. Because of the inequality $x(I_p - S_{11}^{1/2}\hat{\alpha}^r\hat{\alpha}^{r'}S_{11}^{1/2})x' \le xx'$ for any vector x,

$$\begin{aligned}I &\le \sum_{i=1}^p\gamma_i'H_T^{r+'}(\hat{\alpha}^r - \alpha^{r_0}H_T^r)'S_{11}(\hat{\alpha}^r - \alpha^{r_0}H_T^r)H_T^{r+}\gamma_i\\ &= \sum_{i=1}^p\gamma_i'\alpha^{r_0'}\alpha^{r_0}(\hat{\alpha}^{r'}\alpha^{r_0})^{r+'}(\hat{\alpha}^r - \alpha^{r_0}H_T^r)'S_{11}(\hat{\alpha}^r - \alpha^{r_0}H_T^r)(\hat{\alpha}^{r'}\alpha^{r_0})^{r+}\alpha^{r_0'}\alpha^{r_0}\gamma_i\end{aligned}$$

where γ_i' is the ith column of γ. From (2.17), it follows that $(\hat{\alpha}^r - \alpha^{r_0} H_T^r)' S_{11} (\hat{\alpha}^r - \alpha^{r_0} H_T^r) = O_p(\frac{1}{T})$. Additionally, $(\hat{\alpha}^{r\prime} \alpha^{r_0})^{r+} = O_p(1)$ has full rank r_0. Hence, $I = O_p(\frac{1}{T})$. For $I\!I$, we have

$$\begin{aligned}
I\!I =& 2tr\big[\gamma H_T^{r+\prime}[(\hat{\alpha}^r - \alpha^{r_0} H_T^r)' S_{11} \hat{\alpha}^r \frac{1}{T}\sum_{t=1}^T \hat{\alpha}^{r\prime} R_{1t} e_t' - (\hat{\alpha}^r - \alpha^{r_0} H_T^r)' \frac{1}{T}\sum_{t=1}^T R_{1t} e_t']\big] \\
=& 2tr\big[\gamma \alpha_0^{r\prime} \alpha^{r_0} (\hat{\alpha}^{r\prime} \alpha^{r_0})^{r+\prime} [(\hat{\alpha}^r - \alpha^{r_0} H_T^r)' S_{11} \hat{\alpha}^r \frac{1}{T}\sum_{t=1}^T \hat{\alpha}^{r\prime} R_{1t} e_t' \\
&- (\hat{\alpha}^r - \alpha^{r_0} H_T^r)' \frac{1}{T}\sum_{t=1}^T R_{1t} e_t']\big] = O_p(\frac{1}{T}).
\end{aligned}$$

The detail proofs for $(\hat{\alpha}^r - \alpha^{r_0} H_T^r)' S_{11} \hat{\alpha}^r = O_p(\frac{1}{T})$, $(\hat{\alpha}^r - \alpha^{r_0} H_T^r)' \frac{1}{T}\sum_{t=1}^T R_{1t} e_t' = O_p(\frac{1}{T})$ and $\frac{1}{T}\sum_{t=1}^T \hat{\alpha}^{r\prime} R_{1t} e_t' = O_p(\frac{1}{\sqrt{T}})$ can be found in (2.17)$\sim$(2.19). The proof of Lemma 3.2 completes. $\square$

Proof of Theorem 3.1. The objective is to verify that $\lim_{T\to\infty} P(\mathbb{M}(r) - \mathbb{M}(r_0) < 0) = 0$ for all $r \leq p$ and $r \neq r_0$, where

$$\mathbb{M}(r) - \mathbb{M}(r_0) = \mathbb{R}(r, \hat{\alpha}^r) - \mathbb{R}(r_0, \hat{\alpha}^{r_0}) - (n_{r_0} - n_r) g(T).$$

For $r < r_0$, from (3.1), we have $\mathbb{R}(r, \hat{\alpha}^r) - \mathbb{R}(r_0, \hat{\alpha}^{r_0}) = \sum_{i=r+1}^{r_0} \hat{\lambda}_i$, where $\hat{\lambda}_i$ is the ith generalized eigenvalue of $S_{10}S_{01}$ respect to S_{11} in decreasing order. Therefore, if $g(T) \to 0$ as $T \to \infty$,

$$\begin{aligned}
P(\mathbb{M}(r) - \mathbb{M}(r_0) < 0) &= P(\sum_{i=r+1}^{r_0} \hat{\lambda}_i < (r_0 - r)(2p - (r_0 + r)) g(T)) \\
&\to P(\sum_{i=r+1}^{r_0} \lambda_i < 0) = 0
\end{aligned}$$

by Lemma 2.2 (e) that $\hat{\lambda}_i \xrightarrow{P} \lambda_i > 0$.

For $r > r_0$, Lemma 3.2 implies that $\mathbb{R}(r_0, \hat{\alpha}^{r_0}) - \mathbb{R}(r, \hat{\alpha}^r) = O_p(\frac{1}{T})$. Thus, if $Tg(T) \to \infty$ as $T \to \infty$, we have

$$\begin{aligned}
P(\mathbb{M}(r) - \mathbb{M}(r_0) < 0) &= P(\mathbb{R}(r_0, \hat{\alpha}^{r_0}) - \mathbb{R}(r, \hat{\alpha}^r) > (r - r_0)(2p - (r + r_0)) g(T)) \\
&= P(T[\mathbb{R}(r_0, \hat{\alpha}^{r_0}) - \mathbb{R}(r, \hat{\alpha}^r)] > (r - r_0)(2p - (r + r_0)) T g(T)) \\
&\to 0.
\end{aligned}$$

The proof of Theorem 3.1 is completed. $\square$

3.2 Determining the cointegration rank r and the lag order k jointly

One of the important issues in applying ECM is to determine the lag order k. Johansen (1991) adopted a two-step procedure as follows: first the lag order k is determined by either an appropriate information criterion or a sequence of likelihood ratio test, and then the cointegration rank r is determined by an LRT. We proceed differently below and determine both r and k simultaneously by minimizing an appropriate penalized goodness-of-fit criterion.

Put

$$\mathbb{M}(r,k)=\mathbb{R}(r,k,\hat{\alpha}_k^r)+n_{r,k}g(T), \tag{3.3}$$

where $\mathbb{R}(r,k,\hat{\alpha}_k^r)$ and $n_{r,k}$ are the same, respectively, as $\mathbb{R}(r,\hat{\alpha}_k^r)$ and n_r in (3.2) in which k is suppressed. We determine both the cointegration rank and the lag order as follows:

$$(\hat{r},\hat{k})=\arg\min_{0\le r\le p,1\le k\le K}\mathbb{M}(r,k),$$

where K is a prescribed positive integer. Let k_0 be the true lag order of model (2.1). The theorem below ensures that $(\hat{r},\hat{k})$ is a consistent estimator for (r_0,k_0).

Theorem 3.2. *As $T\to\infty$, $(\hat{r},\hat{k})\xrightarrow{P}(r_0,k_0)$ provided that $g(T)\to 0$ and $Tg(T)\to\ \infty$.*

We denote ECM with different lag orders ($k_1<k_2$) as

$$Model_{k_1}:\Delta Y_t=\gamma\alpha' Y_{t-1}+\Theta X_t+e_t \tag{3.41}$$

$$Model_{k_2}:\Delta Y_t=\gamma\alpha' Y_{t-1}+\Theta X_t+\Theta^* Z_t+e_t \tag{3.42}$$

with $\Theta=(\mu,\Gamma_1,\ldots,\Gamma_{k_1-1})$, $\Theta^*=(\Gamma_{k_1},\ldots,\Gamma_{k_2-1})$, $X_t=(1,\Delta Y_{t-1}',\ldots,\Delta Y_{t-k_1+1}')'$, $Z_t=(\Delta Y_{t-k_1}',\ldots,\Delta Y_{t-k_2+1}')'$.

Lemma 3.3. *For any $1\le k_1<k_2$,*

if $Model_{k_1}$ is true, $\mathbb{R}(r_0,k_1,\alpha_{k_1}^{r_0})-\mathbb{R}(r_0,k_2,\alpha_{k_2}^{r_0})=O_p(\frac{1}{T})$;

if $Model_{k_2}$ is true, $\mathrm{p}\lim_{T\to\infty}[\mathbb{R}(r_0,k_1,\alpha_{k_1}^{r_0})-\mathbb{R}(r_0,k_2,\alpha_{k_2}^{r_0})]>0$, where $\mathrm{p}\lim$ denotes the limit in probability.

Proof. From the expression of $\mathbb{R}(r,\hat{\alpha})$ in (3.1) and the following matrix identity

$$\begin{aligned}&\begin{pmatrix}X_1' & X_2'\end{pmatrix}\begin{pmatrix}A & B\\ B' & D\end{pmatrix}^{-1}\begin{pmatrix}Y_1\\ Y_2\end{pmatrix}\\ &=X_1'A^{-1}Y_1+(X_2'-X_1'A^{-1}B)(D-B'A^{-1}B)^{-1}(Y_2-B'A^{-1}Y_1),\end{aligned} \tag{3.5}$$

it can be seen that

$$\begin{aligned}\mathbb{R}(r_0,k_1,\alpha_{k_1}^{r_0}) &= tr(S_{00}-S_{01}\alpha_{k_1}^{r_0}(\alpha_{k_1}^{r_0\prime}S_{11}\alpha_{k_1}^{r_0})^{-1}\alpha_{k_1}^{r_0\prime}S_{10}),\\ \mathbb{R}(r_0,k_2,\alpha_{k_2}^{r_0}) &= tr(S_{00}-\begin{pmatrix}S_{01}\alpha_{k_1}^{r_0} & S_{02}\end{pmatrix}\begin{pmatrix}\alpha_{k_1}^{r_0\prime}S_{11}^{-1}\alpha_{k_1}^{r_0} & \alpha_{k_1}^{r_0\prime}S_{12}\\ S_{21}\alpha_{k_1}^{r_0} & S_{22}\end{pmatrix}^{-1}\begin{pmatrix}\alpha_{k_1}^{r_0\prime}S_{10}\\ S_{20}\end{pmatrix})\end{aligned}$$

where $S_{ij} = \frac{1}{T}\sum_{t=1}^{T} R_{it}R'_{jt}$ for $i,j = 0,1,2$, $R_{2t} = Z_t - \sum_{t=1}^{T} Z_t X'_t(\sum_{t=1}^{T} X_t X'_t)^{-1} X_t$, $R_{1t} = Y_{t-1} - \sum_{t=1}^{T} Y_{t-1}X'_t(\sum_{t=1}^{T} X_t X'_t)^{-1}X_t$, and $R_{0t} = \Delta Y_t - \sum_{t=1}^{T}\Delta Y_t X'_t(\sum_{t=1}^{T} X_t X'_t)^{-1}X_t$. Therefore,

$$\begin{aligned}
&\mathbb{R}(r_0,k_1,\alpha_{k_1}^{r_0}) - \mathbb{R}(r_0,k_2,\alpha_{k_2}^{r_0}) \\
=&tr[(S_{02} - S_{01}\alpha_{k_1}^{r_0}(\alpha_{k_1}^{r_0\prime}S_{11}\alpha_{k_1}^{r_0})^{-1}\alpha_{k_1}^{r_0\prime}S_{12}) \\
&(S_{22} - S_{21}\alpha_{k_1}^{r_0}(\alpha_{k_1}^{r_0\prime}S_{11}\alpha_{k_1}^{r_0})^{-1}\alpha_{k_1}^{r_0\prime}S_{12})^{-1}(S_{20} - S_{21}\alpha_{k_1}^{r_0}(\alpha_{k_1}^{r_0\prime}S_{11}\alpha_{k_1}^{r_0})^{-1}\alpha_{k_1}^{r_0\prime}S_{10})].
\end{aligned} \tag{3.6}$$

If the model with lag order k_1 is true, replacing Θ in $Model_{k_1}$ by (2.3), we obtain that $R_{0t} = \gamma\alpha_{k_1}^{r_0\prime}R_{1t} + e_t^*$, and

$$S_{02} = \gamma\alpha_{k_1}^{r_0\prime}S_{12} + \tfrac{1}{T}\textstyle\sum_{t=1}^{T} e_t R'_{2t} = \gamma\alpha_{k_1}^{r_0\prime}S_{12} + O_p(\frac{1}{\sqrt{T}}),$$

$$S_{02} - S_{01}\alpha_{k_1}^{r_0}(\alpha_{k_1}^{r_0\prime}S_{11}\alpha_{k_1}^{r_0})^{-1}\alpha_{k_1}^{r_0\prime}S_{12} = (\gamma - \hat{\gamma}(\alpha_{k_1}^{r_0}))\alpha_{k_1}^{r_0\prime}S_{12} + O_p(\tfrac{1}{\sqrt{T}}) = O_p(\tfrac{1}{\sqrt{T}}).$$

Since e_t, ΔY_t and $\alpha_{k_1}^{r_0\prime}Y_{t-1}$ are stationary sequences, it follows that $\frac{1}{T}\sum_{t=1}^{T} e_t R'_{2t} = O_p(\frac{1}{\sqrt{T}})$ and $\alpha_{k_1}^{r_0\prime}S_{12} = O_p(1)$ by the similar way to that of Lemma 2.2. For the term $(\gamma - \hat{\gamma}(\alpha_{k_1}^{r_0}))$, we have

$$\begin{aligned}
\gamma - \hat{\gamma}(\alpha_{k_1}^{r_0}) =&\gamma - S_{01}\alpha_{k_1}^{r_0}(\alpha_{k_1}^{r_0\prime}S_{11}\alpha_{k_1}^{r_0})^{-1} \\
=&\gamma - (\gamma\alpha_{k_1}^{r_0\prime}S_{11} + T^{-1}\sum_{t=1}^{T} e_t R'_{1t})\alpha_{k_1}^{r_0}(\alpha_{k_1}^{r_0\prime}S_{11}\alpha_{k_1}^{r_0})^{-1} \\
=& - T^{-1}\sum_{t=1}^{T} e_t R'_{1t}\alpha_{k_1}^{r_0}(\alpha_{k_1}^{r_0\prime}S_{11}\alpha_{k_1}^{r_0})^{-1} = O_p(1/\sqrt{T}).
\end{aligned}$$

The last equality holds by Lemma 2.2 (c) and (d). It is easy to find that $S_{22} = O_p(1)$, and then

$$S_{22} - S_{21}\alpha_{k_1}^{r_0}(\alpha_{k_1}^{r_0\prime}S_{11}\alpha_{k_1}^{r_0})^{-1}\alpha_{k_1}^{r_0\prime}S_{12} = O_p(1).$$

Then, it follows that $\mathbb{R}(r_0,k_1,\alpha_{k_1}^{r_0}) - \mathbb{R}(r_0,k_2,\alpha_{k_2}^{r_0}) = O_p(\frac{1}{T})$.

If the model with lag order k_2 is true, denoting the limits of

$$S_{02} - S_{01}\alpha_{k_1}^{r_0}(\alpha_{k_1}^{r_0\prime}S_{11}\alpha_{k_1}^{r_0})^{-1}\alpha_{k_1}^{r_0\prime}S_{12} \quad \text{and} \quad S_{22} - S_{21}\alpha_{k_1}^{r_0}(\alpha_{k_1}^{r_0\prime}S_{11}\alpha_{k_1}^{r_0})^{-1}\alpha_{k_1}^{r_0\prime}S_{12}$$

by E and G respectively, we argue that $tr(EG^{-1}E') > 0$ by the similar way to that given by Aznar and Salvador (2002). Hence, by (3.6), $\text{p}\lim_{T\to\infty}[\mathbb{R}(r_0,k_1,\alpha_{k_1}^{r_0}) - \mathbb{R}(r_0,k_2,\alpha_{k_2}^{r_0})] > 0$. □

Proof of Theorem 3.2. The goal is to verify that $P(\hat{r} = r_0, \hat{k} = k_0) \to 1$ as $T \to \infty$. Note that we have established the consistency of $\hat{r}$ for any fixed lag order k in Theorem 3.1, which implies that $P(\hat{r} = r_0) \to 1$ as $T \to \infty$. Thus, it remains to prove that $P(\hat{k} = k_0|\hat{r} = r_0) \to 1$, or equivalently, for all $1 \le k \le K$ and $k \ne k_0$,

$$\lim_{T\to\infty} P(\mathbb{M}(r_0,k) - \mathbb{M}(r_0,k_0) < 0) = 0.$$

From the proof of Lemma 3.2, we have $\mathbb{R}(r_0, k, \hat{\alpha}_k^{r_0}) - \mathbb{R}(r_0, k, \alpha_k^{r_0}) = O_p(\frac{1}{T})$ for any $k \geq 1$. Therefore,

$$
\begin{aligned}
&\mathbb{M}(r_0, k) - \mathbb{M}(r_0, k_0)\\
&= \mathbb{R}(r_0, k, \hat{\alpha}_k^{r_0}) - \mathbb{R}(r_0, k_0, \hat{\alpha}_{k_0}^{r_0}) + p^2(k - k_0)g(T)\\
&= \mathbb{R}(r_0, k, \alpha_k^{r_0}) - \mathbb{R}(r_0, k_0, \alpha_{k_0}^{r_0}) + p^2(k - k_0)g(T) + O_p(\frac{1}{T}).
\end{aligned}
$$

For $k < k_0$, it holds that if $g(T) \to 0$ as $T \to \infty$,

$$
\begin{aligned}
&P(\mathbb{M}(r_0, k) - \mathbb{M}(r_0, k_0) < 0)\\
&= P(\mathbb{R}(r_0, k, \alpha_k^{r_0}) - \mathbb{R}(r_0, k_0, \alpha_{k_0}^{r_0}) + O_p(\frac{1}{T}) < p^2(k_0 - k)g(T)) \to 0,
\end{aligned}
$$

because $\mathbb{R}(r_0, k, \alpha_k^{r_0}) - \mathbb{R}(r_0, k_0, \alpha_{k_0}^{r_0})$ has a positive limit by Lemma 3.3.

For $k > k_0$, Lemma 3.3 implies that $\mathbb{R}(r_0, k_0, \alpha_{k_0}^{r_0}) - \mathbb{R}(r_0, k, \alpha_k^{r_0}) = O_p(\frac{1}{T})$. Thus, if $Tg(T) \to \infty$ as $T \to \infty$, we have

$$
\begin{aligned}
&P(\mathbb{M}(r_0, k) - \mathbb{M}(r_0, k_0) < 0)\\
&= P(T[\mathbb{R}(r_0, k_0, \alpha_{k_0}^{r_0}) - \mathbb{R}(r_0, k, \alpha_k^{r_0})] + O_p(1) > p^2(k - k_0)Tg(T)) \to 0.
\end{aligned}
$$

The proof is completed. □

4 Numerical properties

4.1 Simulated examples

Two experiments are conducted to examine the finite sample performance of the proposed criteria (3.2) and (3.3). The comparisons with the LRT approach of Johansen (1991) and the information criterion of Aznar and Salvador (2002) are also made. It is easy to see from Theorems 3.1 and 3.2 that the choice of the penalty function $g(\cdot)$ is flexible. It may take a general form

$$
g(T) = \xi \ln(T)/T + 2\eta \ln(\ln(T))/T, \qquad \xi \geq 0,\ \eta \geq 0, \tag{4.1}
$$

which reduces to the BIC of Schwarz (1978) with $\xi = 1$ and $\eta = 0$, to the HQIC of Hannan and Quinn (1979) with $\xi = 0$ and $\eta = 1$, and to the LCIC of Gonzalo and Pitarakis (1998) with $\xi = \eta = \frac{1}{2}$. The motivation for introducing this criterion is to overcome excessive parsimony or overranking in finite samples. For exposition, we use the three concrete forms in the first experiment:

$$
\begin{aligned}
\mathbb{M}_1(r, k) &= \mathbb{R}(r, k, \hat{\alpha}_k^r) + n_{r,k} \ln(T)/T,\\
\mathbb{M}_2(r, k) &= \mathbb{R}(r, k, \hat{\alpha}_k^r) + 2n_{r,k} \ln\{\ln(T)\}/T,\\
\mathbb{M}_3(r, k) &= \mathbb{R}(r, k, \hat{\alpha}_k^r) + n_{r,k}[\ln(T)/6 + 4\ln\{\ln(T)\}/3]/T.
\end{aligned}
$$

In practice, however, best ξ and η are chosen via replicated simulations and then used to fit real data. We set sample size at $T = 30$, 50, 100, 200, 300 or 400. For each setting, we replicate the simulation 2000 times. The data are generated from the ECM (2.1) with either independent errors following one of the four distributions below

$$e_t \sim N(0, I_p), \tag{4.1a}$$

$$e_t = \varepsilon_t + 10\theta\varepsilon_t, \quad \varepsilon_t \sim N(0, I_p), \quad \theta \sim Poisson(\tau), \tag{4.1b}$$

$$e_{it} \sim t(q), \tag{4.1c}$$

$$e_{it} \sim Cauchy, \tag{4.1d}$$

or uncorrelated but dependent errors

$$e_{it} = h_{it}\varepsilon_{it}, \quad h_{it}^2 = \varphi_0 + \varphi_1 e_{it-1}^2 + \psi_1 h_{it-1}^2, \quad \varepsilon_{it} \sim N(0,1), \tag{4.2}$$

$$\varphi_0 > 0,\ \varphi_1 \geq 0,\ \psi_1 \geq 0,\ \varepsilon_{it} \ \text{are independent for all } i \text{ and } t.$$

Distributions in (4.1b) – (4.1d) are heavy-tailed. In particular, (4.1b) is often used in GARCH-Jump models for modelling asset prices. Note that for $e_{it} \sim t(q)$, $E|e_{it}|^q = \infty$. Furthermore, (4.1d) represents an extreme situation with $E|e_{it}| = \infty$, and therefore it does not fulfill Assumption B. We include it to examine the robustness of the methods against the assumption of the finite fourth moment.

Experiment I.

First we generate data from model

$$y_{1t} = \mu + 0.6y_{2t} + e_{1t}, \qquad \Delta y_{it} = \mu + e_{it} \ \text{ for } \ i = 2, 3. \tag{4.3}$$

The cointegration rank $r = 1$ and the lag order $k = 1$.

Assuming $(r, k) = (1, 1)$ is known, we estimate the cointegration vector $\hat{\alpha}$ by using the new approach suggested in Section 2 and Johansen's MLE respectively. For comparison of the precision in estimation, a measuring rule is given as

$$Er = \|\hat{\alpha} - \tilde{\alpha}\| \tag{4.4}$$

where $\tilde{\alpha} = \alpha(\alpha' S_{11}\alpha)^{-\frac{1}{2}}$, and the true cointegration vector α is normalized to satisfy $\tilde{\alpha}' S_{11}\tilde{\alpha} = I_p$. Hence, a better estimator is supposed to lead smaller Er in (4.4).

For each of different settings (4.1a) – (4.1d), we conduct $n = 2000$ replications from (4.3), the Ers are calculated and portrayed in Figure 1. Each point in one subplot has coordinates

$$(Er_j \cos(\frac{\pi}{2n}j), Er_j \sin(\frac{\pi}{2n}j)) \qquad j = 1, \ldots n.$$

Figure 1 contains five couples of Ers for model (4.3) with $e_t \sim N(0, I_3)$, $e_t|\theta \sim N(0, (1+100\theta^2)I_3)$ and $\theta \sim Poisson(1)$, $e_t \sim t(3)$, $e_t \sim Cauchy$ from the top down. The left-hand (a) – (d) are obtained by using our method described in section 2. The right-hand (a') – (d') are obtained by MLE. The sample size $T = 30$ and $\mu = 0.5$. This figure shows that only when the errors are Gaussian ((a), (a')), MLE outperforms our method in estimating the cointegration vectors α. But for errors with Poisson jumps ((b), (b') or heavy tails ((c), (c') and (d), (d')), our method performs better. Similar conclusions can be drawn from figures[4] with sample size $T = 50$, $T = 100$ and $T = 200$, which supports the robustness of our method.

Only assuming $k = 1$ is known, we estimate r by minimizing $\mathbb{M}_i(r, 1)$ for $i = 1, 2, 3$ and also by the Johansen's LRT approach. We conduct 2000 replications from (4.3), the percentages of the replications resulting the correct estimate (i.e. $\hat{r} = 1$) are listed in Tables 1 – 3. Table 1 shows that even with Gaussian errors, our method based on the criterion $\mathbb{M}_3$ outperforms the LRT based method. When the sample size is small (i.e. $T = 30$ or 50), the methods using $\mathbb{M}_1$ and $\mathbb{M}_2$ perform poorly. However the performance improves when T increases. Also noticeable is the fact that the presence of a linear trend (i.e. $\mu \neq 0$) deteriorates slightly the performance of all the four methods. Tables 2 – 3 show that the method based on $\mathbb{M}_3$ remains to perform better than the others when error distribution is changed to (4.1b), (4.1c) and (4.1d), although the heavy tails of the error distribution impact negatively to the performance of all the methods. Especially with Cauchy errors, the percentages of the correct estimates are low for all the four method with sample size T smaller than 100. But still the method based on $\mathbb{M}_3$ always performs better than the other three. Table 4 indicates that the method based on $\mathbb{M}_3$ also outperforms the others even with dependent ARCH(1) (i.e. $\psi_1 = 0$) or GARCH(1,1) errors (i.e. $\psi_1 \neq 0$).

Experiment II

Our second example concerns the model

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} + \begin{pmatrix} 0.3 & 0 \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} \Delta y_{1t-1} \\ \Delta y_{2t-1} \end{pmatrix} + \begin{pmatrix} 0.4 \\ 0.6 \end{pmatrix} \begin{pmatrix} 1, & -2 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}. \quad (4.5)$$

We assume that all the coefficients in the models are unknown. We now estimate the cointegration rank r(=1) and the lag order k(=2) by minimizing $\mathbb{M}_3(r, k)$ with the five different error distributions specified in (4.1a)-(4.1d) and (4.2). For the comparison purpose, we also compute Aznar and Salvador's estimates (Aznar and Salvador (2002)) obtained by minimizing the information

[4] In consideration of paper length, we do not show them all.

criterion (IC)

$$\mathrm{IC}(r,k) = T\{\ln|S_{00}| + \sum_{i=1}^{r} \ln(1-\lambda_i) + n_{r,k} g(T)\},$$

where $g(T) = [\ln(T)/6 + 4\ln\{\ln(T)\}/3]/T$ and λ_i is the i-th largest generalized eigenvalue of $S_{10}S_{00}^{-1}S_{01}$ with respect to S_{11}. The percentages of the correct estimates (i.e. $(\hat{r},\hat{k}) = (1,2)$) in a simulation with 2000 replications are listed in Table 5. Note that the above IC-criterion is based on a Gaussian likelihood function. It is not surprising that it outperforms our method based on $\mathbb{M}_3$ when the errors are Gaussian. However Table 5 also indicates that this IC-criterion is sensitive to the normality assumption. In fact for all the four other error distributions, our method based on $\mathbb{M}_3$ performed better. When the heaviness of the distribution tails increases, the performance of the both methods decreases. We also note that both methods perform poorly when the sample size is as small as $T = 30$.

4.2 A real data example

We consider the annual records of the GDP per capita, labor productivity per person and labor productivity per hour of the Netherlands from 1950 to 2005[5]. The time plots of the logarithmic GDP (solid lines), the labor productivity per person (dash-dotted lines) and the labor productivity per hour (dotted lines) are presented in Figure 2. It indicates that there may exist a linear cointegrating relationship among the three variables.

We determine the cointegration rank by minimising $\mathbb{M}_3(r,k)$. The surface of $\mathbb{M}_3(r,k)$ is plotted against r and k in Figure 3. The minimal point of the surface is attained at $(r,k) = (1,2)$, leading to a fitted ECM model (2.1) for this data set with the lag order 2 and the cointegrating rank 1. The estimate of the cointegrating vector with the first component normalized to one is $\hat{\alpha} = (1.00, 3.82, -3.28)'$. The other estimated coefficients in model (2.1) are as follows

$$\hat{\mu} = (9.09, 10.09, 2.41)', \quad \hat{\gamma} = -(0.23, 0.25, 0.06)', \quad \hat{\Gamma}_1 = \begin{pmatrix} 0.20 & -0.32 & 0.60 \\ -0.36 & 0.19 & 0.55 \\ -0.48 & 0.32 & 0.46 \end{pmatrix}.$$

References

[1] Ahn, S. K. & Reinsel, G. C. (1988) Nested reduced-rank autoregressive models for multiple time series, *Journal of the American Statistical Association*, 83: 849-856.

[5]Data source: The Conference Board and Groningen Growth and Development Center, Total Economy Database, January 2006, http://www.ggdc.net.

[2] Ahn, S. K. & Reinsel, G. C. (1990) Estimation for partially nonstationary multivariate autoregressive models, *Journal of the American Statistical Association*, 85: 813-823.

[3] Anderson, T. W. (1951) Estimating linear restrictions on regression coffcients for multivariate normal distributions, *Annals of Mathematical Statistics*, 22: 327-351.

[4] Aznar, A. & Salvador, M. (2002) Selecting the rank of the cointegration space and the form of the intercept using an information criterion, *Econometric Theory*, 18: 926-947.

[5] Bai, J. S. (2003) Inferential theory for factor models of large dimensions, *Econometrica*, 71: 926-947.

[6] Bossaerts, P. (1988) Common nonstationary components of asset prices, *Journal of Economic Dynamics and Control*, 12: 347-364.

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

[8] Gonzalo, J. (1994) Five alternative methods of estimating long-run equilibrium relationships, *Journal of Econometrics*, 60: 203-233.

[9] Gonzalo, J. & Lee, T. H. (1998) Pitfalls in testing for long-run relationships, *Journal of Econometrics*, 86: 129-154.

[10] Gonzalo, J. & Pitarakis, J. Y. (1998) Specification via model selection in vector error correction models, *Economics Letters*, 60: 321-328.

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

[12] Granger, C. W. J. & Weiss, A. A. (1983) Time series analysis of error correction models, in *Studies in Econometrics, Time Series and Multivariate Statistics (eds. by Karlin, S., Amemiya T., Goodman L.A.)* New York: Academic Press, 1983, 255-278.

[13] Hannan, E. & Quinn, B. (1979) The determination of the order of an autoregression, *Journal of the Royal Statistical Society, Ser. B*, 41: 190-195.

[14] Hubrich, K., Lütkepohl H. & Saikkonen, P. (2001) A Review of Systems Cointegrating Tests, *Econometric Reviews*, 20: 247-318.

[15] Huang, B. N. & Yang, C. W. (1996) Long-run purchasing power parity revisited: a Monte Carlo simulation, *Applied Economics*, 28: 967-974.

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

[17] Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, 59: 1551-1580.

[18] Kapetanios, G. (2004) The asymptotic distribution of the cointegration rank estimator under the Akaike Information Criterion, *Econometric Theory*, 20: 735-742.

[19] Park, J. Y. & Phillips, P. C. B. (1988) Statistical inference in regressions with integrated processes: Part 1., *Econometric Theory*, 4: 468-497.

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

[21] Phillips, P. C. B & Durlauf, S. N. (1986) Multiple time series regression with integrated processes, *The Review of Economic Studies*, 53: 473-495.

[22] Schwarz, G. (1978) Estimating the dimension of a model, *Annals of Statistics*, 6: 461-464.

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

[24] Stock, J. H.& Watson, M. (1988) Testing for common trends, *Journal of the American Statistical Association*, 83: 1097-1107.

[25] Wang, Z. J. & Bessler, D., A monte carlo study on the selection of cointegrating rank using information criteria, *Econometric Theory*, 21: 593-620.

Table 1: Percentages of the correct estimates for r for model (4.3) with $e_t \sim N(0, I_3)$

	$\mu = 0$				$\mu = 0.5$			
	T=30	T=50	T=100	T=200	T=30	T=50	T=100	T=200
$\mathbb{M}_1$	16.95	34.70	88.05	100	14.05	32.70	86.90	100
$\mathbb{M}_2$	42.05	72.20	99.70	100	32.45	67.60	99.65	100
$\mathbb{M}_3$	85.70	97.45	99.55	99.85	83.35	95.55	99.00	99.40
LRT	81.00	94.90	95.55	95.75	74.20	94.10	95.55	95.70

Table 2: Percentages of the correct estimates for r for model (4.3) with $\mu = 0.5$ and e_t defined as in (4.1b)–(4.1c)

		(4.1b)	with	$\tau = 1$		(4.1c)	with	$q = 3$
	T=30	T=50	T=100	T=200	T=30	T=50	T=100	T=200
$\mathbb{M}_1$	27.55	43.05	84.95	99.85	28.80	43.15	79.45	92.55
$\mathbb{M}_2$	52.45	71.35	97.85	100	50.60	72.80	94.15	95.70
$\mathbb{M}_3$	77.30	89.70	96.95	99.00	77.25	89.35	93.65	94.15
LRT	71.65	87.70	90.05	89.30	71.00	88.90	91.65	92.45

Table 3: Percentages of the correct estimates for r for model (4.3) with $\mu = 0.5$ and Cauchy errors

	T=30	T=50	T=100	T=200	T=300	T=400
$\mathbb{M}_1$	7.15	3.95	2.30	4.45	10.20	25.40
$\mathbb{M}_2$	17.75	15.45	14.70	34.30	62.60	86.65
$\mathbb{M}_3$	43.00	40.70	46.55	71.65	82.25	84.75
LRT	35.35	40.25	43.75	66.40	67.90	67.95

Table 4: Percentages of the correct estimates for r for model (4.3) with $\mu = 0.5$, e_t defined as in (4.2), $\varphi_0 = 0.1$ and $\varphi_1 = 0.6$

$\psi_1 = 0$	T=30	T=50	T=100	T=200	T=300	T=400
$\mathbb{M}_1$	29.20	41.75	75.55	80.20	85.00	91.15
$\mathbb{M}_2$	49.50	68.55	85.15	91.65	93.95	95.70
$\mathbb{M}_3$	75.00	86.90	95.75	96.35	95.20	96.95
LRT	68.30	82.30	84.05	85.40	85.50	86.25
$\psi_1 = 0.2$	T=30	T=50	T=100	T=200	T=300	T=400
$\mathbb{M}_1$	29.80	41.35	71.40	75.25	83.05	89.75
$\mathbb{M}_2$	49.70	64.40	83.75	90.20	91.85	93.70
$\mathbb{M}_3$	74.45	83.65	93.35	94.25	94.80	95.75
LRT	66.50	79.55	82.40	84.90	85.15	85.95

Table 5: Percentages of the correct estimates for (r, k) for model (4.5)

T	30	50	100	200	300	400
Independent $N(0, I_2)$ errors						
$\mathbb{M}_3$	10.20	31.75	67.65	82.35	89.20	93.65
IC	24.35	38.90	71.40	90.75	92.05	94.10
Independent errors (4.1b) with $\tau = 1$						
$\mathbb{M}_3$	10.05	30.10	64.30	81.45	89.55	92.80
IC	9.25	19.80	52.15	74.55	79.80	84.15
Independent t-distributed errors (4.1c) with $q = 3$						
$\mathbb{M}_3$	9.15	28.70	58.95	77.40	86.05	90.00
IC	8.30	21.65	49.60	75.20	81.70	82.10
Independent Cauchy errors						
$\mathbb{M}_3$	6.45	20.80	44.70	70.65	82.30	85.80
IC	5.25	19.75	31.85	59.15	67.50	70.35
ARCH(1) errors (4.2) with $\psi_1 = 0$, $\varphi_0 = 0.1$ and $\varphi_1 = 0.6$						
$\mathbb{M}_3$	12.75	30.45	60.85	81.75	87.90	91.25
IC	8.15	19.20	48.65	71.50	72.30	80.25
GARCH(1,1) errors (4.2) with $\psi_1 = 0.2$, $\varphi_0 = 0.1$ and $\varphi_1 = 0.6$						
$\mathbb{M}_3$	11.35	28.95	57.90	80.15	86.00	90.05
IC	7.00	14.40	42.55	69.10	71.60	78.35

Figure 1: Ers for model (4.3) with $\mu = 0.5$ and sample size $T = 30$

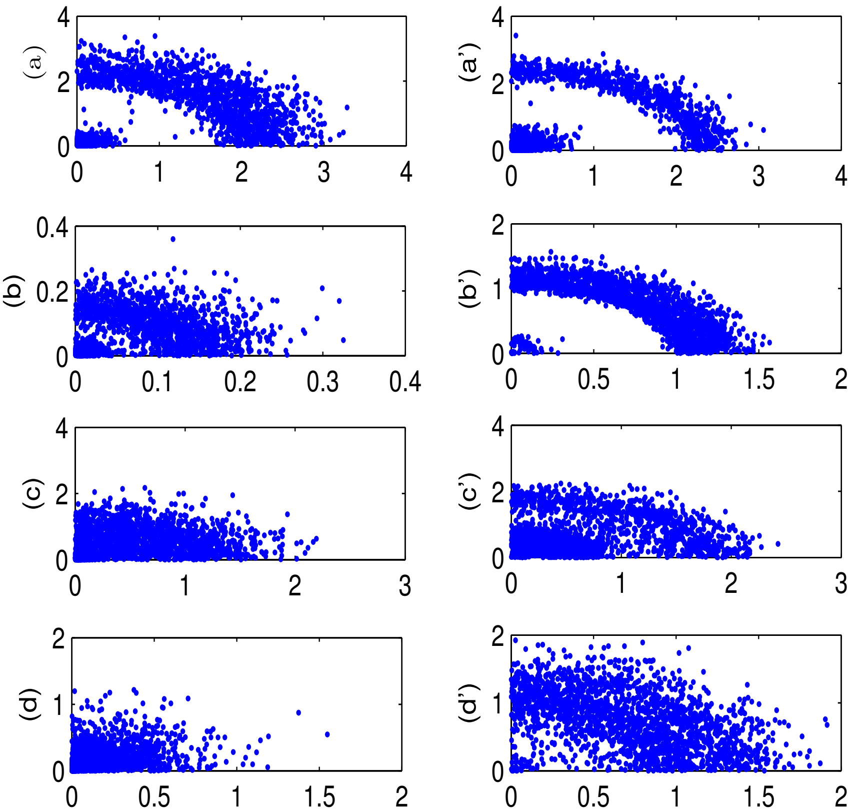

Figure 2: Time plot of logarithmic GDP per capita, labor productivity per person and labor productivity per hour of the Netherlands

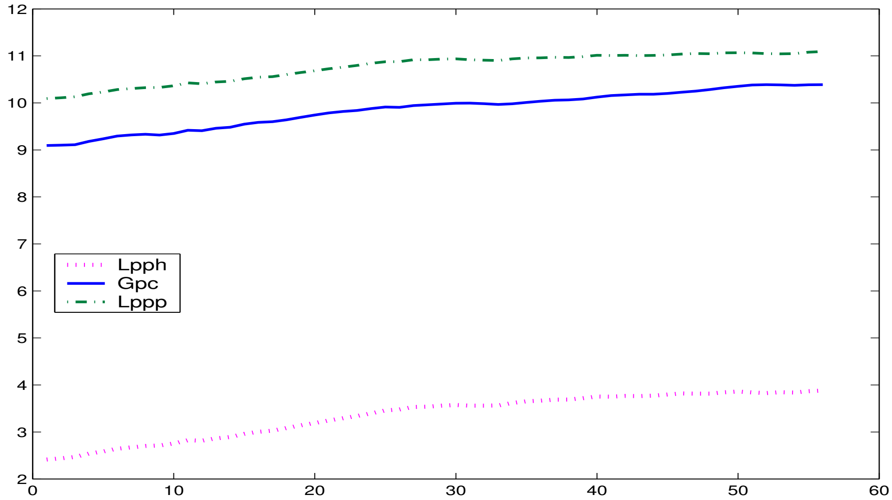

Figure 3: Plot $\mathbb{M}_3(r, k)$ against the cointegration rank r and the lag order k