On Determination of Cointegration Ranks*

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

[1]*Peking University* [2]*London School of Economics*

Abstract

We propose a new method to determine the cointegration rank in the error correction model of Engle and Granger (1987). To this end, we first estimate the cointegration vectors in terms of a *residual-based principal component analysis*. Then the cointegration rank, together with the lag order, is determined by a penalized goodness-of-fit measure. We have shown that the estimated cointegration vectors are asymptotically normal, 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, and 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.

JEL Classification: C32; C51

Keywords: Cointegration; error correction models; residual-based principal components; penalized goodness-of-fit criteria; model selection

*Partially supported by an NNSF research grant of China 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 and 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 (Gonzalo and Lee 1998, and 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 (Huang and Yang 1996). Further developments under the assumption of i.i.d. Gaussion errors include Aznar and Salvador (2002) which proposed to determine the cointegration ranks by minimizing appropriate information criteria. More recently, Kapetanios (2004) established the asymptotic distribution

of the estimate for the cointegration rank obtained by AIC.

In this paper we propose a new approach 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 adopted in the papers cited above. We first estimate the cointegration vectors via a residual-based principle components analysis (RPCA). In fact the RPCA may be viewed as a special case of the reduced rank regression technique introduced by Anderson (1951); see also Ahn and Reinsel (1988, 1990), Johansen (1988, 1991), and Bai (2003). We then determine the cointegration rank by minimising 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 based on RPCA are asymptotically normal with the standard root-T convergence rate. 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 RPCA estimation for cointegration vectors and its asymptotic properties are presented in section 2. Section 3 presents an information 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 of Engle and Granger (1987) 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 $\|A\| = [\mathrm{tr}(A'A)]^{1/2}$ denote the norm of matrix A. Some regularity conditions are now in order.

Assumption A. The process Y_t satisfies the basic assumptions of the Granger representation theorem given by 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 $\{e_t\}$ satisfies the requirements of Multivariate Invariance Principle of Phillips and Durlauf (1986). 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 Estimation via RPCA

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

We may estimate the parameters in model (2.1) by solving the optimization problem

$$\min_{\Theta,\gamma,\alpha} \frac{1}{T}\sum_{t=1}^{T}(\Delta Y_t - \Theta X_t - \gamma\alpha' Y_{t-1})'(\Delta Y_t - \Theta X_t - \gamma\alpha' Y_{t-1}), \tag{2.2}$$

where $\Theta = (\mu, \Gamma_1, \ldots, \Gamma_{k-1})$, $X_t = (1, \Delta Y'_{t-1}, \ldots, \Delta Y'_{t-k+1})'$. Although this can be considered as a standard least square problem, we are unable to derive an explicit solution for α even with the regularity condition to make it identifiable.

To motivate our RPCA approach, we first assume that $\gamma\alpha'$ is given. Then (2.2) reduces to

$$\min_{\Theta(\gamma\alpha')} \frac{1}{T}\sum_{t=1}^{T}(\Delta Y_t - \Theta(\gamma\alpha') X_t - \gamma\alpha' Y_{t-1})'(\Delta Y_t - \Theta(\gamma\alpha') X_t - \gamma\alpha' Y_{t-1}), \tag{2.3}$$

which admits the solution

$$\Theta(\gamma\alpha') = \Theta_1 - \gamma\alpha'\Theta_2, \tag{2.4}$$

where

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

Now replacing Θ in (2.2) by (2.4), (2.2) reduces to

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

where $R_{0t} = \Delta Y_t - \Theta_1 X_t$, $R_{1t} = Y_{t-1} - \Theta_2 X_t$. 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

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

It is easy to see 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

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

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 R_{0t} and R_{1t} are, respectively, the residuals of ΔY_t and Y_{t-1} regressing on X_t, $\hat{\alpha}_1, \cdots, \hat{\alpha}_r$ may be viewed as the residual-based principle components. Also note that $\hat{\gamma} = S_{01}\hat{\alpha}$ is the cointegration loading matrix.

Gonzalo (1994) compared numerically the five different methods for estimating the cointegrating vectors: ordinary least squares (Engle and Granger 1987), non-linear 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 residual-based principal components estimator proposed in this paper tends to overcome these shortcomings.

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

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 $\Psi(1) = I + \Psi_1 + \Psi_2 + \cdots$. If $\{j\Psi_j\}_{j=0}^{\infty}$ is assumed to be absolutely summable, it holds that

$$\alpha'\Psi(1) = 0,$$

and (2.8) implies that

$$Y_t = Y_0 + \delta t + \Psi(L)(e_t + e_{t-1} + \cdots + e_1). \tag{2.9}$$

We introduce two technical lemmas first. The proof of Lemma 2.1 may be found found in Phillips and Durlauf (1986), and Park and Phillips (1988). The notation " $\Rightarrow$ " denotes the week convergence, and " $\xrightarrow{P}$ " denotes the convergence in probability. We always assume that Assumptions A & B hold and Y_t satisfies (2.9) in the sequel of this subsection. Note that (2.9) is implied by Assumption A.

Lemma 2.1. *As $T \to \infty$, it holds that*
(a) $T^{-3/2}\sum_{t=1}^{T}(Y_t - \delta t) \Rightarrow \Psi(1)\int_0^1 W(s)ds$,
(b) $T^{-2}\sum_{t=1}^{T}(Y_t - \delta t)(Y_t - \delta t)' \Rightarrow \Psi(1)\int_0^1 W(s)W(s)'ds\Psi(1)'$,
(c) $T^{-1}\sum_{t=1}^{T}(Y_{t-1} - \delta(t-1))e_{t-j}' \Rightarrow \Psi(1)\int_0^1 W(s)dW(s)' + \Psi(1)\Omega$,
(d) $T^{-5/2}\sum_{t=1}^{T}t(Y_t - \delta t)' \Rightarrow \int_0^1 sW(s)'ds\Psi(1)'$, *and*
(e) $T^{-3/2}\sum_{t=1}^{T}te_{t-j}' \Rightarrow \int_0^1 sdW(s)'$,
where $W(s)$ is a vector Wiener process on $C[0,1]^p$ with covariance matrix $\Omega = E(e_t e_t')$.

Lemma 2.2. *As $T \to \infty$,*
(a) $S_{11}^{-1} = O_p(T^{-2})$, $T^{-3/2}\sum_{t=1}^{T} e_t R_{1t}' = O_p(1)$,
(b) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \alpha' R_{1t} e_t' = O_p(1)$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \hat{\alpha}' R_{1t} e_{it} \Rightarrow N(0, \sigma_i^2 I_r)$,

(c) $\alpha' S_{11}\alpha \xrightarrow{P} \Sigma_{11}$, $\alpha' S_{11}\hat{\alpha} = O_p(1)$, *where* Σ_{11} *is some positive definite matrix, and*
(d) $V_T \xrightarrow{P} V$, *where* $V_T = diag(\lambda_{1T}, \lambda_{1T}, \ldots, \lambda_{rT})$ *and* $V = diag(\lambda_1, \lambda_2, \ldots, \lambda_r)$, $\lambda_{1T} \geq \cdots \geq \lambda_{rT}$ *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.*

Proof. (a) Note that

$$\begin{aligned}\frac{1}{T^2}S_{11} &= \frac{1}{T^3}\sum_{t=1}^{T} Y_{t-1}Y'_{t-1} - \frac{1}{T^2}\sum_{t=1}^{T} Y_{t-1}X'_t(\frac{1}{T}\sum_{t=1}^{T} X_tX'_t)^{-1}\frac{1}{T^2}\sum_{t=1}^{T} X_tY'_{t-1} \\ \frac{1}{T^{3/2}}\sum_{t=1}^{T} R_{1t}e'_t &= \frac{1}{T^{3/2}}\sum_{t=1}^{T} Y_{t-1}e'_t - \frac{1}{T^2}\sum_{t=1}^{T} Y_{t-1}X'_t(\frac{1}{T}\sum_{t=1}^{T} X_tX'_t)^{-1}\frac{1}{T^{1/2}}\sum_{t=1}^{T} X_te'_t.\end{aligned}$$

From Lemma 2.1, it can be easily seen that

$$\begin{aligned}T^{-3}\sum\nolimits_{t=1}^{T} Y_{t-1}Y'_{t-1} = O_p(1), \quad T^{-2}\sum\nolimits_{t=1}^{T} Y_{t-1} = O_p(1) \\ T^{-2}\sum\nolimits_{t=1}^{T} Y_{t-1}\Delta Y'_{t-j} = O_p(1), \quad T^{-3/2}\sum\nolimits_{t=1}^{T} Y_{t-1}e'_t = O_p(1)\end{aligned}$$

hold for $j = 1, \ldots, k-1$ if $\delta \neq 0$, which is a general assumption of the moving average representation (2.8) for a cointegrated system with constant terms[2]. Meanwhile, $\frac{1}{T}\sum_{t=1}^{T} X_tX'_t = O_p(1)$ holds obviously since ΔY_t is stationary with finite fourth moments under Assumption B. Therefore, $S_{11}^{-1} = O_p(T^{-2})$.

Then, let us focus on $T^{-1/2}\sum_{t=1}^{T} X_te'_t$. Note that each column of $\{\Delta Y_{t-j}e'_t\}$ is a vector martingale difference sequence. By the central limit theorem for martingale difference sequences (Hall and Heyde (1980)), $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \Delta Y_{t-j}e_{it} \Rightarrow N(0, \sigma_i^2\Omega_1)$, $j = 1, \ldots, k-1$, where $\Omega_1 = E(\Delta Y_{t-j}\Delta Y'_{t-j})$. But, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_{it} \Rightarrow N(0, \sigma_i^2)$. Then, $T^{-3/2}\sum_{t=1}^{T} e_tR'_{1t} = O_p(1)$.

(b) The ith column of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \alpha' R_{1t}e'_t$ is

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \alpha' R_{1t}e_{it} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \alpha' Y_{t-1}e_{it} - \frac{1}{T}\sum_{t=1}^{T} \alpha' Y_{t-1}X'_t(\frac{1}{T}\sum_{t=1}^{T} X_tX'_t)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_te_{it},$$

where $\{\alpha' Y_{t-1}\}$ is a stationary vector series. The first term on the right hand side of the above equality converges weakly to $N(0, \sigma_i^2\Omega_0)$ with $\Omega_0 = E(\alpha' Y_{t-1}Y'_{t-1}\alpha)$ by the central limit theorem for martingale difference sequences under Assumption

[2]For $\delta = 0$, it can be proved that $T^{-1}S_{11} = O_p(1)$.

B. For the second term, $(\frac{1}{T}\sum_{t=1}^{T} X_t X_t')^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t e_{it} = O_p(1)$ has already been proved above in part (a), and $\frac{1}{T}\sum_{t=1}^{T}\alpha' Y_{t-1} = O_p(1)$. Now we need to verify $\frac{1}{T}\sum_{t=1}^{T}\alpha' Y_{t-1}\Delta Y_{t-j}' = O_p(1), j = 1, \ldots, k-1$. Note that

$$\frac{1}{T}\sum_{t=1}^{T}\alpha' Y_{t-1}\Delta Y_{t-j}' = \frac{1}{T}\sum_{t=1}^{T}\alpha'(Y_{t-1} - \delta(t-1))(\Delta Y_{t-j} - \delta)' + \frac{1}{T}\sum_{t=1}^{T}\alpha'\delta(t-1)(\Delta Y_{t-j} - \delta)' + \frac{1}{T}\sum_{t=1}^{T}\alpha' Y_{t-1}\delta'.$$

It follows from Lemma 2.1(c) and $\alpha'\Psi(1) = 0$ that

$$\frac{1}{T}\sum_{t=1}^{T}\alpha'(Y_{t-1} - \delta(t-1))(\Delta Y_{t-j} - \delta)' \Rightarrow \alpha'\Psi(1)(\int_0^1 W(s)dW(s)' + \Omega)\Psi(1)' = 0.$$

And we have $\frac{1}{T}\sum_{t=1}^{T}\alpha'\delta(t-1)(\Delta Y_{t-j} - \delta)' = 0$ since $\delta = \Psi(1)\mu$. Then, $\frac{1}{T}\sum_{t=1}^{T}\alpha' Y_{t-1}\Delta Y_{t-j}'$ $= O_p(1)$. Therefore, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\alpha' R_{1t} e_t = O_p(1)$.

For the second relation in part (b), we can see that $\{\hat{\alpha}' R_{1t} e_{it}\}$ is a martingale difference sequence, and

$$E[\hat{\alpha}' R_{1t} e_{it}(\hat{\alpha}' R_{1t} e_{it})'] = \sigma_i^2 E(\hat{\alpha}' R_{1t} R_{1t}'\hat{\alpha}) \triangleq \Lambda_t.$$

But, Λ_t is a positive definite matrix satisfying

$$\frac{1}{T}\sum_{t=1}^{T}\Lambda_t = \sigma_i^2 E(\hat{\alpha}' S_{11}\hat{\alpha}) = \sigma_i^2 I_r \triangleq \Lambda,$$

since $\hat{\alpha}' S_{11}\hat{\alpha}$ is an identity matrix with rank r. If we can show that

$$\frac{1}{T}\sum_{t=1}^{T}\hat{\alpha}' R_{1t} e_{it}(\hat{\alpha}' R_{1t} e_{it})' \xrightarrow{P} \Lambda, \tag{2.10}$$

then $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\alpha}' R_{1t} e_{it} \Rightarrow N(0,\Lambda)$ holds. To verify (2.10), we notice that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\hat{\alpha}' R_{1t} e_{it}(\hat{\alpha}' R_{1t} e_{it})' &= \frac{1}{T}\sum_{t=1}^{T} e_{it}^2\hat{\alpha}' R_{1t} R_{1t}'\hat{\alpha} \\ &= \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - \sigma_i^2)\hat{\alpha}' R_{1t} R_{1t}'\hat{\alpha} + \frac{1}{T}\sum_{t=1}^{T}\sigma_i^2\hat{\alpha}' R_{1t} R_{1t}'\hat{\alpha}\end{aligned}$$

where $\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - \sigma_i^2)\hat{\alpha}' R_{1t} R_{1t}'\hat{\alpha} \xrightarrow{P} 0$ by the law of large numbers for martingale difference sequence, and the second term equals to $\sigma_i^2 I_r$. Thus, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\alpha}' R_{1t} e_{it} \Rightarrow$ $N(0,\sigma_i^2 I_r)$.

(c) Since

$$\alpha' S_{11}\alpha = \frac{1}{T}\sum_{t=1}^{T}\alpha' Y_{t-1}Y_{t-1}'\alpha - \frac{1}{T}\sum_{t=1}^{T}\alpha' Y_{t-1}X_t'(\frac{1}{T}\sum_{t=1}^{T}X_tX_t')^{-1}\frac{1}{T}\sum_{t=1}^{T}X_tY_{t-1}'\alpha,$$

using Lemma 2.1 again and noticing that $\alpha'\Psi(1) = 0, \alpha'\delta = 0$, and $\alpha' Y_{t-1}$ is stationary, we have $\alpha' S_{11}\alpha \xrightarrow{P} \Sigma_{11}$.

For the second relation in part (c), letting R_1 be the $T\times p$ matrix $[R_{11}, R_{12}, \cdots, R_{1T}]'$, we have

$$\begin{aligned}\|\alpha' S_{11}\hat{\alpha}\| &= \|\alpha'\frac{R_1'R_1}{T}\hat{\alpha}\| \leq \|\alpha'\frac{R_1'R_1}{T}\alpha\|^{1/2}\|\hat{\alpha}'\frac{R_1'R_1}{T}\hat{\alpha}\|^{1/2} \\ &= \|\alpha' S_{11}\alpha\|^{1/2}\|\hat{\alpha}' S_{11}\hat{\alpha}\|^{1/2} = O_p(1).\end{aligned}$$

(d) The proof of this part is omitted because it is similar to that given by Johansen (1988, 1991). □

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

Theorem 2.1. *There exists a $r\times r$ invertible matrix H_T for which*

$$\sqrt{T}(\hat{\alpha} - \alpha H_T) = O_p(1)$$

as $T\to\infty$. Furthermore, for each $1 \leq i \leq r$,

$$\sqrt{T}(\hat{\alpha}_i - \alpha H_{iT}) = \alpha\gamma'\frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_tR_{1t}'\hat{\alpha}_i\lambda_{iT}^{-1} + o_p(1) \Rightarrow N(0, \lambda_i^{-2}\alpha\gamma'\Omega\gamma\alpha'),$$

where $\Omega = E(e_te_t')$, H_{iT} is the i-th column of H_T and λ_{iT} is the i-th largest generalized eigenvalue of $S_{10}S_{01}$ with respect to S_{11}.

Proof. Recall the definition of $V_T = diag(\lambda_{1T}, \lambda_{2T}, \ldots, \lambda_{rT})$. We have $S_{10}S_{01}\hat{\alpha} = S_{11}\hat{\alpha}V_T$ or equivalently $S_{11}^{-1}S_{10}S_{01}\hat{\alpha}V_T^{-1} = \hat{\alpha}$. Take $H_T = \gamma'\gamma\alpha' S_{11}\hat{\alpha}V_T^{-1}$, which is an invertible matrix. Expanding $S_{10}S_{01}$ by using the fact $S_{01} = \gamma\alpha' S_{11} + \frac{1}{T}\sum_{t=1}^{T}e_tR_{1t}'$, we can obtain

$$\begin{aligned}\hat{\alpha} - \alpha H_T \;\; = \;\; & (\alpha\gamma'\frac{1}{T}\sum_{t=1}^{T}e_tR_{1t}'\hat{\alpha} + S_{11}^{-1}\frac{1}{T}\sum_{t=1}^{T}R_{1t}e_t'\gamma\alpha' S_{11}\hat{\alpha} \\ & + S_{11}^{-1}\frac{1}{T}\sum_{t=1}^{T}R_{1t}e_t'\frac{1}{T}\sum_{t=1}^{T}e_tR_{1t}'\hat{\alpha})V_T^{-1}. \qquad (2.11)\end{aligned}$$

By Lemma 2.2, $\frac{1}{T}\sum_{t=1}^{T} e_t R_{1t}'\hat{\alpha} = O_p(1/\sqrt{T})$, $S_{11}^{-1} = O_p(1/T^2)$, $\frac{1}{T}\sum_{t=1}^{T} R_{1t}e_t' = O_p(\sqrt{T})$, $\alpha' S_{11}\hat{\alpha} = O_p(1)$ and $V_T^{-1} = O_p(1)$, it follows that

$$\hat{\alpha} - \alpha H_T = O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{\sqrt{T^3}}) + O_p(\frac{1}{T^2}).$$

Thus, $\sqrt{T}(\hat{\alpha} - \alpha H_T) = O_p(1)$, and the asymptotic distribution of each cointegrating vector $\hat{\alpha}_i$ is determined by the first term on the right hand side of (2.11). We have

$$\sqrt{T}(\hat{\alpha}_i - \alpha H_{iT}) = \alpha\gamma' \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_t R_{1t}' \hat{\alpha}_i \lambda_{iT}^{-1} + o_p(1) \Rightarrow N(0, \lambda_i^{-2}\alpha\gamma'\Omega\gamma\alpha')$$

as stated, where $\lambda_{iT} \xrightarrow{P} \lambda_i$ by Lemma 2.2(d) and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_t R_{1t}'\hat{\alpha}_i \Rightarrow N(0,\Omega)$ can be proved similarly to the second part of Lemma 2.2(b). $\square$

Theorem 2.1 implies that $\hat{\alpha}$ is a 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

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

The second term is $O_p(1/\sqrt{T})$ by Lemma 2.2(b). But, the first term can be rewritten as $-\gamma(H_T')^{-1}(\hat{\alpha} - \alpha H_T)' S_{11}\hat{\alpha}$. From the expression of $\hat{\alpha} - \alpha H_T$ in (2.11), we have

$$\begin{aligned} \hat{\alpha}' S_{11}(\hat{\alpha} - \alpha H_T) &= (\hat{\alpha}' S_{11}\alpha\gamma' \frac{1}{T}\sum_{t=1}^{T} e_t R_{1t}'\hat{\alpha} + \frac{1}{T}\sum_{t=1}^{T} \hat{\alpha}' R_{1t} e_t' \gamma\alpha' S_{11}\hat{\alpha} \\ &\quad + \frac{1}{T}\sum_{t=1}^{T} \hat{\alpha}' R_{1t} e_t' \frac{1}{T}\sum_{t=1}^{T} e_t R_{1t}'\hat{\alpha}) V_T^{-1}. \end{aligned} \quad (2.12)$$

But, for $H_T = \gamma'\gamma(\alpha' S_{11}\hat{\alpha})V_T^{-1}$, it holds that $H_T^{-1} = O_p(1)$, because V_T converges to a positive definite matrix V and $\alpha' S_{11}\hat{\alpha} = O_p(1)$ has full rank. It follows that $(\hat{\alpha} - \alpha H_T)' S_{11}\hat{\alpha} = O_p(1/\sqrt{T})$. Thus, $\sqrt{T}(\hat{\gamma} - \gamma(H_T')^{-1}) = O_p(1)$.

Consider the second relation now. It holds that

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

This completes the proof of Theorem 2.2. □

Remark. From Theorem 2.1, the asymptotic covariance matrix of $\hat{\alpha}_i$ is $\lambda_i^{-2}\Gamma_0'\Omega\Gamma_0$. It admits a consistent estimator $\lambda_{iT}^{-2}\hat{\Gamma}_0'(\frac{1}{T}\sum_{t=1}^T \hat{e}_t\hat{e}_t')\hat{\Gamma}_0$, where

$$
\hat{e}_t = \frac{1}{T}\sum_{t=1}^{T}(\Delta Y_t - \Theta(\hat{\Gamma}_0)X_t - \hat{\Gamma}_0 Y_{t-1})'(\Delta Y_t - \Theta(\hat{\Gamma}_0)X_t - \hat{\Gamma}_0 Y_{t-1}).
$$

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

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

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 $Tg(T) \to \infty$.*

The theorem above shows 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 $T \to \infty$*

$$\sqrt{T}(\hat{\alpha} - \alpha H_T^r)V_T = O_p(1).$$

Proof. The proof is the same as that of Theorem 2.1 without any modification, except that V_T is not necessarily invertible and $H_T^r V_T = \gamma'\gamma\alpha' S_{11}\hat{\alpha}$ is not a square matrix anymore if $r \ne r_0$. 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

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

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

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

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

Now, for $r \ge r_0$, $H_T^r V_T = \gamma'\gamma\alpha^{r_0'} S_{11}\hat{\alpha}^r$ has rank r_0. Let H_T^{r+} denote the generalized inverse of H_T^r such that $H_T^r H_T^{r+} = I_{r_0}$, then it can be written as $H_T^{r+} = V_T(\alpha^{r_0'} S_{11}\hat{\alpha}^r)^{r+}(\gamma'\gamma)^{-1}$. 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,

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

where γ_i' is the ith column of γ. From the expression of $(\hat{\alpha}^r - \alpha^{r_0}H_T^r)V_T$ in (2.11), it follows that $V_T'(\hat{\alpha}^r - \alpha^{r_0}H_T^r)' S_{11}(\hat{\alpha}^r - \alpha^{r_0}H_T^r)V_T = O_p(\frac{1}{T})$. Additionally, $(\alpha^{r_0'} S_{11}\hat{\alpha}^r)^{r+} =$

$O_p(1)$ because $\alpha^{r_0'}S_{11}\hat{\alpha}^r = O_p(1)$ has full rank[3] r_0 by Lemma 2.2(c). Hence, $I = O_p(\frac{1}{T})$.

For $I\!I$, we have

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

By using the expression of $(\hat{\alpha}^r - \alpha^{r_0}H_T^r)V_T$ in (2.11) and Lemma 2.2 again, we obtain that $V_T'(\hat{\alpha}^r - \alpha^{r_0}H_T^r)'S_{11}\hat{\alpha}^r = O_p(\frac{1}{\sqrt{T}})$ and $V_T'(\hat{\alpha}^r - \alpha^{r_0}H_T^r)'\frac{1}{T}\sum_{t=1}^{T}R_{1t}e_t' = O_p(\frac{1}{T})$. The details are similar to those for (2.12). Finally, the facts $\alpha^{r_0'}S_{11}\hat{\alpha}^r = O_p(1)$ and $\frac{1}{T}\sum_{t=1}^{T}\hat{\alpha}^{r'}R_{1t}e_t' = O_p(\frac{1}{\sqrt{T}})$ imply that $I\!I = O_p(\frac{1}{T})$. This completes the proof of Lemma 3.2. □

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 $r \le p$ and $r \ne r_0$, 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 $\mathbb{R}(r, \hat{\alpha}^r) - \mathbb{R}(r_0, \hat{\alpha}^{r_0}) = \sum_{i=r+1}^{r_0}\lambda_{iT}$, where λ_{iT} is the ith generalized eigenvalue of $S_{10}S_{01}$ respect to S_{11} in decreasing order. Therefore, if $g(T) \to 0$ as $T \to \infty$,

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

by Lemma 2.2 (d) that $\lambda_{iT} \xrightarrow{P} \lambda_i > 0$.

[3]The limit of $\alpha^{r_0'}S_{11}\hat{\alpha}^r$ can be established in a similar way to that of Proposition 1 in Bai (2003) (with p fixed).

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

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

The proof of Theorem 3.1 is completed. □

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 estimtor 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

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

$$Model_{k_2} : \Delta Y_t = \gamma\alpha' Y_{t-1} + \Theta X_t + \Theta_1 Z_t + e_t \tag{3.42}$$

with $\Theta = (\mu, \Gamma_1, \ldots, \Gamma_{k_1-1})$, $\Theta_1 = (\Gamma_{k_1}, \ldots, \Gamma_{k_2-1})$, $X_t = (1, \Delta Y'_{t-1}, \ldots, \Delta Y'_{t-k_1+1})'$, $Z_t = (\Delta Y'_{t-k_1}, \ldots, \Delta Y'_{t-k_2+1})'$.

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

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

if $Model_{k_2}$ is true, $\operatorname{p\lim}_{T\to\infty}[\mathbb{R}(r_0, k_1, \alpha^{r_0}_{k_1}) - \mathbb{R}(r_0, k_2, \alpha^{r_0}_{k_2})] > 0$, where $\operatorname{p\lim}$ 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

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

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,

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

If the model with lag order k_1 is true, putting the estimator of Θ defined in section 2 into $Model_{k_1}$, we obtain that $R_{0t} = \gamma\alpha^{r_0\prime}_{k_1} R_{1t} + e_t$, and

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

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

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 (b) and (c). 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 $1 \leq k \leq K$ and $k \neq k_0$,

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

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 $k < k_0$, we have

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

if $g(T) \to 0$ as $T \to \infty$, because $\mathbb{R}(r_0, k, \alpha_k^{r_0}) - \mathbb{R}(r_0, k_0, \alpha_{k_0}^{r_0}) > 0$ 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

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

The proof is completed. $\square$

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 $\xi = \eta = \frac{1}{2}$. We use the three concrete forms in our experiments:

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

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 dependend errors

$$\begin{aligned} &e_{it} = h_{it}\varepsilon_{it}, \quad h_{it}^2 = \varphi_0 + \varphi_1 e_{it-1}^2 + \psi_1 h_{it-1}^2, \quad \varepsilon_{it} \sim N(0,1), \\ &\varphi_0 > 0,\ \varphi_1 \geq 0,\ \psi_1 \geq 0,\ \varepsilon_{it} \ \text{ are independent for all } i \text{ and } t. \end{aligned} \tag{4.2}$$

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.

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 $k = 1$ is known but both μ and the coefficient 0.6 are unknown, we estimate r by minimizing $\mathbb{M}_i(r, 1)$ for $i = 1, 2, 3$ and also by the Johansen's LRT approach. For each of different settings, we draw 2000 samples from (4.3), the percentages of the samples resulting the correct estimate (i.e. $\hat{r} = 1$) are listed in Tables 1 – 5. 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 noticiable 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 – 4 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 5 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. $\psi_1 \neq 0$).

Our second example concerns the model

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} + \begin{pmatrix} 0.3 & 0 \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} \Delta y_{1t-1} \\ \Delta y_{2t-1} \end{pmatrix} + \begin{pmatrix} 0.4 \\ 0.6 \end{pmatrix} \begin{pmatrix} 1, & -2 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}. \tag{4.4}$$

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 (2002) estimates obtained by minimizing the information 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 6. 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 6 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 the 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[4]. 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 1. 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 2. 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

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

REFERENCES

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

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

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

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

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

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

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

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

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

Gonzalo, J. & J. Y. Pitarakis (1998) Specification via model selection in vector error correction models. *Economics Letters* 60, 321C328.

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

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

Hall, P. & C. C. Heyde (1980). Martingale Limit Theory and Its Applications. Academic Press, New York.

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

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

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

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

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

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

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

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

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

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

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

Wang, Zijun. & D. Bessler (2005) 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 $\mu = 0.5$, $e_t|\theta \sim N(0, (1 + 100\theta^2)I_3)$ and $\theta \sim Poisson(\tau)$

	$\tau = 1$				$\tau = 4$			
	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	26.05	43.25	88.70	99.55
$\mathbb{M}_2$	52.45	71.35	97.85	100	53.65	77.35	99.25	100
$\mathbb{M}_3$	77.30	89.70	96.95	99.00	82.70	93.05	97.85	99.10
LRT	71.65	87.70	90.05	89.30	72.45	87.10	87.30	88.50

Table 3: Percentages of the correct estimates for r for model (4.3) with $\mu = 0.5$ and $e_t \sim t(q)$

	$q = 5$				$q = 3$			
	T=30	T=50	T=100	T=200	T=30	T=50	T=100	T=200
$\mathbb{M}_1$	24.75	44.40	87.50	94.70	28.80	43.15	79.45	92.55
$\mathbb{M}_2$	52.85	76.75	95.15	96.95	50.60	72.80	94.15	95.70
$\mathbb{M}_3$	81.40	93.90	95.60	96.45	77.25	89.35	93.65	94.15
LRT	72.15	90.75	92.80	93.45	71.00	88.90	91.65	92.45

Table 4: 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 5: 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 6: Percentages of the correct estimates for (r, k) for model (4.4)

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 $q = 5$						
$\mathbb{M}_3$	10.25	30.65	65.50	81.65	88.90	93.15
IC	10.20	22.40	63.75	80.30	82.90	84.85
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: Time plot of logarithmic GDP per capita, labor productivity per person and labor productivity per hour of the Netherlands

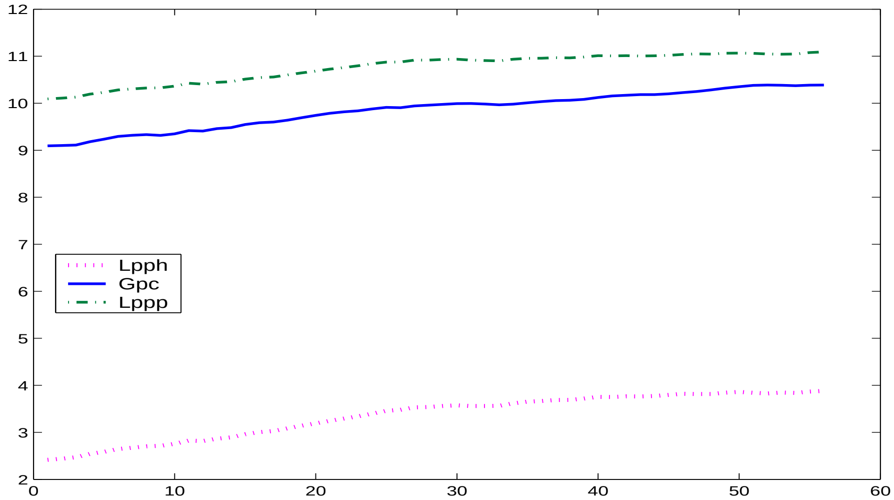

Figure 2: Plot $\mathbb{M}_3(r, k)$ against the cointegration rank r and the lag order k