Iterative Maximum Likelihood Estimation of Cointegrating Vectors[1]

In-Moo Kim
Sungkyunkwan University
and
Joon Y. Park
Rice University and *Sungkyunkwan University*

Abstract

This paper introduces an iterative method to estimate the cointegrating vectors in the error correction models. The method provides the asymptotically efficient estimators for the cointegrating vectors if iterated once or more. If it is iterated until convergence, we may obtain the maximum likelihood estimator by Johansen. For all values of $1 \leq k \leq \infty$, the k-step iterative estimators are asymptotically equivalent, and as efficient as the maximum likelihood estimator. Their finite sample performances are, however, quite different for different values of k, most notably for the two extreme cases $k = 1$ and $k = \infty$. The finite-step iterative estimators generally perform better in small samples than the infinite-step iterative estimator, i.e., the maximum likelihood estimator. In particular, the former are much more robust than the latter, which is known to occasionally yield some extreme outliers in samples of relatively small size. Our iterative procedure indeed can be very useful in detecting the occurrences of outliers for the maximum likelihood estimator, since its realized values tend to deviate largely from those of the finite-step iterative estimators when the extreme outliers are produced. The proposed method is very flexible and can be easily implemented for the cointegrated models that are specified in an arbitrary structural form.

First Draft: December 2003
This Version: January 2005

Key words and phrases: Cointegrating regression, error correction model, maximum likelihood estimation, iterative estimation method, asymptotic efficiency

[1]We are grateful to Yoosoon Chang, Chung-Ming Kuan, Peter Phillips and seminar participants at Academia Sinica, Sungkyunkwan University and Kamp Econometrics, Pusan, Korea, for many useful comments and discussions. Some of the results included in the paper were circulated earlier in 1990 as a working paper at Economics Institute, Aarhus University, under the title, "Maximum Likelihood Estimation of Simultaneous Cointegrated Models," by Park.

1. Introduction

The concept of cointegration has been very intensely explored in the past two decades by numerous authors, since it was first introduced by Granger (1983) and more rigorously formulated later by Engle and Granger (1987). Many economic longrun equilibrium relationships are now routinely modelled as cointegation models, and such models are widely accepted as useful and appropriate formulations. The statistical aspects of the cointegration models have also been throughly investigated, which has resulted in producing various methods for their effective inferences. There are several methods that are still used widely, including the maximum likelihood approach based on the error correction model by Johansen (1988, 1991), the fully modified procedure by Phillips and Hansen (1990), the canonical cointegrating regression by Park (1992) and the dynamic leads and lags method by Saikkonnen (1991) and Stock and Watson (1993). The reader is referred to Watson (1994) for a survey and more related discussions on the subject.

We develop in the paper a new method of inference on cointegrating vectors, which complements the Johansen's maximum likelihood approach based on the Gaussian error correction models. Instead of obtaining the maximum likelihood estimate directly by maximizing the likelihood as in the Johansen's procedure, we suggest an alternative method based on an iterative procedure. Our iterative procedure yields a class of asymptotically efficient estimators for the cointegrating vectors, which are defined in terms of the maximum number k of iterations. If the iteration is done until convergence, or equivalently, if we set $k = \infty$, then the resulting estimator becomes numerically identical to the maximum likelihood estimator obtained earlier by Johansen. For all the values of $1 \leq k \leq \infty$, the k-step iterative estimators for the cointegrating vectors are asymptotically equivalent: They all have the identical asymptotic distribution. Their finite sample performances are, however, quite different. In particular, the infinite-step iterative estimator, which is nothing but the maximum likelihood estimator, behaves quite distinctively in finite samples from the iterative estimators with some fixed numbers of maximum iterations, i.e., the finite-step iterative estimators.

Our simulations clearly demonstrate that the behaviors of the finite-step iterative estimators and the maximum likelihood estimator are rather different in finite samples. In general, the former perform much better than the latter in small samples, though most of their differing behaviors disappear as the sample size increases. When the sample size is small, the maximum likelihood estimator yields extreme outliers that we do not observe for the finite-step iterative estimators. The outliers in the maximum likelihood estimator occur more frequently when the sample size is smaller, and when the order of the fitted error correction model is bigger. The finite sample performance of the maximum likelihood estimator has been studied by numerous authors. See, e.g., Maddala and Kim (1998) for an extensive survey on the relevant studies. In particular, it was noted earlier by Ogaki and Park (1991) through simulation that the maximum likelihood estimator occasionally yields extreme outliers and appears to have undefined sample moments in small samples. Phillips (1994) later showed analytically that the finite sample distribution of the maximum likelihood estimation has Cauchy-like tails and has no finite integral moments. The occurrence of extreme outliers in the maximum likelihood procedure has recently been reconfirmed by

Cappuccio and Lubian (2001).

Our iterative procedure not only provides a new class of asymptotically efficient estimators that are much more robust in small samples than the maximum likelihood estimator, but also makes it possible to detect whether or not the observed maximum likelihood estimator is likely to be an outlier. Whenever an outlier is observed, the maximum likelihood estimate tends to sharply diverge from the finite-step iterative estimates. Therefore, we may just compare their estimates for the outlier detection. If they are far apart from each other, it is likely that the observed maximum likelihood estimate is an outlier. Clearly, we should not rely on the maximum likelihood estimate in this case. If, on the other hand, their values are close to each other, it seems safer and may be better to use the maximum likelihood estimate. It is also possible to use the hybrid estimators, which alternately assign the finite-step iterative estimates and the maximum likelihood estimate depending upon how far they fall apart. Such estimators can easily be defined and practically implemented using the iterative maximum likelihood procedure developed in the paper.

The proposed method is very flexible, and can be readily implemented to estimate the simultaneous cointegrated model specified in general structural form. In the presence of multiple cointegrating relationships, we need to impose restrictions to individually identify each relationship. See, e.g., Park (1990) and Hsiao (1997) for more issues on the identification of cointegrated models. The triangular representation of the cointegrating relationships used by Phillips (1991) and others can be regarded as the simultaneous cointegrated model given in reduced form, which is necessarily just-identified. Of course, the simultaneous cointegrated model can be more generally specified in structural form, possibly with over-identifying restrictions. Our procedure is developed for the simultaneous cointegrated model with general linear identifying restrictions, and in particular allows for the presence of over-identifying restrictions. It can therefore be used to estimate the general simultaneous cointegrated model given in structural form with an arbitrary number of over-identifying restrictions. The Johansen's methodology may deal with such a system with an appropriate reparametrization, but ours provides an alternative that is more comparable with the traditional approach to the inference in the classical simultaneous equations model.

The rest of this paper is organized as follows. In Section 2, we introduce the models and assumptions. Various specifications of cointegrating vectors are considered with required identification conditions. Some existing methods and their asymptotic theories are introduced and compared in Section 3. This is to motivate the methodology proposed in the paper, and to relate our asymptotic theories to those of the existing methods. The new method of computing the maximum likelihood estimator is provided in Section 4. A new class of the iterative estimators that are asymptotically equivalent to the maximum likelihood estimator are then introduced. The maximum likelihood estimator is interpreted as a member of this class. Section 5 reports a rather extensive set of simulation results, which investigate the finite sample performances of the estimators for cointegrating vectors considered in the paper. An example is also given to illustrate the empirical relevancy of our methodology. Some conclusing remarks follow in Section 6, and all the mathematical proofs are given in Appendix.

2. The Models and Assumptions

Let

$$y_t = \Pi' x_t + u_t \tag{1}$$

and assume that (x_t) is integrated and (u_t) is stationary, where (x_t) and (u_t) are, respectively, m- and ℓ-dimensional, and Π is the $m \times \ell$ parameter matrix. Under this specification, the time series (y_t) and (x_t) are cointegrated with the matrix of cointegrating vectors given by Π and the cointegrating errors (u_t). If we set

$$z_t = (y_t', x_t')' \tag{2}$$

then the time series (z_t) becomes an s-dimensional time series, $s = \ell + m$, which has ℓ-cointegrating relationships, or equivalently, m-common trends in the terminologies of Engle and Granger (1987) and Stock and Watson (1988). We may also consider cointegration model given in a more general form. For this purpose, we write

$$B' z_t = u_t \tag{3}$$

where B is an $s \times \ell$ matrix of cointegrating vectors. Of course, the model given in (1) can be regarded as a particular specification of B with $B = (I, -\Pi')'$ in (3). Moreover, (x_t) should not be cointegrated, if (z_t) has ℓ-cointegrating relationships and m-common trends as we assume here.

We now introduce a set of explicit assumptions that are required to develop our asymptotic theories in the paper.

Assumption 2.1 *The process (w_t), which is defined by*

$$w_t = (u_t', \triangle x_t')' \tag{4}$$

satisfies the invariance principle

Assumption 2.2 *Let (z_t) be generated by the error correction model* (ECM)

$$C(L)\triangle z_t = -AB' z_{t-1} + \varepsilon_t \tag{5}$$

where $C(z) = I - C_1 z - \cdots - C_{p-1} z^{p-1}$ and (ε_t) are white noise.

Assumption 2.1 is sufficient to develop our asymptotics in the paper. All the existing literature on the statistical analysis of cointegrating relationships assume that the invariance principle in Assumption 2.1 holds. Assumption 2.2 is introduced to more specifically deal with the shortrun dynamics in the model. The ECM in Assumption 2.2 models the shortrun error correcting mechanisms as well as the longrun equilibrium relationships. Throughout the paper, we assume that both Assumptions 2.1 and 2.2 hold. The p-th order vector autoregression (VAR) consisting of integrated time series that are cointegrated satisfy the assumptions under some extra conditions. The reader is referrred to Johansen (1987, 1991) for details.

If (z_t) is generated by the ECM in (5), then (w_t) defined in (4) can be represented as VAR(p), i.e., the VAR of order p. This was shown in, e.g., Park and Ogaki (1991). For the explicit representation of (z_t) as VAR(p), define an $s \times s$ matrix

$$H = \begin{pmatrix} \mathrm{I} & -\Pi' \\ 0 & \mathrm{I} \end{pmatrix}$$

and let I_a be an $s \times s$ matrix with the $\ell \times \ell$ identity matrix in northwest block and zero elsewhere. Also, define A_a to be an $s \times s$ matrix which is obtained by augmenting $s \times m$ zeros to the $s \times \ell$ matrix A. Then we have

$$w_t = \sum_{k=1}^{p} T_k w_{t-k} + v_t \tag{6}$$

where $T_1 = I_a - HA_a + HC_1H^{-1}$, $T_k = HC_kH^{-1} - HC_{k-1}H^{-1}I_a$ for $2 \leq k \leq p-1$, $T_p = -HC_{p-1}H^{-1}I_a$, and $v_t = H\varepsilon_t$. Consequently, the ECM representation of (z_t) in (5) amounts to assuming that (w_t) in (4) follows VAR(p).

Now we introduce the identification conditions for the cointegrating vectors. As is well known, the cointegrating vectors are identified only up to the space spanned by them, or the so-called the cointegration space. Of course, it is possible to estimate the cointegration space directly as in Johansen (1988, 1991) using a computationally convenient normalization. In the paper, we use an alternative approach relying on a priori identification. We assume that the parameters in the given cointegration model are identified through restrictions, exactly as in the classical simultaneous equations model (SEM). This is to fully develop the asymptotics for both unidentified and identified cointegrating vectors, and also to more explicitly compare the asymptotics for the cointegration model formulated as the multivariate regression and as a longrun component in the error correction model. Both types of cointegration models have been used widely for many theoretical and empirical researchers. If we only need an estimate for the cointegration space, an arbitrary set of identifying restrictions can be imposed. This will be explained in more detail later. In an analogy to SEM, we will refer throughout the paper the system of multiple cointegrating relations to as the simultaneous cointegrated model (SCM).

Just like the SEM, the SCM can be specified in the *reduced form* (RF) as in (1), or in the *structural form* (SF) as in (3). The parameter Π for the RF representation in (1) is clearly identified, as long as there is no cointegrating relationship in (x_t). The identification of the more general SF representation in (3), however, requires more discussions. For the cointegrating vectors in B to be individually identified, we need some identifying restrictions. Without restrictions, the matrix B is only identified up to the space spanned by its column vectors, or up to BT^{-1} for any $\ell \times \ell$ nonsingular transformation matrix T. Note in particular that the joint transformation $A \mapsto AT'$ and $B \mapsto BT^{-1}$ do not change the error correction model in Assumption 2.2. The identification of the SCM is essentially identical to that of the SEM, and naturally, all the standard theories for the identification of the SEM apply. This will be seen more clearly as we develop our theory subsequently.

As in the identification of the standard SEM, we let R and r be the known matrix and vector that are $q \times \ell m$- and q-dimensional, respectively, and assume that the restrictions

are given by

$$R\,\mathrm{vec}B = r$$

Here, q is the number of restrictions, and ℓ and m are the numbers of cointegrating vectors and common trends as specified earlier. The operator "vec" stacks column vectors of a matrix and defines a long vector. For the identification of the cointegrated model in SF introduced in (3), it is necessary and sufficient that

Assumption 2.3 $\mathrm{rank}\, R(I_\ell \otimes B) = \ell^2$.

This is exactly the same as what is known as the rank condition for SEM. The condition in Assumption 2.3 is sufficient, as well as necessary, for the identification of cointegrating vectors and will be referred to as the *rank condition*. The weaker necessary condition is given by $q \geq \ell^2$, i.e., the number of restrictions is at least as large as the number of cointegrating relationships squared. This is the *order condition* for the identification of cointegrating vectors. It can be said that the cointegrating vectors are just (under and over, respectively) identified if $q = \ell^2$ ($q < \ell^2$ and $q > \ell^2$, respectively).

Clearly, it is also possible to discuss the identifiability of the individual cointegrating vector. For this, we let β_i denote the i-th column of the matrix B of cointegrating vectors introduced in (3), and denote by R_i and r_i the pair of the restriction matrix and vector that impose the q_i-number of restrictions of the form $R_i\beta_i = r_i$ for the i-th cointegrating vector β_i for $i = 1, \ldots, \ell$. Then the necessary and sufficient rank condition for the identification of the i-th cointegrating vector β_i is given by $\mathrm{rank}\, R_i B = \ell$. The necessary order condition in this case becomes $q_i \geq \ell$, i.e., we require that the number of restrictions imposed on the i-th cointegrating vector be at least as large as the number of cointegrating relationships. Naturally, we may say that the i-th cointegration vector β_i is just (under and over, respectively) identified if $q_i = \ell$ ($q_i < \ell$ and $q_i > \ell$, respectively), just as in the SEM.

Now we suppose the matrix B of cointegrating vectors is identified, and write the cointegration model as

$$\begin{array}{ccc} y_{1t} &=& \delta_1' z_{1t} + u_{1t} \\ \vdots & \vdots & \vdots \\ y_{\ell t} &=& \delta_\ell' z_{\ell t} + u_{\ell t} \end{array} \tag{7}$$

which represents each of ℓ-cointegrating relationships between (y_t) and (x_t) in the format of seemingly unrelated regression (SUR). Note that we let $y_t = (y_{1t}, \ldots, y_{\ell t})'$ and $u_t = (u_{1t}, \ldots, u_{\ell t})'$, and (z_{it}) is a subset of (y_t) and (x_t) included in the right hand side of the i-th equation. Here we assume that the i-th column β_i of the matrix B of cointegrating vectors can be represented, after an appropriate reparametrization if necessary, by the unknown parameter δ_i for $i = 1, \ldots \ell$. In what follows, we let

$$\delta = (\delta_1', \ldots, \delta_\ell')'$$

which represents the unknown parameters in the matrix B of cointegrating vectors.

In what follows, we will consider the methods of estimating the parameter Π in the RF model (1) or δ in the SF model (7) of the SCM. We will look at several versions of their estimators that are asymptotically equivalent to the maximum likelihood (ML) estimator including the ML estimator itself. Due to the invariance property of the ML estimation, the ML estimator is independent of any specific parametrization. Without over-identifying restrictions, the ML estimator of the SF form parameter is numerically the same regardless of whether it is based directly on any SF model or it is indirectly inferred from the RF model. As a result, the ML estimation of any just identified SF model is essentially identical to that of the RF model. Moreover, the ML estimation of the cointegration space can be done on any just identified SF model, as well as on the RF model. The same arguments apply also to other ML equivalent estimators considered in the paper, but only asymptotically. In finite samples, their performances may well depend on particular specifications.

We will not explicitly consider in the paper the models which include deterministic, as well as stochastic, trends. This is just to simplify our presentation and save the space. All of our subsequent methods can be extended in a rather straightforward manner to such models with only some obvious modifications. The reader is referred to, e.g., Park (1992) for the detailed exposition on the required modifications. The theoretical results on the mixed normality of the limiting distributions for the estimators and the validity of the standard chi-square tests presented in the subsequent sections, in particular, continue to hold for the SCM with deterministic trends, as long as the included deterministic trends satisfy some mild regularity conditions.

3. Existing Methods and Their Asymptotic Theories

In this section, we review the existing methods for the estimation of cointegrating vectors and discuss their asymptotics. This is to relate our methods and theories to those for the other existing estimation procedures of cointegrating vectors. For this purpose, we consider in particular the ML method by Johansen (1988) based on the Gaussian ECM, and other nonparametric methods that are asymptotically equivalent to the Gaussian ECM based ML estimator. For the latter, we only explicitly consider the CCR approach by Park (1992), since in terms of motivation it is believed to be most closely related to our iterative procedure introduced in the paper. As is well known, the CCR method is asymptotically equivalent to the fully modified procedure by Phillips and Hansen (1990) and the method relying on the dynamic regressions augmented with leads and lags by Saikonnen (1991) and Stock and Watson (1993). The procedures by Phillips and Hansen (1990) and Park (1992) are applicable only under Assumption 2.1, and the dynamic regressions proposed by Saikonnen (1991) and Stock and Watson (1993) are also valid under Assumption 2.1 with only very mild extra conditions.

For the processes (w_t) and (ε_t) appeared in Assumptions 2.1 and 2.2, we define

$$\Omega = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left(\sum_{t=1}^{n} w_t\right)\left(\sum_{t=1}^{n} w_t\right)' \tag{8}$$

and

$$\Sigma = \mathbb{E}\left(\varepsilon_t \varepsilon_t'\right) \tag{9}$$

The longrun variance Ω of (w_t) in (8) is further partitioned as

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$$

conformably with $w_t = (u_t', \triangle x_t')'$ in (4), and we define

$$\Omega_{11\cdot 2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21} \tag{10}$$

which is the longrun conditional variance of (u_t) given $(\triangle x_t)$. Note that $\Omega_{22} > 0$, since we assume that there is no cointegration relationship in (x_t).

It can be shown that

Lemma 3.1 *Under* Assumptions 2.1 *and* 2.2*, we have*

$$\Omega_{11\cdot 2} = (A'\Sigma^{-1}A)^{-1}$$

in the notation defined above.

In light of our result in Lemma 3.1, we will use the notation

$$V = \Omega_{11\cdot 2} \quad \text{or} \quad (A'\Sigma^{-1}A)^{-1} \tag{11}$$

in all our subsequent discussions.

We now consider the estimation of the RF model in (1). To introduce the Johansen's estimator $\hat{\Pi}$ for Π, we first let S_{ij} for $i, j = 0, 1$ be the moment matrices of the residuals from the regressions of $(\triangle z_t)$ and (z_{t-1}), respectively for $i, j = 0$ and $i, j = 1$, on the lagged differences $(\triangle z_{t-k})$ for $k = 1, \ldots, p-1$. Then we define $\hat{B}$ to satisfy the optimization problem

$$\hat{B} = \underset{B}{\operatorname{argmin}} \det\left(S_{00} - S_{01}B(B'S_{11}B)^{-1}B'S_{10}\right)$$

subject to the normalization $\hat{B}'S_{11}\hat{B} = I_m$ and $\hat{B}'S_{10}S_{00}^{-1}S_{01}\hat{B} = \hat{\Lambda}_m$, where $\hat{\Lambda}_m$ is the diagonal matrix given by $\hat{\Lambda}_m = \operatorname{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_m)$ with the ordered m-largest generalized eigenvalues of $S_{10}S_{00}^{-1}S_{01}$ with respect to S_{11}, i.e., the ordered m-largest roots of the determinental equation $\det(\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}) = 0$. The ML estimate for Π may now be easily obtained from

$$\hat{\Pi} = -\hat{B}_1^{-1}\hat{B}_2$$

where $\hat{B} = (\hat{B}_1', \hat{B}_2')'$.

The CCR method transforms, nonparametrically, the integrated variables (y_t) and (x_t) in (1) to generate a new system

$$y_t^* = \Pi' x_t^* + u_t^* \tag{12}$$

where (y_t^*) and (x_t^*) are stationary deviations respectively from (y_t) and (x_t), which are given by

$$\begin{aligned} y_t^* &= y_t - \Omega_{12}\Omega_{22}^{-1}\triangle x_t - \Pi'\Lambda\Sigma^{-1}w_t \\ x_t^* &= x_t - \Lambda\Sigma^{-1}w_t \end{aligned}$$

where

$$\Lambda = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left(x_t w_t'\right) \quad \text{and} \quad \Sigma = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left(w_t w_t'\right)$$

and other notations are defined earlier. Note that the transformed regression (12) has the new error term (u_t^*), which is different from the error (u_t) in the original model (1), though it represents the same cointegrating relationships. The CCR estimator $\tilde{\Pi}$ of Π is defined simply as the OLS estimator from the transformed regression (12). Of course, the unknown parameters in the transformation in (12) must be estimated to implement the CCR methodology in practice. The replacement of these unknown parameters by their consistent estimates, however, do not affect the asymptotics of the CCR estimator, as shown in Park (1992).

The ML estimator and the CCR and other nonparametric estimators are asymptotically equivalent, i.e., they all have the same limiting distribution as we demonstrate below.

Theorem 3.2 *Suppose that* Assumptions 2.1 *and* 2.2 *hold. Then we have*

$$n\left(\hat{\Pi}-\Pi\right),\ n\left(\tilde{\Pi}-\Pi\right) \ \to_d\ \mathbb{MN}\left(0, M\otimes V\right)$$

where M is a random matrix given by

$$M = l.i.d.\left(\frac{1}{n^2}X'X\right)^{-1}$$

and V is defined in (11).

In Theorem 3.2 and elsewhere in the paper, we use "$\mathbb{MN}$" to denote the *mixed normal distribution* and "l.i.d." to signify *limit in distribution*. Also, X is the matrix contructed from (x_t) for $t = 1, \ldots, n$ in the usual way. This convention will be used throughout the paper.

The Johansens's approach may be used to estimate the cointegrating vectors given in the SF (7), though he uses different formulations for the restrictions. Instead of representing the cointegrated model in the SF as we do here, he recognizes any over-identifying restrictions and imposes them directly on the cointegrating space, i.e., the range of B. The reader is referred to Johansen and Juselius (1990, 1992) for some examples of such restrictions. For instance, if we denote by $\mathcal{R}(T)$ the range of any matrix T, they consider the estimation of B under restrictions such as $\mathcal{R}(B) \subset \mathcal{R}(R_1)$ and $\mathcal{R}(R_2) \subset \mathcal{R}(B)$, and also under the restriction on the dimensionality of $\mathcal{R}(B) \cap \mathcal{R}(R_3)$, where R_1, R_2 and R_3 are some known matrices used to restrict the cointegrating relations. The first example imposes some linear

restrictions on all the cointegrating relations, while the second examples assumes that some of the cointegrating relations are known. The last example, on the other hand, puts some restrictions on a subset of the cointegrating relations that are chosen in the range of R_3. For the just-identified SCM, the ML estimate of the SF parameter δ can simply be obtained from that of the RF parameter Π, due to the invariance of the ML estimator with respect to reparametrization. Note that we have one-to-one correspondence between δ and Π for the just-identified SCM. In what follows, we will denote by $\hat{\delta}$ the Johansen's ML estimator of the SF parameter δ in the SCM.

The CCR method can also be used to estimate the cointegrating vectors specified in the SF form, following the approach developed in Park and Ogaki (1990) to deal with SUR systems. Their method can be used to estimate the cointegrating vectors specified in SF. More precisely, we construct a new system of equations, instead of (7), given by

$$y_{it}^* = \delta_i' z_{it}^* + u_{it}^* \tag{13}$$

with

$$\begin{aligned} y_{it}^* &= y_{it} - \omega_{i2}\Omega_{22}^{-1}\triangle x_t - \delta_i'\Lambda_i\Sigma^{-1}w_t \\ z_{it}^* &= z_{it} - \Lambda_i\Sigma^{-1}w_t \end{aligned}$$

where

$$\Lambda_i = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left(z_{it}w_t'\right)$$

and ω_{i2} is the long run covariance of (u_{it}) and $(\triangle x_t)$. Other notations are defined earlier. The CCR estimator $\tilde{\delta}$ of δ is simply defined as the SUR estimator from (13). Just as in the CCR estimator for the RF model, the replacement of the unknown parameters by their consistent estimators do not change the limiting distribution of the CCR estimator for the SF model.

For both the CCR's applied to the RF and SF models, we have the error term given by

$$u_t^* = u_t - \Omega_{12}\Omega_{22}^{-1}\triangle x_t \tag{14}$$

which has the longrun variance $\Omega_{11\cdot 2}$, i.e., the longrun conditional covariance of (u_t) given $(\triangle x_t)$ that we denote by V. Due to Lemma 3.1, we also write as before $V = (A'\Sigma^{-1}A)^{-1}$.

Define

$$\underline{\mathrm{Z}} = \mathrm{diag}\left(Z_1, \ldots, Z_\ell\right)$$

where *diag* denotes the block diagonal matrix, having a matrix Z_i which is defined from (z_{it}) in the usual way, as the i-th block diagonal element. Then we have

Theorem 3.3 *Suppose that* Assumptions 2.1 – 2.3 *hold. Then we have*

$$n\left(\hat{\delta} - \delta\right),\ n\left(\tilde{\delta} - \delta\right) \ \to_d\ \mathbb{MN}\left(0, Q\right)$$

where Q is a random matrix given by

$$Q = l.i.d. \left(\frac{1}{n^2} \underline{Z}' (V^{-1} \otimes I) \underline{Z} \right)^{-1}$$

and V is defined in (11).

As in the RF model, the ML estimator and the CCR and other nonparametric estimators for the cointegrating vectors given in the SF are asymptotically equivalent. For the asymptotics of the estimators of the SF parameter, the identification condition in Assumption 2.3 is crucial. As can be easily seen from the proof of Theorem 3.3, Q is invertible a.s. if and only if the rank condition holds.

4. Iterative ML Estimation

In this section, we introduce an iterative procedure for the ML estimation of the cointegrated vectors based on the ECM in (5). Momentarily, we assume that all the shortrun parameters $A, C(z)$ and Σ are known and concentrate on the estimation of the longrun parameters in B. As we will show below, this concentrated likelihood approach yields not only an alternative, very flexible method of computing the ML estimate of B, but also a new class of estimators for B that are asymptotically equivalent to the ML estimator. Define

$$J = \Sigma^{-1} A (A' \Sigma^{-1} A)^{-1} \tag{15}$$

Then we have

Theorem 4.1 *Suppose that* Assumption 2.2 *holds, and that the values of the shortrun parameters in* ECM (5) *are known. Then the concentrated ML estimator for B is given by* GLS *in the regression*

$$B' z_t + J' C(L) \triangle z_{t+1} = J' \varepsilon_{t+1} \tag{16}$$

where J is the matrix defined in (15).

We may now readily derive an iterative method of computing the ML estimate for B from Theorem 4.1, as we explain below.

Note that the ML estimation of the shortrun parameters $A, C(z)$ and Σ in ECM (5) is easy and straightforward, once the ML estimate of B is found. They can be obtained from the OLS regression

$$\triangle z_t = -A(B' z_{t-1}) + C_1 \triangle z_{t-1} + \cdots + C_{p-1} \triangle z_{t-p+1} + \varepsilon_t \tag{17}$$

i.e., the $(p-1)$-th order VAR in differences augmented with the lagged disequilibrium error term $(B' z_{t-1})$. Quite clearly, our results here imply that the ML estimates of the parameters in ECM (5) can be obtained by iterating alternatively two regressions (16) and (17), the former by GLS and the latter by OLS, until convergence. An initial estimate for the longrun parameter B can be obtained by OLS in regression (1) or (7). The iteration is

to be proceeded as follows: We use the initial estimate for B to obtain estimates of $A, C(z)$ and Σ from the OLS estimation of the regression (17), which are in turn used to estimate B by the GLS estimation of the regression (16). The new estimate for B is now plugged into regression (17) to reestimate $A, C(z)$ and Σ, and so on. More discussions will follow on how to implement the iterative ML procedure respectively for the RF and SF models.

For the RF model with $B = (I, -\Pi')'$, Theorem 4.1 implies that we may obtain the concentrated ML estimate for Π in (1) from the multivariate regression

$$y_t^* = \Pi' x_t + u_t^* \tag{18}$$

where

$$y_t^* = y_t + J' C(L) \triangle z_{t+1}$$

and

$$u_t^* = J' \varepsilon_{t+1}$$

Of course, we may estimate Π by OLS in this case, since GLS becomes identical to OLS. The concentrated ML estimation of B is therefore particularly simple for the RF model.

For the SF model with more general specification of B in (7), we may similarly obtain the concentrated ML estimate for δ from the seemingly unrelated regression

$$\begin{array}{ccc} y_{1t}^* & = & \delta_1' z_{1t} + u_{1t}^* \\ \vdots & \vdots & \vdots \\ y_{\ell t}^* & = & \delta_\ell' z_{\ell t} + u_{\ell t}^* \end{array} \tag{19}$$

where, with J_i denoting the i-th column vector of J, we have

$$y_{it}^* = y_t + J_i' C(L) \triangle z_{t+1}$$

and

$$u_{it}^* = J_i' \varepsilon_{t+1}$$

for $i = 1, \ldots, \ell$. For the cointegrating vectors specified in the SF, our procedure here provides a very convenient method to compute the ML estimate.

As we mentioned above, our results here provide a new class of the estimators for the cointegrating vectors, as well as a method of computing the ML estimate. The class consists of the estimators that are asymptotically equivalent to the ML estimator. To introduce this class, we just note that the estimator for B from any "feasible" version of regression (18) or (19), i.e., the regression with transformation using any consistent estimates of $A, C(z)$ and Σ, becomes asymptotically equivalent to the ML estimator. To see this more precisely, let $\bar{A}, \bar{C}(z)$ and $\bar{\Sigma}$ be estimates of $A, C(z)$ and Σ, respectively, and consider

$$\bar{y}_t^* = \Pi' x_t + \bar{u}_t^* \tag{20}$$

where

$$\bar{y}_t^* = y_t + \bar{J}' \bar{C}(L) \triangle z_{t+1}$$

for the RF model and

$$\begin{array}{rcl} \bar{y}_{1t}^* &=& \delta_1' z_{1t} + \bar{u}_{1t}^* \\ \vdots & \vdots & \vdots \\ \bar{y}_{\ell t}^* &=& \delta_\ell' z_{\ell t} + \bar{u}_{\ell t}^* \end{array} \tag{21}$$

where

$$\bar{y}_{it}^* = y_t + \bar{J}_i' \bar{C}(L) \triangle z_{t+1}$$

for the SF model. The error (u_t^*) in (20) or (21) is defined accordingly.

It is straightforward to deduce that

Proposition 4.2 *Let $\bar{A}, \bar{C}(z)$ and $\bar{\Sigma}$ be any consistent estimates of $A, C(z)$ and Σ, respectively. Then the OLS estimator of Π in regression* (20) *is asymptotically equivalent to that in regression* (18) *with known values of $A, C(z)$ and Σ. Similarly, the GLS estimator of δ in regression* (21) *is asymptotically equivalent to that in regression* (19) *with known values of $A, C(z)$ and Σ.*

Now it is obvious that the estimators for B from the OLS and GLS regressions respectively in (20) and (21) with any consistent estimates of the shortrun parameters $A, C(z)$ and Σ are asymptotically equivalent to the ML estimators.

Our formulation of the concentrated ML estimation in Theorem 4.1 and the subsequent development of our theory make it very clear how the ML method is related to the other nonparametric procedures. In particular, the ML estimator can be regarded as a specially designed form of the CCR estimator by Park (1992) or the fully modified estimator by Phillips and Hansen (1990). The ML estimator is different from the CCR estimator or the fully modified estimator only in the way how they treat the shortrun dynamics of cointegration model. The former exploits the structure of ECM in (5), while the latter does not presume any specific dynamics and relies on a nonparametric correction for the asymptotic efficiency. It may now be seen very clearly why the ML estimator should have the limiting distribution identical to the other existing nonparametric estimators, which we show in Theorems 3.2 and 3.3.

For the asymptotically equivalent class of estimators for B that are generated by the feasible OLS and GLS regressions in (20) and (21), it appears to be particularly interesting to consider the *k-step iterative estimators*. They are the estimators given in the k-th step of the iterations that are required to compute the ML estimate based on the alternating regressions (16) and (17), starting from the OLS estimate of B. We set the zero step iterative estimator to be the initial OLS estimator of B. The first step iterative estimator for B is the estimator obtained from regression (20) or (21) using the shortrun parameters estimated by OLS from regression (17) run with the zero step estimator for B. Likewise, the k-step iterative estimator for B is obtained from regression (20) or (21) formulated with the shortrun parameters estimated by regression (17) run with the $(k-1)$-st step estimator for B. Of course, the ML estimator can be defined to be the infinite-step iterative estimator,

and in this context, the k-step iterative estimators may be viewed as the ones with the maximum number of iterations restricted by k, for each $k = 1, 2, \ldots$.

The iterative estimators that we propose here can be quite useful in practical applications. The ML estimator, though it performs quite well for reasonably large size samples, tends to yield extreme outliers in small samples. As we show through extensive simulations reported in the next section, outliers are often observed for the ML estimation when the sample size is 100 and the order of the underlying ECM model is relatively large. The occurrence of outliers becomes more frequent as the sample size gets small and the model is fitted by higher orders of ECM. The inclusion of the constant or time trend terms makes the problem much worse. The extreme outliers do occur occasionally in these cases. Even in the aforementioned case, the outliers appear frequently enough to make the mean squared errors of the ML estimator uncomparably larger than any other estimator considered in the paper. Fortunately, the k-step iterative estimators can be used to avoid using nonsensical estimates given by the outliers. The k-step iterative estimators are pretty robust even when the sample size is small and a higher order ECM is used to fit the model. Moreover, virtually at all times when the outliers occur for the ML estimation, the k-step iterative estimators yield values that are very distinctive of the ML estimate. We may thus simply compare the ML estimate with the k-step iterative estimates to judge whether it is to be seen as an outlier.

5. Simulations and Empirical Illustrations

In this section, Monte Carlo methods are used to examine the finite sample performances of the estimators considered in the paper. For the simulations, we use the bivariate model

$$\triangle z_t = -\alpha\beta' z_{t-1} + C\triangle z_{t-1} + e_t \tag{22}$$

where $z_t = (y_t, x_t)'$ and $\beta = (1, -\pi)'$. The errors (e_t) are generated by

$$e_t = \varepsilon_t + D\varepsilon_{t-1} \tag{23}$$

where (ε_t) are standard normals with covariance matrix Σ. Here and elsewhere in this section, we use the lower case letters to denote the parameters defined in the previous sections, whenever they are vectors or scalars.

More explicitly, the models that we use for simulations are given by

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \tag{24}$$

for α in (22), where we let $\alpha_1 = 0.2$ and $\alpha_2 = 0, -0.2, -0.4$ or -0.6. We denote the DGP's corresponding to each of these values of α_2 by (i), (ii), (iii) and (iv). We set $\pi = 1$. Also, the covariance matrix Σ of (ε_t) in (23) is specified as

$$\Sigma = \begin{pmatrix} 1 & \sigma \\ \sigma & 1 \end{pmatrix} \tag{25}$$

where $\sigma = -0.3, 0, 0.3$. Similarly as before, we denote the models for each of these values of σ by a, b and c. Taken (24) and (25) together, the DGP's are signified as (kf) with k = i, ii, iii and f = a, b, c.

We consider three different types of models: VAR(1), VAR(2) and VARMA(1,1). For the VAR(1) model, we let $C = D = 0$ with other parameters given as above. The coefficient matrix C in the VAR(2) model is set

$$C = \begin{pmatrix} -0.4 & 0 \\ 0 & 0.2 \end{pmatrix}$$

with $D = 0$. The VARMA(1,1) model is specified by $C = 0$ and

$$D = \begin{pmatrix} 1 & 0.2 \\ -0.4 & 1 \end{pmatrix}$$

The parameters C and D in the VAR(2) and VARMA(1,1) models are chosen in such ways that the theoretical asymptotic variances of the estimators in these models are largely the same as those in the VAR(1) model.

As shown in earlier sections, the CCR and ML estimators are asymptotically equivalent, and have the same limiting distribution. Let (w_t) be defined by $w_t = (u_t, \triangle x_t)'$, where $u_t = y_t - \pi x_t$, as in (4), and write the longrun variance Ω of (w_t) as $\Omega = (\omega_{ij})$ for $i, j = 1, 2$. If we denote by $\bar{\pi}$ the CCR or ML estimator of the cointegrating coefficient π, then we may readily deduce from Theorem 3.2 that

$$n\,(\bar{\pi} - \pi) \to_d \mathbb{MN}\left(0, \omega_*^2 \left(\int_0^1 W(s)^2 ds\right)^{-1}\right)$$

where

$$\omega_*^2 = \frac{1}{\omega_{22}}\left(\omega_{11} - \frac{\omega_{12}^2}{\omega_{22}}\right) = \frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{12}^2}{\omega_{22}^2} \tag{26}$$

and W is the standard Brownian motion. The asymptotic variance of $\bar{\pi}$ is therefore given by

$$\frac{1}{n^2}\,\omega_*^2\,\mathbb{E}\left(\int_0^1 W(s)^2 ds\right)^{-1} \tag{27}$$

with ω_*^2 defined in (26).

The longrun variance Ω of (w_t) can be easily obtained from (6) for our models given by (22) and (23). Using the notations I_a, A_a and H introduced in (6), we may write

$$w_t = T_1 w_{t-1} + T_2 w_{t-2} + v_t$$

where $T_1 = I_a - HA_a + HCH^{-1}$, $T_2 = -HCH^{-1}I_a$ and $v_t = H(\varepsilon_t + D\varepsilon_{t-1})$. Therefore, we may easily deduce that

$$\Omega = (I - T_1 - T_2)^{-1} H(I + D)\Sigma(I + D)'H'(I - T_1 - T_2)^{-1\prime}$$

Moreover, we found through simulation that

$$\mathbb{E}\left(\int_0^1 W(s)^2 ds\right)^{-1} \approx 5.58$$

The asymptotic variance of $\bar{\pi}$ can now be computed from (27) for each of the DGP's considered in our simulations.

The design of our simulations and the choice of estimators are made to serve several purposes. The main objective of our simulation study is to assess the finite sample performance of the ML estimator with an emphasis on its tail behavior, when the sample size is relatively small. Especially, we like to evaluate the finite sample behavior of the ML estimator in comparisons with the finite-step iterative estimators and to the nonparametric CCR estimator. This is done for the misspecified or overparametrized models, as well as the correctly specified models, to investigate the effects of various types of misspecification and overparametrization. The finite-step iterative estimators use the information on the shortrun dynamics, while the nonparametric CCR estimator does not. Therefore, we may also see how important it is to use the shortrun information for the estimation of the longrun cointegrated model. However, since the relative finite sample performances of the ML estimator and other nonparametric estimators are well known due to many previous simulation studies, we will make very brief the discussions on these issues. The reader is referred to Cappuccio and Lubian (2001) for a recent study and to Maddala and Kim (1998) for a survey.

Our simulation results are summarized in Tables 1 and 2. Each of Tables 1 and 2 has three panels, A,B and C, respectively for the results of the DGP's given by the VAR(1), VAR(2) and VARMA(1,1) models. All our simulations are based on the samples of size 100 for the 10,000 times of iterations. The iterative ML procedures are investigated for various combinations of the maximum number k of iteration and the order p of the fitted VAR model. In the tables, the k-step iterative estimator based on the VAR(p) model is denoted by ML_p^k. The CCR estimator is implemented with the longrun variance estimated nonparametrically using the Parzen kernel and the automatic bandwidth selection rule given by Andrews (1991). It is signified by CCR in the tables. In addition to those reported in Tables 1 and 2, we also consider other sets of models, such as the models with fitted intercept and time trend, with different parameter values and varying sizes of samples. Though we do not report the details to save the space, we will make comments on them whenever they are relevant in the discussions below.

Table 1 reports the bias and mean squared error (MSE) with asymptotic variance. First, it is clearly seen in our simulation results that the use of the information on shortrun dynamics, if available, is extremely important. The k-step iterative estimator with $k = 1$ or 2 outperforms the nonparametric CCR estimator in a majority of cases that we consider in our simulations. In particular, the iterative estimators yield significantly smaller biases in almost all cases. The comparison is also much favorable for the iterative estimators in terms of MSE in predominantly many cases. This is largely so, even when the model is misspecified or overparametrized. Across all the DGP's given by VAR(1), VAR(2) and VARMA(1,1) that we consider here, the finite-step iterative estimators generally do better,

both in terms bias and MSE, than the nonparametric CCR estimator, regardless of the order of the fitted VAR model on which the ML procedure is based. The relative superiority of the finite-step iterative estimators, however, is somewhat smaller in magnitude, if it is based on the misspecified or overparametrized VAR models. In general, misspecification has a more significant adverse effect on the bias, while overparametrization is more detrimental to the MSE.

The finite sample performance of the ML estimator, i.e., the infinite-step iterative estimator, is already well known, and our results here are consistent with many previous studies. It usually performs well, but does yield the extreme outliers. See, e.g., the recent simulation study by Cappuccio and Lubian (2001) and the references cited there for similar results. The outlier occurs infrequently, but if it does it takes rather extreme values. The extreme outliers do appear in many cases we consider, and because of this the ML estimator is noticeably outperformed by the nonparametric CCR estimator, as well as the finite-step iterative estimators, both in terms of bias and MSE. The occurrence of the extreme outliers is a finite sample phenominon. It disappears rapidly as the sample size increases. We can hardly observe any extreme outlier in any of the DGP's used in our simulations, if the sample size gets as large as 300. Given the sample size, the occurrence of the outlier becomes more frequent, if we include the constant or time trend term in the models, and if the dimension of the model gets large.

To further investigate the tail behaviors of the estimators that we consider in the simulations, we look at the ranges of the estimates and the frequences of outlier occurrences. They are reported in Table 2 for the same DGP's used for the results in Table 1. In almost every case we examine, the ML estimate has wider range than the finite-step iterative estimates. In general, the range becomes wider as the order of the fitted VAR model increases. In many cases, the ML estimator based on the fourth order VAR model has the ranges that are unacceptably large. It seems clear that the ML estimate yields the extreme outliers much more often than the finite-step iterative estimates. The same conclusion is reached when we count the number of outlier occurrences. The reported results are the counted numbers of observations out of 10,000 simulation iterations, which fall outside the confidence bands defined by the twenty and thirty times of the asymptotic standard error. As expected, the nonparametric CCR estimator seems most robust and does not yield any extreme outliers. The finite-step iterative estimators also rarely generate the extreme outliers. In sharp contrast, the ML estimator does yield the extreme outliers. Though generally not frequent, the extreme outliers consistently appear in our simulations. The appearance of the extreme outliers becomes more frequent as the order of the fitted VAR increases.

To show the empirical relevancy of our simulation results, we analyze the simple cointegrating relationship between the SP500 index and index futures. Of course, the presumed value of the coefficient in the index/index futures cointegrating regression is unity. In Table 3, we present the estimated cointegrating coefficients in the regression of the SP500 index against the index futures. Here the running estimates are obtained from the eleven sets of samples for the 101 business days using the data set that spans the period from October 13, 1998 to March 23, 1999. It is clearly seen that an extreme outlier occurs for the ML estimator based on the samples for the period 10/20/1998 - 03/16/1999. The estimated coefficient in this case is unacceptably large, with the error that appears to be even larger

than 10,000 times the standard error. In all the other cases, the estimated coefficients are very close to unity. In particular, the finite-step iterative estimates seem to outperform the ML and CCR estimates. In this illustrative example, the 1- or 2-step iterative estimator yields the coefficients that are somewhat closer to the true value, compared with the other two estimates. The 2-step iterative estimator does not appear to do any better than the 1-step iterative estimator.

6. Concluding Remarks

This paper introduces an iterative method to estimate the cointegrating vectors in error correction models. The method provides a class of estimators for the cointegrating vectors that are asymptotically equivalent to the ML estimator, which we call the k-step iterative estimators. The ML estimator itself also belongs to this class and amounts to the infinite-step iterative estimator. The k-step iterative estimator performs very well in finite samples. It is as efficient as the ML estimator, yet it is quite robust and does not yield outliers. The ML estimator, on the other hand, often yields extreme outliers, especially when the sample size is small and a higher order ECM is used to fit the data. The k-step iterative estimator can be used jointly with the ML estimator. Whenever the ML procedure produces outliers, it diverges far away from the k-step estimate. Therefore, we may tell whether or not the observed ML estimate is an outlier by comparing the ML estimate with the k-step iterative estimator. We may also propose an estimator, which combines the ML estimate and the k-step iterative estimate.

We also make it clear how the ML estimator is related to the existing nonparametric methods. It is shown, in particular, that the ML estimator can be regarded as a parametric CCR estimator by Park (1992) or fully modified OLS estimator by Phillips and Hansen (1990), which utilizes the information on shortrun dynamic structure of the model in correcting for endogeneity. The asymptotic equivalence of the ML estimation and other existing efficient nonparametric methods can be more clearly seen in our approach. The k-step iterative estimator, like the ML estimator, uses the specific shortrun dynamics implied by ECM. Our simulation shows that using the explicit shortrun dynamics is quite important in small samples, if the true data generating mechanism is given by ECM. In this case, the k-step iterative estimators perform better, unambiguously and ubiquitously, in finite samples than the nonparametric estimators. This is true for many different data generating processes.

Appendix: Mathematical Proofs

Proof of Lemma 3.1 . Partition the matrix lag polynomial $C(z)$ in (5) as

$$C(z) = (C_a(z), C_b(z))$$

conformably with the partition of (z_t) in (2), and define

$$F(z) = C_a(z)\Pi' + C_b(z)$$

where Π is the RF parameter in (1). We have, after some trivial algebra,

$$\varepsilon_t = (A, F(L))w_t + (C_a(L) - A)\triangle u_t \tag{A1}$$

where (ε_t) is the ECM error in (5) and (w_t) is defined in (4).

Comparing the longrun variances of the two sides of (A1), we may easily deduce that

$$\Sigma = G\Omega G'$$

where Ω and Σ are given respectively in (8) and (9), and

$$G = (A, F(1)) \tag{A2}$$

Therefore, we have

$$A'\Sigma^{-1}A = A'G'^{-1}\Omega^{-1}G^{-1}A \tag{A3}$$

It is, however, obvious from (A2) that

$$G^{-1}A = I_1$$

where $I_1 = (I_{\ell\times\ell}, 0)$. The result stated in Lemma 1 follows directly from (A3). Notice that

$$\Omega_{11\cdot 2} = (I_1'\Omega^{-1}I_1)^{-1}$$

as is well known.

Proof of Theorem 3.2 The result for $\tilde{\Pi}$ is given in Park (1992). The limiting distribution of $\hat{\Pi}$ may be easily found following Johansen (1987). Since, however, his result is not directly comparable to ours, we will briefly sketch the proof for our result. In what follows, the consistency of the ML estimate is assumed. We let $\hat{B} = (I, -\hat{\Pi}')'$ and $B = (I, -\Pi')'$.

Let S_{aa}, S_{ab} and S_{bb} be defined similarly as in Johansen (1987), with the convention that our subscripts a and b correspond to his 0 and k, respectively. The ML estimate $\hat{B}$ satisfies the equation

$$S_{bb}\hat{B}\hat{T} = S_{ba}S_{aa}^{-1}S_{ab}\hat{B} \tag{A4}$$

where $\hat{T}$ is an $\ell \times \ell$ invertible matrix. It follows directly from ECM (5) that

$$S_{ab} = -AB'S_{bb} + \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t z_{t-1}' + o_p(1) \tag{A5}$$

Also, if we define $\Sigma_{aa}, \Sigma_{ab}B$ and $B'\Sigma_{bb}B$ to be the probability limits of $S_{aa}, S_{ab}B$ and $B'S_{bb}B$, respectively, we have

$$\Sigma_{aa} = A(B'\Sigma_{bb}B)A' + \Sigma \tag{A6}$$

$$\Sigma_{ab}B = -A(B'\Sigma_{bb}B) \tag{A7}$$

where Σ is defined in (9).

Let T be the probability limit of $\hat{T}$, which is given by

$$T = -A'\Sigma_{aa}^{-1}\Sigma_{ab}B \tag{A8}$$

as can be easily deduced from (A4) and (A7). We now have from (A4) – (A8) that

$$\begin{aligned} S_{bb}(\hat{B} - B) &= S_{ba}S_{aa}^{-1}S_{ab}\hat{B}\hat{T}^{-1} - S_{bb}B \\ &= \left(-S_{bb}BA' + \frac{1}{n}\sum_{t=1}^{n} z_{t-1}\varepsilon_t'\right)\Sigma_{aa}^{-1}\Sigma_{ab}BT^{-1} - S_{bb}B + o_p(1) \\ &= \left(\frac{1}{n}\sum_{t=1}^{n} z_{t-1}\varepsilon_t'\right)\Sigma_{aa}^{-1}\Sigma_{ab}BT^{-1} + o_p(1) \\ &= \left(\frac{1}{n}\sum_{t=1}^{n} z_{t-1}\varepsilon_t'\right)\Sigma^{-1}A(A'\Sigma^{-1}A)^{-1} + o_p(1) \end{aligned}$$

Notice from (A7) and (A8) that

$$\begin{aligned} \Sigma_{ab}BT^{-1} &= \Sigma_{ab}B(A'\Sigma_{aa}^{-1}\Sigma_{ab}B)^{-1} \\ &= A(A'\Sigma_{aa}^{-1}A)^{-1} \end{aligned}$$

Also,

$$(A'\Sigma_{aa}^{-1}A)^{-1}A'\Sigma_{aa}^{-1} = (A'\Sigma^{-1}A)^{-1}A'\Sigma^{-1}$$

since, from (A6),

$$\mathcal{R}(\Sigma_{aa}\mathcal{R}(A)^{\perp}) = \mathcal{R}(\Sigma\mathcal{R}(A)^{\perp})$$

i.e., the equivalence of the subspaces that are Σ - and Σ_{aa} - conjugate to $\mathcal{R}(A)$. The stated result directly follows from premultiplying the above equation by $I_2 = (0, I_{m\times m})$.

Proof of Theorem 3.3 For the CCR estimator $\tilde{\delta}$, we first observe that:

1. (z_{it}^*) and (u_t^*) are asymptotically uncorrelated, i.e.,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\,(z_{it}^* u_t^{*\prime}) = 0$$

   for all $i = 1, \ldots, \ell$, and

2. $(\triangle z_{it}^*)$ and (u_t^*) have zero longrun convariance.

The above two conditions are clearly satisfied by construction. Now it suffices to show that Q is invertible a.s., due to the result in Park and Phillips (1988) for the regression with strictly exogenous regressors and that for the CCR in Park (1992).

Proof of Theorem 4.1 . Define J_c to be a $s \times m$ matrix such that $J_c'A = 0$. Note that J_c is Σ-conjugate to J, i.e.,

$$J_c'\Sigma J = 0 \tag{A9}$$

where J is the matrix introduced in (15). We write ECM (5) as two subsystems as

$$\begin{aligned} J'C(L)\triangle z_t &= -B'z_{t-1} + J'\varepsilon_t \\ J_c'C(L)\triangle z_t &= J_c'\varepsilon_t \end{aligned}$$

Due to the Σ-conjugacy of J and J_c in (A9), $J'\varepsilon_t$ and $J_c'\varepsilon_t$ are uncorrelated. Therefore, under normality, the errors in the two subsystems are independent. Moreover, the second subsystem does not include B. It is therefore clear that maximizing likelihood of the entire system with respect to B is identical to that of the first subsystem. When $A, C(z)$ and Σ are known, however, the maximization of the likelihood of the first subsystem with respect to B is given by GLS on the regression presented in the theorem. The proof is therefore complete.

Proof of Proposition 4.2 Note that

$$\bar{u}_t^* = \bar{J}'\varepsilon_{t+1} + \left(\bar{J}'\bar{C}(L) - J'C(L)\right)\triangle z_{t+1} - \left(\bar{J}' - J'\right)\varepsilon_{t+1}$$

where we have

$$\bar{J}' - J',\ \bar{J}'\bar{C}(L) - J'C(L) = o_p(1)$$

However,

$$\frac{1}{n}\sum_{t=1}^{n} z_t\triangle z_{t+1}',\ \frac{1}{n}\sum_{t=1}^{n} z_t\varepsilon_{t+1}' = O_p(1)$$

as shown in, e.g., Park and Phillips (1988), and we therefore have

$$\frac{1}{n}\sum_{t=1}^{n} z_t\bar{u}_t^* = \frac{1}{n}\sum_{t=1}^{n} z_t u_t^* + o_p(1)$$

The rest of the proof is trivial and omitted.

References

Andrews, D.W.K. (1991). "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817-858.

Cappuccio, N. and D. Lubian (2001). "Estimation and inference on long-run equilibria: A simulation study," *Econometric Reviews*, 20, 61-84.

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

Granger, C.W.J. (1983). "Co-integrated variables and error-correcting models," Mimeographed, Department of Economics, University of California, San Diego.

Hsiao, C. (1997). "Cointegration and dynamic simultaneous equations model," *Econometrica*, 65, 647-670.

Johansen, S. (1988). "Statistical analysis of cointegrating 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.

Johansen, S. and K. Juselius (1990). "Maximum likelihood estimation and inference on cointegration – with applications to the demand for money," *Oxford Bulletin of Economics and Statistics*, 52, 169-210.

Johansen, S. and K. Juselius (1992). "Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK," *Journal of Econometrics*, 53, 211-244.

Maddala, G.S. and I.-M. Kim (1998). *Unit Roots, Cointegration, and Structural Change*, Cambridge University Press, Cambridge, UK.

Park, J.Y. (1990). "Maximum likelihood estimation of simultaneous cointegrated models," Mimeographed, Institute of Economics, Aarhus University.

Park, J.Y. (1992). "Canonical cointegrating regressions," *Econometrica*, 60, 119-143.

Park, J.Y. and M. Ogaki (1990). "Seemingly unrelated canonical cointegrating regressions," Mimeographed, Department of Economics, Cornell University.

Park, J.Y. and M. Ogaki (1991). "Inference in cointegrated models using VAR prewhitening to estimate shortrun dynamics," Mimeographed, Department of Economics, Cornell University.

Park, J.Y. and 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. (1994). "Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models," *Econometrica*, 62, 73-93.

Phillips, P.C.B. and B.E. Hansen (1990). "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies*, 57, 99-126.

Saikkonen, P. (1991). "Asymptotically efficient estimation of cointegrating regression," *Econometric Theory*, 7, 1-21.

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

Stock, J.H. and M.W. Watson (1993). "A simple estimator of cointegrating vectors in higher order integrated systems," *Econometrica*, 61, 783-820.

Watson, M.W. (1994). "Vector autoregressions and cointegration," In R.F. Engle and D. McFadden, eds., *Handbook of Econometrics*, Vol. 4, 2844-2915, Elsevier: Amsterdam.

Table 1A: *Finite Sample Bias, MSE and Asymptotic Variance*

DGP: VAR(1)

Bias

DGP	CCR	ML_1^1	ML_1^2	ML_1	ML_2^1	ML_2^2	ML_2	ML_4^1	ML_4^2	ML_4
(ia)	.0493	.0086	.0028	–.0001	.0133	.0047	–.0002	.0223	.0096	–.0011
(iia)	.0543	.0122	.0009	–.0074	.0188	.0042	–.0044	.0306	.0113	–.0083
(iiia)	.0273	.0056	.0006	–.0016	.0104	.0023	–.0018	.0189	.0065	–.0021
(iva)	.0184	.0034	.0004	–.0007	.0075	.0018	–.0006	.0154	.0055	–.0009
(ib)	.0382	.0093	.0039	–.0041	.0125	.0052	–.0034	.0190	.0090	–.0080
(iib)	.0277	.0064	.0000	–.0036	.0100	.0016	–.0043	.0161	.0050	–.0073
(iiib)	.0159	.0036	.0002	–.0010	.0064	.0012	–.0011	.0114	.0036	–.0015
(ivb)	.0117	.0023	.0002	–.0004	.0049	.0010	–.0004	.0095	.0030	–.0006
(ic)	.0272	.0074	.0026	–.0036	.0096	.0036	.0019	.0139	.0062	–.0057
(iic)	.0140	.0035	.0000	–.0016	.0053	.0008	–.0015	.0083	.0022	–.0021
(iiic)	.0088	.0021	.0000	–.0006	.0037	.0006	–.0006	.0064	.0018	–.0011
(ivc)	.0070	.0015	.0001	–.0002	.0030	.0006	–.0001	.0055	.0016	–.0003

MSE

DGP	CCR	ML_1^1	ML_1^2	ML_1	ML_2^1	ML_2^2	ML_2	ML_4^1	ML_4^2	ML_4	Avar
(ia)	.0146	.0080	.0078	.0085	.0090	.0087	.0097	.0107	.0099	.0193	.0127
(iia)	.0232	.0155	.0149	.0205	.0170	.0159	.1343	.0207	.0187	.0355	.0259
(iiia)	.0076	.0043	.0041	.0043	.0051	.0044	.0047	.0071	.0054	.0059	.0079
(iva)	.0034	.0017	.0016	.0016	.0022	.0017	.0017	.0038	.0025	.0021	.0030
(ib)	.0119	.0086	.0089	.1707	.0094	.0097	.0827	.0109	.0112	.2982	.0140
(iib)	.0101	.0080	.0079	.0091	.0086	.0084	.0110	.0100	.0097	.1224	.0140
(iiib)	.0038	.0027	.0026	.0027	.0031	.0028	.0029	.0039	.0033	.0038	.0050
(ivb)	.0018	.0012	.0011	.0012	.0014	.0012	.0012	.0021	.0015	.0015	.0022
(ic)	.0086	.0076	.0081	.0939	.0081	.0088	.1253	.0089	.0097	.2211	.0127
(iic)	.0045	.0041	.0042	.0046	.0044	.0045	.0053	.0048	.0050	.0083	.0075
(iiic)	.0019	.0015	.0015	.0016	.0017	.0016	.0017	.0020	.0019	.0023	.0030
(ivc)	.0010	.0008	.0007	.0008	.0008	.0008	.0008	.0011	.0009	.0009	.0015

Note: CCR denotes the CCR estimator, and ML_p^k and ML_p signify respectively the k-step iterative estimator and the ML estimator based on VAR(p). Asymptotic variance is abbreviated by Avar.

Table 1B: *Finite Sample Bias, MSE and Asymptotic Variance*

DGP: VAR(2)

Bias

DGP	CCR	ML_1^1	ML_1^2	ML_1	ML_2^1	ML_2^2	ML_2	ML_4^1	ML_4^2	ML_4
(ia)	.0645	.0481	.0451	.0442	.0129	.0054	–.0003	.0246	.0117	–.0297
(iia)	.0885	.0759	.0671	.0633	.0255	.0077	–.0137	.0424	.0182	.0327
(iiia)	.0527	.0436	.0380	.0361	.0139	.0038	–.0034	.0279	.0107	–.0041
(iva)	.0407	.0282	.0242	.0229	.0093	.0026	–.0013	.0236	.0092	–.0018
(ib)	.0554	.0472	.0446	.0437	.0135	.0061	.0074	.0228	.0113	–.0037
(iib)	.0535	.0539	.0487	.0469	.0150	.0035	–.0076	.0252	.0095	–.0079
(iiib)	.0355	.0346	.0308	.0298	.0097	.0023	–.0022	.0189	.0068	–.0029
(ivb)	.0300	.0239	.0209	.0202	.0070	.0016	–.0009	.0162	.0057	–.0012
(ic)	.0472	.0465	.0443	.0424	.0123	.0051	–.0096	.0195	.0090	–.0041
(iic)	.0342	.0415	.0386	.0377	.0093	.0020	–.0018	.0157	.0052	–.0047
(iiic)	.0246	.0285	.0261	.0255	.0065	.0013	–.0014	.0126	.0042	–.0021
(ivc)	.0228	.0208	.0188	.0183	.0050	.0011	–.0005	.0109	.0036	–.0007

MSE

DGP	CCR	ML_1^1	ML_1^2	ML_1	ML_2^1	ML_2^2	ML_2	ML_4^1	ML_4^2	ML_4	Avar
(ia)	.0157	.0104	.0095	.0093	.0066	.0059	.0067	.0088	.0073	7.1417	.0081
(iia)	.0341	.0282	.0248	.0236	.0212	.0189	.0424	.0269	.0227	25.2069	.0313
(iiia)	.0146	.0112	.0095	.0089	.0075	.0062	.0069	.0114	.0080	.0086	.0114
(iva)	.0081	.0050	.0041	.0038	.0034	.0026	.0026	.0070	.0042	.0036	.0047
(ib)	.0128	.0104	.0099	.0099	.0069	.0067	.6713	.0087	.0080	.0745	.0089
(iib)	.0159	.0151	.0137	.0132	.0108	.0100	.0188	.0133	.0118	.1817	.0169
(iiib)	.0077	.0069	.0060	.0057	.0045	.0039	.0042	.0065	.0049	.0054	.0072
(ivb)	.0047	.0034	.0029	.0027	.0023	.0019	.0019	.0040	.0026	.0024	.0035
(ic)	.0099	.0095	.0092	.0165	.0060	.0060	.6663	.0070	.0066	.0199	.0081
(iic)	.0075	.0083	.0077	.0075	.0054	.0053	.0220	.0064	.0059	.0148	.0091
(iiic)	.0040	.0042	.0037	.0036	.0025	.0023	.0025	.0034	.0028	.0038	.0043
(ivc)	.0027	.0023	.0020	.0019	.0014	.0012	.0012	.0021	.0015	.0015	.0023

Note: CCR denotes the CCR estimator, and ML_p^k and ML_p signify respectively the k-step iterative estimator and the ML estimator based on VAR(p). Asymptotic variance is abbreviated by Avar.

Table 1C: *Finite Sample Bias, MSE and Asymptotic Variance*

DGP: VARMA(1,1)

Bias

DGP	CCR	ML_1^1	ML_1^2	ML_1	ML_2^1	ML_2^2	ML_2	ML_4^1	ML_4^2	ML_4
(ia)	.0536	.0027	–.0024	–.0054	.0198	.0102	.0054	.0247	.0123	.0558
(iia)	.0626	.0028	–.0100	–.0061	.0239	.0093	–.0051	.0348	.0162	.0541
(iiia)	.0290	.0089	.0032	–.0005	.0103	.0028	–.0022	.0187	.0077	–.0005
(iva)	.0157	.0063	.0035	.0024	.0063	.0015	–.0004	.0122	.0049	.0004
(ib)	.0433	.0005	–.0045	–.0079	.0192	.0109	.0097	.0224	.0122	.0205
(iib)	.0322	.0013	–.0062	–.0280	.0143	.0059	.0007	.0209	.0106	–.0026
(iiib)	.0147	.0053	.0021	.0008	.0064	.0021	.0002	.0111	.0049	.0008
(ivb)	.0086	.0042	.0024	.0019	.0041	.0012	.0002	.0077	.0034	.0009
(ic)	.0318	–.0012	–.0054	–.0067	.0182	.0124	.0080	.0200	.0131	.0033
(iic)	.0146	.0001	–.0036	–.0123	.0083	.0042	.0019	.0117	.0067	.0005
(iiic)	.0073	.0030	.0013	.0037	.0042	.0019	.0010	.0068	.0036	.0017
(ivc)	.0045	.0025	.0015	.0013	.0028	.0012	.0007	.0048	.0026	.0011

MSE

DGP	CCR	ML_1^1	ML_1^2	ML_1	ML_2^1	ML_2^2	ML_2	ML_4^1	ML_4^2	ML_4	Avar
(ia)	.0143	.0066	.0067	.0085	.0081	.0077	.0303	.0096	.0088	31.0410	.0098
(iia)	.0279	.0205	.0212	1.6220	.0187	.0183	.2054	.0220	.0212	42.7945	.0302
(iiia)	.0100	.0064	.0064	.0094	.0066	.0061	.0141	.0081	.0070	.0091	.0104
(iva)	.0037	.0023	.0023	.0025	.0026	.0023	.0023	.0034	.0027	.0028	.0039
(ib)	.0131	.0088	.0090	.0103	.0101	.0103	.1618	.0110	.0113	3.0035	.0134
(iib)	.0131	.0117	.0123	3.1828	.0106	.0108	.0162	.0122	.0124	.1385	.0170
(iiib)	.0044	.0035	.0037	.0042	.0035	.0034	.0037	.0042	.0039	.0047	.0060
(ivb)	.0019	.0015	.0015	.0016	.0016	.0015	.0015	.0019	.0017	.0019	.0026
(ic)	.0111	.0099	.0110	.0836	.0109	.0119	.0308	.0120	.0135	.3783	.0156
(iic)	.0061	.0062	.0067	.3597	.0057	.0058	.0068	.0064	.0069	.0210	.0093
(iiic)	.0022	.0019	.0020	.0917	.0020	.0020	.0020	.0023	.0023	.0037	.0033
(ivc)	.0010	.0009	.0009	.0009	.0009	.0009	.0009	.0010	.0010	.0013	.0015

Note: CCR denotes the CCR estimator, and ML_p^k and ML_p signify respectively the k-step iterative estimator and the ML estimator based on VAR(p). Asymptotic variance is abbreviated by Avar.

Table 2A: *Range of Estimate and Frequency of Outlier Occurrence*

DGP: VAR(1)

Range of Estimate

DGP	CCR	ML_1^1	ML_1^2	ML_1	ML_2^1	ML_2^2	ML_2	ML_4^1	ML_4^2	ML_4
(ia)	1.7277	1.4814	1.4557	2.0295	1.6754	1.7889	2.4172	1.5831	1.8279	11.1170
(iia)	1.8565	1.9282	1.8618	6.1753	2.1994	2.4678	37.0216	2.5310	3.1106	11.8804
(iiia)	1.1879	0.9708	0.9787	1.2967	1.0695	1.0418	1.3556	1.3024	1.3811	2.5200
(iva)	0.9000	0.6837	0.6518	0.6905	0.7687	0.6400	0.8586	0.9507	0.9424	1.5629
(ib)	1.6350	1.7479	2.1922	42.0132	1.9364	2.2400	23.4079	2.3381	2.6375	53.3985
(iib)	1.5744	1.8565	2.0718	3.4341	1.5209	1.7120	4.0655	1.7082	1.9290	40.9779
(iiib)	0.8801	0.8281	0.8116	0.8448	0.9387	0.9566	1.0497	1.1491	1.2529	2.1574
(ivb)	0.6109	0.5426	0.5335	0.5435	0.5850	0.5181	0.7107	0.6870	0.7335	1.1426
(ic)	1.1457	1.5232	1.7828	30.3453	1.5395	1.9530	35.2815	1.5033	1.6506	49.4316
(iic)	1.0315	1.1679	1.2773	1.7933	1.1521	1.3587	2.7724	1.1339	1.3077	5.2906
(iiic)	0.6634	0.5804	0.6547	0.8071	0.6106	0.5876	0.8445	0.6356	0.7018	1.6314
(ivc)	0.4130	0.4172	0.4237	0.4270	0.4119	0.4471	0.4827	0.4982	0.5396	0.6326

Frequency of Outlier Occurrence

DGP	CCR	ML_1^1	ML_1^2	ML_1	ML_2^1	ML_2^2	ML_2	ML_4^1	ML_4^2	ML_4
(ia)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	2, 2
(iia)	0, 0	0, 0	0, 0	1, 1	0, 0	0, 0	3, 1	0, 0	0, 0	2, 2
(iiia)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0
(iva)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	1, 0
(ib)	0, 0	0, 0	0, 0	4, 3	0, 0	0, 0	2, 2	0, 0	0, 0	9, 5
(iib)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	1, 0	0, 0	0, 0	5, 3
(iiib)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0
(ivb)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0
(ic)	0, 0	0, 0	0, 0	1, 1	0, 0	0, 0	1, 1	0, 0	0, 0	10, 8
(iic)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	2, 1
(iiic)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0
(ivc)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0

Note: CCR denotes the CCR estimator, and ML_p^k and ML_p signify respectively the k-step iterative estimator and the ML estimator based on VAR(p). In the first panel, the ranges of the estimates are obtained simply by subtracting the minimum from the maximum estimates. In the second panel, the pairs of numbers are the counts of the estimates that fall outside the confidence intervals given by twenty and thirty times of the standard errors.

Table 2B: *Range of Estimate and Frequency of Outlier Occurrence*

DGP: VAR(2)

Range of Estimate

DGP	CCR	ML_1^1	ML_1^2	ML_1	ML_2^1	ML_2^2	ML_2	ML_4^1	ML_4^2	ML_4
(ia)	1.3594	1.2095	1.2036	1.2504	1.4315	1.4406	1.7101	1.6474	1.9283	270.9605
(iia)	1.8865	1.6788	1.6747	1.7912	2.5343	3.0937	12.8223	2.5296	3.1171	520.7314
(iiia)	1.1941	1.1553	1.1245	1.1288	1.2695	1.1793	1.7737	1.4140	1.5389	2.5670
(iva)	1.0010	0.9152	0.8561	0.8108	0.9689	0.7810	1.0311	1.2827	1.5239	2.5318
(ib)	1.3842	1.3309	1.5622	2.0529	1.5510	1.8141	89.1504	1.9513	2.3165	27.7724
(iib)	1.4911	1.5389	1.6137	1.7595	1.6808	1.6505	7.4601	1.7635	1.8627	50.6010
(iiib)	1.0402	0.9807	0.9503	0.9480	1.1691	1.0687	1.2259	1.3679	1.4291	2.1656
(ivb)	0.7428	0.7346	0.6797	0.6430	0.7512	0.6365	0.8713	0.8924	0.8890	1.5321
(ic)	1.1772	1.1035	1.3112	9.5302	1.4467	1.4754	81.7136	1.2957	1.4537	9.0034
(iic)	1.0083	1.0321	1.0829	1.2699	1.2051	1.2636	14.1876	1.2022	1.3847	7.7970
(iiic)	0.8046	0.8302	0.7952	0.7637	0.7764	0.7038	1.2543	0.9046	0.9788	3.0189
(ivc)	0.5185	0.5077	0.4791	0.4723	0.5308	0.5244	0.5952	0.6382	0.6529	0.8154

Frequency of Outlier Occurrence

DGP	CCR	ML_1^1	ML_1^2	ML_1	ML_2^1	ML_2^2	ML_2	ML_4^1	ML_4^2	ML_4
(ia)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	3, 3
(iia)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	4, 2	0, 0	0, 0	7, 4
(iiia)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0
(iva)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	1, 1
(ib)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	5, 3	0, 0	0, 0	9, 7
(iib)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	3, 1	0, 0	0, 0	8, 5
(iiib)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0
(ivb)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	1, 0
(ic)	0, 0	0, 0	0, 0	1, 1	0, 0	0, 0	1, 1	0, 0	0, 0	6, 4
(iic)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	1, 1	0, 0	0, 0	4, 3
(iiic)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	1, 0
(ivc)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0

Note: CCR denotes the CCR estimator, and ML_p^k and ML_p signify respectively the k-step iterative estimator and the ML estimator based on VAR(p). In the first panel, the ranges of the estimates are obtained simply by subtracting the minimum from the maximum estimates. In the second panel, the pairs of numbers are the counts of the estimates that fall outside the confidence intervals given by twenty and thirty times of the standard errors.

Table 2C: *Range of Estimate and Frequency of Outlier Occurrence*

DGP: VARMA(1,1)

Range of Estimate

DGP	CCR	ML_1^1	ML_1^2	ML_1	ML_2^1	ML_2^2	ML_2	ML_4^1	ML_4^2	ML_4
(ia)	1.3455	1.5978	2.1731	3.6093	1.2926	1.9707	17.3348	1.5432	1.6432	586.5321
(iia)	1.9387	2.2690	2.2648	119.9999	2.2476	2.4504	51.2471	2.6155	3.4402	712.2868
(iiia)	1.4547	1.4677	1.4192	5.4320	1.5041	1.5201	9.2573	1.5353	1.7292	3.7831
(iva)	0.9188	0.7523	0.7520	0.8055	0.8947	0.8653	0.9665	1.1351	1.2003	1.4154
(ib)	1.7109	1.8612	1.8655	2.2744	1.6469	1.7915	38.5454	1.5508	1.8209	180.1772
(iib)	1.4765	1.6751	1.7678	202.1409	1.6280	1.9100	7.8114	2.0613	2.4116	34.2359
(iiib)	0.9580	1.0215	1.1561	1.5916	1.0138	1.1137	1.5689	1.0345	1.0524	1.9694
(ivb)	0.7340	0.5528	0.5840	0.6970	0.6165	0.6093	0.6587	0.7824	0.8437	1.8622
(ic)	1.7545	1.7947	2.1650	29.8126	1.9168	2.1703	12.2952	1.9938	2.5088	71.8730
(iic)	1.2635	1.5388	1.6198	61.8021	1.3542	1.4390	2.4819	1.4819	1.7355	9.1092
(iiic)	0.8032	0.7075	0.7553	30.3574	0.7314	0.8426	0.9284	0.9627	1.0209	4.3405
(ivc)	0.4872	0.4351	0.4715	0.4991	0.4401	0.4856	0.6687	0.4821	0.5241	1.2119

Frequency of Outlier Occurrence

DGP	CCR	ML_1^1	ML_1^2	ML_1	ML_2^1	ML_2^2	ML_2	ML_4^1	ML_4^2	ML_4
(ia)	0, 0	0, 0	0, 0	2, 1	0, 0	0, 0	2, 2	0, 0	0, 0	8, 5
(iia)	0, 0	0, 0	0, 0	10, 9	0, 0	0, 0	6, 5	0, 0	0, 0	11, 8
(iiia)	0, 0	0, 0	0, 0	1, 1	0, 0	0, 0	1, 1	0, 0	0, 0	0, 0
(iva)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0
(ib)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	4, 3	0, 0	0, 0	10, 6
(iib)	0, 0	0, 0	0, 0	4, 4	0, 0	0, 0	1, 1	0, 0	0, 0	7, 4
(iiib)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0
(ivb)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	1, 0
(ic)	0, 0	0, 0	0, 0	6, 3	0, 0	0, 0	7, 4	0, 0	0, 0	17, 12
(iic)	0, 0	0, 0	0, 0	4, 1	0, 0	0, 0	0, 0	0, 0	0, 0	7, 3
(iiic)	0, 0	0, 0	0, 0	1, 1	0, 0	0, 0	0, 0	0, 0	0, 0	1, 1
(ivc)	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	3, 0	3, 0	2, 0

Note: CCR denotes the CCR estimator, and ML_p^k and ML_p signify respectively the k-step iterative estimator and the ML estimator based on VAR(p). In the first panel, the ranges of the estimates are obtained simply by subtracting the minimum from the maximum estimates. In the second panel, the pairs of numbers are the counts of the estimates that fall outside the confidence intervals given by twenty and thirty times of the standard errors.

Table 3: *Running Estimates for SP500 Index/Index Futures Cointegrating Coefficient*

Size: 101, Period: 10/13/1998 - 03/23/1999

From	To	CCR	ML_9^1	ML_9^2	ML_9
10/13/1998	03/09/1999	.9961	.9967	.9978	1.0014
		(.0007)	(.0020)	(.0018)	(.0023)
10/14/1998	03/10/1999	.9961	.9965	.9976	1.0025
		(.0007)	(.0020)	(.0019)	(.0027)
10/15/1998	03/11/1999	.9960	.9961	.9966	1.0058
		(.0007)	(.0020)	(.0019)	(.0039)
10/16/1998	03/12/1999	.9961	.9960	.9968	1.0134
		(.0007)	(.0024)	(.0024)	(.0069)
10/19/1998	03/15/1999	.9962	.9957	.9954	1.0376
		(.0007)	(.0022)	(.0021)	(.0171)
10/20/1998	03/16/1999	.9963	.9958	.9954	**37.6418**
		(.0007)	(.0021)	(.0020)	**(15.4529)**
10/21/1998	03/17/1999	.9964	.9960	.9959	1.1477
		(.0007)	(.0021)	(.0021)	(.0644)
10/22/1998	03/18/1999	.9962	.9958	.9954	.9795
		(.0007)	(.0018)	(.0018)	(.0070)
10/23/1998	03/19/1999	.9962	.9957	.9953	.9850
		(.0007)	(.0018)	(.0017)	(.0050)
10/26/1998	03/22/1999	.9960	.9950	.9943	.9915
		(.0008)	(.0015)	(.0014)	(.0022)
10/27/1998	03/23/1999	.9957	.9950	.9945	.9933
		(.0008)	(.0014)	(.0013)	(.0017)

Note: CCR denotes the CCR estimator, and ML_9^k and ML_9 signify respectively the k-step iterative estimator and the ML estimator based on VAR(9). The presented estimates are obtained in the cointegrating regression of the SP500 index against the index futures using eleven sets of samples, each of which consists of the data for the 101 consecutive business days. The estimated standard errors are given in parenthesis.