Asymptotic Properties of the Efficient Estimators for Cointegrating Regression Models with Serially Dependent Errors[1]

EIJI KUROZUMI[2] KAZUHIKO HAYAKAWA[3]

Department of Economics

Hitotsubashi University

December 6, 2006

This version: October 3, 2008

Abstract

In this paper, we analytically investigate three efficient estimators for cointegrating regression models: Phillips and Hansen's (1990) fully modified OLS estimator, Park's (1992) canonical cointegrating regression estimator, and Saikkonen's (1991) dynamic OLS estimator. We consider the case where the regression errors are moderately serially correlated and the AR coefficient in the regression errors approaches 1 at a rate slower than $1/T$, where T represents the sample size. We derive the limiting distributions of the efficient estimators under this system and find that they depend on the approaching rate of the AR coefficient. If the rate is slow enough, efficiency is established for the three estimators; however, if the approaching rate is relatively faster, the estimators will have the same limiting distribution as the OLS estimator. For the intermediate case, the second-order bias of the OLS estimator is partially eliminated by the efficient methods. This result expl! ains why, in finite samples, the effect of the efficient methods diminishes as the serial correlation in the regression errors becomes stronger. We also propose to modify the existing efficient estimators in order to eliminate the second-order bias, which possibly remains in the efficient estimators. Using Monte Carlo simulations, we demonstrate that our modification is effective when the regression errors are moderately serially correlated and the simultaneous correlation is relatively strong.

JEL classification: C13; C22

Key words: Cointegration; second-order bias; fully modified regressions; canonical cointegrating regressions; dynamic ordinary least squares regressions

[1]We are grateful to Professors In Choi, Kotaro Hitomi, Mitsuru Nakagawa, Yoshihiko Nishiyama, Peter Phillips, Katsuto Tanaka, an associate editor and two anonymous referees for their helpful comments. All errors are our responsibility. Correspondence: Eiji Kurozumi, Department of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo, 186-8601, Japan.

[2]Kurozumi's research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology under Grants-in-Aid Nos. 17203016 and 18730142.

[3]Hayakawa is a JSPS research fellow and acknowledges its financial support.

1. Introduction

Since the seminal work of Engle and Granger (1987), cointegrating regressions have become one of the standard tools for analyzing integrated variables. With regard to the estimation of cointegrating regression models, it is well known that the ordinary least squares (OLS) estimator contains the second-order bias, comprising the endogeneity bias and the non-centrality bias, when the I(1) regressors are endogenous and/or the regression errors are serially correlated. Thus, several efficient methods for estimating cointegrating regressions have been proposed in the literature. Phillips and Hansen (1990) proposed a nonparametric correction for the OLS estimator; their method is known as the fully modified regression (FMR) method and it was further developed by Phillips (1995) and Kitamura and Phillips (1997). Park (1992) proposed the canonical cointegrating regression (CCR) method, which is also based on a nonparametric correction that is similar to the FMR method. However, the ! CCR method eliminates the non-centrality bias in a different manner. On the other hand, Phillips and Loretan (1991), Saikkonen (1991), and Stock and Watson (1993) considered a parametric correction by adding leads and lags of the first difference of the I(1) variables as regressors; this method is known as the dynamic ordinary least squares (DOLS) method. These three efficient estimators—FMR, CCR, and DOLS —are asymptotically equivalent, and as proved by Saikkonen (1991), they are efficient.

However, the finite sample behavior of these estimators is fairly different as reported by, for example, Inder (1993), Montalvo (1995), Cappuccio and Lubian (2001), and Christou and Pittis (2002) using Monte Carlo simulations. The first two papers recommend the use of the DOLS type approach to eliminate the second-order bias of the OLS estimator, whereas the last paper demonstrated that the FMR estimator outperforms the DOLS estimator in terms of the bias; thus, the answer to the question of which estimator performs best in finite samples remains inconclusive. It appears that the performance of the three efficient estimators is fairly dependent on the data generating process used in Monte Carlo simulations, as pointed out by Cappuccio and Lubian (2001). However, these Monte Carlo simulations commonly suggest that the efficient estimation methods break down and perform very poorly when the cointegrating regression errors are strongly serially correlated. Although the finite s! ample performance of the FMR and CCR estimators may improve if the prewhitening method by Andrews and Monahan (1992), which has been further modified by Sul, Phillips, and Choi (2005), is used to estimate the long-run variance, a large bias still remains in the estimator as shown in the discussion paper version of this paper.

In this paper, we analytically explain the poor performance of the three efficient estimators with a moderate serial correlation. We introduce the local-to-unity system in which the AR coefficient approaches 1 at a rate slower than $1/T$, where T represents the sample size. This type of local-to-unity system is considered by Phillips and Magdalinos (2007a, b) and Giraitis and Phillips (2006). We will demonstrate that the limiting distributions of the efficient estimators change depending on the approaching speed of the AR coefficient. Intuitively, the three efficient methods can eliminate the second-order bias of the OLS estimator if the AR coefficient approaches 1 slowly enough; however, these methods no longer work well when the approaching speed is very fast. For the intermediate case, the second-order bias of the OLS estimator is partially eliminated by these efficient methods; however, a part of the bias still remains. This result explains why the effect of the effici! ent methods diminishes as the serial correlation in the regression errors becomes stronger. We will demonstrate that the result depends on a relation between the approaching speed of the AR coefficient and the diverging rate of the bandwidth parameter used for the estimation of the long-run variance in the FMR and CCR methods or the diverging rate of the lead-lag truncation parameter used in the DOLS method. We also propose to modify the FMR and CCR estimators in order to eliminate the second-order bias when the regression errors are moderately serially correlated. The estimation of the localizing parameter plays a key role in our modification.

The remainder of this paper is organized as follows. In Section 2, we briefly review the FMR, CCR, and DOLS methods. Section 3 investigates the asymptotic properties of the three efficient estimators as well as the OLS estimator under the local-to-unity system in which the AR coefficient approaches 1 at a rate slower than $1/T$. In section 4, we propose the modified FMR and CCR estimators, and in section 5, the finite sample properties of the estimators are investigated using Monte Carlo simulations. Concluding remarks are provided in Section 6.

2. Review of the Efficient Estimation Methods

This section reviews the three efficient estimators for cointegration regression models. Let us consider the following model:

$$\begin{aligned} y_t &= \mu + \beta' x_t + u_{1,t} = \theta' z_t + u_{1,t} \\ \Delta x_t &= u_{2,t} \end{aligned} \tag{1}$$

for $t = 1, \cdots, T$, where $\theta = (\mu, \beta')'$, $z_t = (1, x_t')'$, and y_t and x_t are observed time series with 1 and n dimensions, respectively. For $u_t = [u_{1,t}, u_{2,t}']'$, we assume that the functional central limit theorem (FCLT) can be applied as follows:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t \Rightarrow B(r) = \begin{bmatrix} B_1(r) \\ B_2(r) \end{bmatrix} \tag{2}$$

for $0 \leq r \leq 1$, where $B(\cdot)$ is a Brownian motion on $[0, 1]$ with a variance-covariance matrix Ω ($B(\cdot) \sim BM(\Omega)$) and $\Rightarrow$ signifies weak convergence of the associated probability measures. We assume that Ω is positive definite. Note that the long-run variance of u_t and its one-sided version can be expressed as

$$\Omega = \Sigma + \Phi + \Phi' \quad \text{and} \quad \Lambda = \Sigma + \Phi,$$

$$\text{where} \quad \Sigma = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E(u_t u_t') \quad \text{and} \quad \Phi = \lim_{T \to \infty} \frac{1}{T} \sum_{j=1}^{T-1} \sum_{t=1}^{T-j} E(u_t u_{t+j}').$$

We partition Ω and Λ conformably with u_t as

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{bmatrix} \qquad \Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \Lambda_{22} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \Lambda_2 \end{bmatrix}. \tag{3}$$

It is known that the OLS estimator of θ, denoted by $\hat{\theta}$, is consistent but inefficient in general. The centered OLS estimator with a normalizing matrix $D_T = \text{diag}\{\sqrt{T}, T I_n\}$ weakly converges to

$$D_T(\hat{\theta} - \theta) \Rightarrow \left(\int_0^1 \underline{B}_2(r) \underline{B}_2'(r) dr \right)^{-1} \left(\int_0^1 \underline{B}_2(r) dB_1(r) + [0, \lambda_{21}']' \right) \tag{4}$$

where $\underline{B}_2(r) = [1, B_2'(r)]'$ and we can observe that this limiting distribution contains the second-order bias from the correlation between $B_1(\cdot)$ and $B_2(\cdot)$ and the non-centrality parameter λ_{21}. As explained in Phillips and Hansen (1990) and Phillips (1995), the former bias arises from the endogeneity of the I(1) regressor x_t while the non-centrality bias comes from the fact that the regression errors are serially correlated.

In order to eliminate the second-order bias, Phillips and Hansen (1990) propose the FMR estimator, which is defined as

$$\hat{\theta}_{FMR} = \left(\sum_{t=1}^{T} z_t z_t'\right)^{-1} \left(\sum_{t=1}^{T} z_t y_t^+ - T\hat{J}^+\right), \tag{5}$$

$$\text{where} \quad y_t^+ = y_t - \hat{\omega}_{12}\hat{\Omega}_{22}^{-1}u_{2t} \quad \text{and} \quad \hat{J}^+ = \begin{bmatrix} 0 \\ \hat{\lambda}_{21} - \hat{\Lambda}_{22}\hat{\Omega}_{22}^{-1}\hat{\omega}_{21} \end{bmatrix},$$

with $\hat{\omega}_{12}$, $\hat{\Omega}_{22}$, $\hat{\lambda}_{21}$, and $\hat{\Lambda}_{22}$ being consistent estimators of ω_{12}, Ω_{22}, λ_{21}, and Λ_{22}, respectively. It can be shown that the correction term for y_t is associated with the correction for the endogeneity bias while $\hat{J}^+$ eliminates the non-centrality bias.

In order to define the CCR estimator, we first modify y_t and x_t such that

$$y_t^* = y_t - \left(\hat{\beta}'\hat{\Lambda}_2\hat{\Sigma}^{-1} + [0, \hat{\omega}_{12}\hat{\Omega}_{22}^{-1}]\right)\hat{u}_t \quad \text{and} \quad z_t^* = (1, x_t^{*'})' \quad \text{with} \quad x_t^* = x_t - \hat{\Lambda}_2\hat{\Sigma}^{-1}\hat{u}_t,$$

where $\hat{\beta}$ is the OLS estimator of β and $\hat{u}_t = [\hat{u}_{1t}, \Delta x_t']'$ consists of the OLS residuals and the first difference of the I(1) regressors. Then, the CCR estimator proposed by Park (1992) is defined as

$$\hat{\theta}_{CCR} = \left(\sum_{t=1}^{T} z_t^* z_t^{*'}\right)^{-1} \left(\sum_{t=1}^{T} z_t^* y_t^*\right). \tag{6}$$

The CCR method uses the same principle as the FMR method in order to eliminate the endogeneity bias, while it deals with the non-centrality parameter in a different manner.

Contrary to the non-parametric approaches adopted by the FMR and CCR methods, the DOLS method is based on parametric regressions. Phillips and Loretan (1991), Saikkonen (1991), and Stock and Watson (1993) propose to augment the leads and lags of the first difference of x_t as regressors and to estimate

$$y_t = \theta' z_t + \sum_{j=-K}^{K} \pi_j' \Delta x_{t-j} + \eta_t. \tag{7}$$

The DOLS estimator is defined as the OLS estimator of θ for (7):

$$\hat{\theta}_{DOLS} = \left(\sum_{t=K+1}^{T-K} \tilde{z}_t \tilde{z}_t' \right)^{-1} \left(\sum_{t=K+1}^{T-K} \tilde{z}_t \tilde{y}_t \right), \tag{8}$$

where $\tilde{z}_t$ and $\tilde{y}_t$ are regression residuals of z_t and y_t on $w_t = (u_{2,t+K}', ..., u_{2,t-K}')'$, respectively. The regression form (7) is based on the fact that under some regularity conditions, the regression errors $u_{1,t}$ in (1) can be expressed as

$$u_{1,t} = v_t + \sum_{j=-\infty}^{\infty} \pi_j' u_{2,t-j} = v_t + r_t, \tag{9}$$

where $r_t = \sum_{j=-\infty}^{\infty} \pi_j' u_{2,t-j}$ and $\sum_{j=-\infty}^{\infty} ||\pi_j|| < \infty$ with $|| \cdot ||$ being the standard Euclidian norm; further, v_t is uncorrelated with u_{2t-j} for all j. For details, refer to Brillinger (1981). From (9), we observe that

$$\eta_t = v_t + \sum_{|j|>K} \pi_j' u_{2,t-j}. \tag{10}$$

The uncorrelatedness of v_t with all the leads and lags of $u_{2,t}$ is an important property to prove that the DOLS method successfully eliminates the second-order bias of the OLS estimator.

As explained in Phillips and Hansen (1990), Saikkonen (1991), and Park (1992), these three efficient estimators have an identical limiting distribution that is given by

$$D_T(\hat{\theta}_E - \theta) \Rightarrow \left(\int_0^1 \underline{B}_2(r) \underline{B}_2'(r) dr \right)^{-1} \int_0^1 \underline{B}_2(r) dB_{1\cdot 2}(r), \tag{11}$$

where $\hat{\theta}_E = \hat{\theta}_{FMR}$, $\hat{\theta}_{CCR}$, and $\hat{\theta}_{DOLS}$ and $B_{1\cdot 2}(\cdot) \sim BM(\omega_{1\cdot 2})$ is independent of $B_2(\cdot)$ with $\omega_{1\cdot 2} = \omega_{11} - \omega_{12} \Omega_{22}^{-1} \omega_{21}$. Then, we observe that the three efficient methods can eliminate both the endogenous bias and the non-centrality parameter. Moreover, Saikkonen (1991) showed that this limiting distribution is efficient in a certain class of estimators.

3. Asymptotic Properties of the Estimators with Moderately Serially Correlated Errors

3.1. Model and assumptions

This section investigates the asymptotic properties of the three efficient estimators as well as the OLS estimator when the cointegrating regression errors are moderately serially correlated. As explained in Section 1, the finite sample performance of the efficient estimators gradually becomes poorer as the serial correlation in the regression errors becomes stronger. In order to explain this finite sample evidence using asymptotic theory, we consider model (1) with the following structure in the error term:

$$y_t = \theta' z_t + \dot{u}_{1,t}, \quad \dot{u}_{1,t} = \rho \dot{u}_{1,t-1} + u_{1,t}, \quad \Delta x_t = u_{2,t}, \tag{12}$$

$$u_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}, \quad \sum_{j=1}^{\infty} j^2 \|\Psi_j\| < \infty \quad \text{with} \quad \Psi_0 = I_{n+1}, \tag{13}$$

where $u_{1,0}$ is some constant independent of T, $u_t = [u_{1,t}, u_{2,t}']'$, $\varepsilon_t = [\varepsilon_{1,t}, \varepsilon_{2,t}']' \sim i.i.d.(0, \Sigma_\varepsilon)$ with a spectral density function bounded away from 0 and above and with finite δ-th moment for some $\delta > 4$, and $\|\cdot\|$ is a matrix norm defined by $\|A\| = (tr(A'A))^{1/2}$ for a given matrix A. Note that for a given sequence of $\{\Psi_j\}$, the strength of the serial correlation in the error term can be changed by the AR coefficient ρ.

Following Phillips and Solo (1992), let us decompose u_t into

$$u_t = \Psi_L \varepsilon_t + \tilde{u}_{t-1} - \tilde{u}_t, \quad \text{where} \quad \Psi_L = \sum_{j=0}^{\infty} \Psi_j, \quad \tilde{u}_t = \sum_{j=0}^{\infty} \tilde{\Psi}_j \varepsilon_{t-j}, \quad \tilde{\Psi}_j = \sum_{i=j+1}^{\infty} \Psi_i. \tag{14}$$

Note that $\tilde{u}_t$ is stationary with $\sum_{j=1}^{\infty} j \|\tilde{\Psi}_j\| < \infty$. The short- and long-run variance matrices of u_t are defined in a manner similar to that in the previous section:

$$\Sigma = E(u_t u_t') = \sum_{j=0}^{\infty} \Psi_j \Sigma_\varepsilon \Psi_j', \quad \Omega = \sum_{j=-\infty}^{\infty} E[u_t u_{t+j}'] = \Psi_L \Sigma_\varepsilon \Psi_L',$$

$$\Phi = \sum_{j=1}^{\infty} E[u_t u_{t+j}'] = \sum_{j=1}^{\infty} \sum_{i=0}^{\infty} \Psi_i \Sigma_\varepsilon \Psi_{i+j}' = \sum_{i=0}^{\infty} \Psi_i \Sigma \tilde{\Psi}_i', \quad \Lambda = \Sigma + \Phi.$$

We assume that Ω is positive definite. Note that the weak convergence (2) holds under the above assumption. We can also observe by Theorems 3.8.3 and 8.3.1 of Brillinger (1981) that the relation (9) holds with $\sum_{j=-\infty}^{\infty} j^2 \|\pi_j\| < \infty$.

With regard to the FMR and CCR estimators, we focus on the case where the long-run variances are estimated by the kernel method as follows:

$$\hat{\Omega} = \hat{\Sigma} + \hat{\Phi} + \hat{\Phi}' \quad \text{and} \quad \hat{\Lambda} = \hat{\Sigma} + \hat{\Phi},$$

$$\text{where} \quad \hat{\Sigma} = \hat{\Gamma}(0), \quad \hat{\Phi} = \sum_{j=1}^{T-1} k\left(\frac{j}{M}\right)\hat{\Gamma}(j), \quad \text{and} \quad \hat{\Gamma}(j) = \frac{1}{T}\sum_{t=1}^{T-j} \hat{u}_t \hat{u}_{t+j}'$$

with $\hat{u}_t = [\hat{u}_{1,t}, \Delta x_t']'$, where $\hat{u}_{1,t}$ is the regression residual of y_t on z_t and $k(\cdot)$ is a kernel function.

Before proceeding with the asymptotic analysis, it is necessary to state two assumptions concerning the kernel method and the lead-lag truncation parameter.

Assumption 1 *(a) The kernel* $k(\cdot) : R \to [-1,1]$ *is continuous at zero,* $k(0) = 1$, $\sup_{r \geq 0} |k(r)| < \infty$, *and* $\int_{[0,\infty)} \bar{k}(r)dr < \infty$, *where* $\bar{k}(r) = \sup_{s \geq r} |k(s)|$. *(b) The bandwidth parameter* M *goes to infinity as* $T \to \infty$ *and* $M = o(T^{1/2})$.

Assumption 2 *(a)* $K = o(T^{1/2})$. *(b)* $K \sum_{|j|>K} ||\pi_j|| \to 0$.

Assumption 1 is sufficient to consistently estimate the long-run variances. See Jansson (2002). Assumption 2 is provided by Kejriwal and Perron (2008) and is different from that in Saikkonen (1991) in which it is assumed that (a') $K = o(T^{1/3})$ and (b') $T^{1/2} \sum_{|j|>K} ||\pi_j|| \to 0$. The assumptions (a) and (a') are associated with the upper bound condition on K; Kejriwal and Perron (2008) proved that K can increase at a faster rate than that considered by Saikkonen (1991). The assumptions (b) and (b') provide the lower bound condition on K; the matter of importance is that (b') excludes the case where K is chosen by an information criterion, while (b) is sufficiently general for K to increase at a logarithmic rate and then allow an information criterion for the selection of K. Note that Assumption 2 (b) is automatically satisfied in our model because $K\sum_{|j|>K}^{\infty}\|\pi_j\| \leq \sum_! |j| > K^{\infty} j\|\pi_j\| \to 0$ as $K \to \infty$, which is guaranteed by $\sum_{j=-\infty}^{\infty} j^2\|\pi_j\| < \infty$.

Under Assumptions 1 and 2, it is shown that when ρ is fixed, the centered estimators have the same limiting distributions as given by (4) and (11). However, the assumption of the fixed ρ is not necessarily appropriate when the error term is moderately serially correlated. In the following subsection, we consider a local-to-unity system such that ρ approaches 1 as T goes to infinity.

3.2. Asymptotic properties with the N local-to-unity system

In this section, we consider the case where the AR coefficient ρ is moderately or relatively strongly close to 1. Such a case can be modeled by the N local-to-unity system, which is defined as

$$\rho = \rho_N = 1 - \frac{c}{N}, \quad N \to \infty \quad \text{and} \quad \frac{N}{T} \to 0. \tag{15}$$

Note that the N local-to-unity system is different from the conventional local-to-unity system with $\rho = 1 - c/T$, which we refer to as the T local-to-unity system. The T local-to-unity system has often been assumed in the literature in order to investigate the asymptotic local power of unit root/cointegration tests. For example, see Phillips (1987), Tanaka (1996), and Saikkonen and Lütkepohl (1999) among others. Note that we usually test for cointegration before estimating cointegrating regression models and that tests for cointegration do not necessarily detect the existence of cointegration with probability 1 even asymptotically when $\rho = 1 - c/T$. This is because the T local-to-unity system corresponds to the local alternative for cointegration tests. In this sense, the T local-to-unity system is not attractive for investigating the cointegrating relation. On the other hand, with the N local-to-unity system, tests for cointegration detect the cointegrating re! lation with an asymptotic probability 1 because $\rho = \rho_N$ approaches 1 at a slower rate than does the T local-to-unity system. The aim in considering the N local-to-unity system is to investigate the behavior of the estimators when the AR coefficient is close to 1 but not too close. This type of local-to-unity system is also considered by Phillips and Magdalinos (2007a, b) and Giraitis and Phillips

(2006).

In order to obtain the limiting distributions of the efficient estimators it is necessary to make the following assumption.

Assumption 3 *(a)* $M/N^2 \to 0$. *(b)* $K/N^2 \to 0$.

This assumption restricts the upper bounds on the bandwidth parameter M and the lead-lag truncation parameter K; as N gets larger or ρ gets closer to 1, we can choose larger M and K, although they must satisfy $M = o(T^{1/2})$ and $K = o(T^{1/2})$, as is the case in Assumptions 1 and 2. Assumption 3 will not be required in a special case where u_t is an i.i.d. sequence, as investigated by the discussion paper version of this paper. For a general linear process, we need Assumption 3 in order for the long-run variance estimator to converge and for the remaining term $\sum_{|j|>K} \pi_j' u_{2,t-j}$ in (10) to be negligible.

The following theorem provides the asymptotic distributions of the estimators.

Theorem 1 *Under Assumptions 1–3 with the N local-to-unity system,*

$$\frac{1}{N} D_T(\hat{\theta} - \theta) \Rightarrow H^{-1} h_{OLS}, \tag{16}$$

$$\frac{1}{N} D_T(\hat{\theta}_{FMR} - \theta) \Rightarrow H^{-1} h_{FMR}, \tag{17}$$

$$\frac{1}{N} D_T(\hat{\theta}_{CCR} - \theta) \Rightarrow H^{-1} h_{FMR}, \tag{18}$$

$$\frac{1}{N} D_T(\hat{\theta}_{DOLS} - \theta) \Rightarrow H^{-1} h_{DOLS}, \tag{19}$$

where $H = \int_0^1 \underline{B}_2(r) \underline{B}_2'(r) dr$ *and*

$$h_{OLS} = \begin{bmatrix} \frac{1}{c} B_1(1) \\ \frac{1}{c} \left(\int_0^1 B_2(r) dB_1(r) + \omega_{21} \right) \end{bmatrix},$$

$$h_{FMR} = \begin{cases} \begin{bmatrix} \frac{1}{c}B_{1\cdot 2}(1) \\ \frac{1}{c}\int_0^1 B_2(r)dB_{1\cdot 2}(r) \end{bmatrix}, & \frac{M}{N} \to \infty, \\ \begin{bmatrix} \frac{1}{c}\left(B_1(1) - B_2'(1)\Omega_{22}^{-1}\omega_{21}\kappa\right) \\ \left[\frac{1}{c}\left(\int_0^1 B_2(r)dB_1(r) - \int_0^1 B_2(r)dB_2'(r)\Omega_{22}^{-1}\omega_{21}\kappa\right)\right. \\ \left. +\frac{1}{c}(1-\kappa)\omega_{21}\right] \end{bmatrix}, & \frac{M}{N} \to d_M, \\ \begin{bmatrix} \frac{1}{c}B_1(1) \\ \frac{1}{c}\left(\int_0^1 B_2(r)dB_1(r) + \omega_{21}\right) \end{bmatrix}, & \frac{M}{N} \to 0, \end{cases}$$

$$h_{DOLS} = \begin{cases} \begin{bmatrix} \frac{1}{c}B_{1\cdot 2}(1) \\ \frac{1}{c}\int_0^1 B_2(r)dB_{1\cdot 2}(r) \end{bmatrix}, & \frac{K}{N} \to \infty, \\ \begin{bmatrix} \frac{1}{c}\left\{(1-e^{-cd_K})B_{1\cdot 2}(1) + e^{-cd_K}B_1(1)\right\} \\ \left[\frac{1}{c}\left\{(1-e^{-cd_K})\int_0^1 B_2(r)dB_{1\cdot 2}(r)\right.\right. \\ \left.\left. +e^{-cd_K}\left(\int_0^1 B_2(r)dB_1(r) + \omega_{21}\right)\right\}\right] \end{bmatrix}, & \frac{K}{N} \to d_K, \\ \begin{bmatrix} \frac{1}{c}B_1(1) \\ \frac{1}{c}\left(\int_0^1 B_2(r)dB_1(r) + \omega_{21}\right) \end{bmatrix}, & \frac{K}{N} \to 0, \end{cases}$$

with $\kappa = (cd_M)\int_0^\infty k(r)e^{-(cd_M)r}dr$, d_M *and* d_K *being fixed positive values and* $B_{1\cdot 2}(r) = B_1(r) - \omega_{12}\Omega_{22}^{-1}B_2(r)$.

Remark 1: All the estimators are consistent under the N local-to-unity system; however, they are not T-consistent but the convergence rate is slower than T.

Remark 2: Since M and K must be slower than $T^{1/2}$ from Assumptions 1(b) and 2(a), we cannot expect an efficiency gain by using the efficient estimation methods for the case where N grows at a faster rate than $T^{1/2}$. The cases where $M/N \to \infty$ or d_M and where $K/N \to \infty$ or d_K can be considered only when $N = o(T^{1/2})$. In other words, the three efficient methods possibly work well only in the case where ρ is moderately close to 1.

Remark 3: When M/N or $K/N \to \infty$, the distributions of the centered efficient estimators normalized by D_T can be approximated by

$$D_T(\hat{\theta}_E - \theta) \simeq_d \frac{N}{c}\left(\int_0^1 \underline{B}_2(r)\underline{B}_2'(r)dr\right)^{-1}\int_0^1 \underline{B}_2(r)dB_{1\cdot 2}(r). \quad (20)$$

Then, we can see that the distribution under the N local-to-unity system is approximately N/c times the efficient distribution (11). For example, since $c/N = 1 - \rho$, the standard deviation of (20) when $\rho = 0.7$, 0.8, and 0.9 is approximately 3.3, 5, and 10 times larger than that of (11). It is apparent that the standard deviation becomes larger as ρ is closer to 1.

From Theorem 1, we observe that the three efficient estimators do not have the second-order bias when N is sufficiently slow as compared with the bandwidth parameter M or the lead-lag truncation parameter K. This implies that compared with the OLS estimator, the FMR, CCR, and DOLS estimators are efficient when ρ approaches 1 slowly or ρ is sufficiently away from 1. On the other hand, when ρ approaches 1 rapidly or when ρ is very close to 1, these three estimators have the same asymptotic distribution as the OLS estimator and hence suffer from the second-order bias. For the intermediate case where N is of the same order as M or K, the second-order bias persists in the efficient estimators; the bias is only partially eliminated by the efficient methods. For example, the endogeneity bias of $\hat{\beta}$ is partially eliminated from the FMR and CCR methods by observing the corresponding term of h_{FMR} when $M/N \to d_M$,

$$dB_1(r) - dB_2'(r)\Omega_{22}^{-1}\omega_{21}\kappa,$$

while the noncentrality is adjusted by the term $(1-\kappa)\,\omega_{21}$. Note that

$$0 \le \kappa = cd_M \int_0^\infty k(r)e^{-(cd_M)r}dr \le cd_M \int_0^\infty e^{-(cd_M)r}dr = 1 \tag{21}$$

if $k(r) \ge 0$ for positive r, which is satisfied by, for example, the Bartlett and Parzen kernels. Inequality (21) is also satisfied by the quadratic spectral (QS) kernel. Since $\kappa \to 0$ as $c \to 0$, we can see that the adjustment for the second-order bias decreases as ρ approaches 1. Since $B_{1\cdot 2}(r) = B_1(r) - \omega_{12}\Omega_{22}^{-1}B_2(r)$, we can also express h_{FMR} when $M/N \to d_M$ as

$$h_{FMR} = \begin{bmatrix} \frac{1}{c}\left\{\kappa B_{1\cdot 2}(1) + (1-\kappa)B_1(1)\right\} \\ \frac{1}{c}\left\{\kappa \int_0^1 B_2(r)dB_{1\cdot 2}(r) + (1-\kappa)\left(\int_0^1 B_2(r)dB_1(r) + \omega_{21}\right)\right\} \end{bmatrix};$$

thereafter the limiting distribution in this case can be observed as the weighted sum of the efficient and inefficient distributions. This implies that the corresponding distribution is located between the efficient and inefficient distributions. Similar effects can be observed for the limiting distribution of the DOLS estimator when $K/N \to d_K$. Thus, Theorem 1 implies that when ρ is relatively further away from 1, the three efficient methods are effective compared to the OLS estimator. However, as ρ approaches 1, the difference between the efficient estimators and the OLS estimator reduces, and eventually, when ρ is sufficiently close to 1, the difference becomes negligible. This is consistent with the finite sample behavior of the estimators observed in Section 5 and in previous literature. In other words, the N local-to-unity system can adequately explain the finite sample evidence that the effect of the efficient methods gradually! diminishes as ρ approaches 1.

We demonstrate the probability density functions (pdf) of the distributions provided in Theorem 1. Figure 1 illustrates the pdfs[4] for $\omega_{21} = 0.4$ and 0.8, $c = 1/2$ and 1, and $d_K = d_M = 1$. These are obtained from 100,000 replications from the distribution of the discrete approximation based on 2,000 steps to the limiting distribution provided in Theorem 1. We can observe that the limiting distribution for a slow N is centered at and symmetric around the origin, whereas the limiting distribution of the OLS estimator is shifted and skewed toward the right-hand side. In addition, by observing the limiting distributions corresponding to the cases where $M/N = 1$! and $K/N = 1$, we observe that the efficient methods partially eliminate the second-order bias. Overall, the second-order bias of the OLS estimator increases for a larger ω_{21} and a smaller c.

4. Modification of the Efficient Methods

As shown in the previous section, we need to carefully choose the bandwidth parameter and the lead-lag truncation parameter in order for the FMR, CCR, and DOLS methods to work appropriately. Theorem 1 suggests that both M and K should be as large as possible. This implies that a bandwidth selected by an existing data dependent rule is not necessarily appropriate when the serial correlation in the error term is moderately strong; we need to seek the selection rule for a bandwidth and a truncation parameter. In this case, the

[4]These densities are drawn for the range of 1% to 99% points by the kernel method with a Gaussian kernel. The smoothing parameter, h, is decided by equation (3.31) in Silverman (1986): $h = 0.9AT^{-1/5}$ where $A = \min(\text{standard deviation, interquartile range}/1.34)$.

difficulty lies in the fact that although M and K should be as large as possible, they must also be strictly slower than N^2 based on Assumption 3. As a result, even if we prespecify the growing rate of N, it appears practically difficult to decide the theoretical optimal rate of the bandwidth parameter and the lead-lag truncation parameter.

Instead, we focus on the case where $M/N \to d_M$ with $N = o(T^{1/2})$ and consider modifying the FMR and CCR estimators[5] such that the remaining second-order bias can be eliminated. From Theorem 1 and Lemma A.3(c) and (d), we can observe that the partial adjustment of the FMR and CCR estimators emerges from the asymptotic behavior of the long-run variance estimator:

$$\frac{1}{N}\hat{\omega}_{21},\ \frac{1}{N}\hat{\lambda}_{21} \xrightarrow{p} d_M \int_0^\infty k(r)e^{-cd_M r}dr\omega_{21} = \frac{\kappa}{c}\omega_{21}.$$

See also (42) in the Appendix. Hence, if κ can be consistently estimated by, say, $\hat{\kappa}$, the long-run variance estimator should be modified such that

$$\tilde{\omega}_{21} = \frac{1}{\hat{\kappa}}\hat{\omega}_{21},\ \tilde{\lambda}_{21} = \frac{1}{\hat{\kappa}}\hat{\lambda}_{21} \quad \text{and} \quad \frac{1}{N}\tilde{\omega}_{21},\ \frac{1}{N}\tilde{\lambda}_{21} \xrightarrow{p} \frac{1}{c}\omega_{21}.$$

Therefore, if $\hat{\omega}_{21}$ and $\hat{\lambda}_{21}$ are replaced by $\tilde{\omega}_{21}$ and $\tilde{\lambda}_{21}$, respectively, it can be shown in exactly the same way as the Appendix that the FMR and CCR estimators have the same efficient limiting distribution as in the case where $M/N \to \infty$, even if $M/N \to d_M$. Hence, we need to find the consistent estimator of κ. In this case, note that κ is determined only by the localizing parameter c for given $k(\cdot)$ and d_M because κ depends on c, $k(\cdot)$, and d_M. Subsequently, we can consistently estimate κ once we obtain the consistent estimator of c.

Let us suppose that a researcher has specified $k(\cdot)$ and d_M from the outset. Since Lemma A.3(a) shows that

$$\frac{1}{N}\hat{\sigma}_{11} = \frac{1}{NT}\sum_{t=1}^{T}\hat{\dot{u}}_{1,t}^2 \xrightarrow{p} \frac{\omega_{11}}{2c} \tag{22}$$

where $\hat{\dot{u}}_{1,t}$ is the regression residual of y_t on z_t, as previously mentioned, we need to consistently estimate ω_{11}, the long-run variance of $u_{1,t}$. From the definition of $\dot{u}_{1,t}$, we can observe

[5] Since such a modification can be applied only for the FMR and CCR estimators, we do not consider the DOLS estimator in this section.

that

$$\Delta \dot{u}_{1,t} = u_{1,t} - \frac{c}{N}\dot{u}_{1,t-1}.$$

Intuitively, the long-run variance ω_{11} can be estimated by the kernel method using $\Delta \dot{u}_{1,t}$, noting that the first term on the right-hand side dominates the second term, which is $O_p(1/N^{1/2})$ from Lemma A.1(d). Then, the estimator of ω_{11} is given by

$$\hat{\omega}_{\Delta 11} = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{M_c}\right)\frac{1}{T}\sum_{1\le t,t+j\le T} \Delta\hat{\dot{u}}_{1,t}\Delta\hat{\dot{u}}_{1,t+j}. \tag{23}$$

From (22) and (23), we consider the following estimators of the localizing parameter c and κ:

$$\hat{c} = \frac{N}{2}\frac{\hat{\omega}_{\Delta 11}}{\hat{\sigma}_{11}} \quad \text{and} \quad \hat{\kappa} = \hat{c}d_M \int_0^\infty k(r)e^{-\hat{c}d_M r}dr. \tag{24}$$

Theorem 2 *Suppose that Assumption 1 holds and that M and N grow at a known rate slower than $T^{1/2}$ such that $M/N \to d_M$ where d_M is known. If $M_c/N \to 0$,*
(i) $\hat{c} \xrightarrow{p} c$ and $\hat{\kappa} \xrightarrow{p} \kappa$,
(ii) if the FMR and CCR estimators are constructed as in Section 3 with $\hat{\omega}_{21}$ and $\hat{\lambda}_{21}$ replaced by $\tilde{\omega}_{21}$ and $\tilde{\lambda}_{21}$, respectively, they have the same efficient limiting distribution as that stated in Theorem 1 for the case where $M/N \to \infty$.

As is shown in Theorem 2, we can still construct the efficient estimators by modifying the FMR and CCR estimators even if the bandwidth parameter M grows at the same rate as N. However, to construct the estimator, we need to prespecify M, N, d_M and M_c. One of the possible selection rules is to (i) choose the bandwidth parameter M by an existing data dependent rule, (ii) set $N = M$ and then $d_M = 1$, and (iii) set M_c slower than N, such as $M_c = N^{2/3}$. This selection rule is sufficient to obtain the efficient estimators, at least in the case of the model with the N local-to-unity system is concerned.

However, the N local-to-unity system does not seem appropriate for approximating the error process when ρ is not close to 1; ρ should be considered as fixed in such a case. Thus, if we want the modified estimators to accommodate both the N local-to-unity system and the stationary system with a fixed ρ, we need to carefully choose the above parameters.

Note that when ρ is fixed, $\dot{u}_{1,t}$ is stationary and $\Delta\dot{u}_{1,t}$ is an over-differenced series; this implies that the spectral density of $\Delta\dot{u}_{1,t}$ at zero frequency equals 0. Therefore, $\hat{\omega}_{\Delta 11}$ in (23) converges to 0 in probability. On the other hand, it is easy to observe that $\hat{\sigma}_{11}$ converges to σ_{11} in probability, which is supposed to be positive. Then, from the definition of $\hat{c}$ in (24), the asymptotic behavior of $\hat{c}$ depends on the divergence rate of N and the convergence rate of $\hat{\omega}_{\Delta 11}$ to 0. The following corollary provides the conditions under which the modified estimators remains efficient even when ρ is fixed.

Corollary 1 *Suppose that the assumptions in Theorem 2 hold. Further, assume that* $M_c N^2/T \to \infty$*. Then, if* ρ *is fixed,*
(i) $\hat{c} \xrightarrow{p} \infty$ *and* $\hat{\kappa} \xrightarrow{p} 1$,
(ii) the modified FMR and CCR estimators have the same efficient limiting distribution as the original estimators.

Note that the condition in Corollary 1 becomes

$$\frac{M_c N^2}{T} = \frac{M_c}{N}\frac{N^3}{T} = o(1) \times \frac{N^3}{T} \to \infty,$$

which implies that N must go to infinity at a faster rate than $T^{1/3}$. This condition is satisfied by, for example, $M = N = T^{0.4}$ and $M_c = N^{2/3}$. Note that the bandwidth parameter selected by an existing data dependent rule does not necessarily satisfy this condition. We will investigate the finite sample performance of the estimators in the next section.

5. Finite Sample Evidence

This section investigates the finite sample performance of the original and modified efficient estimates as well as the OLS estimate using Monte Carlo simulations. In the simulations, we focus on the effect of the serial correlation in the cointegrating regression errors and thereby, we consider the following simple data generating process:

$$y_t = \mu + \beta x_t + u_{1t}, \quad x_t = x_{t-1} + u_{2t},$$

where x_t is a scalar unit root process. The error term $u_t = [u_{1t}, u_{2t}]'$ is generated from

$$u_{1t} = \rho u_{1t-1} + \varepsilon_{1t} \quad \text{and} \quad u_{2t} = \varepsilon_{2t},$$

$$\text{where} \quad \varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \sim i.i.d.N(0, \Sigma) \quad \text{with} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{21} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}.$$

We set $\rho = 0.1, 0.3, 0.5, 0.7, 0.8, 0.85, 0.9, 0.95$ and $T = 100$ and 300, whereas μ and β are set as 1 throughout the simulations. The variances σ_{11} and σ_{22} are set as 1 and the covariance σ_{21} is 0.4 and 0.8. The number of replications is 10,000 and all computations are carried out by using the GAUSS matrix language.

The long-run variances are estimated by employing the kernel method[6]. We use either the Bartlett or the QS kernel with the bandwidth parameter chosen by either Andrews' (1991, hereafter AN) automatic bandwidth selection method or Newey and West's (1994, hereafter NW) method. Thus, there are four versions of the long-run variance estimates for the FMR and CCR estimates.

With regard to the selection of the lead-lag truncation parameter, we choose K by either the Akaike information criterion (AIC), Bayesian information criterion (BIC), or the general-to-specific rule as proposed by Ng and Perron (1995) at the 1% or 5% significance level. We set the maximum of K to be $[12(T/100)^{1/4}]$.

For the modified FMR and CCR estimates, we set d_M, M, M_c, and N as explained in Section 4; $d_M = 1$, $N = M$, $M_c = M^{2/3}$, and M is chosen either by the AN method or NW method, or from $T^{0.4}$ (labeled "FX").

Tables 1a–1d present the bias and the mean squared error (MSE) of the estimates of β. For the original FMR and CCR estimates, we report only the case where the QS kernel is used with the AN bandwidth because these estimates with other combinations of a kernel and a bandwidth perform in as similar manner. For the same reason, in the case of the modified FMR and CCR estimates (labeled "FMR-BC" and "CCR-BC," respectively), we

[6]We also employed Andrews and Monahan's (1992) prewhitening method with the specification of a first-order vector autoregression (VAR(1)) for u_t. Since the VAR(1) specification is accurate in our simulations, the prewhitening method works better than the kernel method when ρ is not close to 1. However, when ρ approaches 1, the MSE of the estimates becomes larger and the estimates using the prewhitening method performed as poorly as those using the kernel method.

report only the case with the QS kernel. We first summarize the simulation result for the OLS and the original efficient estimates as follows:

(i) Both the bias and the MSE become larger for all the estimates as ρ approaches 1.

(ii) All the original three efficient methods eliminate the bias of the OLS estimate more or less for all the values of ρ considered in the simulations. However, the effect of the efficient methods when ρ is close to 1 is not pronounced to the same extent as it is when ρ is relatively small.

(iii) When ρ is close to 1, the MSE of the original efficient estimates is not necessarily smaller than that of the OLS estimate.

(iv) The performance of the FMR estimate is similar to that of the CCR estimate, although the bias of the former tends to be slightly smaller than that of the latter, whereas the MSE of the former seems to be larger than that of the latter.

(v) The performance of the DOLS estimate depends considerably on the selection of the lead-lag truncation method.

Points (i)–(iii) are related to the case in which ρ is close to 1. We observe that all the three efficient methods have some drawbacks when the regression errors are moderately serially correlated. Point (iv) may be expected because both the FMR and CCR methods eliminate the endogeneity bias in the same manner. Point (v) is a natural result because the lead-lag truncation parameter K must diverge to infinity as T goes to infinity.

Next, we summarize the finite sample performance of the modified FMR and CCR estimates:

(i') The modified estimation method successfully eliminates the bias of the estimates compared with the original three efficient methods when ρ is between 0.5 and 0.9. However, when ρ is 0.95, the effect of our method is not pronounced to a similar extent as when ρ is moderately close to 1.

(ii') When ρ is small, the bias of the original efficient estimates is smaller than that of the modified estimates.

(iii') The MSE of the modified estimates is smaller than that of the original estimates when ρ is moderately close to 1 and σ_{21} is 0.8; this relation is reversed in other cases.

(iv') The bias of the modified FMR estimate is smaller than that of the modified CCR estimates, whereas the MSE of the former estimate is larger than that of the latter.

(v') The modified estimates with the bandwidth selected by the AN rule are more biased than those with the NW bandwidth.

Point (i') implies that as far as the bias is concerned, our method is effective when ρ is moderately close to 1, as expected from Theorem 2. Point (ii') is a natural result because our method is established under the assumption that ρ is moderately close to 1. However, in almost all the cases, the modified estimates are less biased than the OLS estimate. With regard to the MSE, we observe from point (iii') that our method is effective when ρ is moderately close to 1 and the endogeneity is relatively strong. For example, in Table 1b, the MSE of the modified FMR estimate with the fixed M is 0.01871 when ρ is 0.8, whereas the MSE of the original FMR estimate is 0.02978. However, when $\sigma_{21} = 0.4$, the former MSE is 0.03243, while the latter is 0.02714, as is observed in Table 1a; then, the MSE of the original estimate is smaller than that of the modified estimate in the latter case. Point (v') is related to the method to select the! bandwidth parameter M; it appears that the AN rule selects an extremely long bandwidth parameter in finite samples, as is explained below; this leads to point (v').

From the above observation, it can be seen that there is a trade-off between the bias and the MSE when ρ is moderately close to 1 and the simultaneous correlation is not strong. In this case, the modified estimates are less biased but have a larger MSE than the original efficient estimates. However, when the endogeneity is relatively strong, our modified estimates dominate the original ones in terms of both the bias and the MSE.

We finally investigate the effect of the length of the bandwidth parameter and the lead-lag truncation parameter on the finite sample performance of the original FMR and DOLS estimates. Figure 2 depicts the bias and the MSE of the FMR estimate using the QS kernel with a fixed bandwidth parameter when $\rho = 0.7$, 0.8, and 0.9 (labeled "fixed"). The fixed bandwidth assumes values from 1 to 26 when $T = 100$ and from 1 to 35 when $T = 300$. To draw a comparison between the fixed bandwidth case and the automatic selection rule, we plot the averaged length of the automatic bandwidth versus the bias (and the MSE) of the FMR estimate with either the AN rule (labeled "auto (AN)") or the NW rule (labeled "auto (NW)"). For example, when $T = 100$ and $\sigma_{21} = 0.4$ (Figure 2(i-a)), the averaged AN bandwidth and the bias are $(9.1, 0.029)$, $(12.7, 0.048)$, and $(20.2, 0.100)$ for $\rho = 0.7$, 0.8, and 0.9, respectively, whereas they are $(3.5, 0.031)$, $(3.8, 0.054)$, and $(4.0, 0.112)$! for the NW rule. According to the figure, the bandwidth selected by the NW rule is too short, whereas that selected by the AN rule is too long. Thus, although our simulation setting is limited, it seems that we should choose a bandwidth parameter that is longer than the one selected by the NW rule or shorter than the AN rule, as long as the error term is moderately serially correlated. One possible choice is to select an average of the AN and NW bandwidths.

Figure 3 illustrates the bias and the MSE of the DOLS estimate with a fixed lead-lag truncation parameter. As illustrated in Figure 2, we also plot the averaged length of the plug-in truncation parameter adopted in this paper versus the bias and the MSE. With regard to K considered in this paper, the bias monotonically decreases as K increases, whereas there is a point at which the MSE is minimized. In general, when we use BIC, the truncation parameter is very short and both the bias and MSE tend to be larger than those obtained by the other plug-in methods. On the other hand, when we use the general-to-specific rule with the 5% significance level, it selects a long K and the MSE becomes relatively large. Thereafter, as far as ρ is moderately close to 1, AIC or the general-to-specific rule with the 1% significance level may be recommended for finite samples.

6. Conclusion

In this paper, we theoretically investigated three efficient estimators for cointegrating regression models: the FMR, CCR, and DOLS. We showed that under the N local-to-unity system where the AR coefficient approaches 1, the asymptotic behavior of the efficient estimators depends on the approaching speed of the AR coefficient; these estimators are efficient in some cases but the bias remains in others. We then proposed the modified FMR and CCR estimators that have the efficient limiting distribution. We also investigated the finite sample properties of the estimators and found that our modified method is effective when the regression errors are moderately serially correlated and the endogeneity is relatively strong. Overall, the analytical investigation in this paper can adequately explain the poor finite sample performance of the three efficient estimators when the regression errors are serially correlated.

Appendix

We denote some constant that is independent of both T and the subscript j as C in general. We also assume that $u_{1,0} = 0$ without loss of generality to simplify the proof.

Proof of Theorem 1: By partitioning $\Psi_L = [\psi'_{L1}, \Psi'_{L2}]'$, $\dot{u}_{1,t}$ is expressed using (14) as

$$\begin{aligned}
\dot{u}_{1,t} &= \sum_{l=1}^{t} \rho^{t-l} \left(\psi_{L1}\varepsilon_l + \tilde{u}_{1,l-1} - \tilde{u}_{1,l}\right) \\
&= \psi_{L1} \sum_{l=1}^{t} \rho^{t-l}\varepsilon_l + \sum_{l=1}^{t} \left(\rho^{t-l+1}\tilde{u}_{1,l-1} - \rho^{t-l}\tilde{u}_{1,l}\right) + \sum_{l=1}^{t} \left(\rho^{t-l} - \rho^{t-l+1}\right) \tilde{u}_{1,l-1} \\
&= \xi_{1,t} + \xi_{2,t} + \xi_{3,t}, \qquad (25)
\end{aligned}$$

where $\xi_{1,t} = \psi_{L1}\dot{\varepsilon}_t$ with $\dot{\varepsilon}_t = \sum_{l=1}^{t} \rho^{t-l}\varepsilon_l$, $\xi_{2,t} = \left(\rho^t \tilde{u}_{1,0} - \tilde{u}_{1,t}\right)$ and $\xi_{3,t} = (1-\rho)\dot{\tilde{u}}_{1,t-1}$ with $\dot{\tilde{u}}_{1,t-1} = \sum_{l=1}^{t} \rho^{t-l}\tilde{u}_{1,l-1}$. Note that $\dot{\varepsilon}_t$ and $\dot{\tilde{u}}_{1,t}$ essentially have the same structure as $\dot{u}_{1,t}$.

Lemma A.1 *For $\rho = \rho_N = 1 - c/N$ and for any given $1 \leq t \leq T$,*

$$\begin{aligned}
&(a) \quad Var(\xi_{1,t}) \leq CN \quad and \quad \xi_{1,t} = O_p(\sqrt{N}), \\
&(b) \quad Var(\xi_{2,t}) \leq C \quad and \quad \xi_{2,t} = O_p(1), \\
&(c) \quad Var(\xi_{3,t}) \leq \frac{C}{N} \quad and \quad \xi_{3,t} = O_p\left(\frac{1}{\sqrt{N}}\right), \\
&(d) \quad Var(\dot{u}_{1,t}) \leq CN \quad and \quad \dot{u}_{1,t} = O_p(\sqrt{N}).
\end{aligned}$$

Proof of Lemma A.1: (a) is proved from direct calculation and (b) is obvious because $\tilde{u}_{1,t}$ is stationary. For (c) and (d), we have, after some algebra,

$$\xi_{3,t} = \frac{c}{N} \sum_{l=1}^{t} \sum_{j=0}^{\infty} \rho^{t-l} \tilde{\psi}_{1,j} \varepsilon_{l-j-1} = \frac{c}{N} \sum_{l=0}^{t-1} \alpha_{1,l}\varepsilon_l + \frac{c}{N} \sum_{l=1}^{\infty} \alpha_{2,l}\varepsilon_{-l}, \qquad (26)$$

where $\alpha_{1,l} = \sum_{j=0}^{t-l-1} \rho^{t-j-l-1}\tilde{\psi}_{1,j}$ and $\alpha_{2,l} = \sum_{j=1}^{t} \rho^{t-j}\tilde{\psi}_{1,j+l-1}$. Since it can be shown that $\sum_{l=0}^{\infty} \|\alpha_{1,l}\|/N$ is bounded above by some constant that is independent of N, $\sum_{l=0}^{\infty} \|\alpha_{1,l}\|^2/N^2$ is also bounded above. In exactly the same way, we can also show that $\sum_{l=1}^{\infty} \|\alpha_{2,l}\|^2/N^2 \leq C$. Hence, it is observed that $Var(\xi_{3,t}) \leq C$, which implies $\xi_{3,t} = O_p(1)$. (d) is obtained by noting that $Var(\dot{u}_{1,t}) \leq 3\,(Var(\xi_{1,t}) + Var(\xi_{2,t}) + Var(\xi_{3,t})) \leq CN$. Moreover, since $\sum_{j=1}^{\infty} j\|\tilde{\Psi}_j\| < \infty$, we can decompose $\tilde{u}_{1,t}$ in the same manner as (14) and then it is shown that $Var(\xi_{3,t}) \leq C/N$ because $Var(\dot{\tilde{u}}_{1,t-1}) \leq CN$ is deduced by noting that $\dot{\tilde{u}}_{1,t-1}$ has the same structure as $\dot{u}_{1,t}$.□

Lemma A.2 *For $\rho = \rho_N = 1 - c/N$,*

$$(a) \quad \frac{1}{NT}\sum_{t=1}^{T} \dot{u}_{1,t}^2 \stackrel{p}{\longrightarrow} \frac{\omega_{11}}{2c},$$

$$(b) \quad \frac{1}{T}\sum_{t=1}^{T-j} u_{2,t}\dot{u}_{1,t+j} \stackrel{p}{\longrightarrow} \rho^j\omega_{21} - E(u_{2,t}\tilde{u}_{1,t+j}) \quad \textit{for a given} \quad j \geq 0,$$

$$(c) \quad \frac{1}{T}\sum_{t=1}^{T-j} u_{2,t+j}\dot{u}_{1,t} \stackrel{p}{\longrightarrow} E(\tilde{u}_{2,t+j-1}u_{1,t}), \quad \textit{for a given} \quad j \geq 1,$$

$$(d) \quad \frac{1}{N\sqrt{T}}\sum_{t=1}^{T} \dot{u}_{1,t} \Rightarrow \frac{1}{c}B_1(1),$$

$$(e) \quad \frac{1}{NT}\sum_{t=1}^{T} x_t\dot{u}_{1,t} \Rightarrow \frac{1}{c}\left(\int_0^1 B_2(r)dB_1(r) + \omega_{21}\right).$$

Proof of Lemma A.2: (a) Using expression (25) we have

$$\frac{1}{NT}\sum_{t=1}^{T} \dot{u}_{1,t}^2 = \frac{1}{NT}\sum_{t=1}^{T} \xi_{1,t}^2 + \sum_{i=j\neq 1} \frac{1}{NT}\sum_{t=1}^{T} \xi_{i,t}\xi_{j,t}.$$

The second term on the right-hand side is shown to be $O_p(1/N^{1/2})$ using the Cauchy-Schwarz inequality and Lemma A.1(a)–(c). On the other hand,

$$\sum_{t=1}^{T} \xi_{1,t}^2 = \sum_{t=1}^{T}\sum_{l=1}^{t} \rho^{2(t-l)}(\psi_{L1}\varepsilon_l)^2 + 2\sum_{t=2}^{T}\sum_{l=1}^{t-1}\sum_{k=l+1}^{t} \rho^{2t-k-l}(\psi_{L1}\varepsilon_k)(\psi_{L1}\varepsilon_l). \tag{27}$$

The first term on the right-hand side of (27) can be expressed as

$$\begin{aligned}\sum_{t=1}^{T}\sum_{l=1}^{t} \rho^{2(t-l)}(\psi_{L1}\varepsilon_l)^2 &= \sum_{t=1}^{T}\sum_{l=0}^{T-t} \rho^{2l}(\psi_{L1}\varepsilon_t)^2 \\ &= \frac{1}{1-\rho^2}\sum_{t=1}^{T}(\psi_{L1}\varepsilon_t)^2 - \frac{1}{1-\rho^2}\sum_{t=1}^{T}\rho^{2(T-t+1)}(\psi_{L1}\varepsilon_t)^2. \end{aligned} \tag{28}$$

Since $1-\rho^2 = 2c/N - c^2/N^2$, the first term of (28) divided by NT converges in probability to $\psi_{L1}\Sigma_\varepsilon\psi'_{L1}/(2c) = \omega_{11}/(2c)$ by the weak law of large numbers (WLLN), whereas the second term of (28) divided by NT is easily shown to be $O_p(N/T)$. Then, we have $(NT)^{-1}\sum_{t=1}^{T}\sum_{l=1}^{t}\rho^{2(t-l)}(\psi_{L1}\varepsilon_l)^2 \xrightarrow{p} \frac{\omega_{11}}{2c}$.

On the other hand, after some algebra, the second term on the right-hand side of (27) is shown to equal

$$\sum_{l=1}^{T-1}\sum_{k=l+1}^{T}\sum_{t=0}^{T-k}\rho^{k-l+2t}(\psi_{L1}\varepsilon_k)(\psi_{L1}\varepsilon_l) = \frac{1}{1-\rho^2}\sum_{l=1}^{T-1}\sum_{k=l+1}^{T}(\rho^{k-l}-\rho^{2T-k-l+2})(\psi_{L1}\varepsilon_k)(\psi_{L1}\varepsilon_l).$$

Note that the right-hand side has mean zero and its variance is shown to be $O(N^3T)$. Hence, the second term on the right-hand side of (27) divided by NT is $O_p(N^{1/2}/T^{1/2})$. Then, we have

$$\frac{1}{NT}\sum_{t=1}^{T}\xi_{1,t}^2 \xrightarrow{p} \frac{\omega_{11}}{2c} \tag{29}$$

and hence, (a) is obtained.

(b) Using expression (25),

$$\frac{1}{T}\sum_{t=1}^{T-j}u_{2,t}\dot{u}_{1,t+j} = \frac{1}{T}\sum_{t=1}^{T-j}u_{2,t}\xi_{1,t+j} + \frac{1}{T}\sum_{t=1}^{T-j}u_{2,t}\xi_{2,t+j} + \frac{1}{T}\sum_{t=1}^{T-j}u_{2,t}\xi_{3,t+j}. \tag{30}$$

The third term on the right-hand side of (30) is $O_p(1/N^{1/2})$ because $\left\|T^{-1}\sum_t u_{2,t}\xi_{3,t+j}\right\| \le T^{-1}\left(\sum_t\|u_{2,t}\|^2\right)^{1/2}\left(\sum_t \xi_{3,t+j}^2\right)^{1/2} = O_p\left(1/N^{1/2}\right)$ based on stationarity of $u_{2,t}$ and Lemma A.1(c).

Using (14), the first term on the right-hand side of (30) becomes

$$\frac{1}{T}\sum_{t=1}^{T-j}u_{2,t}\xi_{1,t+j} = \frac{1}{T}\Psi_{L2}\sum_{t=1}^{T-j}\varepsilon_t\xi_{1,t+j} + \frac{1}{T}\sum_{t=1}^{T-j}(\tilde{u}_{2,t-1}-\tilde{u}_{2,t})\xi_{1,t+j}. \tag{31}$$

Since $E\left(T^{-1}\Psi_{L2}\sum_{t=1}^{T-j}\varepsilon_t\xi_{1,t+j}\right) = \rho^j\Psi_{L2}\Sigma_\varepsilon\psi'_{L1} + o(1) = \rho^j\omega_{21} + o(1)$ while its variance is $O(N/T)$, the first term on the right-hand side of (31) converges in probability to $\rho^j\omega_{21}$ for a given $j \ge 0$ by the WLLN. On the other hand, the second term of (31) is expressed as

$$\frac{1}{T}\sum_{t=1}^{T-j}\{(\tilde{u}_{2,t-1}\xi_{1,t+j-1} - \tilde{u}_{2,t}\xi_{1,t+j}) + \tilde{u}_{2,t-1}(\xi_{1,t+j}-\xi_{1,t+j-1})\}$$

$$= \frac{1}{T}(\tilde{u}_{2,0}\xi_{1,j} - \tilde{u}_{2,T-j}\xi_{1,T}) + \frac{1}{T}\sum_{t=1}^{T-j}\tilde{u}_{2,t-1}\varepsilon'_{t+j}\psi'_{L1} - \frac{c}{NT}\sum_{t=1}^{T-j}\tilde{u}_{2,t-1}\xi_{1,t+j-1}, \quad (32)$$

where we used the relation

$$\xi_{1,t} - \xi_{1,t-1} = \psi_{L1}\varepsilon_t + (\rho - 1)\psi_{L1}\sum_{l=1}^{t-1}\rho^{t-l-1}\varepsilon_l = \psi_{L1}\varepsilon_t - \frac{c}{N}\xi_{1,t-1}. \quad (33)$$

Thus, from Lemma A.1(a), we can observe that the first two terms of (32) are $O_p(N^{1/2}/T)$ and $O_p(1/T^{1/2})$, while the third term is

$$\left\|\frac{c}{NT}\sum_{t=1}^{T-j}\tilde{u}_{2,t-1}\xi_{1,t+j-1}\right\| \leq \frac{c}{NT}\left(\sum_{t=1}^{T-j}\|\tilde{u}_{2,t-1}\|^2\right)^{1/2}\left(\sum_{t=1}^{T-j}\xi_{1,t+j-1}^2\right)^{1/2} = O_p\left(\frac{1}{\sqrt{N}}\right)$$

using (29). Based on these results, we can see that the first term on the right-hand side of (30) converges in probability to $\rho^j\omega_{21}$.

For the second term on the right-hand side of (30), we have

$$\frac{1}{T}\sum_{t=1}^{T-j}u_{2,t}\xi_{2,t+j} = \frac{1}{T}\sum_{t=1}^{T-j}u_{2,t}\rho^{t+j}\tilde{u}_{1,0} - \frac{1}{T}\sum_{t=1}^{T-j}u_{2,t}\tilde{u}_{1,t+j} \xrightarrow{p} -E(u_{2t}\tilde{u}_{1,t+j})$$

by the WLLN. We then obtain (b).

Here, note that since $\dot{\varepsilon}_t$ and $\dot{\tilde{u}}_{1,t}$ have the same structure as $\dot{u}_{1,t}$, the third terms on the right-hand side of (30) and (32) can be shown to be $O_p(1/N)$, although they were only proved to be $O_p(1/N^{1/2})$. These relations imply that

$$\frac{1}{T}\sum_{t=1}^{T-j}u_{2,t}\dot{u}_{1,t+j} = \frac{\rho^j}{T}\Psi_{L2}\sum_{t=1}^{T-j}\varepsilon_t\varepsilon'_t\psi'_{L1} - \frac{1}{T}\sum_{t=1}^{T-j}u_{2,t}\tilde{u}_{1,t+j} + O_p\left(\frac{1}{N}\right) + O_p\left(\sqrt{\frac{N}{T}}\right), \quad (34)$$

which will be used in the proof of Lemma A.3(c).

(c) Using (14) and (25),

$$\frac{1}{T}\sum_{t=1}^{T-j}u_{2,t+j}\dot{u}_{1,t} = \frac{1}{T}\Psi_{L2}\sum_{t=1}^{T-j}\varepsilon_{t+j}(\xi_{1,t} + \xi_{2,t} + \xi_{3,t}) + \frac{1}{T}\sum_{t=1}^{T-j}(\tilde{u}_{2,t+j-1} - \tilde{u}_{2,t+j})\dot{u}_{1,t}. \quad (35)$$

Using Lemma A.1(a), we obtain $E\left\|T^{-1}\sum_{t=1}^{T-j}\varepsilon_{t+j}\xi_{1,t}\right\|^2 \leq CT^{-2}\sum_{t=1}^{T-j}E(\xi_{1,t}^2) = O\left(N/T\right)$, which implies $T^{-1}\sum\varepsilon_{t+j}\xi_{1,t} = O_p(N^{1/2}/T^{1/2})$. We can also observe that $T^{-1}\sum_{t=1}^{T-j}\varepsilon_{t+j}\xi_{3,t} =$

$O_p(1/(NT)^{1/2})$ using Lemma A.1 (c) and that $T^{-1}\sum_{t=1}^{T-j}\varepsilon_{t+j}\xi_{2,t} = O_p\left(1/T^{1/2}\right)$. Thus, the first term on the right-hand side of (35) is $O_p(N^{1/2}/T^{1/2})$.

On the other hand, the second term on the right-hand side of (35) becomes equal to

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T-j}\{\tilde{u}_{2,t+j-1}(\dot{u}_{1,t}-\dot{u}_{1,t-1})+(\tilde{u}_{2,t+j-1}\dot{u}_{1,t-1}-\tilde{u}_{2,t+j}\dot{u}_{1,t})\} \\
&= \frac{1}{T}\sum_{t=1}^{T-j}\tilde{u}_{2,t+j-1}(\dot{u}_{1,t}-\dot{u}_{1,t-1})+\frac{1}{T}(\tilde{u}_{2,j}\dot{u}_{1,0}-\tilde{u}_{2,T}\dot{u}_{1,T-j}) \\
&= \frac{1}{T}\sum_{t=1}^{T-j}\tilde{u}_{2,t+j-1}u_{1,t}-\frac{c}{NT}\sum_{t=1}^{T-j}\tilde{u}_{2,t+j-1}\dot{u}_{1,t-1}+O_p\left(\frac{\sqrt{N}}{T}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T-j}\tilde{u}_{2,t+j-1}u_{1,t}+O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{\sqrt{N}}{T}\right), \qquad (36)
\end{aligned}$$

where the second equation holds because of the relation $\dot{u}_{1,t}-\dot{u}_{1,t-1} = u_{1,t}-(c/N)\dot{u}_{1,t-1}$ and Lemma A.1(d), while the last equality is established by

$$\left\|\frac{1}{NT}\sum_{t=1}^{T-j}\tilde{u}_{2,t+j-1}\dot{u}_{1,t-1}\right\| \leq \frac{1}{NT}\left(\sum_{t=1}^{T-j}\|\tilde{u}_{2,t+j-1}\|^2\sum_{t=1}^{T-j}\dot{u}_{1,t-1}^2\right)^{1/2} = O_p\left(\frac{1}{\sqrt{N}}\right), \qquad (37)$$

which holds from Lemma A.2(a). Then, (c) is obtained by the WLLN.

In this case, note that since $\tilde{u}_{2,t}$ is stationary with 1-summable coefficients, we can similarly show that $(NT)^{-1}\sum\tilde{u}_{2,t+j-1}\dot{u}_{1,t}$ is $O_p(1/N)$, although it was shown only to be $O_p(1/N^{1/2})$ in (37). Then, we have

$$\frac{1}{T}\sum_{t=1}^{T-j}u_{2,t+j}\dot{u}_{1,t} = \frac{1}{T}\sum_{t=1}^{T-j}\tilde{u}_{2,t+j-1}u_{1,t}+O_p\left(\frac{1}{N}\right)+O_p\left(\sqrt{\frac{N}{T}}\right), \qquad (38)$$

which will be used in the proof of Lemma A.3(d) and Lemma A.4(b).

(d) From the definition of $\dot{u}_{1,t}$,

$$\frac{1}{N\sqrt{T}}\sum_{t=1}^{T}\dot{u}_{1,t} = \frac{1}{N\sqrt{T}}\sum_{t=1}^{T}\sum_{l=0}^{T-t}\rho^l u_{1,t} = \frac{1}{c\sqrt{T}}\sum_{t=1}^{T}u_{1,t}-\frac{\rho}{c\sqrt{T}}\dot{u}_{1,T}.$$

We can see that the right-hand side weakly converges to $(1/c)B_1(1)$ by the FCLT and Lemma A.1(d).

(e) Using the identity $(1/N)\dot{u}_{1,t-1} = (1/c)u_{1,t} + (1/c)(\dot{u}_{1,t-1} - \dot{u}_{1,t})$, we have

$$
\begin{aligned}
\frac{1}{NT}\sum_{t=1}^{T} x_t \dot{u}_{1,t} &= \frac{1}{cT}\sum_{t=1}^{T} x_t u_{1,t+1} + \frac{1}{cT}\sum_{t=1}^{T} x_t(\dot{u}_{1,t} - \dot{u}_{1,t+1}) \\
&= \frac{1}{cT}\sum_{t=1}^{T} x_t u_{1,t+1} + \frac{1}{cT}\sum_{t=1}^{T} \{(x_t - x_{t-1})\dot{u}_{1,t} + (x_{t-1}\dot{u}_{1,t} - x_t\dot{u}_{1,t+1})\} \\
&= \frac{1}{cT}\sum_{t=1}^{T} x_t u_{1,t+1} + \frac{1}{cT}\sum_{t=1}^{T} u_{2,t}\dot{u}_{1,t} + \frac{1}{cT}(x_0\dot{u}_{1,1} - x_T\dot{u}_{1,T+1}) \\
&\Rightarrow \frac{1}{c}\left(\int_0^1 B_2(r)dB_1(r) + \sum_{j=1}^{\infty} E(u_{2,t}u_{1,t+j})\right) + \frac{1}{c}\left(\omega_{21} - E(u_{2,t}\tilde{u}_{1,t})\right),
\end{aligned}
$$

where the last convergence holds by the FCLT, Lemma A.1(d), and Lemma A.2(b). The result is established by noting that

$$
\begin{aligned}
\sum_{j=1}^{\infty} E(u_{2,t}u_{1,t+j}) - E(u_{2,t}\tilde{u}_{1,t}) &= \sum_{i=0}^{\infty} \Psi_{2,j}\Sigma_\varepsilon \sum_{j=1}^{\infty} \psi'_{1,i+j} - \sum_{i=0}^{\infty} \Psi_{2,j}\Sigma_\varepsilon \tilde{\psi}'_{1,j} \\
&= \sum_{i=0}^{\infty} \Psi_{2,j}\Sigma_\varepsilon \sum_{j=i+1}^{\infty} \psi'_{1,j} - \sum_{i=0}^{\infty} \Psi_{2,j}\Sigma_\varepsilon \sum_{j=i+1}^{\infty} \psi'_{1,j} = 0. \square
\end{aligned}
$$

Using Lemma A.2 (d) and (e), the FCLT, and the continuous mapping theorem (CMT), the limiting distribution of the OLS estimator is obtained as given by (16).

Next, we present the following lemma, which is used to derive the limiting distributions of the FMR and CCR estimators.

Lemma A.3 *Let* $\kappa = cd_M \int_0^\infty k(r)e^{-cd_M r}dr$. *For* $\rho = \rho_N = 1 - c/N$,

$$(a) \quad \frac{1}{N}\hat{\sigma}_{11} = \frac{1}{NT}\sum_{t=1}^{T} \hat{\dot{u}}_{1,t}^2 \xrightarrow{p} \frac{\omega_{11}}{2c},$$

$$(b) \quad \hat{\sigma}_{21} = \frac{1}{T}\sum_{t=1}^{T} u_{2,t}\hat{\dot{u}}_{1,t} \xrightarrow{p} \omega_{21} - E(u_{2,t}\tilde{u}_{1,t}),$$

$$(c) \quad \frac{1}{N}\hat{\lambda}_{21} \xrightarrow{p} \bar{\lambda}_{21} \equiv \begin{cases} \frac{1}{c}\omega_{21}, & \frac{M}{N} \to \infty, \\ \frac{\kappa}{c}\omega_{21}, & \frac{M}{N} \to d_M, \\ 0, & \frac{M}{N} \to 0, \end{cases}$$

$$(d) \quad \frac{1}{N}\hat{\omega}_{21} \xrightarrow{p} \bar{\lambda}_{21}.$$

Proof of Lemma A.3: (a) Since $\hat{u}_{1,t} = \dot{u}_{1,t} - (\hat{\theta} - \theta)' z_t$ and $(1/N) D_T(\hat{\theta} - \theta) = O_p(1)$, we can see Lemma A.2 (a) that

$$\frac{1}{NT}\sum_{t=1}^{T}\hat{u}_{1,t}^2 = \frac{1}{NT}\sum_{t=1}^{T}\dot{u}_{1,t}^2 + O_p\left(\frac{N}{T}\right) \xrightarrow{p} \frac{1}{2c}\omega_{11}.$$

(b) Noting that

$$\frac{1}{T}\sum_{t=1}^{T-j} u_{2,t}\hat{u}_{1,t+j} = \frac{1}{T}\sum_{t=1}^{T-j} u_{2,t}\dot{u}_{1,t+j} + O_p\left(\frac{N}{T}\right) \tag{39}$$

for any given j, (b) is obtained by Lemma A.2 (b) for $j = 0$.

(c) From (39) and (34),

$$\begin{aligned}\frac{1}{N}\hat{\lambda}_{21} &= \frac{1}{N}\sum_{j=0}^{T-1} k\left(\frac{j}{M}\right)\frac{1}{T}\sum_{t=1}^{T-j} u_{2t}\dot{u}_{1,t+j} + O_p\left(\frac{M}{T}\right) \\ &= \frac{1}{N}\left(\sum_{j=0}^{N^*} + \sum_{j=N^*+1}^{T-1}\right) k\left(\frac{j}{M}\right)\frac{\rho^j}{T}\Psi_{L2}\sum_{t=1}^{T-j}\varepsilon_t\varepsilon_t'\psi_{L1}' \\ &\quad + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{M}{N^2}\right) + O_p\left(\frac{M}{\sqrt{NT}}\right) + O_p\left(\frac{M}{T}\right),\end{aligned} \tag{40}$$

where N^* satisfies $N^*/N \to \infty$ and $N^*/T \to 0$. The last equality holds because $u_{2,t}\tilde{u}_{1,t+j}$ has a bounded spectral density, and then $N^{-1}\sum_j k(j/M) T^{-1}\sum_t u_{2,t}\tilde{u}_{1,t+j} = O_p(1/N)$. Note that the last four terms in the last expression converge in probability to 0 from the condition of M, N, and T. Since $T^{-1}\sum_{t=1}^{T-j}\varepsilon_t\varepsilon_t'$ is shown to be $O_p(1)$ for a given $j > N^*$ and $|k(\cdot)| \le 1$, the second summation in the first term of (40) becomes

$$\left\| \frac{1}{N}\sum_{j=N^*+1}^{T-1} k\left(\frac{j}{M}\right)\frac{\rho^j}{T}\Psi_{L2}\sum_{t=1}^{T-j}\varepsilon_t\varepsilon_t'\psi_{L1}' \right\| \le \frac{C}{N}\sum_{j=N^*+1}^{T-1}\rho^j \times O_p(1) = \frac{\rho^{N^*+1} - \rho^T}{N(1-\rho)} \times O_p(1),$$

which converges in probability to 0 because $N(1-\rho) = c$, $\rho^{N^*+1} = (1 - c/N)^{N^*+1} \to 0$ and $\rho^T = (1 - c/N)^T \to 0$ since $c > 0$, $N^*/N \to \infty$, and $T/N \to \infty$. In the following, we derive the limit of the first summation in the first term of (40), depending on the rate of N.

When $M/N \to \infty$, we can select N^* such that $N^*/M \to 0$ while $N^*/N \to \infty$. Since $k(0) = 1$, $k(\cdot)$ is continuous at 0, $j/M \to 0$ over $0 \le j \le N^*$, and $T^{-1}\sum_{t=1}^{T-j}\varepsilon_t\varepsilon_t'$ converges in probability to Σ_ε for $0 \leq j \leq N^*$, we have

$$\begin{aligned}\frac{1}{N}\sum_{j=0}^{N^*} k\left(\frac{j}{M}\right)\frac{\rho^j}{T}\Psi_{L2}\sum_{t=1}^{T-j}\varepsilon_t\varepsilon_t'\psi_{L1}' &= (\Psi_{L2}\Sigma_\varepsilon\psi_{L1}' + o_p(1))\frac{1}{N}\sum_{j=0}^{N^*}(1+o(1))\rho^j \\ &\xrightarrow{p} \omega_{21}\lim_{N\to\infty}\frac{1-\rho^{N^*+1}}{N(1-\rho)} = \frac{1}{c}\omega_{21}.\end{aligned}$$

Thus, we obtain (c) for $M/N \to \infty$.

When $M/N \to d_M$, N^*/M must go to infinity. Thus, we have

$$\begin{aligned}\frac{1}{N}\sum_{j=0}^{N^*} k\left(\frac{j}{M}\right)\frac{\rho^j}{T}\Psi_{L2}\sum_{t=1}^{T-j}\varepsilon_t\varepsilon_t'\psi_{L1}' &= (\omega_{21} + o_p(1))\frac{M}{N}\frac{1}{M}\sum_{j=0}^{N^*} k\left(\frac{j}{M}\right)\left(1-\frac{M}{N}\frac{c}{M}\right)^j \\ &\xrightarrow{p} d_M\int_0^\infty k(r)e^{-cd_M r}dr\omega_{21}.\end{aligned}$$

When $M/N \to 0$, we can see that

$$\left\|\frac{1}{N}\sum_{j=0}^{N^*} k\left(\frac{j}{M}\right)\frac{\rho^j}{T}\Psi_{L2}\sum_{t=1}^{T-j}\varepsilon_t\varepsilon_t'\psi_{L1}'\right\| \leq C\frac{M}{N}\frac{1}{M}\sum_{j=0}^{N^*}\left|k\left(\frac{j}{M}\right)\right| \times O_p(1) \xrightarrow{p} 0.$$

(d) Since

$$\frac{1}{N}\hat{\omega}_{21} = \frac{1}{N}\sum_{j=1}^{T-1} k\left(\frac{j}{M}\right)\frac{1}{T}\sum_{t=1}^{T-j} u_{2,t+j}\hat{\dot{u}}_{1,t} + \frac{1}{N}\hat{\lambda}_{21}, \tag{41}$$

it suffices to show that the first term on the right-hand side of (41) converges in probability to 0. From (39) and (38), we have

$$\begin{aligned}\frac{1}{N}\sum_{j=1}^{T-1} k\left(\frac{j}{M}\right)\frac{1}{T}\sum_{t=1}^{T-j} u_{2,t+j}\hat{\dot{u}}_{1,t} &= \frac{1}{N}\sum_{j=1}^{T-1} k\left(\frac{j}{M}\right)\frac{1}{T}\sum_{t=1}^{T-j} u_{2,t+j}\dot{u}_{1,t} + O_p\left(\frac{M}{T}\right) \\ &= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{M}{N^2}\right) + O_p\left(\frac{M}{\sqrt{NT}}\right) + O_p\left(\frac{M}{T}\right). \square\end{aligned}$$

We are now in a position to derive the limiting distributions of the FMR and CCR estimators. Note that $N^{-1}D_T(\hat{\theta}_{FMR} - \theta) = H_T^{-1}h_{FMR,T}$ where $H_T = D_T^{-1}\sum_{t=1}^T z_t z_t' D_T^{-1}$ and $h_{FMR} = N^{-1}D_T^{-1}\sum_{t=1}^T z_t\dot{u}_{1,t} - D_T^{-1}\sum_{t=1}^T z_t u_{2,t}'\hat{\Omega}_{22}^{-1}N^{-1}\hat{\omega}_{21} - N^{-1}\hat{J}^+$. We can see that $H_T^{-1} \Rightarrow H^{-1}$ as in the case of the OLS estimator, whereas from the FCLT, the CMT, and

Lemmas A.2 and A.3,

$$
\begin{aligned}
h_{FMR,T} &= \begin{bmatrix} \frac{1}{N\sqrt{T}}\sum_{t=1}^{T}\dot{u}_{1,t} - \frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{2,t}'\hat{\Omega}_{22}^{-1}\frac{1}{N}\hat{\omega}_{21} \\ \frac{1}{NT}\sum_{t=1}^{T}x_t\dot{u}_{1,t} - \frac{1}{T}\sum_{t=1}^{T}x_t u_{2,t}'\hat{\Omega}_{22}^{-1}\frac{1}{N}\hat{\omega}_{21} - \frac{1}{N}\left(\hat{\lambda}_{21} - \hat{\Lambda}_{22}\hat{\Omega}_{22}^{-1}\hat{\omega}_{21}\right) \end{bmatrix} \\
&\Rightarrow \begin{bmatrix} \frac{1}{c}B_1(1) - B_2'(1)\Omega_{22}^{-1}\bar{\lambda}_{21} \\ \left(\frac{1}{c}\int_0^1 B_2(r)dB_1(r) - \int_0^1 B_2(r)dB_2'(r)\Omega_{22}^{-1}\bar{\lambda}_{21}\right) + \left(\frac{1}{c}\omega_{21} - \bar{\lambda}_{21}\right) \end{bmatrix}. \qquad (42)
\end{aligned}
$$

For the CCR estimator, note that

$$
\frac{1}{N}D_T(\hat{\theta}_{CCR} - \theta) = \begin{bmatrix} 1 & \frac{1}{T\sqrt{T}}\sum_{t=1}^{T}x_t^{*\prime} \\ \frac{1}{T\sqrt{T}}\sum_{t=1}^{T}x_t^* & \frac{1}{T^2}\sum_{t=1}^{T}x_t^* x_t^{*\prime} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{N\sqrt{T}}\sum_{t=1}^{T}\dot{u}_{1,t}^* \\ \frac{1}{NT}\sum_{t=1}^{T}x_t^*\dot{u}_{1,t}^* \end{bmatrix},
$$

where $\dot{u}_{1,t}^* = \dot{u}_{1,t} - (\hat{\beta}-\beta)'\hat{\Lambda}_2\hat{\Sigma}^{-1}\hat{u}_t - \hat{\omega}_{12}\hat{\Omega}_{22}^{-1}u_{2t}$ with $\hat{\beta}$ being the OLS estimator of β. From Lemma A.3, we obtain

$$
\hat{\Lambda}_2\hat{\Sigma}^{-1} \xrightarrow{p} \left[\bar{\lambda}_{21}\frac{2c}{\omega_{11}}, -\bar{\lambda}_{21}\frac{2c}{\omega_{11}}\{\omega_{12} - E(\tilde{u}_{1,t}u_{2,t}')\}\Sigma_{22}^{-1} + \Lambda_{22}\Sigma_{22}^{-1}\right]. \qquad (43)
$$

Using (43), we can observe that

$$
\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}x_t^* = \frac{1}{T\sqrt{T}}\sum_{t=1}^{T}x_t - \hat{\Lambda}_2\hat{\Sigma}^{-1}\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\hat{u}_t \Rightarrow \int_0^1 B_2(r)dr, \qquad (44)
$$

$$
\frac{1}{T^2}\sum_{t=1}^{T}x_t^* x_t^{*\prime} \Rightarrow \int_0^1 B_2(r)B_2'(r)dr. \qquad (45)
$$

Further, using Lemmas A.2 and A.3 and because $(\hat{\beta} - \beta) = O_p(N/T)$, it is shown that

$$
\frac{1}{N\sqrt{T}}\sum_{t=1}^{T}\dot{u}_{1,t}^* \Rightarrow \frac{1}{c}B_1(1) - \bar{\lambda}_{12}\Omega_{22}^{-1}B_2(1), \qquad (46)
$$

$$
\frac{1}{NT}\sum_{t=1}^{T}x_t^*\dot{u}_{1,t}^* \Rightarrow \left(\frac{1}{c}\int_0^1 B_2(r)dB_1(r) - \int_0^1 B_2(r)dB_2'(r)\Omega_{22}^{-1}\bar{\lambda}_{21}\right) + \left(\frac{1}{c}\omega_{21} - \bar{\lambda}_{21}\right) \qquad (47)
$$

after some algebra. The limiting distribution is obtained using (44)–(47).

For the DOLS estimator, we need to express the error term $\dot{u}_{1,t}$ by v_t and r_t. Note that from Theorems 3.8.3 and 8.3.1 of Brillinger (1981),

$$
u_{1,t} = v_t + r_t, \quad \text{where} \quad r_t = \sum_{j=-\infty}^{\infty}\pi_j' u_{2,t-j} \quad \text{with} \quad \sum_{j=-\infty}^{\infty} j^2\|\pi_j\| < \infty,
$$

$E(u_{2,s}v_t) = 0$ for all s and t, the long-run variance of v_t is given by $\omega_{1\cdot 2} = \omega_{11} - \omega_{12}\Omega_{22}^{-1}\omega_{21}$. From Theorem 3.8.4 of Brillinger (1981), v_t has an MA(∞) expression with the same summability condition of the coefficients as r_t. Then, we have

$$\dot{u}_{1,t} = \sum_{l=1}^{t} \rho^{t-l} u_{1,l} = \dot{v}_t + \rho^{K+1} \dot{r}_{t-K-1} + \sum_{l=t-K}^{t} \rho^{t-l} r_l, \tag{48}$$

where $\dot{v}_t = \sum_{l=1}^{t} \rho^{t-l} v_l$ and $\dot{r}_t = \sum_{l=1}^{t} \rho^{t-l} r_l$. After some algebra, the third term of (48) can be expressed as

$$\sum_{l=t-K}^{t} \rho^{t-l} r_l = \sum_{l=t-K}^{t} \rho^{t-l} \sum_{j=-\infty}^{\infty} \pi_j' u_{2,l-j} = \sum_{j=-K}^{K} \pi_j^{*\prime} u_{2,t-j} + e_{1,t} + e_{2,t}, \tag{49}$$

where $\pi_j^* = \sum_{i=0}^{K} \rho^i \pi_{j-i}$, $e_{1,t} = \sum_{j=K+1}^{\infty} \pi_j^{*\prime} u_{2,t-j}$ and $e_{2,t} = \sum_{j=-\infty}^{-K-1} \pi_j^{*\prime} u_{2,t-j}$. From (48) and (49), $\dot{u}_{1,t}$ is expressed as

$$\dot{u}_{1,t} = \sum_{j=-K}^{K} \pi_j^{*\prime} u_{2,t-j} + \dot{\eta}_t \quad \text{where} \quad \dot{\eta}_t = \dot{v}_t + \rho^{K+1} \dot{r}_{t-K-1} + e_{1,t} + e_{2,t}.$$

Using this expression, model (12) becomes $y_t = \theta' z_t + \sum_{j=-K}^{K} \pi_j^{*\prime} \Delta x_{t-j} + \dot{\eta}_t$. Then, from the inverse formula of a partitioned matrix, the DOLS estimator can be expressed as

$$\frac{1}{N} D_T(\hat{\theta}_{DOLS} - \theta) = \left(D_T^{-1} \sum_{t=K+1}^{T-K} z_t z_t' D_T^{-1} - G_1 G_2^{-1} G_1' \right)^{-1} \left(\frac{1}{N} D_T^{-1} \sum_{t=K+1}^{T-K} z_t \dot{\eta}_t - G_1 G_2^{-1} G_3 \right)$$

$$\text{where} \quad G_1 = \frac{1}{\sqrt{T}} D_T^{-1} \sum_{t=K+1}^{T-K} z_t w_t', \quad G_2 = \frac{1}{T} \sum_{t=K+1}^{T-K} w_t w_t', \quad \text{and} \quad G_3 = \frac{1}{N\sqrt{T}} \sum_{t=K+1}^{T-K} w_t \dot{\eta}_t.$$

Lemma A.4 *For $\rho = \rho_N = 1 - c/N$,*

(a) $\|G_1\|^2 = O_p\left(\frac{K}{T}\right)$,

(b) $\|G_3\|^2 = O_p\left(\frac{T}{N^2}\right) + O_p\left(\frac{K}{N}\right) + O_p\left(\frac{KN}{T}\right)$,

(c) $$\frac{1}{N\sqrt{T}} \sum_{t=K+1}^{T-K} \dot{\eta}_t \Rightarrow \begin{cases} \frac{1}{c} B_{1\cdot 2}(1) & : \frac{K}{N} \to \infty \\ \frac{1}{c} \left\{ (1 - e^{-cd_K}) B_{1\cdot 2}(1) + e^{-cd_K} B_1(1) \right\} & : \frac{K}{N} \to d_K \\ \frac{1}{c} B_1(1) & : \frac{K}{N} \to 0, \end{cases}$$

$$
(d) \quad \frac{1}{NT}\sum_{t=K+1}^{T-K} x_t \dot{\eta}_t \Rightarrow \begin{cases} \frac{1}{c}\int_0^1 B_2(r)dB_{1\cdot 2}(r)dr & : \frac{K}{N}\to\infty \\ \frac{1}{c}\Big\{(1-e^{-cd_K})\int_0^1 B_2(r)dB_{1\cdot 2}(r) & \\ \qquad +e^{-cd_K}\left(\int_0^1 B_2(r)dB_1(r)+\omega_{21}\right)\Big\} & : \frac{K}{N}\to d_K \\ \frac{1}{c}\left(\int_0^1 B_2(r)dB_1(r)+\omega_{21}\right) & : \frac{K}{N}\to 0. \end{cases}
$$

Proof of Lemma A.4: (a) is shown by

$$
\begin{aligned}
\|G_1\|^2 &= \frac{1}{T}\left[\left\|\frac{1}{\sqrt{T}}\sum_{t=K+1}^{T-K} w_t'\right\|^2 + \left\|\frac{1}{T}\sum_{t=K+1}^{T-K} x_t w_t'\right\|^2\right] \\
&= \frac{1}{T}\left[\sum_{j=-K}^{K}\left\|\frac{1}{\sqrt{T}}\sum_{t=K+1}^{T-K} u_{2,t-j}'\right\|^2 + \sum_{j=-K}^{K}\left\|\frac{1}{T}\sum_{t=K+1}^{T-K} x_t u_{2,t-j}'\right\|^2\right] = O_p\left(\frac{K}{T}\right).
\end{aligned}
$$

(b) Since $\dot{\eta}_t = \dot{v}_t + \rho^{K+1}\dot{r}_{t-K-1} + e_{1,t} + e_{2,t}$,

$$
\begin{aligned}
\|G_3\|^2 \leq C\sum_{j=-K}^{K}&\left[\left\|\frac{1}{N\sqrt{T}}\sum_{t=K+1}^{T-K} u_{2,t-j}\dot{v}_t\right\|^2 + \left\|\frac{\rho^{K+1}}{N\sqrt{T}}\sum_{t=K+1}^{T-K} u_{2,t-j}\dot{r}_{t-K-1}\right\|^2\right. \\
&\left.\left\|\frac{1}{N\sqrt{T}}\sum_{t=K+1}^{T-K} u_{2,t-j}e_{1,t}\right\|^2 + \left\|\frac{1}{N\sqrt{T}}\sum_{t=K+1}^{T-K} u_{2,t-j}e_{2,t}\right\|^2\right]. \qquad (50)
\end{aligned}
$$

In the following, we evaluate each term of (50). First, since $E(u_{2,s}v_t) = 0$ for all s and t, we can observe that

$$
E\left\|\frac{1}{N\sqrt{T}}\sum_{t=K+1}^{T-K} u_{2,t-j}\dot{v}_t\right\|^2 \leq \frac{1}{N^2T}\sum_{s=K+1}^{T-K}\sum_{t=K+1}^{T-K} E(u_{2,s-j}'u_{2,t-j})E(\dot{v}_s\dot{v}_t).
$$

Since $\dot{v}_t$ has the same structure as $\dot{u}_{1,t}$, we can observe that $Var(\dot{v}_t) = O_p(N)$ in the same manner as Lemma A.1(d), so that

$$
E(\dot{v}_s\dot{v}_t) \leq (Var(\dot{v}_s)Var(\dot{e}_t))^{1/2} = O(N). \qquad (51)
$$

Since the autocovariances of $u_{2,t}$ are absolutely summable, we have

$$
E\left\|\frac{1}{N\sqrt{T}}\sum_{t=K+1}^{T-K} u_{2,t-j}\dot{v}_t\right\|^2 \leq \frac{T}{N^2T}\sum_{j=-\infty}^{\infty} E(u_{2,t}u_{2,t-j}')O(N) = O\left(\frac{1}{N}\right), \qquad (52)
$$

which implies that

$$\sum_{j=-K}^{K}\left\|\frac{1}{N\sqrt{T}}\sum_{t=K+1}^{T-K}u_{2,t-j}\dot{v}_t\right\|^2 = O_p\left(\frac{K}{N}\right). \tag{53}$$

The second term of (50) can be expressed as

$$\frac{1}{T}\sum_{t=K+1}^{T-K}u_{2,t-j}\dot{r}_{t-K-1} = \frac{1}{T}\sum_{t=K+1}^{T-K}u_{2,t-j}\dot{u}_{1,t-K-1} - \frac{1}{T}\sum_{t=K+1}^{T-K}u_{2,t-j}\dot{v}_{t-K-1},$$

because $\dot{r}_{t-K-1} = \dot{u}_{1,t-K-1} - \dot{v}_{t-K-1}$. The second term on the right-hand side is $O_p(N^{1/2}/T^{1/2})$ from (52), while the first term is expressed as (38). Note that

$$\frac{1}{T}\sum_{t=K+1}^{T-K}\tilde{u}_{2,t-j}u_{1,t-K-1} = \gamma_{\tilde{2}1,j-K-1} + O_p\left(\frac{1}{\sqrt{T}}\right), \tag{54}$$

where $\gamma_{\tilde{2}1,l} = E(\tilde{u}_{2,t}u_{1,t+l})$ because the variance on the left-hand side is $O(1/T)$ uniformly over $-K \le j \le K$. From (38) and (54) and using $\|a+b\|^2 \le \|a\|^2 + 2\|a\|\|b\| + \|b\|^2$,

$$\begin{aligned}
&\sum_{j=-K}^{K}\left\|\frac{1}{N\sqrt{T}}\sum_{t=K+1}^{T-K}u_{2,t-j}\dot{u}_{1,t-K-1}\right\|^2 \\
&\le \frac{T}{N^2}\sum_{j=-K}^{K}\left\|\gamma_{\tilde{2}1,j-K-1} + O_p\left(\sqrt{\frac{N}{T}}\right) + O_p\left(\frac{1}{N}\right)\right\|^2 \\
&\le \frac{T}{N^2}\sum_{j=-K}^{K}\left\|\gamma_{\tilde{2}1,j-K-1}\right\|^2 + \frac{2T}{N^2}\left\{O_p\left(\sqrt{\frac{N}{T}}\right) + O_p\left(\frac{1}{N}\right)\right\}\sum_{j=-K}^{K}\left\|\gamma_{\tilde{2}1,j-K-1}\right\| \\
&\quad + O_p\left(\frac{K}{N}\right) + O_p\left(\frac{KT}{N^4}\right) + O_p\left(\frac{K\sqrt{T}}{N^2\sqrt{N}}\right) \\
&= O_p\left(\frac{T}{N^2}\right) + O_p\left(\frac{K}{N}\right),
\end{aligned}$$

because $\{\gamma_{\tilde{2}1,K-j}\}$ is an absolutely summable sequence, where the slower terms are absorbed into the faster terms like $O_p(KT/N^4) = O_p(T/N^2)$. Thus, we have

$$\sum_{j=-K}^{K}\left\|\frac{\rho^{K+1}}{N\sqrt{T}}\sum_{t=K+1}^{T-K}u_{2,t-j}\dot{r}_{t-K-1}\right\|^2 = O_p\left(\frac{T}{N^2}\right) + O_p\left(\frac{K}{N}\right) + O_p\left(\frac{KN}{T}\right). \tag{55}$$

For the third term of (50), we first note that $T^{-1}\sum u_{2,t}u_{2,t+j} = E(u_{2,t}u_{2,t+j}) + O_p(1/T^{1/2}) = \Gamma_{22,j} + O_p(1/T^{1/2})$, say, for the same reason as (54). Then,

$$\begin{aligned}
\left\| \frac{1}{N\sqrt{T}} \sum_{t=K+1}^{T-K} u_{2,t-j} e_{1,t} \right\|^2 &= \frac{T}{N^2} \left\| \sum_{l=K+1}^{\infty} \pi_l^{*\prime} \frac{1}{T} \sum_{t=K+1}^{T-K} u_{2,t-l} u_{2,t-j}^{\prime} \right\|^2 \\
&= \frac{T}{N^2} \left\| \sum_{l=K+1}^{\infty} \pi_l^{*\prime} \left(\Gamma_{22,l-j} + O_p\left(\frac{1}{\sqrt{T}}\right)\right) \right\|^2 \\
&\le C\frac{T}{N^2} \sum_{l=K+1}^{\infty} \|\pi_l^*\|^2 \sum_{l=K+1}^{\infty} \|\Gamma_{22,l-j}\|^2 + \sum_{l=K+1}^{\infty} \|\pi_l^*\|^2 O_p\left(\frac{1}{N^2}\right).
\end{aligned}$$

By noting that $\sum_{j=-K}^{K}\sum_{l=K+1}^{\infty}\|\Gamma_{22,l-j}\|^2 \le \sum_{j=1}^{\infty} j\|\Gamma_{22,j}\|^2 < \infty$ and

$$\sum_{l=K+1}^{\infty} \|\pi_l^*\| \le \sum_{l=K+1}^{\infty}\sum_{i=0}^{K} \rho^i \|\pi_{l-i}\| \le \sum_{l=K+1}^{\infty} l\|\pi_l\| = o(1), \tag{56}$$

we can observe that

$$\sum_{j=-K}^{K} \left\| \frac{1}{N\sqrt{T}} \sum_{t=K+1}^{T-K} u_{2,t-j} e_{1,t} \right\|^2 = o_p\left(\frac{T}{N^2}\right) + o_p\left(\frac{K}{N^2}\right) = o_p\left(\frac{T}{N^2}\right). \tag{57}$$

Since the fourth term of (50) has the same structure as the third term, we obtain (b) from (53), (55), and (57).

(c) Note that

$$\frac{1}{N\sqrt{T}} \sum_{t=K+1}^{T-K} \dot{\eta}_t = \frac{1}{N\sqrt{T}} \sum_{t=K+1}^{T-K} \dot{v}_t + \frac{\rho^{K+1}}{N\sqrt{T}} \sum_{t=K+1}^{T-K} (\dot{u}_{1,t-K-1} - \dot{v}_{t-K-1}) + o_p\left(\frac{1}{N}\right) \tag{58}$$

where the order in the last term is obtained by observing that

$$\left\| \frac{1}{N\sqrt{T}} \sum_{t=K+1}^{T-K} e_{1,t} \right\| \le \frac{1}{N} \sum_{j=K+1}^{\infty} \|\pi_j^*\| \sup_j \left\| \frac{1}{\sqrt{T}} \sum_{t=K+1}^{T-K} u_{2,t-j} \right\| = o_p\left(\frac{1}{N}\right),$$

because $\{\pi_j^*\}$ is an absolutely summable sequence as shown in (56) and $N^{-1}T^{-1/2}\sum e_{2,t} = o_p(N^{-1})$ is similarly shown. For the first term of (58), it is proved that

$$\frac{1}{N\sqrt{T}} \sum_{t=K+1}^{T-K} \dot{v}_t \Rightarrow \frac{1}{c} B_{1\cdot 2}(1) \tag{59}$$

in exactly the same manner as Lemma A.2(d). On the other hand, the convergence of the second term of (58) depends on the relation between K and N. Since

$$\rho^{K+1} = \left(1 - \frac{c}{N}\right)^{N\cdot(K/N)} \to \begin{cases} 0 & : \ \frac{K}{N} \to \infty \\ e^{-cd_K} & : \ \frac{K}{N} \to d_K \\ 1 & : \ \frac{K}{N} \to 0, \end{cases} \tag{60}$$

we obtain the result by Lemma A.2(d), (58), and (59).

(d) As in the proof of (c), we first observe that

$$\frac{1}{NT}\sum_{t=K+1}^{T-K} x_t \dot{\eta}_t = \frac{1}{NT}\sum_{t=K+1}^{T-K} x_t \dot{v}_t + \frac{\rho^{K+1}}{NT}\sum_{t=K+1}^{T-K} x_t(\dot{u}_{1,t-K-1} - \dot{v}_{t-K-1}) + o_p\left(\frac{1}{N}\right) \tag{61}$$

because

$$\left\|\frac{1}{NT}\sum_{t=K+1}^{T-K} x_t e_{1,t}\right\| \le \frac{1}{N}\sum_{j=K+1}^{\infty}\|\pi_j^*\| \sup_j \left\|\frac{1}{T}\sum_{t=K+1}^{T-K} x_t u_{2,t-j}\right\| = o_p\left(\frac{1}{N}\right),$$

and $N^{-1}T^{-1}\sum x_t e_{2,t} = o_p(1/N)$ is similarly shown.

Similar to the proof of Lemma A.2 (e), we can observe that

$$\frac{1}{NT}\sum_{t=K+1}^{T-K} x_t \dot{v}_t \Rightarrow \frac{1}{c}\int_0^1 B_2(r)dB_{1\cdot 2}(r) \tag{62}$$

because $u_{2,s}$ and v_t are uncorrelated. We also observe that

$$\frac{1}{NT}\sum_{t=K+1}^{T-K} x_t \dot{u}_{1,t-K-1} = \frac{1}{NT}\sum_{t=K+1}^{T-K} x_t \dot{u}_{1,t} - \frac{1}{NT}\sum_{t=K+1}^{T-K} x_t(\dot{u}_{1,t} - \dot{u}_{1,t-K-1}).$$

The second term on the right-hand side is expressed as

$$\begin{aligned} &\frac{1}{NT}\sum_{t=K+1}^{T-K}\{(x_t\dot{u}_{1,t} - x_{t-K-1}\dot{u}_{1,t-K-1}) - (x_t - x_{t-K-1})\dot{u}_{1,t-K-1}\} \\ &= \frac{1}{NT}\left(\sum_{t=T-2K}^{T-K} - \sum_{t=0}^{K}\right) x_t \dot{u}_{1,t} - \frac{1}{NT}\sum_{j=0}^{K}\sum_{t=K+1}^{T-K} u_{2,t-j}\dot{u}_{1,t-K-1}. \end{aligned}$$

The first term of the last expression is $O_p(K/(NT)^{1/2})$ because $x_t = O_p(T^{1/2})$ and $\dot{u}_{1,t} = O_p(N^{1/2})$ by Lemma A.1(d), while using (38) and (54) we can observe that

$$\begin{aligned} \left\|\frac{1}{NT}\sum_{j=0}^{K}\sum_{t=K+1}^{T-K} u_{2,t-j}\dot{u}_{1,t-K-1}\right\| &\le \frac{1}{N}\sum_{j=0}^{K}\left\|\gamma_{\tilde{2}1,K-j} + O_p\left(\sqrt{\frac{N}{T}}\right) + O_p\left(\frac{1}{N}\right)\right\| \\ &\le O\left(\frac{1}{N}\right) + O_p\left(\frac{K}{\sqrt{NT}}\right) + O_p\left(\frac{K}{N^2}\right) = o_p(1). \end{aligned}$$

Then, we can observe that

$$\frac{1}{NT}\sum_{t=K+1}^{T-K} x_t \dot{u}_{1,t-K-1} = \frac{1}{NT}\sum_{t=K+1}^{T-K} x_t \dot{u}_{1,t} + o_p(1). \tag{63}$$

we obtain the result from Lemma A.2(e), (60), (62), and(63).□

The limiting distribution of the DOLS estimator can be obtained using Lemma A.4, noting that $\|G_1 G_2^{-1} G_1'\|^2 \le O_p\left(K^2/T^2\right) = o_p(1)$ and $\|G_1 G_2^{-1} G_3\|^2 \le O_p\left(K/N^2\right) + O_p\left(K^2/NT\right) + O_p\left(K^2 N/T^2\right) = o_p(1)$, where we used $\|G_2^{-1}\|_1 = O_p(1)$ that was proved in Saikkonen (1991).■

Proof of Theorem 2: (i) Since $\hat{u}_{1,t} = \dot{u}_{1,t} - (\hat{\theta}-\theta)' z_t$, we can observe that

$$\Delta \hat{u}_{1,t} = \Delta \dot{u}_{1,t} - (\hat{\theta}-\theta)'\Delta z_t = u_{1,t} - \frac{c}{N}\dot{u}_{1,t-1} - (\hat{\beta}-\beta)' u_{2,t}.$$

Since $T^{-1}\sum u_{1,t}\dot{u}_{1,t-1}$ can be shown to be $O_p(1)$ in the same manner as Lemma A.2(c), it is shown from Lemma A.2(a) and (c) and $(\hat{\beta}-\beta) = O_p(N/T)$ that

$$\frac{1}{T}\sum_{1\le t,t+j\le T} \Delta\hat{u}_{1,t}\Delta\hat{u}_{1,t+j} = \frac{1}{T}\sum_{1\le t,t+j\le T} u_{1,t}u_{1,t+j} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{N}{T}\right).$$

This implies that

$$\hat{\omega}_{\Delta 11} = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{M_c}\right)\frac{1}{T}\sum_{1\le t,t+j\le T} u_{1,t}u_{1,t+j} + O_p\left(\frac{M_c}{N}\right) + O_p\left(\frac{M_c N}{T}\right) \xrightarrow{p} \omega_{11},$$

where the last convergence holds because $M_c/N \to 0$ and $N/T^{1/2} \to 0$ by assumption. The consistency of $\hat{c}$ and $\hat{\kappa}$ is obtained using Lemma A.3 (a).

(ii) can be obtained directly in the same manner as the proof of Theorem 1.■

Proof of Corollary 1: (i) Since the spectral density of $\Delta\dot{u}_{1,t}$ at zero frequency equals 0 under the assumption of a fixed ρ, we can see that $\hat{\omega}_{\Delta 11} = O_p(M_c^{1/2}/T^{1/2})$ from Chapter 9 of Anderson (1971). This proves the first part of (i).

To prove the second part of (i), let us choose r^* such that $r^* \to 0$ and $cr^* \to \infty$ as $c \to \infty$ and express κ as

$$\kappa = cd_M \int_0^{r^*} k(r)e^{-cd_M r}dr + cd_M \int_{r^*}^{\infty} k(r)e^{-cd_M r}dr.$$

Since $k(0) = 1$ and it is continuous at 0, the first term on the right-hand side becomes

$$cd_M \int_0^{r^*} k(r)e^{-cd_M r}dr = cd_M \int_0^{r^*} (1 + o(1))e^{-cd_M r}dr \to 1,$$

while for the second term,

$$\left| cd_M \int_{r^*}^{\infty} k(r)e^{-cd_M r}dr \right| \leq \left| cd_M \int_{r^*}^{\infty} e^{-cd_M r}dr \right| \leq e^{-cd_M r^*} \to 0.$$

We then obtain the second part of (i).

(ii) is obtained because $\tilde{\omega}_{21} - \hat{\omega}_{21} \xrightarrow{p} 0$ and $\tilde{\lambda}_{21} - \hat{\lambda}_{21} \xrightarrow{p} 0$, using $\hat{\kappa} \xrightarrow{p} 1$.■

References

[1] Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

[2] Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* 59, 817-858.

[3] Andrews, D. W. K. and J. C. Monahan (1992). An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica* 60, 953-966.

[4] Brillinger, D. R. (1981). *Time Series Data Analysis and Theory*. Holden-Day, San Francisco.

[5] Cappuccio, N. and D. Lubian (2001). Estimation and Inference on Long-Run Equilibria: A Simulation Study. *Econometric Reviews* 20, 61-84.

[6] Christou, C. and N. Pittis (2002). Kernel and Bandwidth Selection, Prewhitening, and the Performance of the Fully Modified Least Squares Estimation Method. *Econometric Theory* 18, 948-961.

[7] Engle, R. F. and C. W. J. Granger (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* 55, 251-276.

[8] Giraitis, L. and P. C. B. Phillips (2006). Uniform Limit Theory for Stationary Autoregression. *Journal of Time Series Analysis* 27, 51-60.

[9] Hansen, B. E. (1992). Convergence to Stochastic Integrals for Dependent Heterogeneous Processes. *Econometric Theory* 8, 489-500.

[10] Inder, B. (1993). Estimating Long-Run Relationships in Economics. *Journal of Econometrics* 57, 53-68.

[11] Jansson, M. (2002). Consistent Covariance Matrix Estimation for Linear Processes. *Econometric Theory* 18, 1449-1459.

[12] Kitamura, Y. and P. C. B. Phillips (1997). Fully Modified IV, GIVE and GMM Estimation with Possibly Non-Stationary Regressors and Instruments. *Journal of Econometrics* 80, 85-123.

[13] Kejriwal, M. and P. Perron (2008). Data Dependent Rules for the Selection of the Number of Leads and Lags in the Dynamic OLS Cointegrating Regression, *Econometric Theory* 24, 1425–1441.

[14] Montalvo, J. G. (1995). Comparing Cointegrating Regression Estimators: Some Additional Monte Carlo Results. *Economics Letters* 48, 229-234.

[15] Newey, W. K. and K. D. West (1994). Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies* 61, 631-653.

[16] Ng, S. and P. Perron (1995). Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag. *Journal of the American Statistical Society* 90, 268-281.

[17] Park, J. Y. (1992) Canonical Cointegrating Regressions. *Econometrica* 60, 119-143.

[18] Phillips, P. C. B. (1987). Towards a Unified Asymptotic Theory for Autoregression. *Biometrika* 74, 535-547.

[19] Phillips, P. C. B. (1995). Fully Modified Least Squares and Vector Autoregression. *Econometrica* 63, 1023-1078.

[20] Phillips, P. C. B. and B. E. Hansen (1990). Statistical Inference in Instrumental Variables Regression with I(1) Processes. *Review of Economic Studies* 57, 99-125.

[21] Phillips, P. C. B. and M. Loretan (1991). Estimating Long-Run Economic Equilibria. *Review of Economic Studies* 58, 407-436.

[22] Phillips, P. C. B. and T. Magdalinos (2007a). Limit Theory for Moderate Deviations from a Unit Root. *Journal of Econometrics* 136, 115-130.

[23] Phillips, P. C. B. and T. Magdalinos (2007b). Limit Theory for Moderate Deviations from a Unit Root under Weak Dependence, in G. D. A. Phillips and E. Tzavalis, ed., *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*. Cambridge University Press, Cambridge.

[24] Phillips, P. C. B. and V. Solo (1992). Asymptotics for Linear Processes. *Annals of Statistics* 20, 971-1001.

[25] Saikkonen, P. (1991) Asymptotically Efficient Estimation of Cointegration Regressions. *Econometric Theory* 7, 1-21.

[26] Saikkonen, P. and H. Lütkepohl (1999). Local Power of Likelihood Ratio Tests for the Cointegrating Rank of a VAR Process. *Econometric Theory* 15, 50-78.

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

[28] Stock, J. H. and M. W. Watson (1993). A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica* 61, 783-820.

[29] Sul, D., P. C. B. Phillips and C. Y. Choi (2005). Prewhitening Bias in HAC Estimation. *Oxford Bulletin of Economics and Statistics* 67, 517-546.

[30] Tanaka, K. (1996). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. Wiley, New York.