

Singapore Management University

## Institutional Knowledge at Singapore Management University

---

Research Collection School Of Economics

School of Economics

---

2-2023

### Fully modified least squares cointegrating parameter estimation in multicointegrated systems

Igor L. KHEIFETS

Peter C. B. PHILLIPS

Follow this and additional works at: [https://ink.library.smu.edu.sg/soe\\_research](https://ink.library.smu.edu.sg/soe_research)



Part of the [Econometrics Commons](#)

---

This Journal Article is brought to you for free and open access by the School of Economics at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Economics by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [cherylds@smu.edu.sg](mailto:cherylds@smu.edu.sg).

# Fully Modified Least Squares Cointegrating Parameter Estimation in Multicointegrated Systems

Igor L. Kheifets<sup>a,b</sup>Peter C. B. Phillips<sup>c,d,e,f</sup><sup>a</sup>HSE University, Russia<sup>b</sup>Center for Econometrics and Business Analytics, St. Petersburg State University, Russia<sup>c</sup>Yale University, USA<sup>d</sup>University of Auckland, New Zealand<sup>e</sup>University of Southampton, UK<sup>f</sup>Singapore Management University, Singapore

August 10, 2021

## Abstract

Multicointegration is traditionally defined as a particular long run relationship among variables in a parametric vector autoregressive model that introduces additional cointegrating links between these variables and partial sums of the equilibrium errors. This paper departs from the parametric model, using a semiparametric formulation that reveals the explicit role that singularity of the long run conditional covariance matrix plays in determining multicointegration. The semiparametric framework has the advantage that short run dynamics do not need to be modeled and estimation by standard techniques such as fully modified least squares (FM-OLS) on the original  $I(1)$  system is straightforward. The paper derives FM-OLS limit theory in the multicointegrated setting, showing how faster rates of convergence are achieved in the direction of singularity and that the limit distribution depends on the distribution of the conditional one-sided long run covariance estimator used in FM-OLS estimation. Wald tests of restrictions on the regression coefficients have nonstandard limit theory which depends on nuisance parameters in general. The usual tests are shown to be conservative when the restrictions are isolated to the directions of singularity and, under certain conditions, are invariant to singularity otherwise. Simulations show that approximations derived in the paper work well in finite samples. The findings are illustrated empirically in an analysis of fiscal sustainability of the US government over the post-war period.

*Keywords:* Cointegration, Multicointegration, Fully modified regression, Singular long run variance matrix, Degenerate Wald test, Fiscal sustainability.

*JEL Codes:* C12, C13, C22

## 1 Introduction

Many economic time series are non-stationary and contain stochastic trends, which are naturally modeled using cointegration. For example, two  $I(1)$  variables  $y_t$  and  $x_t$  are *cointegrated* if for some  $A$ ,  $u_{0t} = y_t - Ax_t$  is  $I(0)$ . Granger and Lee (1990) call *multicointegration* a situation when

---

\*© 2021 This manuscript version is made available under the CC-BY-NC-ND 4.0 license <https://creativecommons.org/licenses/by-nc-nd/4.0/>

the cumulative error  $U_{0t} = \sum_{s=1}^t u_{0s}$  is cointegrated with  $x_t$  or  $y_t$ . They analyze a case where  $(y_t, x_t, u_{0t})$  are production, sales and inventory investment,  $A = 1$  and  $U_{0t}$  is the level of inventories. Inventory stock  $U_{0t}$  may then be cointegrated with production via an adjustment mechanism that captures firm decision making on inventory investment, as well as satisfying an identity arising from the aggregation of the defining relationship  $y_t = x_t + u_{0t}$ .

It is important to take into account the presence of multicointegration in a cointegrated system: on one hand it can invalidate usual procedures of estimation and testing in cointegrated systems by affecting asymptotic properties; and on the other it may lead to advantages in improved forecasting performance. Multicointegration has so far been analyzed only in a VAR framework<sup>1</sup> and naturally involves implicit restrictions on the model induced by the extra layer of cointegration. Engsted and Johansen (1999), for example, show that if the process is generated by a VAR model for  $I(k)$  variables, multicointegration may occur if  $k = 2$  but not if  $k = 1$ . Likelihood-based estimators of cointegration parameters in  $I(2)$  VAR multicointegrated systems have mixed normal limit distributions and likelihood ratio statistics for hypothesis testing about the parameters generally have asymptotic  $\chi^2$  null distributions under conditions of correct specification, as shown for example in Johansen (1997, 2006), Boswijk (2000, 2010), and Paruolo (2000). Berenguer-Rico and Carrion-i-Silvestre (2011) provide an application of this approach that examines government debt sustainability.

In contrast to these studies, the present paper studies  $I(1)$  cointegrated models that are possibly multicointegrated in a semiparametric framework with specific focus on the use of fully modified least squares (FM-OLS) estimation. In related work, the authors (Phillips and Kheifets, 2019) explore the concept of multicointegration in a general  $I(1)$  triangular cointegrated system with weakly dependent errors, showing how multicointegration emerges naturally from singularity of the long run covariance matrix. This formulation gives an explicit mechanism generating multicointegration as a property of a general triangular  $I(1)$  system, as opposed to imposing multicointegration subsequently on a parametric system like a VAR. The contrast lies in the capacity of a general  $I(1)$  triangular system to implicitly involve the effects of multicointegration without changing or restricting the cointegrating coefficients. The implicit effects propagate from the nonparametric treatment of the equation errors and are therefore typically unknown to the investigator. This property is one of the primary motivations of our study. A second motivation is to show that FM-OLS estimation of the cointegration coefficients has some useful robustness properties to the possible unknown presence of multicointegration.

More specifically, the present paper contributes by developing asymptotic theory for FM-OLS estimation and testing in cointegrating relationships that involve multicointegration in a semiparametric setting. The analysis of triangular cointegrated systems under singularity that is developed is of some independent interest. The results show that cointegrated system estimation may proceed under certain conditions in a general  $I(1)$  cointegrated system in the presence (and without prior knowledge) of multicointegration.

To define multicointegration for weakly dependent data, we take the triangular representation

---

<sup>1</sup>It is of course possible to write stationary VARs in moving average form and vice versa under invertibility conditions. There is now a large literature describing such explicit representations for cointegrated time series. For a general approach based of Laurent series representations, see Franchi and Paruolo (2019) and the references therein.

of a linear cointegrating relationship. In the cointegrating regression model

$$y_t = Ax_t + u_{0t}, \quad x_t = x_{t-1} + u_{xt}, \quad t = 1, \dots, T, \quad (1)$$

$A$  is a  $m_0 \times m_x$  cointegrating coefficient matrix,  $x_t$  is initialized at  $t = 0$  by  $x_0 = O_p(1)$ , and the combined error vector  $u_t = (u'_{0t}, u'_{xt})'$  follows the linear process

$$u_t = D(L)\eta_t = \sum_{j=0}^{\infty} D_j \eta_{t-j}, \quad \eta_t \sim iid(0, I_m), \quad \text{with } \sum_{j=0}^{\infty} j^\nu \|D_j\| < \infty, \quad (2)$$

for some  $\nu > 2$ , finite fourth order cumulants of  $\eta_t$ , and where  $m = m_0 + m_x$ . It is common in the literature to consider such time series with an additional assumption  $|D(1)| \neq 0$  (e.g. Phillips, 1995) that assures nonsingularity of the long run variance matrix of  $u_t$ , which we relax here.

Let  $\Gamma_{u,u}(h) = \mathbb{E}u_{t+h}u'_t$ . The linear operator  $D(L)$ , the long run covariance matrix  $\Omega = \sum_{h=-\infty}^{\infty} \Gamma_{u,u}(h) = D(1)D(1)' = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} D_j D'_k$  of  $u_t$  and one-sided long run covariance matrix  $\Gamma^+ = \sum_{h=0}^{\infty} \Gamma_{u,u}(h) = \sum_{k=0}^{\infty} \sum_{j=0}^k D_j D'_k$  of  $u_t$  are partitioned conformably with  $u_t$  as

$$D(L) = \begin{bmatrix} D_{00}(L) & D_{0x}(L) \\ D_{x0}(L) & D_{xx}(L) \end{bmatrix}, \quad \Omega = \begin{bmatrix} \Omega_{00} & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx} \end{bmatrix}, \quad \Gamma^+ = \begin{bmatrix} \Delta_{00} & \Delta_{0x} \\ \Delta_{x0} & \Delta_{xx} \end{bmatrix},$$

where  $\Omega_{xx} > 0$  is positive definite so that  $x_t$  is a full rank  $I(1)$  regressor vector, as commonly assumed in triangular systems such (1) and (2) following Phillips(1991)<sup>2</sup>. The conditional long run covariance matrix, defined as the Schur complement of the block  $\Omega_{xx}$ , is  $\Omega_{00.x} = \Omega_{00} - \Omega_{0x}\Omega_{xx}^{-1}\Omega_{x0}$  and is positive (semi-) definite if and only if  $\Omega$  is positive (semi-) definite (by virtue of the Guttman rank additivity formula). In this paper we consider a situation when the long run variance matrix is singular, or, equivalently, when the conditional long run covariance matrix is singular. It corresponds to a case where partial sums of  $y_t$  and  $x_t$  are cointegrated with an  $I(0)$  error in some unknown direction, i.e. when there is a multicointegration in the spirit of Granger and Lee (1990), but is semiparametric in the sense that the short run dynamics are left unspecified. We therefore introduce the following definition.

**Definition 1.** *The process generated by a triangular cointegrating system is called multicointegrated if its long run error covariance matrix is singular.*

The advantage of this framework is that it provides the explicit origin from which the multicointegrating relationship arises in an  $I(1)$  system. Thus, if we take partial sums of the augmented regression form (Phillips, 1991)

$$y_t = Ax_t + F(1 - L)x_t + u_{0.x,t}, \quad (3)$$

where  $F = \Omega_{0x}\Omega_{xx}^{-1}$  is the long run regression coefficient of  $u_{0t}$  on  $x_t$  and  $u_{0.x,t} = u_{0t} - \Omega_{0x}\Omega_{xx}^{-1}u_{xt}$ , giving (using capitals with time index  $t$  for partial sums)

$$Y_t = AX_t + Fx_t + U_{0.x,t}. \quad (4)$$

---

<sup>2</sup>The case where the regressors  $x_t$  are themselves cointegrated is considered in Phillips (1995) but is not considered in this paper.

It becomes clear that in the direction of singularity of  $\Omega_{00.x}$  we have an exact long run relationship that links the time series  $Y_t$ ,  $X_t$ , and  $x_t$  and this relationship is prescribed in terms of the coefficients  $A$ ,  $F$  and the singular direction of  $\Omega_{00.x}$ , which is estimable. In earlier work on multicointegration, the hypothesis about multicointegration is imposed a priori, directly and explicitly as in Granger and Lee (1990) or through rank conditions in VAR analyses. What our approach does is: (i) show that multicointegration may exist in a triangular  $I(1)$  system such as (1); (ii) reveal the leading and intuitively simple role that the singularity of the long run conditional error covariance matrix  $\Omega_{00.x}$  plays in giving rise to multicointegration, a feature of the model that may be unknown to the investigator; (iii) allow for both cointegration and multicointegration within the same specification; and (iv) use a nonparametric formulation to provide a general setting for the analysis, for the form of the cointegrating and multicointegrating coefficients, and for practical work.

In a VAR framework Engsted and Johansen (1999) show that multicointegration, as defined in Granger and Lee (1990) of a linear  $I(1)$  process<sup>3</sup>  $(y'_t, x'_t)' = (1 - L)^{-1} C(L)\eta_t$  where the roots of  $|C(z)| = 0$  satisfy  $|z| > 1$  or  $z = 1$ , occurs when  $z = 1$  is a root<sup>4</sup>, so that  $C(1) = \xi\epsilon'$  has reduced rank and  $\xi'_\perp \dot{C}(1)\epsilon_\perp$  is singular (explicit forms of the matrices  $\{\xi, \epsilon\}$  and their orthogonal complements  $\{\xi_\perp, \epsilon_\perp\}$  are given in the proof of the Proposition 1). This is the case when  $\Omega$  is singular or more specifically in the present context when  $\Omega_{00.x}$  is singular, as shown below.

**Proposition 1.** *A linear process  $(y'_t, x'_t)'$  generated by (1)-(2) with  $\Omega_{xx} > 0$  is multicointegrated, i.e.,  $\Omega$  is singular, if and only if it satisfies the multicointegration condition of Engsted and Johansen (1999). The rank of the multicointegrating relation equals  $m - \text{rank}(\Omega) = m_0 - \text{rank}(\Omega_{00.xx})$ .*

In what follows data matrices are denoted by upper case letters without indexes, e.g.,  $Y' = [y_1, \dots, y_T]$ . The OLS estimator  $\hat{A} = Y'X(X'X)^{-1}$  is consistent at the rate at least  $O(T)$ . The FM-OLS estimator (Phillips and Hansen 1990) has the form  $\hat{A}^+ = (\hat{Y}^{+'}X - T\hat{\Delta}_{0x}^+) (X'X)^{-1}$  and employs corrections for endogeneity in the regressor  $x_t$ , leading to the transformed dependent variable  $\hat{y}_t^+ = y_t - \hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}(x_t - x_{t-1})$  and a bias correction term involving  $\hat{\Delta}_{0x}^+ = \hat{\Delta}_{0x} - \hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}\hat{\Delta}_{xx}$ , which is constructed in the usual way using consistent nonparametric estimators of submatrices of the long run and one sided long run quantities  $\Omega$  and  $\Gamma^+$ . Compared with OLS, the FM-OLS estimator removes asymptotic bias and increases efficiency by correcting both the long run serial correlation in  $u_t$  and endogeneity in  $x_t$  caused by the long run correlation between  $u_{0t}$  and  $u_{xt}$ . The properties of FM-OLS in general regressions as well as VARs are studied in Phillips (1995). Here we advance the analysis by allowing for the possibility of a singular conditional long run variance matrix  $\Omega_{00.x}$ . When  $\Omega_{00.x}$  is singular, i.e. when modified  $y_t$  is cointegrated and in some direction the errors in the cointegrating equation are  $I(-1)$ , the limit theory of the FM-OLS estimator is degenerate at the usual  $O(T)$  rate and the faster convergence rate affects both estimation and inference.

The paper makes the following contributions. First, we derive the new rates of convergence

---

<sup>3</sup>The summability condition in the specification (2) imposes a restriction on  $C(z)$ . This is because the matrix moving average power series  $D(z)$  of a linear process generated by (1)-(2) does not have poles at  $z = 1$ , which implies that, when the system is written in the form  $(y'_t, x'_t)' = (1 - L)^{-1} C(L)\eta_t$ , it must be that  $\xi_1 C(1) = 0$ , where  $\xi_1 = (I_{m_0}, -A)$ . Indeed, the upper block of  $D(z)$  of such a system satisfies  $(D_{00}(z), D_{0x}(z)) = \xi_1(1 - z)^{-1} C(z) = (1 - z)^{-1} \xi_1 C(z)$  and does not have poles at  $z = 1$  if and only if  $\xi_1 C(1) = 0$ . We thank a referee for this clarifying observation and suggestions to improve the statement and the proof of Proposition 1.

<sup>4</sup>The order  $m$  of zeros of  $C(z)$  at  $z = 1$  or, equivalently, the order of poles of  $C(z)^{-1}$  at  $z = 1$ , is not restricted to  $m = 1$  and is unknown. A unified treatment of the different representations of cointegrated systems for known  $m$  is given in a recent paper by Franchi and Paruolo (2019).

and limit distribution of the FM-OLS estimator in the case of a null conditional long run variance matrix. The new rate exceeds  $O(T)$  and depends on the bandwidth used in estimating the long run covariance matrix quantities that are employed in making corrections for endogeneity and serial correlation in FM-OLS. The resulting limit distribution is no longer mixed normal and depends on nuisance parameters. Similar properties hold in the direction of singularity in the case of a singular long run variance matrix. Second, under certain conditions, the limit distribution of Wald statistics for testing restrictions on the cointegrating space and cointegrating parameters is  $\chi^2$  and is invariant to the presence of singularity. Third, we show that when those restrictions fail, the Wald test is conservative. Monte Carlo simulations reveal that the empirical level of the test can be far below the nominal 1%, 5% and 10% levels in singular and near singular cases.

As an application of our methods we analyze fiscal sustainability of the US government over the period 1947-2019 by testing the null hypothesis that the cointegration relationship between government revenue and expenditure has the parametric form  $(1, -1)$ . Multicointegration between government revenue and expenditure naturally arises if bounds are imposed on deviations of debt from revenue. We reject the null hypothesis and, as our theoretical results show, this conclusion is not affected by the presence of multicointegration. The finding is important for practical purposes, as a separate treatment of the multicointegration case is not necessary (cf., Quintos, 1995, and Berenguer-Rico and Carrion-i-Silvestre, 2011).

The paper is organized as follows. In Section 2 we derive the rates of convergence of elements of  $\widehat{A}^+$  and establish its limit distribution. After some preliminary observations we begin our discussion with the null case where  $\Omega_{00.x} = 0$ , then move on to a case of a general singular matrix. The implications of singularity for hypothesis testing are discussed in Section 3. The finite sample properties of the FM-OLS and Wald test statistics are explored in Section 4. The application to government fiscal sustainability is considered in Section 5. Section 6 concludes. Proofs are given in the Appendix.

## 2 Fully Modified OLS

Under the stated conditions the functional law  $T^{-1/2} \sum_{t=1}^{[T\cdot]} u_t \rightarrow_d B(\cdot) \equiv BM(\Omega)$  holds for partial sums of  $u_t$  (e.g., Phillips and Solo, 1992). Define the partition  $B = (B'_0, B'_x)'$  into the first  $m_0$  and the final  $m_x$  subvectors of the Brownian motion, conformably with  $u_t$ . Introducing the matrices

$$\mathcal{L}_{0.\Omega} = \begin{bmatrix} I_{m_0} & -\Omega_{0x}\Omega_{xx}^{-1} \end{bmatrix}, \quad \mathcal{L}_\Omega = \begin{bmatrix} I_{m_0} & -\Omega_{0x}\Omega_{xx}^{-1} \\ 0 & I_{m_x} \end{bmatrix}$$

and the Brownian motion  $B_{0.x} = \mathcal{L}_{0.\Omega}B$ , we have

$$\begin{bmatrix} B_{0.x} \\ B_x \end{bmatrix} = \mathcal{L}_\Omega \begin{bmatrix} B_0 \\ B_x \end{bmatrix} = BM(\mathcal{L}_\Omega \Omega \mathcal{L}'_\Omega) = BM \left( \begin{bmatrix} \Omega_{00.x} & 0 \\ 0 & \Omega_{xx} \end{bmatrix} \right),$$

where  $B_{0.x} \equiv BM(\Omega_{00.x})$  is orthogonal to  $B_x$ . Note that  $\Omega_{00.x}$  is the long run variance of  $u_{0.x,t} = \mathcal{L}_{0.\Omega}u_t = D_{0.x}(L)\eta_t$ , where  $D_{0.x}(L) = \mathcal{L}_{0.\Omega}D(L)$ . It is well known that the OLS estimator of  $A$  in (1) is  $O(T)$  consistent with a limit distribution that depends on the nuisance parameters  $\Omega$  and  $\Gamma^+$ ,

viz.,

$$\begin{aligned}
T \left( \widehat{A} - A \right) \rightarrow_d & \left( \int_0^1 dB_0 B'_x + \Delta_{0x} \right) \left( \int_0^1 B_x B'_x \right)^{-1} = \left( \int_0^1 dB_{0.x} B'_x \right) \left( \int_0^1 B_x B'_x \right)^{-1} \\
& + \Omega_{0x} \Omega_{xx}^{-1} \left( \int_0^1 dB_x B'_x \right) \left( \int_0^1 B_x B'_x \right)^{-1} + \Delta_{0x} \left( \int_0^1 B_x B'_x \right)^{-1}. \quad (5)
\end{aligned}$$

The last two terms of (5) are the endogeneity and serial correlation biases that FM-OLS seeks to remove.

Suppose  $\Omega$  and  $\Gamma^+$  are estimated in the usual way (e.g., Priestley 1981; Hannan, 1970) as

$$\widehat{\Omega} = \sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{\widehat{u}, \widehat{u}}(j) \quad \text{and} \quad \widehat{\Delta} = \sum_{j=0}^{T-1} w(j/K) \widehat{\Gamma}_{\widehat{u}, \widehat{u}}(j),$$

where  $w(\cdot)$  is a kernel function,  $K$  is a bandwidth parameter and the sample covariances are  $\widehat{\Gamma}_{\widehat{u}, \widehat{u}}(j) = T^{-1} \sum_{1 \leq t, t+j \leq T} \widehat{u}_{t+j} \widehat{u}'_t$ ,  $\widehat{u}_t = (\widehat{u}'_{0t}, u'_{xt})'$ , where  $\widehat{u}_{0t} = y_t - \widehat{A}x_t$ . Similar to Phillips (1995), we consider the following kernels and bandwidth rates.

**Assumption K** (Kernel Condition) For given  $k \in (0, 1)$ , the bandwidth parameter  $K$  has the rate  $K \sim c_T T^k$  as  $T \rightarrow \infty$ , where  $c_T$  is slowly varying at infinity, i.e.  $c_{xT}/c_T \rightarrow 1$  for  $x > 0$  and  $T \rightarrow \infty$ . The kernel function  $w(\cdot) : \mathbb{R} \rightarrow [-1, 1]$  is a twice continuously differentiable even function with

- (a)  $w(0) = 1, w'(0) = 0, w''(0) \neq 0$ , and
- (b)  $w(x) = 0, |x| \geq 1$ , with  $\lim_{|x| \rightarrow 1} w(x)/(1 - |x|)^2 = \text{const.}$

Parzen and Tukey–Hanning kernels satisfy Assumption K. The Bartlett–Priestley or quadratic spectral kernels do not satisfy Assumption K but to use them in the following development these kernels need to satisfy

$$(b') \quad w(x) = O(x^{-2}), \text{ as } |x| \rightarrow \infty$$

and (a) with support  $(-\infty, \infty)$ . Under Assumption K, with  $0 < k < 1$ , and any consistent estimator  $\widehat{A}$  we have  $\widehat{\Gamma} \rightarrow_p \Gamma$ ,  $\widehat{\Omega} \rightarrow_p \Omega$ ,  $\widehat{\Delta} \rightarrow_p \Gamma^+$ .

**Proposition 2.** *Under Assumption K with  $0 < k < 1$ ,*

$$T \left( \widehat{A}^+ - A \right) \rightarrow_d \left( \int_0^1 dB_{0.x} B'_x \right) \left( \int_0^1 B_x B'_x \right)^{-1}.$$

For the nonsingular case this result appears in Corollary 4.3 in Phillips (1995). The proof reveals that singularity does not alter the above convergence but makes the limit distribution degenerate. If  $\Omega_{00.x}$  has full rank, the rate of convergence of the FM-OLS estimator is determined by the rates of weak convergence of the sample covariances and the rate of nonparametric estimation of  $\Omega$  and  $\Gamma^+$  does not play any role. We will show that in case  $\Omega_{00.x}$  is singular, the rate of convergence of

the FM-OLS estimator along the null direction of  $\Omega_{00.x}$  increases by  $\delta(T)$ , where

$$\delta(T) = \begin{cases} T^{1+2k}, & k \in (0, 1/4), \\ T^{3/2}, & k \in [1/4, 1/2], \\ T^{2-k}, & k \in (1/2, 1). \end{cases}$$

The fastest rate of convergence of FM-OLS in the null direction is  $T^{3/2}$  when bandwidth expansion rate is  $k \in [1/4, 1/2]$ .

For example, in the case where  $\Omega_{00.x} = 0$ , we have  $T(\widehat{A}^+ - A) \rightarrow_p 0$  and the precise rate of convergence of the FM-OLS estimator depends on the bandwidth parameter expansion rate  $k$  in kernel estimation of the nonparametric components. Parameter dependencies may then be present in the resulting asymptotic theory, arising from first order terms in the limit behavior of the long run covariances that influence the asymptotics in this degenerate case. In particular, when  $\Omega_{00.x} = 0$  the component  $D_{0.x}(1) = 0$  in the Beveridge-Nelson decomposition of  $u_{0.x,t}$ . In this case,  $u_{0.x,t} = D_{0.x}(1)\eta_t + \widetilde{D}_{0.x}(L)\eta_{t-1} - \widetilde{D}_{0.x}(L)\eta_t$  reduces to a first difference  $u_{0.x,t} = -(e_t - e_{t-1})$ , which is  $I(-1)^5$  with  $e_t = \widetilde{D}_{0.x}(L)\eta_t = \sum_{j=0}^{\infty} \widetilde{D}_{0.x,j}\eta_{t-j}$ , with  $\widetilde{D}_{0.x,j} = \mathcal{L}_{0,\Omega} \sum_{t=j+1}^{\infty} D_t$ , and  $e_t$  has long run variance matrix  $\Omega_{ee} = \widetilde{D}_{0.x}(1)\widetilde{D}_{0.x}(1)'$ .

The next proposition establishes convergence properties of FM-OLS for such time series. It is particularly useful for the case of a single cointegration relationship with  $m_0 = 1$  (e.g., Phillips and Loretan, 1991), because singularity implies that the conditional long run variance is zero. This reduction makes explicit the effect of singularity on the convergence rates and serves as the basis of a general result.

**Proposition 3.** *Suppose  $\Omega_{00.x} = 0$ . Under Assumption K with  $0 < k < 1$ ,*

$$\delta(T) \left( \widehat{A}^+ - A \right) = O_p(1).$$

As the proof of Proposition 3 reveals, the limit distribution of the restandardized estimation error  $\delta(T)(\widehat{A}^+ - A)$  depends on nuisance parameters associated with the nonparametric estimation of the long run covariance matrices. For kernel estimators, the limit depends on the covariance structure of the errors, on the bandwidth growth rate, and on the second derivative of the kernel function. For illustration, consider the case when the bandwidth  $K$  grows slower than  $T^{1/4}$ , which includes the typical optimal bandwidth rate  $T^{1/5}$  for long run variance estimation. Under these conditions, we have the following limit theory.

**Proposition 4.** *Suppose  $\Omega_{00.x} = 0$ . Under Assumption K with  $k < 1/4$ ,*

$$T^{1+2k} \left( \widehat{A}^+ - A \right) \rightarrow_d w''(0) \left( \Phi_0 + \Phi_{-\infty} \Omega_{xx}^{-1} \int_0^1 dB_x B_x' \right) \left( \int_0^1 B_x B_x' \right)^{-1},$$

where  $\Phi_h = \sum_{j=h}^{\infty} (j + 1/2) \mathbb{E} e_{t+j} u_{xt}'$ .

Unlike the corresponding limit theory in the nonsingular case (Phillips, 1995; Phillips and Hansen,

---

<sup>5</sup>The fact that  $u_{0.x,t}$  is  $I(-1)$  includes the possibility that in some or all directions  $u_{0.x,t}$  may be  $I(-d)$ ,  $d > 1$ , but that possibility does not affect the convergence properties of the FM-OLS estimator as the next proposition shows.



1990; Phillips, 2014), the limit distribution of FM-OLS now depends on the covariance structure of the errors  $u_{xt}$  and  $e_t$  and on the second derivative of the kernel function.

Next, consider a general case of singular  $\Omega_{00.x}$  with rank  $r < m_0$ , so that  $\Omega$  has rank  $r + m_x$ . To isolate nondegenerate directions decompose  $\Omega_{00.x} = RR'$ , where  $R$  is an  $m_0 \times r$  matrix of rank  $r$ . Then  $R'R$  has full rank,  $R'u_{0.x,t}$  has full rank long run variance matrix and Proposition 2 applies in this direction. In the orthogonal direction<sup>6</sup>  $R_\perp$ ,  $R'_\perp u_{0.x,t} = -(e_t - e_{t-1})$  is  $I(-1)$ <sup>7</sup>, where  $e_t = R'_\perp \tilde{D}_{0,x}(L)\eta_t$  has long run variance  $\Omega_{ee} = R'_\perp \tilde{D}_{0,x}(1)\tilde{D}_{0,x}(1)'R_\perp$  and Proposition 3 applies, showing that elements in this direction  $R'_\perp A$  are estimated at a faster rate than  $O(T)$ .

We now state our first main result.

**Theorem 1.** *Suppose  $\Omega_{00.x} = RR'$ , where  $R$  is an  $m_0 \times r$  matrix with  $\text{rank}(R) = r < m_0$ . Then under Assumption K*

$$T \left( \hat{A}^+ - A \right) \rightarrow_d \left( \int_0^1 dB_{0.x} B'_x \right) \left( \int_0^1 B_x B'_x \right)^{-1},$$

which is degenerate mixed normal. The limit distribution is not degenerate and has full rank in direction  $R$  with

$$TR' \left( \hat{A}^+ - A \right) \rightarrow_d \left( \int_0^1 dB_{f.x} B'_x \right) \left( \int_0^1 B_x B'_x \right)^{-1},$$

where  $B_{f.x} \equiv BM(\Omega_{ff.x})$  and  $\Omega_{ff.x} = R'RR'R$  is the full rank  $r \times r$  conditional long run variance matrix of  $R'U_{0.x}$ . In the direction  $R_\perp$  orthogonal to  $R$  the convergence of  $\hat{A}^+$  is at the faster rate  $O(\delta(T))$  and  $\delta(T)R'_\perp (\hat{A}^+ - A) = O_p(1)$ .

The FM-OLS estimator of a singular triangular system with multicointegration therefore has the following properties: (i) FM-OLS is consistent; (ii) the limit distribution is degenerate in the original coordinates; and (iii) rates of convergence are  $O(T)$  in nondegenerate directions and  $O(\delta(T))$  in degenerate directions. In the degenerate direction the limit distribution is of the type shown in Proposition 4. Singularity of the limit distribution means that care is needed when undertaking inference and these matters are considered in the next section. The situation is in some ways analogous to that of causality testing in cointegrated VAR regressions, as analyzed in Toda and Phillips (1993), and cointegrating regressions with cointegrated regressors, as analyzed in Phillips (1995). In the present case, it is necessary to analyze the directions of singularity of the long run covariance structure and the behavior of the estimates in these directions.

### 3 Testing

We consider the following hypothesis involving functions  $\phi \in C_q^1(-\infty, \infty)$ , the space of  $q$  dimensional continuously differentiable functions  $\mathcal{H}_0 : \phi(\text{vec}(A)) = 0$ , where  $\text{vec}(\cdot)$  is row vectorization. Suppose

<sup>6</sup>By the usual eigenvalue decomposition for symmetric matrices there is a set of orthonormal eigenvectors  $\{q_i\}_{i=1}^m$  of  $\Omega_{00.x}$ , stacked as an orthogonal matrix  $C$  and real eigenvalues  $\lambda_i$  in decreasing order on diagonal matrix  $\Lambda$ , such that  $\Omega_{00.x} = C\Lambda C' = \sum_{i=1}^r \lambda_i q_i q_i'$ . In this notation,  $C\Lambda^{1/2} = (R, 0)$  and  $R_\perp$  spans the space of eigenvectors corresponding to zero eigenvalues.

<sup>7</sup>This representation allows for  $R'_\perp u_{0.x,t}$  being  $I(-d)$  with  $d > 1$  in some directions.

$\Omega_{00.x} = RR'$ , where  $R$  is an  $m_0 \times r$  matrix of  $\text{rank}(R) = r < m_0$ , so that  $R'R$  is  $r \times r$  nonsingular. Then under Assumption K

$$T \left( \widehat{A}^+ - A \right) \rightarrow_d \left( \int_0^1 dB_{0.x} B_x' \right) \left( \int_0^1 B_x B_x' \right)^{-1} \equiv \mathcal{MN} \left( 0, \Omega_{00.x} \otimes \left( \int_0^1 B_x B_x' \right)^{-1} \right).$$

The limit distribution is mixed normal ( $\mathcal{MN}$ ) and standard inference methods can be applied. The usual Wald statistic for testing  $\mathcal{H}_0$  is

$$W = \phi(\widehat{a}^+)' \left\{ \Phi(\widehat{a}^+) \left( \widehat{\Omega}_{00.x} \otimes (X'X)^{-1} \right) \Phi(\widehat{a}^+)' \right\}^{-1} \phi(\widehat{a}^+),$$

where  $\widehat{a}^+ = \text{vec}(\widehat{A}^+)$ ,  $\Phi(a) = \partial\phi(a)/\partial a'$  and  $a = \text{vec}(A)$  is row vectorization. Suppose that the following rank condition holds

$$\text{rank} \left\{ \Phi(a) \left( \Omega_{00.x} \otimes \left( \int_0^1 B_x B_x' \right)^{-1} \right) \Phi(a)' \right\} = q. \quad (6)$$

Under Assumption K,  $W \rightarrow_d \chi_q^2$ . So, under the rank condition (6), the limit distribution of the Wald statistics is invariant to the presence of singularity.

### 3.1 Violation of the rank condition

Consider the linear hypothesis

$$\mathcal{H}_0 : Q \text{vec}(A) = r_0,$$

with restriction  $q \times m_0 m_x$  matrix  $Q = R_1 \otimes R_2$  of rank  $q = q_1 q_2$  and component matrices  $R_1$ ,  $m_0 \times q_1$ , of rank  $q_1$  and  $R_2$ ,  $m_0 \times q_2$ , of rank  $q_2$ . Then  $\mathcal{H}_0$  has a tensor form with matrix representation  $\mathcal{H}_0 : R_1 A R_2 = R_3$ , with  $\text{vec} R_3 = r_0$ , and

$$Q \left( \Omega_{00.x} \otimes \left( \int_0^1 B_x B_x' \right)^{-1} \right) Q' = R_1 \Omega_{00.x} R_1' \otimes R_2' \left( \int_0^1 B_x B_x' \right) R_2.$$

If the rank of  $R_1 \Omega_{00.x} R_1' = R_1 R R' R_1'$  is  $\tilde{q}_1 < q_1$ , then the rank condition (6) fails as  $\tilde{q}_1 q_2 < q_1 q_2 = q$ . This is the case where some of the restrictions isolate directions in which FM-OLS is hyperconsistent with rate exceeding  $O(T)$ . The distribution of the Wald test statistic is then nonstandard and depends on nuisance parameters. In general, failure of estimator mixed normality in the direction of faster convergence produces a non chi-squared limit in the Wald statistic as the faster convergence of the estimator is balanced in the Wald statistic weighting. A related phenomenon arises in Toda and Phillips (1993), who describe situations where Wald tests of Granger causality in cointegrated VAR systems do not follow asymptotically chi-squared distributions. For another example, see Phillips (2016), where singularity in the signal matrix leads to nonstandard inference.

To illustrate the consequences of singularity consider testing  $\mathcal{H}_0 : A = A^0$ . Then  $Q = I_{m_0 m_x}$ ,  $r_0 = \text{vec}(A^0)$ ,  $R_1 = I_{m_0}$ ,  $R_2 = I_{m_x}$ ,  $R_3 = A^0$  and their ranks are  $q = m_0 m_x$ ,  $q_1 = m_0$  and  $q_2 = m_x$ . The

Wald test statistic then simplifies to

$$\begin{aligned} W_I &= \text{vec} \left( \widehat{A}^+ - A^0 \right)' \left( \widehat{\Omega}_{00.x} \otimes (X'X)^{-1} \right)^{-1} \text{vec} \left( \widehat{A}^+ - A^0 \right) \\ &= \text{tr} \left\{ (X'X) \left( \widehat{A}^+ - A^0 \right)' \widehat{\Omega}_{00.x}^{-1} \left( \widehat{A}^+ - A^0 \right) \right\}. \end{aligned}$$

The notational change to  $W_I$  emphasizes that the following analysis only considers the full dimensional restriction structures above. The rank of  $\Omega_{00.x} \otimes \int_0^1 B_x B_x'$  equals the rank of the conditional long run variance times  $m_x$ , i.e. the null hypothesis restrictions isolate ‘all directions’ and the rank condition is satisfied if and only if the conditional long run variance matrix is nonsingular. If the conditional long run variance is nonsingular, the rank condition holds and  $W_I \rightarrow \chi_q^2$ .

Singularity alters the rate of convergence and the limit of  $\widehat{\Omega}_{00.x}$ , which is used in the construction of the Wald test statistics. We proceed to derive the rate of convergence for this quantity. As the proof of the following proposition reveals, the rate of convergence is  $T^{2k}$  if  $k < 1/3$  and  $T^{1-k}$  if  $k \geq 1/3$ . As in the case of FM-OLS estimation, the limit distribution of  $\widehat{\Omega}_{00.x}$  depends on nuisance parameters and on the implementation of the nonparametric estimates of the long run covariance matrices. As an illustration, consider a case when the bandwidth  $K$  grows slower than  $T^{1/3}$ , which includes the usual optimal bandwidth rate  $T^{1/5}$  for long run variance estimation.

**Proposition 5.** *Suppose  $\Omega_{00.x} = 0$  and Assumption K holds with  $0 < k < 1/3$ . Then*

(a)  $T^{2k} \widehat{\Omega}_{00.x} \rightarrow_p -w''(0) \Omega_{ee}$ .

(b) *If  $\Omega_{ee}$  is nonsingular, then  $W_I = O_p(T^{-2k})$  if  $k \in (0, 1/4]$  and  $W_I = O_p(T^{2k-1})$  if  $k \in (1/4, 1/3)$ .*

Nonsingularity of  $\Omega_{ee}$  means that there is no further level of cointegration (system (1) is singular of first order, see Park 1992) and guarantees that the limit in Part (a) is nondegenerate, so that the rate of convergence of  $\widehat{\Omega}_{00.x}$  is sharp. In the more general case where  $\Omega_{00.x}$  is positive semi-definite but not a null matrix we have the following result.

**Theorem 2.** *Suppose  $\Omega_{00.x} = RR'$ , where  $R$  is an  $m_0 \times r$  matrix with  $\text{rank}(R) = r < m_0$ ,  $\Omega_{ee}$  is nonsingular, and Assumption K holds with  $0 < k < 1/3$ , then under the null  $W_I \rightarrow_d \chi_{rm_x}^2$ .*

The proof of the above result reveals that the limit distribution of the Wald test statistic involves the sum of two major components. The first component arises from the limit in the nonsingular direction, which is  $\chi_{rm_x}^2$ , and the second involves the limit in the direction where the conditional long run variance matrix is zero, which is nonstandard, depends on nuisance parameters and decays at the speed established in Proposition 5 above. Therefore, for  $k < 1/3$  the  $\chi_{rm_x}^2$  limit distribution of the  $W_I$  statistic has thinner tails than the distribution of  $\chi_{m_0 m_x}^2$ , so that tests based on the usual degrees of freedom are asymptotically conservative.

## 4 Finite sample performance

The following analysis of finite sample performance is based on 10,000 simulations with different sample sizes.<sup>8</sup> The long run variances are estimated using the Parzen kernel

$$w(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & -1/2 \leq x \leq 1/2, \\ 2(1 - |x|)^3, & 1/2 \leq |x| \leq 1, \\ 0, & 1 \leq |x|, \end{cases}$$

and bandwidth is set to  $K = T^{1/4}$ , if not specified otherwise. The data generating process (DGP) has the form (1) with scalar cointegrating coefficient  $A = 2$  and with a combined error vector  $u_t = (u_{0t}, u'_{xt})'$  that follows the linear process

$$u_t = \eta_t + D_1 \eta_{t-1}, \quad \text{with } \eta_t \sim iid \mathcal{N}(0, \Sigma).$$

We consider estimation of  $A$  and hypothesis testing for the null  $\mathcal{H}_0 : A = 2$ . We look at two classes of bivariate ( $m = 2$ ) DGPs.

DGP1 For parameter choices  $p \in \{0.0, -0.1, \dots, -1.0\}$  define

$$D_1 = \begin{bmatrix} p & 0 \\ 0 & 0 \end{bmatrix}, \quad \Sigma = I_2.$$

DGP2 For parameter choices  $p \in \{0.8, 5.2\}$  define

$$D_1 = \begin{bmatrix} 0.3 & 0.4 \\ p & 0.6 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

The finite sample performance for DGP1 is shown in Figure 1 and Table 1; results for DGP2 are given in Figure 2 and Table 2.

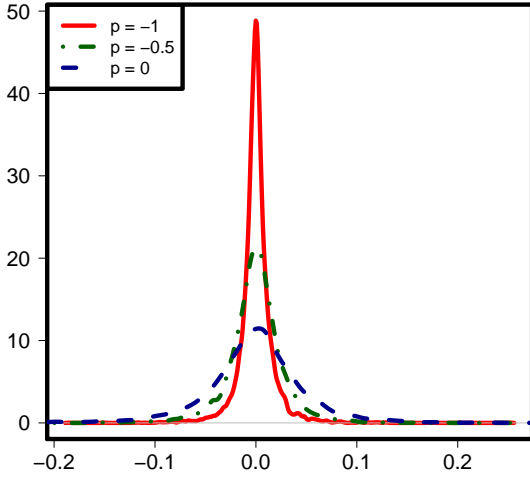
### Discussion of results for DGP1.

When  $p = -1$  we have  $\Omega_{00.x} = 0$  and a singular system. The limit theory generalizes results on estimation and testing to this case. With matrix  $D_1$  diagonal, the errors  $u_{0t}$  and  $u_{xt}$  are independent and the effect of singularity in the long run variance can be studied separately from the effect of long run dependence. When  $p = 0$ , the long run variance is the identity  $I_2$  and the conditional long run variance is 1, giving a standard nonsingular case. For values of  $p$  between 0 and  $-1$  the system is still in the nonsingular case but in finite samples for smaller values of  $p$  the limit theory for the singular case may lead to better approximations than the nonsingular case and simulations help to guide this assessment.

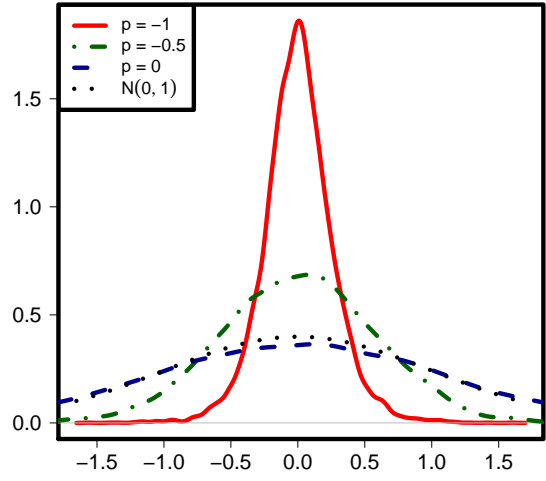
In Figure 1, Panel (a), the densities of the bias  $\hat{A}^+ - A$  are shown for sample size  $T = 50$ . We compare the densities in the singular case ( $p = -1$ ) with two nonsingular cases ( $p = 0$  and  $p = -0.5$ ). The figure shows that the bias in the singular case is much smaller than the bias in the

---

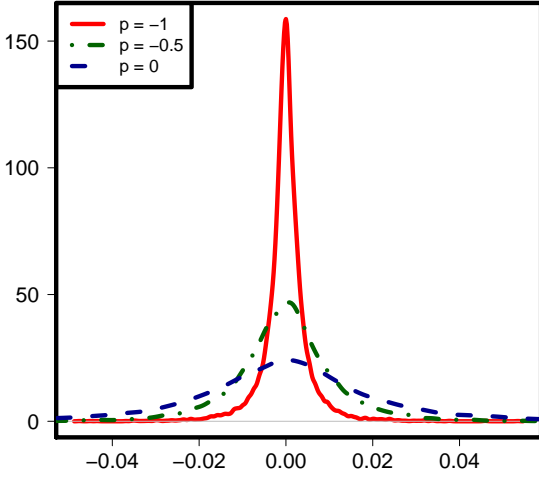
<sup>8</sup>Statistical computing in this paper uses R version 3.4.4.



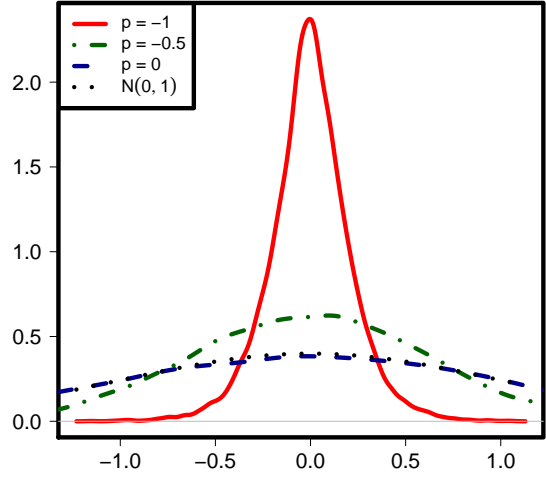
(a) Bias,  $T = 50$



(b)  $t$ -statistics,  $T = 50$



(c) Bias,  $T = 100$

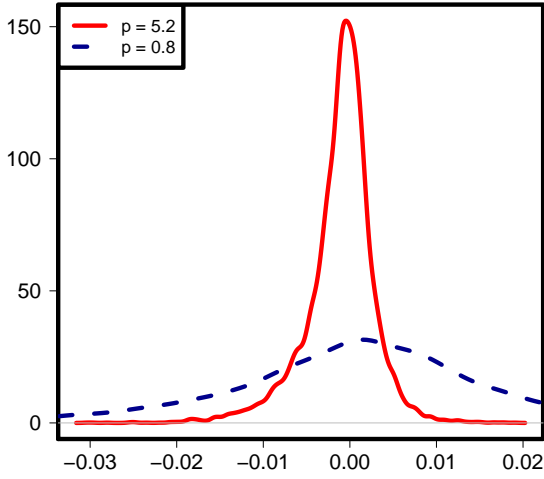


(d)  $t$ -statistics,  $T = 100$

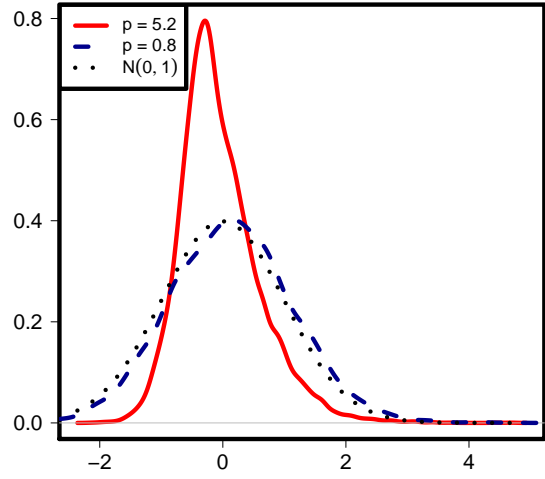
Figure 1: Kernel estimates of the densities of the FM-OLS estimation error  $\hat{A}^+ - A$  and  $t$ -statistic  $t_{\hat{A}^+}$  for sample sizes  $T = 50$  and  $T = 100$  and DGP1 with parameters  $p \in \{0.0, -0.5, -1.0\}$ . Multicointegration occurs if  $p = -1.0$ .

nonsingular cases. A more pronounced effect is observed for  $T = 100$  in Panel (c) confirming the higher convergence rates established for FM-OLS under singularity.

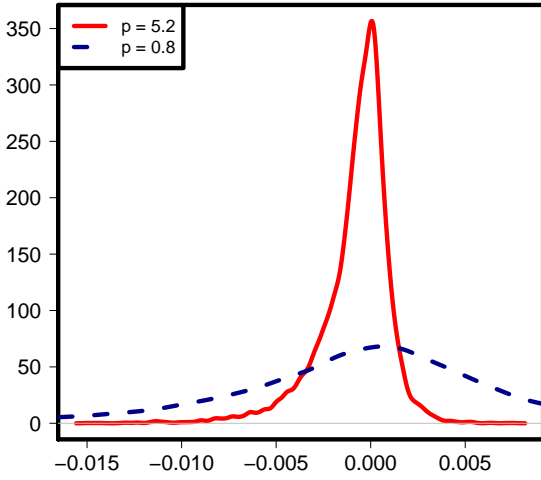
We use the  $t$ -statistic  $t_{\hat{A}^+} = (\hat{A}^+ - A^0) / \{\hat{\Omega}_{00.x} / \sum_{t=1}^T x_t^2\}^{1/2}$  for testing the hypothesis  $\mathcal{H}_0 : A = 2$ . In Figure 1, Panel (b) the densities of the  $t$ -statistic are shown for sample size  $T = 50$ . We compare the densities in the singular case ( $p = -1$ ) with two nonsingular cases ( $p = 0$  and  $p = -0.5$ ). Theory predicts that in nonsingular cases the test statistics is asymptotically standard normal, whose density is also plotted. This approximation is quite accurate for  $p = 0$ . However, the



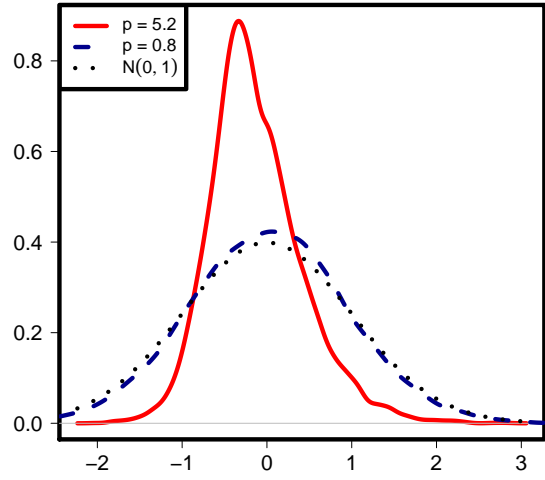
(a) Bias,  $T = 50$



(b)  $t$ -statistic,  $T = 50$



(c) Bias,  $T = 100$



(d)  $t$ -statistic,  $T = 100$

Figure 2: Kernel estimates of the densities of the FM-OLS estimation error  $\hat{A}^+ - A$  and  $t$ -statistic  $t_{\hat{A}^+}$  for sample sizes  $T = 50$  and  $T = 100$  and DGP2 with parameters  $p = 5.2$  and  $p = 0.8$ . Multicointegration occurs if  $p = 5.2$ .

density of the test statistic for  $p = -0.5$  has thinner tails, so that the test based on standard normal approximation is conservative as our theory predicts for singular case. Results for the sample size  $T = 100$  are plotted in Panel (d), and it is evident that the test statistic for  $p = -0.5$  still has thin tails.

Further simulation results are in Table 1. We vary sample sizes from 50 to 100 and the bandwidth is set to  $K = T^{1/4}$ . The bias is zero up to the 3d digit in all cases. When the sample size increases from  $T = 50$  to  $T = 100$  the precision of the FM-OLS measured by the standard deviation of the

Table 1: The mean and standard deviation of the bias and  $t$ -statistics, the rejection rates for the nominal 0.01, 0.05, and 0.10 levels of the  $t$ -statistics based on the FM-OLS and the mean and standard deviation of the bias of the OLS estimator are shown for DGP1 for various values of  $p$ , sample sizes and bandwidths. Multicointegration occurs when  $p = -1$ .

	$T$	$p$	$K$	Bias-OLS	SD-OLS	Bias	SD	t-Bias	t-SD	0.10	0.05	0.01
1	50	-1.0	2	0.0001	0.0238	0.0003	0.0175	0.0044	0.2598	0.000	0.000	0.000
2	100	-1.0	3	-0.0001	0.0088	-0.0000	0.0055	-0.0023	0.2124	0.000	0.000	0.000
3	50	-0.9	2	0.0000	0.0231	0.0003	0.0173	0.0047	0.2752	0.000	0.000	0.000
4	100	-0.9	3	-0.0001	0.0087	-0.0001	0.0057	-0.0031	0.2379	0.000	0.000	0.000
5	50	-0.8	2	-0.0000	0.0234	0.0002	0.0186	0.0050	0.3214	0.000	0.000	0.000
6	100	-0.8	3	-0.0001	0.0091	-0.0001	0.0068	-0.0041	0.3108	0.000	0.000	0.000
7	50	-0.7	2	-0.0001	0.0247	0.0001	0.0210	0.0053	0.3937	0.000	0.000	0.000
8	100	-0.7	3	-0.0001	0.0101	-0.0001	0.0085	-0.0050	0.4112	0.000	0.000	0.000
9	50	-0.6	2	-0.0001	0.0267	0.0001	0.0242	0.0056	0.4846	0.001	0.000	0.000
10	100	-0.6	3	-0.0001	0.0115	-0.0001	0.0105	-0.0059	0.5232	0.002	0.000	0.000
11	50	-0.5	2	-0.0002	0.0294	0.0000	0.0280	0.0060	0.5875	0.007	0.002	0.000
12	100	-0.5	3	-0.0001	0.0132	-0.0001	0.0126	-0.0068	0.6363	0.010	0.003	0.000
13	50	-0.4	2	-0.0003	0.0326	-0.0001	0.0321	0.0063	0.6967	0.019	0.007	0.000
14	100	-0.4	3	-0.0001	0.0151	-0.0001	0.0149	-0.0077	0.7433	0.027	0.009	0.001
15	50	-0.3	2	-0.0003	0.0361	-0.0001	0.0364	0.0066	0.8073	0.041	0.017	0.003
16	100	-0.3	3	-0.0001	0.0172	-0.0001	0.0171	-0.0084	0.8392	0.050	0.019	0.002
17	50	-0.2	2	-0.0004	0.0399	-0.0002	0.0409	0.0069	0.9146	0.071	0.031	0.007
18	100	-0.2	3	-0.0001	0.0192	-0.0001	0.0195	-0.0090	0.9215	0.075	0.034	0.006
19	50	-0.1	2	-0.0005	0.0439	-0.0003	0.0455	0.0071	1.0146	0.106	0.053	0.013
20	100	-0.1	3	-0.0001	0.0214	-0.0001	0.0218	-0.0094	0.9894	0.097	0.049	0.009
21	50	0.0	2	-0.0005	0.0481	-0.0003	0.0502	0.0073	1.1043	0.134	0.073	0.021
22	100	0.0	3	-0.0002	0.0236	-0.0001	0.0242	-0.0098	1.0438	0.116	0.061	0.014

Table 2: The mean and standard deviation of the bias and  $t$ -statistics, the rejection rates for the nominal 0.01, 0.05, and 0.10 levels of the  $t$ -statistics based on FM-OLS and the mean and standard deviation of the bias of the OLS estimator are shown for DGP2 for various values of  $p$ , sample sizes and bandwidths. Multicointegration occurs when  $p = 5.2$ .

	$T$	$p$	$K$	Bias-OLS	SD-OLS	Bias	SD	t-Bias	t-SD	0.10	0.05	0.01
1	50	0.8	3	0.0207	0.0223	0.0013	0.0184	0.1368	1.0091	0.102	0.054	0.014
2	50	5.2	3	0.0026	0.0054	-0.0011	0.0040	-0.0442	0.6474	0.018	0.009	0.003
3	50	0.8	5	0.0207	0.0223	0.0039	0.0189	0.2638	1.0806	0.135	0.078	0.023
4	50	5.2	5	0.0026	0.0054	0.0003	0.0029	0.1242	0.6440	0.026	0.014	0.004
5	50	0.8	7	0.0207	0.0223	0.0056	0.0197	0.3659	1.1444	0.161	0.100	0.035
6	50	5.2	7	0.0026	0.0054	0.0008	0.0029	0.2425	0.7146	0.044	0.025	0.008
7	50	0.8	10	0.0207	0.0223	0.0076	0.0206	0.4959	1.2321	0.195	0.130	0.055
8	50	5.2	10	0.0026	0.0054	0.0012	0.0031	0.3657	0.8392	0.074	0.045	0.018
9	100	0.8	3	0.0104	0.0111	-0.0003	0.0086	0.0146	0.9389	0.081	0.038	0.008
10	100	5.2	3	0.0013	0.0027	-0.0008	0.0019	-0.1256	0.5461	0.007	0.003	0.000
11	100	0.8	5	0.0104	0.0111	0.0007	0.0087	0.1036	0.9928	0.098	0.052	0.013
12	100	5.2	5	0.0013	0.0027	-0.0002	0.0011	0.0058	0.4907	0.007	0.003	0.000
13	100	0.8	7	0.0104	0.0111	0.0014	0.0089	0.1700	1.0314	0.114	0.060	0.017
14	100	5.2	7	0.0013	0.0027	0.0001	0.0010	0.0977	0.5234	0.013	0.005	0.001
15	100	0.8	10	0.0104	0.0111	0.0020	0.0092	0.2500	1.0780	0.134	0.078	0.024
16	100	5.2	10	0.0013	0.0027	0.0003	0.0010	0.1937	0.5980	0.024	0.012	0.003

bias term increases by  $0.0175/0.0055 = 3.18$  in the singular case and by  $0.0502/0.0242 = 2.07$  in nonsingular case  $p = 0$  corroborating the hyperconsistency of FM-OLS in the singular case and superconsistency in nonsingular case. When we compare FM-OLS with OLS, the former is more

precise in the singular and near-singular cases, while in the nonsingular case both estimators are comparable as removing second order bias effects that do not exist under DGP1 in OLS does not give an advantage to FM-OLS.

Interestingly, the rejection rates show that for  $T = 50$  the test is conservative in all cases except  $p = 0$  and  $p = -0.1$ . Even when the sample size is raised to the high level  $T = 10,000$  (unreported here) the test is still conservative for values of  $p = 0.5$  and below. Thus, the phenomenon described in this paper extends far beyond the pure singular case.

### Discussion of results for DGP2.

Since Phillips and Loretan (1991) numerous simulation studies have considered this DGP with different degrees of endogeneity, by varying values for  $p$  between  $-0.8$  and  $0.8$ . We consider  $p = 5.2$ , for which  $\Omega_{00.x} = 0$  and compare these results to the nearest nonsingular case with  $p = 0.8$ .

In Figure 2 we again observe thinner tails in the bias and  $t$ -statistic densities in the singular case compared to the nonsingular case, as well as thinner tails in the density of the  $t$ -statistic compared to the standard normal density already for  $T = 50$ . Unlike the situation with DGP1, second order biases are evident due to long run covariance between  $e_t$  and  $u_{xt}$  as predicted by theory and indicated in the proofs in the Appendix.

Table 2 confirms the above findings and allows analysis of the effect of bandwidth choice. Unlike previous simulations, the bandwidth is here fixed for  $K$  to a range between 3 and 10. This range includes many typical rules for bandwidth choice for the sample sizes considered. We see that test statistics for the singular case control size even for bandwidth choices for which the nonsingular case results in over-rejection ( $T = 50, K = 10$ ). In all other cases singularity in the model results in a conservative test for all bandwidth choices under consideration.

## 5 Evaluating Fiscal Sustainability

Soaring government debt in many countries calls for better economic understanding of fiscal sustainability for which improved methods of econometric analysis may be helpful given the presence of nonstationarity and endogeneities in the relevant data. Econometric analysis of sustainability has a long tradition, going back to early work by Hamilton and Flavin (1986) who suggested to test stationarity of the discounted debt. Hakkio and Rush (1991), Huag (1991), Trehan and Walsh (1991), and Quintos (1995) were among the first to test cointegration between revenues and expenditures. Quintos (1995) calls sustainability ‘strong’ when revenues and expenditures cointegrate with the explicit coefficients  $(1, -1)$  and tests the later using FM-OLS based  $t$ -statistics. A recent discussion of other approaches to evaluate fiscal sustainability is given in the chapter by D’Erasmus, Mendoza and Zhang (2016) in the Handbook of Macroeconomics.

Two remarks concerning the cointegration approach are relevant to our following analysis. First, cointegration between revenues and expenditures is only a sufficient condition for an intertemporal budget constraint (IBC) to hold and there are many other data generating processes consistent with IBC. This means that rejecting cointegration does not imply that IBC does not hold. Following



Bohn (2007), consider

$$B_t = B_{t-1} + G_t - R_t = G_t^0 - R_t + (1 + r_t)B_{t-1}, \quad \text{Budget Identity (BI),}$$

where  $B_t$  is government debt,  $R_t$  is government revenue,  $r_t$  is the interest rate, which is assumed to be stationary with mean  $r > 0$ ,  $G_t$  is government expenditure,  $G_t^0$  is government expenditure excluding interest on debt, and  $G_t^a = G_t^0 + (r_t - r)B_{t-1}$  is adjusted expenditure. These variables can be defined in nominal or real terms, possibly deflated by GDP or population. For example, Quintos (1995) constructed real variables by deflating nominal variables by the GNP price deflator and by population. BI implies

$$B_t = \frac{1}{1+r} \mathbb{E}_t G_t^0 + (1+r_t)B_{t-1}, \quad \text{Difference Equation (DE),}$$

which together with

$$B_t = \lim_{j \rightarrow \infty} \frac{1}{(1+r)^j} \mathbb{E}_t B_{t+j} = 0, \quad (m.s.), \quad \text{Transversality Condition (TC),}$$

where the limit is in the mean square sense, implies

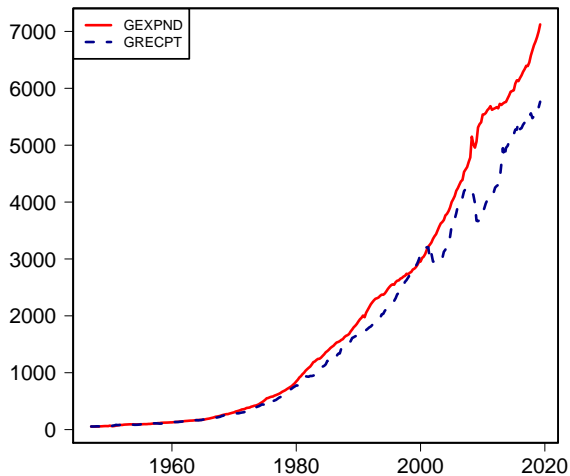
$$B_t = \sum_{j=1}^{\infty} \frac{1}{(1+r)^j} \mathbb{E}_t (R_{t+j} - G_{t+j}^a), \quad \text{Intertemporal Budget Constraint (IBC).}$$

IBC holds when the debt matches the expected present discounted value of the future surplus, a desirable requirement for sustainability. Bohn (2007) shows that if  $B_t \sim I(m)$  for some finite  $m \geq 0$ , then  $B_t$  satisfies TC and IBC holds. Therefore, the Quintos (1995) concept of strong sustainability, defined as  $B_t \sim I(1)$ , while intuitively appealing, is one of many possibilities of data generating processes satisfying IBC.

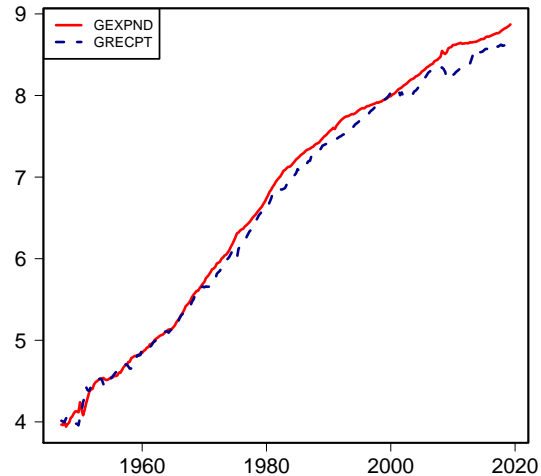
Second, there are economic considerations that restrict the DGP, besides IBC. For example, fiscal sustainability may involve bounds or restrictions on the deficit  $\Delta B_t$  that can be formulated as  $\Delta B_t \sim I(0)$ , which corresponds to strong sustainability by Quintos (1995), and  $G_t - R_t \sim I(0)$  if  $G_t, R_t \sim I(1)$ . Furthermore, there could be bounds on deviations of debt from revenue, that can be formulated as cointegration between  $B_t$  and  $R_t$ . In that case  $G_t$  and  $R_t$  are multicointegrated and the conditions for the asymptotic result in Phillips and Hansen (1990) employed in Quintos (1995) are not met. To allow for multicointegration, Berenguer-Rico and Carrion-i-Silvestre (2011) model the revenue-expenditure relationship in an  $I(2)$  VAR system, as suggested by Haldrup (1994) and Engsted et al (1997). The results of the present paper show that it is not necessary to work in an  $I(2)$  system and it is possible to go beyond a VAR specification. In particular: (i) multicointegration can be allowed directly in the  $I(1)$  system considered in Equation (6) in Quintos (1995); (ii) multicointegration invalidates the normal approximation of the test statistics  $t^+$  used in Section 3.1.2 in Quintos (1995); and (iii) multicointegration does not alter the conclusion that the null hypothesis of cointegration between  $G_t$  and  $R_t$  with coefficient  $(1, -1)$  is rejected. We explore these points and provide revised estimates and tests based on an updated dataset.

The data are provided by the US Bureau of Economic Analysis and retrieved from FRED, Federal Reserve Bank of St. Louis on November 17, 2019. We consider two series:  $x_t = \text{Government}$

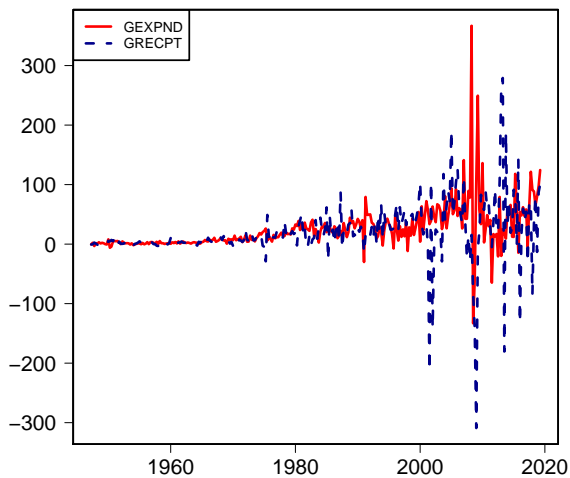
Current Expenditures (GEXPND), inclusive of interest payments, and  $y_t =$  Government Current Receipts (GRECPT). Both series are in billions of dollars, seasonally adjusted annual rate, at quarterly frequency from 1947:Q1 to 2019:Q1,  $T = 291$  observations.



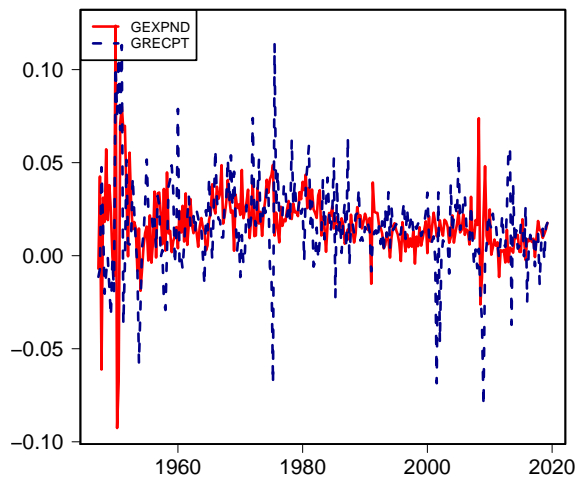
(a) Levels



(b) Logs



(c) First differences



(d) First differences of logs

Figure 3: US Government expenditures and receipts, billions of dollars, seasonally adjusted annual rate, quarterly frequency.

The series are plotted in Figure 3(a). We see that the series start to diverge in the mid 1990s and even more so after year 2000. We estimate the equation  $y_t = Ax_t + u_{0t}$  and test the null hypothesis of strong sustainability, viz.,  $\mathcal{H}_0 : A = 1$ . FM-OLS estimation of the full sample gives  $\hat{A}^+ = 0.83$  with standard error 0.01 and  $t$ -statistic  $(0.83 - 1)/0.01 = -17$ , rejecting the null hypothesis. The result is similar if we include the constant and for bandwidth  $T^{1/5}$  in place of  $3T^{1/5}$ .

The divergence of the series in mid 1990s in Figure 3(a) may signify a structural break in the relationship. In fact, several studies (e.g. Berenguer-Rico and Carrion-i-Silvestre, 2011) found a break in the 4th quarter of 1996, which could be attributed to the 1997 Clinton tax cut. The study of the properties of the FM-OLS under multicointegration in the presence of structural breaks we leave for future research. But we do estimate the model for the period from 1947:Q1 to 1996:Q4 ( $T = 200$ ) finding that  $\hat{A}^+ = 0.87$  with standard error 0.005 and  $t$ -statistic  $(0.87 - 1)/0.005 = -26$ , so the cointegrating coefficient is closer to but still statistically different from  $(1, -1)$ .

From the Campbell–Shiller work on log-linearization of present value identities we may expect that linear time series models provide better approximations in logarithms of the time series, which has the further advantage of stabilizing variances. The series in logs are plotted in Figure 3(b). We also plot the first differences in levels and in logs in Figure 3. The first differences of logs (Figure 3(d)) show less heteroskedasticity than first differences in levels (Figure 3(c)) so our theory results seem better suited for specification in logs<sup>9</sup>. FM-OLS estimation of the full sample in logs gives  $\hat{A}^+ = 0.98$  with standard error 0.002 and  $t$ -statistic  $(0.98 - 1)/0.002 = -10$ , rejecting the null hypothesis. The value of the  $t$ -statistic is similar for the period from 1947:Q1 to 1996:Q4.

We also estimate the cointegration relationship between real revenue and expenditure constructed using the GDP deflator. We take the same data series<sup>10</sup> as in Berenguer-Rico and Carrion-i-Silvestre (2011), but instead of looking at  $I(2)$  systems (which means working with  $\sum_{j=0}^t R_j$ ,  $\sum_{j=0}^t G_j$  and  $G_t$ ) we again run FM-OLS  $R_t$  on  $G_t$  and obtain  $\hat{A}^+ = 0.92$  with standard error 0.01 and  $t$ -statistic  $(0.92 - 1)/0.01 = -8$ , rejecting the null hypothesis that revenue and expenditure are cointegrated with coefficient  $(1, -1)$ .

## 6 Conclusion

In a semiparametric triangular representation of  $I(1)$  cointegrated time series the presence of multicointegration results in a singular long run error variance matrix which has decisive effects on standard methods of estimation and inference in such models. The consequences are higher rates of convergence and non pivotal limit theory in certain directions for estimators such as FM-OLS. Notwithstanding these effects, we show that FM-OLS Wald tests are invariant to singularity under well defined rank conditions and, when those conditions fail, the tests are conservative in certain cases. In particular, simulation experiments show that in such situations the test rejection rates under the null hypothesis are far below nominal levels based on standard asymptotics in singular and near singular cases. We illustrate our methods by analyzing the fiscal sustainability of the US government, testing the hypothesis that government revenue and expenditure are strongly cointegrated with coefficient  $(1, -1)$ , where multicointegration naturally arises if bounds are imposed on deviations of debt from revenue.

The results obtained here motivate the development of new robust approaches to estimating cointegrating relationships that allow for the possible presence of multicointegration and that are pivotal in the presence of the singularity it produces. This is an ongoing area of research by the authors.

---

<sup>9</sup>We thank a referee for this suggestion.

<sup>10</sup>Available at the *Journal of Applied Econometrics* Data Archive, <http://qed.econ.queensu.ca/jae/2011-v26.2/>

## 7 Acknowledgments

This paper has origins in a 2011 Yale Take Home Examination. For various and non-overlapping parts of this research Kheifets acknowledges support from the Russian Science Foundation under project 20-78-10113 (Monte Carlo simulations and the fiscal sustainability evaluation) and the Spanish Ministerio de Ciencia, Innovacion y Universidades under Grant ECO2017-86009-P (econometric theory). Phillips acknowledges support from the National Science Foundation under Grant SES 18-50860 and a Kelly Fellowship at the University of Auckland. This research was supported in part through computational resources of HPC facilities at HSE University.

## A Appendix

### A.1 Preliminary Lemmas

We start by stating and proving some results for the rates of convergence and limits of the kernel estimator of the long run variance when  $\Omega_{00.x} = 0$ . These results are of independent interest and are formulated as separate lemmas. The discussion proceeds and results are stated under the assumptions made in the body of the paper.

The results contribute to general literature on the asymptotic bias and variance of spectral estimates, see e.g. Section V in Hannan (1970). We use ideas from Lemma 8.1 (a), (b), and (g) in Phillips (1995), although that lemma does not directly apply to our case. In particular, the  $I(-1)$  errors that appear in Lemma 8.1 in Phillips (1995) arise from a different source: if the regressor vector  $x_t$  is cointegrated, but the cointegrating relationship is unknown, FM-OLS uses the first differences of the full vector  $x_t$  in making nonparametric adjustments to OLS, thereby producing linear combination of the first differences of stationary errors which are  $I(-1)$ . In our case it is assumed that  $\Omega_{xx}$  is positive definite, i.e. that  $x_t$  are full rank nonstationary  $I(1)$  and  $\Delta x_t$  are full rank stationary  $I(0)$ . Instead, multicointegration induces singularity in the augmented regression equation error  $u_{0.x,t} = u_{0t} - \Omega_{0x}\Omega_{xx}^{-1}u_{xt}$  in (3) so that  $u_{0.x,t}$  is  $I(-1)$  with consequential effects on the estimation of the long run covariance matrix  $\Omega_{00.x}$ .

Consider the case where  $\Omega_{00.x} = 0$ . Define

$$\begin{aligned} \check{\Delta}_{\Delta e,u} &= \sum_{j=0}^{T-1} w(j/K) \hat{\Gamma}_{\Delta e,u}(j), & \Delta_{\Delta e,u} &= \sum_{j=0}^{\infty} \Gamma_{\Delta e,u}(j), \\ \check{\Omega}_{\Delta e,u} &= \sum_{j=-T+1}^{T-1} w(j/K) \hat{\Gamma}_{\Delta e,u}(j), & \Omega_{\Delta e,u} &= \sum_{j=-\infty}^{\infty} \Gamma_{\Delta e,u}(j), \\ \text{and } \check{\Omega}_{\Delta e,\Delta e} &= \sum_{j=-T+1}^{T-1} w(j/K) \hat{\Gamma}_{\Delta e,\Delta e}(j), \end{aligned}$$

where  $w(\cdot)$  is a kernel function,  $K$  is a bandwidth parameter,  $\Gamma_{a,b}(j) = \mathbb{E} a_{t+j} b'_t$ ,  $\hat{\Gamma}_{a,b}(j) = T^{-1} \sum_{1 \leq t, t+j \leq T} a_{t+j} b'_t$ , and we use  $\check{\cdot}$  in place of  $\hat{\cdot}$  to stress that the kernel estimators are calculated with the true errors  $e_t$  and  $u_t$ .

**Lemma 1.** *Suppose  $\Omega_{00.x} = 0$ . Then*

$$(a) \quad \check{\Delta}_{\Delta e, u} + \widehat{\Gamma}_{e, u}(-1) = -K^{-2}w''(0) \sum_{j=0}^{\infty} (j+1/2) \Gamma_{e, u}(j) \\ + O_p\left((KT)^{-1/2}\right) + o_p(K^{-2}), \text{ and} \\ \check{\Delta}_{\Delta e, u} - \Delta_{\Delta e, u} = -K^{-2}w''(0) \sum_{j=0}^{\infty} (j+1/2) \Gamma_{e, u}(j) \\ + O_p(T^{-1/2}) + o_p(K^{-2}),$$

$$(b) \quad \check{\Omega}_{\Delta e, u} = -K^{-2}w''(0) \sum_{j=-\infty}^{\infty} (j+1/2) \Gamma_{e, u}(j) + O_p\left((KT)^{-1/2}\right) + o_p(K^{-2}),$$

$$(c) \quad \check{\Omega}_{\Delta e, \Delta e} = -K^{-2}w''(0) \sum_{j=-\infty}^{\infty} \Gamma_{e, e}(j) + o_p(K^{-2}).$$

Next consider the case when the kernel estimators are based on regression residuals  $(\widehat{u}'_{0t}, u'_{xt})$ , where  $\widehat{u}'_{0t} = y_t - \widehat{A}x_t = u_{0t} - (\widehat{A} - A)x_t$ , and using the true transform  $\Omega_{0x}\Omega_{xx}^{-1}$  define  $\widetilde{u}_{0.x, t} = \widehat{u}_{0t} - \Omega_{0x}\Omega_{xx}^{-1}u_{xt}$ , whereas  $\widetilde{u}_{0.x, t} = \widehat{u}_{0t} - \widehat{\Omega}_{0x}\widehat{\Omega}_{xx}^{-1}u_{xt}$ . Define  $\Delta_{0x}^+ = \sum_{j=0}^{\infty} \Gamma_{u_{0.x}, u_x}(j)$ ,

$$\begin{aligned} \widetilde{\Delta}_{0x}^+ &= \sum_{j=0}^{T-1} w(j/K) \widehat{\Gamma}_{\widetilde{u}_{0.x}, u_x}(j), & \check{\Delta}_{0x}^+ &= \sum_{j=0}^{T-1} w(j/K) \widehat{\Gamma}_{u_{0.x}, u_x}(j), \\ \widetilde{\Omega}_{0x}^+ &= \sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{\widetilde{u}_{0.x}, u_x}(j), & \check{\Omega}_{0x}^+ &= \sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{u_{0.x}, u_x}(j), \\ \widetilde{\Omega}_{00.x} &= \sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{\widetilde{u}_{0.x}, \widetilde{u}_{0.x}}(j), & \check{\Omega}_{00.x} &= \sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{u_{0.x}, u_{0.x}}(j), \end{aligned}$$

where we use  $\widetilde{\cdot}$  in place of  $\check{\cdot}$  to stress that the kernel estimators use residuals  $(\widetilde{u}'_{0.x, t}, u'_{xt})'$  instead of the true errors  $(u'_{0.x, t}, u'_{xt})'$ .

The following lemmas hold irrespective of the singularity of  $\Omega_{00.x}$ .

**Lemma 2.**  $\widetilde{\Delta}_{0x}^+ - \check{\Delta}_{0x}^+, \widetilde{\Omega}_{0x}^+ - \check{\Omega}_{0x}^+, \widetilde{\Omega}_{00.x} - \check{\Omega}_{00.x}$  are  $O_p(K/T)$ .

**Lemma 3.** *If  $\Omega_{00.x} = 0$ , then for  $u_{0.x, t} = -\Delta e_t$ ,*

$$(a) \quad T^{-1}U'_{0.x}X - \widetilde{\Delta}_{0x}^+ = T^{-1}e_Tx'_T + K^{-2}w''(0) \sum_{j=0}^{\infty} (j+1/2) \Gamma_{e, u_x}(j) \\ + O_p\left((KT)^{-1/2}\right) + O_p(K/T) + o_p(K^{-2}),$$

$$(b) \quad \widetilde{\Omega}_{0x}^+ = K^{-2}w''(0) \sum_{j=-\infty}^{\infty} (j+1/2) \Gamma_{e, u_x}(j) \\ + O_p\left((KT)^{-1/2}\right) + O_p(K/T) + o_p(K^{-2}),$$

$$(c) \quad \widetilde{\Omega}_{00.x} = -K^{-2}w''(0) \sum_{j=-\infty}^{\infty} \Gamma_{e, e}(j) + O_p(K/T) + o_p(K^{-2}).$$

## A.2 Proofs of Lemmas 1 - 3

*Proof of Lemma 1(a).* By Assumption K,  $w(0) = 1$  and  $w(x) = 0$  for  $x \geq 0$ , and for some  $K^* = K^b$  with  $0 < b < 1$

$$\check{\Delta}_{\Delta e, u} = \sum_{j=0}^{T-1} w(j/K) \widehat{\Gamma}_{\Delta e, u}(j),$$

$$\begin{aligned}
&= \sum_{j=0}^{K-1} w(j/K) \left( \widehat{\Gamma}_{e,u}(j) - \widehat{\Gamma}_{e,u}(j-1) \right) \\
&= \left( \sum_{j=0}^{K^*} + \sum_{j=K^*+1}^{K-2} \right) (w(j/K) - w((j+1)/K)) \widehat{\Gamma}_{e,u}(j) - \widehat{\Gamma}_{e,u}(-1) \\
&\quad + w((K-1)/K) \widehat{\Gamma}_{e,u}(K-1) = \sum_{k=1}^4 R_k, \text{ say.}
\end{aligned}$$

We show that  $R_1 + R_2 = -K^{-2}w''(0) \sum_{j=0}^{\infty} (j+1/2) \Gamma_{e,u}(j) + O_p\left((KT)^{-1/2}\right) + o_p(K^{-2})$ ,  $R_3 = -\widehat{\Gamma}_{e,u}(-1) = \Delta_{\Delta e,u} + O_p(T^{-1/2})$ , and  $R_4 = O_p(K^{-2}T^{-1/2})$ .

**Mean of  $R_1$ .** Applying the second order Taylor expansion of function  $w(\cdot)$  at arguments  $(j+1)/K$  around  $j/K$ ,

$$w((j+1)/K) - w(j/K) = K^{-1}w'(j/K) + 1/2 K^{-2}w''(j/K)[1 + o(1)],$$

and for  $j \leq K^*$  we can apply the Taylor expansion of function  $w'(\cdot)$  at arguments  $j/K$  around 0, where  $w'(0) = 0$ , giving  $w'(j/K) = w''(0)(j/K)[1 + o(1)]$ , and then  $w((j+1)/K) - w(j/K) = K^{-2}w''(0)(j+1/2)[1 + o(1)]$ . Hence,

$$K^2 R_1 = K^2 \sum_{j=0}^{K^*} [w(j/K) - w((j+1)/K)] \widehat{\Gamma}_{e,u}(j) = -w''(0) \sum_{j=0}^{K^*} (j+1/2) \widehat{\Gamma}_{e,u}(j) [1 + o(1)],$$

and

$$\mathbb{E} \sum_{j=0}^{K^*} (j+1/2) \widehat{\Gamma}_{e,u}(j) = \sum_{j=0}^{K^*} (j+1/2) (1-j/T) \Gamma_{e,u}(j) \rightarrow \sum_{j=0}^{\infty} (j+1/2) \Gamma_{e,u}(j).$$

**Mean of  $R_2$ .** By the mean value theorem there exists  $x_{j,K} \in (j/K, (j+1)/K)$ , such that  $R_2 = K^{-1} \sum_{j=K^*+1}^{K-2} w'(x_{j,K}) \widehat{\Gamma}_{e,u}(j)$  with mean

$$\mathbb{E} R_2 = K^{-1} \sum_{j=K^*+1}^{K-2} w'(x_{j,K}) (1-j/T) \Gamma_{e,u}(j).$$

whose modulus is dominated by

$$\begin{aligned}
&\sup_x |w'(x)| K^{-1} \sum_{j=K^*+1}^{K-2} \|\Gamma_{e,u}(j)\| \\
&\leq \text{const } K^{-1} \sum_{j>K^*} \sum_{s=0}^{\infty} \|D_s\| \|\widetilde{D}_{s+j}\| \\
&\leq \text{const } K^{-1} K^{*\nu} \sum_{j>K^*} \sum_{s=0}^{\infty} (s+j)^\nu \|D_s\| \|D_{s+j}\|
\end{aligned}$$

$$\leq \text{const } K^{-1}K^{-\nu b} \sum_{s=0}^{\infty} \|D_s\| \sum_{r=0}^{\infty} r^{\nu} \|D_r\| = O(K^{-1-\nu b}) = o(K^{-2}),$$

for  $1/\nu < b < 1$ .

**Variance of  $\mathbf{R}_1 + \mathbf{R}_2$ .**

$$R_1 + R_2 = \sum_{j=0}^{K-2} \left( w \left( \frac{j}{K} \right) - w \left( \frac{j+1}{K} \right) \right) \widehat{\Gamma}_{e,u}(j) = K^{-1} \sum_{j=0}^{K-2} w' \left( \frac{j}{K} \right) \widehat{\Gamma}_{e,u}(j) [1 + O(K^{-1})]$$

and  $\text{Var} \left[ \text{vec} \left[ \sum_{j=0}^{K-2} w'(j/K) \widehat{\Gamma}_{e,u}(j) \right] \right] = O(KT^{-1})$  from Theorem 9 in Hannan (1970, p. 280). So the variance of the dominant term in  $R_1 + R_2$  is  $O(K^{-1}T^{-1})$ .

**Term  $\mathbf{R}_3$ .** Note that  $\Delta_{\Delta e,u} = \sum_{j=0}^{\infty} \Gamma_{\Delta e,u}(j) = -\Gamma_{e,u}(-1)$ .  $\mathbb{E} R_3 = -\mathbb{E} \widehat{\Gamma}_{e,u}(-1) = -(1 - 1/T)\Gamma_{e,u}(-1) = (1 - 1/T)\Delta_{\Delta e,u}$ , and  $\text{Var}(\widehat{\Gamma}_{e,u}(-1)) = O(T^{-1})$ . Therefore,  $-\widehat{\Gamma}_{e,u}(-1) = \Delta_{\Delta e,u} + O(T^{-1/2})$ .

**Term  $\mathbf{R}_4$ .** By Assumption K (b),  $w((K-1)/K) = O(K^{-2})$  when  $K \rightarrow \infty$ . By summability,  $\mathbb{E} \widehat{\Gamma}_{e,u}(K-1) = (1 - (K-1)/T)\Gamma_{e,u}(K-1) = o(1)$  and  $\text{Var}(\widehat{\Gamma}_{e,u}(K-1)) = O(T^{-1})$ . Therefore,  $\widehat{\Gamma}_{e,u}(K-1) = O_p(T^{-1/2})$  and  $R_4 = O_p(K^{-2}T^{-1/2})$ .  $\square$

*Proof of Lemma 1(b).* As above,

$$\begin{aligned} \check{\Omega}_{\Delta e,u} &= \sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{\Delta e,u}(j), = \sum_{j=-K+1}^{K-1} w(j/K) \left( \widehat{\Gamma}_{e,u}(j) - \widehat{\Gamma}_{e,u}(j-1) \right) \\ &= \left( \sum_{j=-K^*}^{K^*} + \sum_{|j|=K^*+1}^{K-2} \right) (w(j/K) - w((j+1)/K)) \widehat{\Gamma}_{e,u}(j) \\ &\quad - w((-K+1)/K) \widehat{\Gamma}_{\Delta e,u}(-K) + w((K-1)/K) \widehat{\Gamma}_{\Delta e,u}(K-1) \\ &= \sum_{k=1}^4 R_k, \text{ say.} \end{aligned}$$

and  $R_1 + R_2 = K^{-2}w''(0) \sum_{j=-\infty}^{\infty} (j+1/2) \Gamma_{e,u}(j) + O_p((KT)^{-1/2}) + o_p(K^{-2})$  while  $R_k = O_p(K^{-2}T^{-1/2})$ , for  $k = 3, 4$  by the same argument as in (a).  $\square$

*Proof of Lemma 1(c).* Similar to (b), for some  $K^* = K^b$  with  $0 < b < 1$

$$\begin{aligned} \check{\Omega}_{\Delta e,\Delta e} &= \sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{\Delta e,\Delta e}(j), = \sum_{j=-K+1}^{K-1} w(j/K) \left( \widehat{\Gamma}_{e,\Delta e}(j) - \widehat{\Gamma}_{e,\Delta e}(j-1) \right) \\ &= \left( \sum_{j=-K^*}^{K^*} + \sum_{|j|=K^*+1}^{K-2} \right) (w(j/K) - w((j+1)/K)) \widehat{\Gamma}_{e,\Delta e}(j) \\ &\quad - w((-K+1)/K) \widehat{\Gamma}_{\Delta e,\Delta e}(-K) + w((K-1)/K) \widehat{\Gamma}_{\Delta e,\Delta e}(K-1) \\ &= \sum_{k=1}^4 R_k, \text{ say.} \end{aligned}$$

By the same argument as above,  $\mathbb{E} R_2 = o(K^{-2})$  and  $R_k = O_p(K^{-2}T^{-1/2})$ , for  $k = 3, 4$ , while we show below that the limit of  $R_1$  can be simplified and the variance bound of  $R_1 + R_2$  can be improved from  $O(K^{-1}T^{-1})$  in (a) and (b) to  $O(K^{-3}T^{-1})$  in the present case. Thus, the contribution from the variance is of smaller order than  $K^{-2}$  for any  $K = T^k$ , with  $k \in (0, 1)$ .

**Mean of  $\mathbf{R}_1$ .**

$$\begin{aligned} K^2 R_1 &= K^2 \sum_{|j| < K^*} [w(j/K) - w((j+1)/K)] \widehat{\Gamma}_{e, \Delta e}(j) \\ &= -w''(0) \sum_{|j| < K^*} (j+1/2) \widehat{\Gamma}_{e, \Delta e}(j) [1 + o(1)], \end{aligned}$$

and

$$\begin{aligned} \mathbb{E} \sum_{|j| < K^*} (j+1/2) \widehat{\Gamma}_{e, \Delta e}(j) &= \sum_{|j| < K^*} (j+1/2) (1 - j/T) \Gamma_{e, \Delta e}(j) \\ &\rightarrow \sum_{j=-\infty}^{\infty} (j+1/2) \Gamma_{e, \Delta e}(j) = \sum_{j=-\infty}^{\infty} \Gamma_{e, e}(j). \end{aligned}$$

because  $\sum_{j=-\infty}^{\infty} j \Gamma_{e, \Delta e}(j) = \sum_{j=-\infty}^{\infty} j \Gamma_{e, e}(j) - \sum_{j=-\infty}^{\infty} j \Gamma_{e, e}(j+1) = \sum_{j=-\infty}^{\infty} \Gamma_{e, e}(j)$  and  $\sum_{j=-\infty}^{\infty} \Gamma_{e, \Delta e}(j) = 0$ . Note that under stricter conditions on the bandwidth rate this limit could be obtained from Theorem 10 in Hannan (1970, p. 283).

**Variance of  $\mathbf{R}_1 + \mathbf{R}_2$ .**

$$\begin{aligned} R_1 + R_2 &= \sum_{j=-K+1}^{K-2} (w(j/K) - w((j+1)/K)) \widehat{\Gamma}_{e, \Delta e}(j) \\ &= -K^{-1} \sum_{j=-K+1}^{K-2} w'(j/K) \widehat{\Gamma}_{e, \Delta e}(j) [1 + O(K^{-1})] \\ &= -K^{-2} \sum_{j=-K+2}^{K-2} w''((j-1)/K) \widehat{\Gamma}_{e, e}(j) + o_p(K^{-2}) \end{aligned}$$

and  $\text{Var} \left[ \text{vec} \left[ \sum_{j=-K+2}^{K-2} w''((j-1)/K) \widehat{\Gamma}_{e, e}(j) \right] \right] = O(KT^{-1})$  from Theorem 9 in Hannan (1970, p.280). The variance of the dominant term in  $R_1 + R_2$  is therefore  $O(K^{-4}KT^{-1}) = O(K^{-3}T^{-1})$ .  $\square$

*Proof of Lemma 2.* Because  $\tilde{u}_{0,x,t} - u_{0,x,t} = \hat{u}_{0t} - u_{0t} = -(\hat{A} - A)x_t$

$$\begin{aligned} \tilde{\Delta}_{0x}^+ - \check{\Delta}_{0x}^+ &= -(\hat{A} - A) \sum_{j=0}^{T-1} w(j/K) \widehat{\Gamma}_{x, u_x}(j) = O_p(K/T), \\ \tilde{\Omega}_{0x}^+ - \check{\Omega}_{0x}^+ &= -(\hat{A} - A) \sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{x, u_x}(j) = O_p(K/T), \\ \tilde{\Omega}_{00.x} - \check{\Omega}_{00.x} &= (\hat{A} - A) \sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{x,x}(j) (\hat{A} - A)' \end{aligned}$$



$$\begin{aligned}
& - (\widehat{A} - A) \sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{x, u_{0,x}}(j) - \sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{u_{0,x}, x}(j) (\widehat{A} - A)' \\
& = O_p(K/T),
\end{aligned}$$

because  $\widehat{A} - A = O_p(T^{-1})$  and it follows from the proof of Theorem 3.1 of Phillips (1991b, pp. 432–433) that

$$\begin{aligned}
\sum_{j=0}^{T-1} w(j/K) \widehat{\Gamma}_{x, u_x}(j) &= O_p(K), & \sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{x, u_x}(j) &= O_p(K), \\
\sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{x, x}(j) &= O_p(KT), & \sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{u_{0,x}, x}(j) &= O_p(K).
\end{aligned}$$

If  $\Omega_{00,x} = 0$ , the last bound can be improved to  $O_p(1)$

$$\sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{u_{0,x}, x}(j) = \sum_{j=-T+1}^{T-1} w(j/K) \widehat{\Gamma}_{\Delta_{e,x}}(j) = O_p(1)$$

similar to the proof of Lemma 8.1(c) of Phillips (1995, pp. 1064–1065).  $\square$

*Proof of Lemma 3.* We apply Lemma 1 to the covariance between  $U_{0,x}$  and  $U_x$  and adjust the limits for the use of residual  $\widehat{U}_0$  in place of the true error  $U_0$  with an additional term of  $O_p(K/T)$  according to Lemma 2, and obtain parts (b) and (c). Part (a) follows by partial summation formula and applying the above procedure to the term in the brackets:

$$\begin{aligned}
T^{-1} U'_{0,x} X - \widetilde{\Delta}_{0x}^+ &= T^{-1} e_T x'_T - \left( \widetilde{\Delta}_{0x}^+ + \widehat{\Gamma}_{e, u_x}(-1) \right) \\
&= T^{-1} e_T x'_T + K^{-2} w''(0) \sum_{j=0}^{\infty} (j+1/2) \Gamma_{e, u_x}(j) \\
&\quad + O_p\left((KT)^{-1/2}\right) + O_p(K/T) + o_p(K^{-2}).
\end{aligned}$$

$\square$

### A.3 Proofs of Propositions

*Proof of Proposition 1.* We can write  $(y'_t, x'_t)' = (1-L)^{-1} C(L) \eta_t$ , where the roots of  $|C(z)| = 0$  satisfy  $|z| > 1$  or  $z = 1$ . Multicointegration of such a linear  $I(1)$  process occurs (see Johansen 1992, Engsted and Johansen, 1999) when  $z = 1$  is a root, so that  $C(1) = \xi \epsilon'$  has reduced rank and  $\xi'_\perp \dot{C}(1) \epsilon_\perp$  is singular. The submatrices  $\{\xi, \epsilon\}$  and their orthogonal complements  $\{\xi_\perp, \epsilon_\perp\}$  are given explicitly below for the present context of the triangular system (1) and (2). In particular, write  $C(L)$  as

$$C(L) = \begin{bmatrix} (1-L)I_{m_0} & A \\ 0 & I_{m_x} \end{bmatrix} D(L),$$

and its derivative

$$\dot{C}(L) = \begin{bmatrix} -I_{m_0} & 0 \\ 0 & 0 \end{bmatrix} D(L) + \begin{bmatrix} (1-L)I_{m_0} & A \\ 0 & I_{m_x} \end{bmatrix} \dot{D}(L). \quad (7)$$

We can write

$$C(1) = \begin{bmatrix} 0 & A \\ 0 & I_{m_x} \end{bmatrix} D(1) = \begin{bmatrix} A \\ I_{m_x} \end{bmatrix} [D_{x0}(1), D_{xx}(1)] = \xi \epsilon', \quad \text{taking } \xi = \begin{bmatrix} A \\ I_{m_x} \end{bmatrix},$$

$\xi'_\perp = [I_{m_x}, -A]$  and  $\epsilon' = [D_{x0}(1), D_{xx}(1)]$ , which is an  $m_x \times m$  matrix of full rank  $m_x$  whenever  $\epsilon' \epsilon = \Omega_{xx} > 0$ . Multiplication of  $D(1)$  by the full rank  $m \times m$  matrix  $\epsilon^F = [\epsilon_\perp, \epsilon]$  yields

$$\text{rank}(D(1)) = \text{rank}(D(1)\epsilon^F) = \text{rank}(\xi'_\perp \dot{C}(1)\epsilon_\perp) + \text{rank}(\Omega_{xx}),$$

because

$$D(1)\epsilon^F = \begin{bmatrix} [D_{00}(1), D_{0x}(1)]\epsilon_\perp & [D_{00}(1), D_{0x}(1)]\epsilon \\ 0 & \Omega_{xx} \end{bmatrix},$$

and from (7)

$$\xi'_\perp \dot{C}(1)\epsilon_\perp = \begin{bmatrix} -I_{m_0} & 0 \end{bmatrix} D(1)\epsilon_\perp = -[D_{00}(1), D_{0x}(1)]\epsilon_\perp,$$

since

$$\xi'_\perp \begin{bmatrix} -I_{m_0} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -I_{m_0} & 0 \end{bmatrix} \quad \text{and} \quad \xi'_\perp \begin{bmatrix} 0 & A \\ 0 & I_{m_x} \end{bmatrix} = 0.$$

It follows that when  $\text{rank}(\Omega_{xx}) = m_x$ , we have the rank equivalence  $m_0 - \text{rank}(\xi'_\perp \dot{C}(1)\epsilon_\perp) = m - \text{rank}(\Omega) = m_0 - \text{rank}(\Omega_{00.x})$ , which is the multicointegrating rank in the system defined by (1) and (2). More simply,  $\text{rank}(\xi'_\perp \dot{C}(1)\epsilon_\perp) = \text{rank}(\Omega_{00.x})$ .  $\square$

*Proof of Proposition 2.* Recall that  $\tilde{\Omega}_{0x}^+ = \hat{\Omega}_{0x} - \Omega_{0x}\Omega_{xx}^{-1}\hat{\Omega}_{xx}$ , so

$$\hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1} - \Omega_{0x}\Omega_{xx}^{-1} = \left(\hat{\Omega}_{0x} - \Omega_{0x}\Omega_{xx}^{-1}\hat{\Omega}_{xx}\right)\hat{\Omega}_{xx}^{-1} = \tilde{\Omega}_{0x}^+\hat{\Omega}_{xx}^{-1}.$$

Also,  $\tilde{\Delta}_{0x}^+ = \hat{\Delta}_{0x} - \Omega_{0x}\Omega_{xx}^{-1}\hat{\Delta}_{xx}$ , so

$$\hat{\Delta}_{0x}^+ = \tilde{\Delta}_{0x}^+ - \left(\hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1} - \Omega_{0x}\Omega_{xx}^{-1}\right)\hat{\Delta}_{xx} = \tilde{\Delta}_{0x}^+ - \tilde{\Omega}_{0x}^+\hat{\Omega}_{xx}^{-1}\hat{\Delta}_{xx}.$$

The numerator matrix of FM-OLS is

$$\begin{aligned} \hat{Y}^{+'}X - T\hat{\Delta}_{0x}^+ &= \left(Y' - \hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}U'_x\right)X - T\hat{\Delta}_{0x}^+ \\ &= AX'X + U'_0X - \hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}U'_xX - T\hat{\Delta}_{0x}^+ \\ &= AX'X + U'_{0.x}X - T\tilde{\Delta}_{0x}^+ - \left(\hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1} - \Omega_{0x}\Omega_{xx}^{-1}\right)\left(U'_xX - T\hat{\Delta}_{xx}\right) \\ &= AX'X + U'_{0.x}X - T\tilde{\Delta}_{0x}^+ - \tilde{\Omega}_{0x}^+\hat{\Omega}_{xx}^{-1}\left(U'_xX - T\hat{\Delta}_{xx}\right). \end{aligned}$$

Therefore,

$$T \left( \widehat{A}^+ - A \right) = \left( T^{-1} U'_{0.x} X - \widetilde{\Delta}_{0x}^+ \right) \left( T^{-2} X' X \right)^{-1} \\ - \widetilde{\Omega}_{0x}^+ \widehat{\Omega}_{xx}^{-1} \left( T^{-1} U'_x X - \widehat{\Delta}_{xx} \right) \left( T^{-2} X' X \right)^{-1}. \quad (8)$$

From standard weak convergence theory for sample covariances in Phillips and Durlauf (1986) and Phillips (1989)

$$T^{-2} X' X \rightarrow_d \int_0^1 B_x B'_x, \quad (9)$$

$$T^{-1} U'_x X \rightarrow_d \int_0^1 dB_x B'_x + \Delta_{xx}, \quad (10)$$

$$T^{-1} U'_0 X \rightarrow_d \int_0^1 dB_0 B'_x + \Delta_{0x}, \quad (11)$$

giving

$$T^{-1} U'_{0.x} X \rightarrow_d \int_0^1 dB_{0.x} B'_x + \Delta_{0x}^+.$$

By construction,  $u_{0.x,t}$  has zero long run covariance with the errors  $u_{xt}$  that drive the nonstationary component  $x_t$ , thereby eliminating the endogeneity from  $x_t$  in the long run. Therefore, with any consistent estimators of  $\Omega$  and  $\Delta$ , under Assumption K with  $0 < k < 1$ , and using Lemma 2, we have  $\widetilde{\Delta}_{0x}^+ \rightarrow_p \Delta_{0x}^+$  and  $\widetilde{\Omega}_{0x}^+ \rightarrow_p 0$ , so that by continuous mapping and joint convergence of the components we have

$$T \left( \widehat{A}^+ - A \right) \rightarrow_d \left( \int_0^1 dB_{0.x} B'_x \right) \left( \int_0^1 B_x B'_x \right)^{-1}.$$

□

*Proof of Proposition 3.* The proof follows from expansion (8) by using the consistency of the kernel estimates, the convergences (9) and (10), and the rates of convergence established in Lemma 3(a) and (b). □

*Proof of Proposition 4.* For  $k < 1/4$ ,  $O_p(K^{-2})$  dominates  $O_p(K^{-2}) + O_p((KT)^{-1/2}) + O_p(K/T)$  and the limits in Lemma 3(a) and (b) have the form

$$T^{-1} U'_{0.x} X - \widetilde{\Delta}_{0x}^+ = K^{-2} w''(0) \sum_{j=0}^{\infty} (j+1/2) \Gamma_{e,u_x}(j) + o_p(K^{-2}), \\ \widetilde{\Omega}_{0x}^+ = K^{-2} w''(0) \sum_{j=-\infty}^{\infty} (j+1/2) \Gamma_{e,u_x}(j) + o_p(K^{-2}),$$

which together with (9) and (10) and expansion (8) give the limit distribution of  $K^2 T \left( \widehat{A}^+ - A \right)$ . □

*Proof of Proposition 5. Part (a).* The rate of convergence and limit behavior of  $\widetilde{\Omega}_{00.x}$  is established

in Lemma 3(c), while the difference  $\widehat{\Omega}_{00.x} - \widetilde{\Omega}_{00.x}$  is of smaller order and can be neglected, viz.,

$$\begin{aligned}\widehat{\Omega}_{00.x} - \widetilde{\Omega}_{00.x} &= - \left( \widehat{\Omega}_{0x} \widehat{\Omega}_{xx}^{-1} - \Omega_{0x} \Omega_{xx}^{-1} \right) \widehat{\Omega}_{x0} - \left[ \left( \widehat{\Omega}_{0x} \widehat{\Omega}_{xx}^{-1} - \Omega_{0x} \Omega_{xx}^{-1} \right) \widehat{\Omega}_{x0} \right]' \\ &+ \widehat{\Omega}_{0x} \left( \widehat{\Omega}_{0x} \widehat{\Omega}_{xx}^{-1} - \Omega_{0x} \Omega_{xx}^{-1} \right)' + \left( \widehat{\Omega}_{0x} \widehat{\Omega}_{xx}^{-1} - \Omega_{0x} \Omega_{xx}^{-1} \right) \widehat{\Omega}_{xx}^{-1} \Omega_{xx}^{-1} \Omega'_{0x} \\ &= \left( \widehat{\Omega}_{0x} \widehat{\Omega}_{xx}^{-1} - \Omega_{0x} \Omega_{xx}^{-1} \right) \left( -\widehat{\Omega}_{0x} + \widehat{\Omega}_{xx}^{-1} \Omega_{xx}^{-1} \Omega'_{0x} \right) = -\widetilde{\Omega}_{0x}^+ \widehat{\Omega}_{xx}^{-1} \widetilde{\Omega}_{0x}^+.\end{aligned}$$

By comparing the rates in Part (b) and Part (c) in Lemma 3 we see that the square of the decay rate of  $\widetilde{\Omega}_{0x}^+$  exceeds that of  $\widetilde{\Omega}_{00.x}$ , giving the required result.

*Part (b).* Because of the rates of the FM-OLS estimator established in Proposition 3 and the rate of  $\widehat{\Omega}_{00.x}$  in Part (a), under the null hypothesis we have for  $k < 1/3$

$$\begin{aligned}W_I &= T^{2k} T^2 \delta(T)^{-2} \text{tr} \left\{ (T^{-2} X' X) \delta(T) (\widehat{A}^+ - A^0)' (T^{2k} \widehat{\Omega}_{00.x})^{-1} \delta(T) (\widehat{A}^+ - A^0) \right\} \\ &= O_p \left( T^{2+2k} \min(T^{1+2k}, T^{3/2})^{-2} \right) = O_p \left( T^{-2k} \mathbf{1}_{0 < k \leq 1/4} + T^{2k-1} \mathbf{1}_{1/4 < k < 1/3} \right) = o_p(1),\end{aligned}$$

where  $\mathbf{1}_A$  is the indicator of  $A$ . □

## A.4 Proofs of the Theorems

*Proof of Theorem 1.* We have

$$\begin{bmatrix} B_{f.x} \\ 0 \\ B_x \end{bmatrix} = \mathcal{L}_R \begin{bmatrix} B_{0.x} \\ B_x \end{bmatrix} \equiv BM(\mathcal{L}_R \mathcal{L}_\Omega \Omega \mathcal{L}'_\Omega \mathcal{L}'_R) = BM \left( \begin{bmatrix} \Omega_{ff.x} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Omega_{xx} \end{bmatrix} \right),$$

where

$$\mathcal{L}_R = \begin{bmatrix} R' & 0 \\ R'_\perp & 0 \\ 0 & I_{m_x} \end{bmatrix}.$$

The matrix  $(R, R_\perp)$  rotates  $u_{0.x,t}$  to  $(u'_{f.x,t}, u'_{s.x,t})'$ , where  $u_{f.x,t} = R' u_{0.x,t}$  is  $I(0)$  and  $u_{s.x,t} = R'_\perp u_{0.x,t}$  is  $I(-1)$ . Therefore  $\mathcal{L}_R \mathcal{L}_\Omega$  keeps the nonstationary regressors  $x_t$  and transforms the original cointegration relationship  $y_t = Ax_t + u_{0t}$  to a system of two equations with orthogonal long run errors: (i) an equation with  $I(0)$  errors that has a nonsingular long run variance matrix  $\Omega_{ff.x}$ , and for this equation,  $R' y_t^+ = R' Ax_t + u_{f.x,t}$ , Proposition 2 applies; and (ii) an equation with  $I(-1)$  errors  $R'_\perp y_t^+ = R'_\perp Ax_t + u_{s.x,t}$ , for which Proposition 3 applies. □

*Proof of Theorem 2.* Using coordinate rotation, we write the Wald statistic as a sum of several components, corresponding to the nondegenerate and degenerate directions and their cross products. Recall the partitioned matrix inversion formula

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11.2}^{-1} & -A_{11}^{-1} A_{12} A_{22.1}^{-1} \\ -A_{22.1}^{-1} A_{21} A_{11}^{-1} & A_{22.1}^{-1} \end{bmatrix},$$

where the Shur complement is defined as  $A_{ii.j} = (A_{ii} - A_{ij}A_{jj}^{-1}A_{ji})$ . We apply the above formula to the variance matrix metric in the Wald test statistic giving

$$\begin{aligned} \left( (R, R_{\perp})' \widehat{\Omega}_{00.x} (R, R_{\perp}) \right)^{-1} &= \begin{bmatrix} R' \widehat{\Omega}_{00.x} R & R' \widehat{\Omega}_{00.x} R_{\perp} \\ R'_{\perp} \widehat{\Omega}_{00.x} R & R'_{\perp} \widehat{\Omega}_{00.x} R_{\perp} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \Omega_{ff.x}^{-1} + o_p(1) & O_p(1) \\ O_p(1) & -K^2 \Omega_{e,e}^{-1} / w''(0) + o_p(K^2) \end{bmatrix}, \end{aligned}$$

where we take into account that

- (i)  $R' \widehat{\Omega}_{00.x} R \rightarrow_p \Omega_{ff.x}$ , because  $R$  isolates the nondegenerate direction;
- (ii)  $R'_{\perp} \widehat{\Omega}_{00.x} R = O_p(K^{-2})$ , which is obtained similar to the proof of Lemma 1(b), and switching from residuals  $\widehat{u}_t$  to errors  $u_t$  involves differences of order  $O_p(K/T)$ , as in Lemma 2, which are of smaller order for  $k < 1/3$ ;
- (iii)  $K^2 R'_{\perp} \widehat{\Omega}_{00.x} R_{\perp} \rightarrow_p -w''(0) \Omega_{e,e}$ , which can be obtained similar to the proofs of Lemma 1(c) and Proposition 5(a).

Then

$$\begin{aligned} W_I &= tr \left\{ (T^{-2} X' X) (\widehat{A}^+ - A^0)' (TR, TR_{\perp}) \left( (R, R_{\perp})' \widehat{\Omega}_{00.x} (R, R_{\perp}) \right)^{-1} (TR, TR_{\perp})' (\widehat{A}^+ - A^0) \right\} \\ &= tr \left\{ (T^{-2} X' X) (\widehat{A}^+ - A^0)' TR \left( R' \widehat{\Omega}_{00.x} R \right)^{-1} TR' (\widehat{A}^+ - A^0) \right\} \\ &\quad + T^{2+2k} \delta(T)^{-2} tr \left\{ (T^{-2} X' X) \delta(T) (\widehat{A}^+ - A^0)' R_{\perp} \times O_p(1) \times \delta(T) R'_{\perp} (\widehat{A}^+ - A^0) \right\} \\ &\quad + T \delta(T)^{-1} tr \left\{ (T^{-2} X' X) \delta(T) (\widehat{A}^+ - A^0)' R_{\perp} \times O_p(1) \times TR' (\widehat{A}^+ - A^0) \right\} \\ &\quad + T \delta(T)^{-1} tr \left\{ (T^{-2} X' X) T (\widehat{A}^+ - A^0)' R \times O_p(1) \times \delta(T) R'_{\perp} (\widehat{A}^+ - A^0) \right\} \\ &= \chi_{rm_x}^2 + O_p \left( T^{2+2k} \delta(T)^{-2} \right) + O_p \left( T \delta(T)^{-1} \right), \end{aligned}$$

using the fact that  $TR' (\widehat{A}^+ - A^0) = O_p(1)$  and  $\delta(T) R'_{\perp} (\widehat{A}^+ - A^0) = O_p(1)$  from Theorem 1. For  $k < 1/3$ , we have  $O_p \left( T^{2+2k} \delta(T)^{-2} \right) = o_p(1)$ , and therefore  $W_I \rightarrow_d \chi_{rm_x}^2$ .  $\square$

## References

- [1] Berenguer-Rico, V. and J.L. Carrion-i-Silvestre (2011) Regime shifts in stock-flow  $I(2)$ - $I(1)$  systems: the case of US fiscal sustainability, *Journal of Applied Econometrics* 26, 298–321.
- [2] Bohn, H. (2007) Are stationarity and cointegration restrictions really necessary for the intertemporal budget constraint? *Journal of Monetary Economics* 54, 1837–1847.
- [3] Boswijk, H.P. (2000) Mixed normality and ancillarity in  $I(2)$  systems. *Econometric Theory* 16, 878–904.

- [4] Boswijk, H.P. (2010) Mixed normal inference on multicointegration, *Econometric Theory* 26, 1565–1576.
- [5] D’Erasmus, P.D., Mendoza, E.G. and J. Zhang (2016) What is a Sustainable Public Debt? *Handbook of Macroeconomics* 2B, 2493–2597.
- [6] Engsted, T., Gonzalo, J. and N. Haldrup (1997) Testing for multicointegration. *Economics Letters*, 56, 259–266.
- [7] Engsted, T. and S. Johansen (1999) Granger’s representation theorem and multicointegration. In R.F. Engle and H. White (eds.), *Cointegration, Causality and Forecasting: Festschrift in Honour of Clive Granger*, 200–212. Oxford University Press.
- [8] Franchi, M. and P. Paruolo (2019) A general inversion theorem for cointegration, *Econometric Reviews*, 38, 1176–1201.
- [9] Granger, C.W.J. and T.-H. Lee (1990) Multicointegration. In G.F. Rhodes Jr. and T.B. Fomby (eds.), *Advances in Econometrics: Cointegration, Spurious Regressions and Unit Roots*. 71–84. JAI Press.
- [10] Hakkio, C. S. and M. Rush (1991) Is the budget deficit too large? *Economic Inquiry* 29, 429–445.
- [11] Haldrup, N. (1994) The asymptotics of single-equation cointegration regressions with  $I(1)$  and  $I(2)$  variables. *Journal of Econometrics* 63, 151–181.
- [12] Hamilton, J. D. and M. A. Flavin (1986) On the limitations of government borrowing: a framework for empirical testing. *American Economic Review* 76, 808–816.
- [13] Hannan, E. J. (1970) *Multiple Time Series*. New York: John Wiley & Sons.
- [14] Haug, A. A. (1991) Cointegration and government borrowing constraints: evidence for the United States. *Journal of Business and Economic Statistics* 9, 97–101.
- [15] Johansen, S. (1992) A representation of vector autoregressive processes integrated of order 2. *Econometric Theory* 8, 188–202.
- [16] Johansen, S. (1997) Likelihood analysis of the  $I(2)$  model. *Scandinavian Journal of Statistics* 24, 433–462.
- [17] Johansen, S. (2006) Statistical analysis of hypotheses on the cointegrating relations in the  $I(2)$  model. *Journal of Econometrics* 132, 81–115.
- [18] Park, J. Y. (1992) Canonical Cointegrating Regressions, *Econometrica* 60, 119–143.
- [19] Paruolo, P. (2000) Asymptotic efficiency of the two stage estimator in  $I(2)$  systems. *Econometric Theory* 16, 524–550.
- [20] Phillips, P. C. B. (1989) Partially Identified Econometric Models, *Econometric Theory* 5, 181–240.

- [21] Phillips, P. C. B. (1991) Optimal Inference in Cointegrated Systems, *Econometrica* 59, 283-306.
- [22] Phillips, P. C. B. (1991b) Spectral Regression for Cointegrated Time Series, in *Nonparametric and Semiparametric Methods in Economics and Statistics*, ed. by W. Barnett, J. Powell, and G. Tauchen. New York: Cambridge University Press.
- [23] Phillips, P. C. B. (1995) Fully Modified Least Squares and Vector Autoregression, *Econometrica* 63, 1023–1078.
- [24] Phillips, P.C.B. (2014) Optimal estimation of cointegrated systems with irrelevant instruments. *Journal of Econometrics* 178, 210–224.
- [25] Phillips, P. C. B. (2016) Inference in near-singular regression. *Advances in Econometrics* 36, 461–486.
- [26] Phillips, P. C. B. and S.N. Durlauf (1986) Multiple Time Series Regression with Integrated Processes, *Review of Economic Studies* 53, 473–495.
- [27] 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.
- [28] Phillips, P. C. B. and I. L. Kheifets (2019) On Multicointegration. *Working Paper, Yale University*. Presented at the Rimini Time Series Conference, Larnica, Cyprus, June 2019.
- [29] Phillips, P. C. B. and M. Loretan (1991) Estimating Long-Run Economic Equilibria. *The Review of Economic Studies* 58, 407–436.
- [30] Phillips, P. C. B. and V. Solo (1992) Asymptotics for Linear Processes. *Annals of Statistics* 20, 971–1001.
- [31] Priestley, M.B. (1981) Spectral Analysis and Time Series. Academic Press Inc, New York.
- [32] Quintos, C.E. (1995) Sustainability of the deficit process with structural shifts. *Journal of Business and Economic Statistics* 13, 409–417.
- [33] Toda, H. Y. and P.C.B. Phillips (1993) Vector Autoregressions and Causality *Econometrica* 61, 1367–1393.
- [34] Trehan, B., and C. Walsh (1991) Testing intertemporal budget constraints: theory and applications to U.S. Federal budget and current account deficits. *Journal of Money, Credit and Banking* 23, 210–223.