FULLY MODIFIED LEAST SQUARES FOR MULTICOINTEGRATED SYSTEMS

By

Igor L. Kheifets and Peter C. B. Phillips

November 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 2210

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Fully Modified Least Squares for Multicointegrated Systems*

Igor L. Kheifets[a] and Peter C. B. Phillips[b,c,d,e]

[a]ITAM, Mexico City [b]Yale University, [c]University of Auckland,

[d]University of Southampton & [e]Singapore Management University

November 24, 2019

Link to the latest version

Abstract

Multicointegration is traditionally defined as a particular long run relationship among variables in a parametric vector autoregressive model that introduces 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. We illustrate our findings by analyzing 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

*This paper has origins in a 2011 Yale Take Home Examination. Kheifets acknowledges support from the Spanish Ministerio de Ciencia, Innovacion y Universidades under Grant ECO2017-86009-P. Phillips acknowledges support from the NSF under Grant SES 18-50860 and a Kelly Fellowship at the University of Auckland

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 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 defined only in a VAR framework and naturally involves implicit restrictions on the model induced by the extra layer of cointegration. Engsted and Johansen (1997), 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 $k = 1$.

This paper studies cointegrated-multicointegrated models 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 in a general way as a property of the system, as opposed to imposing multicointegration subsequently on a parametric system like a VAR. The present paper contributes by developing asymptotic theory for FM-OLS estimation and testing in cointegrating relationships that involve multicointegration. The analysis of triangular cointegrated systems under singularity that is developed is of some independent interest.

To define multicointegration for weakly dependent data, we take the triangular representation of a linear cointegrating relationship. In the cointegrating regression model

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

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

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

for some $\nu > 2$, finite fourth order cumulants of η_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 $D(L)$, long run variance 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 variance matrix $\Delta = \sum_{h=0}^{\infty} \Gamma_{u,u}(h) = \sum_{k=0}^{\infty} \sum_{j=0}^{k} D_j D'_k$ of u_t are partitioned conformably with u_t as

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

where $\Omega_{xx} > 0$ is positive definite. The conditional long run covariance matrix, defined as the Schur complement of the block Ω_{xx}, $\Omega_{00.x} = \Omega_{00} - \Omega_{0x}\Omega_{xx}^{-1}\Omega_{x0}$, is positive (semi-) definite if and only if Ω 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 spirit of Granger and Lee (1990), but is semiparametric in the sense that the short run dynamics is 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 covariance is singular.*

The advantage of our framework is that it provides the explicit relationship from which the multicointegration arises. Thus, if we take partial sums of the augmented regression form

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

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 for partial sums)

$$Y_t = AX_t + Fx_t + U_{0.x,t}.$$

It becomes clear that in the direction of singularity of $\Omega_{00.x}$ we have an exact long run relationship that links Y_t, X_t, and x_t and this is known 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 simply imposed a priori, as in the Granger and Lee (1990) paper. What our approach does is to reveal the leading role that the singularity of the long run conditional covariance matrix $\Omega_{00.x}$ plays in determining multicointegration. And this is a nonparametric formulation.

Engsted and Johansen (1997) show that multicointegration as defined in Engle and Lee (1990) of a linear $I(1)$ process $(y_t', x_t')' = (1-L)^{-1}C(L)\eta_t$, where the roots of $|C(z)| = 0$ satisfy $|z| > 1$ or $z = 1$, occurs when $z = 1$ is a root, so that $C(1) = \xi\epsilon'$ has reduced rank, and $\xi_\perp'\dot{C}(1)\epsilon_\perp$ is singular, see also Johansen (1992). This is exactly the case when Ω is singular, as shown below.

Proposition 1. *A linear process $(y_t', x_t')'$ is multicointegrated, i.e Ω is singular, if and only if it satisfies the multicointegration condition of Engsted and Jo-*

hansen (1997). The rank of the multicointegrating relation equals $m - rank(\Omega)$.

Data matrices are denoted by upper case letters without indexes, e.g., $Y' = [y_1, \ldots, y_T]$. The OLS estimator $\widehat{A} = Y'X\left(X'X\right)^{-1}$ is consistent at the rate at least $O(T)$. The FM-OLS estimator (Phillips and Hansen 1990) has the form $\widehat{A}^{+} = \left(\widehat{Y}^{+\prime}X - T\widehat{\Delta}_{0x}^{+}\right)\left(X'X\right)^{-1}$ and employs corrections for endogeneity in the regressor x_t, leading to the transformed dependent variable $\widehat{y}_t^{+} = y_t - \widehat{\Omega}_{0x}\widehat{\Omega}_{xx}^{-1}(x_t - x_{t-1})$ and a bias correction term involving $\widehat{\Delta}_{0x}^{+} = \widehat{\Delta}_{0x} - \widehat{\Omega}_{0x}\widehat{\Omega}_{xx}^{-1}\widehat{\Delta}_{xx}$, which is constructed using consistent nonparametric estimators of submatrices of the long run and one sided long run quantities Ω and Δ. 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 allow 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 distribution of the FM-OLS estimator is degenerate and may have unsatisfactory properties in testing.

The paper makes the following contributions. First, we derive the rates of convergence and the limiting distribution of the FM-OLS estimator in case of a null conditional long run variance matrix. The rate of convergence of FM-OLS is faster than the usual $O(T)$ rate for cointegration and the rate depends on the bandwidth used in estimating the long run quantities that are employed in making corrections for endogeneity and serial correlation in the FM-OLS formula. The resulting limit distribution of FM-OLS 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 χ^2 and is invariant to the presence of singularity. Third, we show that when those restrictions fail, the Wald test is conservative. Using Monte Carlo simulations, we show 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 we analyse fiscal sustainability of the US government over the period 1947-2019 by testing the null hypothesis that the cointegraton relationship between government revenue and expenditure is $(1, -1)$. Multi-cointegration 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. This is important for practical purposes, as a separate treatment of the multicointegration case is not necessary (c.f., 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. We then discuss the implications of singularity for hypothesis testing 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

Let $(B_0', B_x')' \equiv B \equiv BM(\Omega)$ be the first m_0 and the last m_x subvectors of the Brownian motion. Let

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

Let $B_{0.x} = L_{0.\Omega}B$. Then

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

so $B_{0.x} \equiv BM\left(\Omega_{00.x}\right)$ and is orthogonal to B_x. Note that $\Omega_{00.x}$ is the long run variance of $u_{0.x,t} = L_{0.\Omega} u_t = D_{0.x}(L)\eta_t$, where $D_{0.x}(L) = L_{0.\Omega} D(L)$. It is well known that

$$T^{-1/2} \sum_{t=1}^{[T\cdot]} u_t \to_d B(\cdot) \equiv BM(\Omega)$$

and the OLS estimator is $O(T)$ consistent and the limit distribution depends on the nuisance parameters Ω and Δ:

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

The last two terms are the endogeneity and serial correlation biases, which FM-OLS seeks to remove.

Suppose Ω and Δ are estimated as

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

where $w(\cdot)$ is a kernel function, K is a bandwidth parameter (see e.g. Priestley (1981) and Hannan (1970)), and the sample covariances are

$$\widehat{\Gamma}_{\widehat{u},\widehat{u}}(j) = T^{-1} \sum_{1 \le t, t+j \le T}^{T-1} \widehat{u}_{t+j}\widehat{u}_t', \quad \widehat{u}_t = (\widehat{u}_{0t}', u_{xt}')',$$

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 $T \to \infty$, where c_T is slowly varying at infinity, i.e. $c_{xT}/c_T \to 1$ for $x > 0$ and $T \to \infty$. The kernel function $w(\cdot) : \mathbb{R} \to [-1, 1]$

is a twice continuously differentiable even function with:

(a) $w(0) = 1, w'(0) = 0, w''(0) \neq 0$

(b) $w(x) = 0, |x| \geq 1$, with $\lim_{|x| \to 1} w(x)/(1-|x|)^2 = const.$

For example, Parzen and Tukey-Hanning kernels satisfy Assumption K. The Bartlett-Priestley or quadratic spectral kernels do not satisfy Assumption K; in order to use them, one can extend our results for kernels satisfying Assumption K(a) and

(b') $w(x) = O(x^{-2})$, as $|x| \to 1$.

Under Assumption K, with $0 < k < 1$, and for any consistent estimator $\widehat{A}$,

$$\widehat{\Gamma} \to_p \Gamma,\ \widehat{\Omega} \to_p \Omega,\ \widehat{\Delta} \to_p \Delta.$$

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

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

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 limiting 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 Ω and Δ 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 $\Omega_{00.x}$ will increase by δ_{KT}, where $\delta_{KT} = \min(K^2, T^{1/2})$.

For example, in case $\Omega_{00.x} = 0$, $T\left(\widehat{A}^+ - A\right) \to_p 0$, and the precise rate of convergence of the FM-OLS depends on the bandwidth parameter expansion rate k in the kernel estimation in Assumption K, and the first order term in the nonparametric approximation of long run covariances may show up in the limit. If $\Omega_{00.x} = 0$, then $D_{0.x}(1) = 0$ and the Beveridge-Nelson decomposition of $u_{0.x,t}$ reduces to its transient component so that $u_{0.x,t} = \Delta\tilde{\eta}_t$ is $I(-1)$. The next proposition establishes convergence properties of FM-OLS for such time series. It is particularly useful for the case of single one-dimensional cointegration relationship, $m_0 = 1$ (see e.g. Phillips and Loretan, 1991), because singular long run variance implies that the conditional long run variance is zero. It also makes explicit the effect of singularity on the convergence rates and serve as a basis for our general result. Singularity also alters the rate of convergence of $\widehat{\Omega}_{00.x}$, which is used in the construction of the Wald test statistics. Therefore we also derive the rate of convergence for this quantity.

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

$$\delta_{KT} T \left(\widehat{A}^+ - A\right) = O_p(1),$$
$$K^2 \widehat{\Omega}_{00.x} = O_p(1).$$

As the proof reveals, the limit distribution of the above quantities depends on nuisance parameters and on the implementation of the nonparametric estimation of 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. As an illustration, consider a case when the bandwidth K grows slower than $T^{1/4}$, which includes the optimal bandwidth $T^{1/5}$ for long run variance estimation. Under these conditions, we have the following limit theory.

Proposition 4. *Suppose* $\Omega_{00.x} = 0$ *and* $D_{0.x}(L) = (1-L)\widetilde{D}_{0.x}(L)$, *with* $\widetilde{D}_{0.x}(1) \neq 0$. *Under Assumption K with* $k < 1/4$

$$K^2 T \left(\widehat{A}^+ - A\right) \to_d w''(0) \left(\sum_{h=0}^{\infty} (h + 1/2)\, \Gamma_{\tilde{\eta},u_x}(h) + \sum_{h=-\infty}^{\infty} (h + 1/2)\, \Gamma_{\tilde{\eta},u_x}(h) \Omega_{xx}^{-1} \int_0^1 dB_x B_x' \right) \left(\int_0^1 B_x B_x'\right)^{-1},$$

$$K^2 \widehat{\Omega}_{00.x} \to_p -w''(0) \sum_{h=-\infty}^{\infty} \Gamma_{\tilde{\eta},\tilde{\eta}}(h).$$

Unlike the result in the nonsingular case (Phillips, 1995; Phillips and Hansen, 1990), the limit distribution of FM-OLS depends on the covariance structure of the errors u_x and $u_{0.x,t}$ and on the second derivative of the kernel function.

Consider a general case of singular $\Omega_{00.x}$, of rank $r < m_0$. Thus, Ω 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 $R_\perp$, Proposition 3 applies and elements $R'_\perp A$ are estimated at a faster rate.

Alternatively, by the eigenvalue decomposition (singular value decomposition for symmetric matrices), there is a set of orthonormal eigenvectors of $\Omega_{00.x}$, q_i stacked in an orthogonal matrix Q and real eigenvalues λ_i in decreasing order on diagonal matrix Λ, such that $\Omega_{00.x} = Q\Lambda Q' = \sum_{i=1}^r \lambda q_i q_i'$. In this notation, $Q\Lambda^{1/2} = (R, 0)$ and $R_\perp$ spans the space of eigenvectors corresponding to zero eigenvalues.

We now state our first main result.

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

$$T\left(\widehat{A}^+ - A\right) \to_d \left(\int_0^1 dB_{0.x}B_x'\right)\left(\int_0^1 B_xB_x'\right)^{-1},$$

which is degenerate mixed normal. But the limit distribution is not degenerate in the direction R as

$$TR'\left(\widehat{A}^+ - A\right) \to_d \left(\int_0^1 dB_{f.x}B_x'\right)\left(\int_0^1 B_xB_x'\right)^{-1},$$

where $B_{f.x} \equiv BM\left(\Omega_{ff.x}\right)$ and $\Omega_{ff.x} = R'RR'R$ is a full rank $r \times r$ matrix which is the conditional long run variance of $R'U_{0.x}$. In the direction $R_\perp$ orthogonal to R the convergence of $\widehat{A}^+$ is at the faster rate $O(\delta_{KT}T)$ and

$$\delta_{KT}TR_\perp'\left(\widehat{A}^+ - A\right) = O_p(1).$$

It is apparent that the FM-OLS estimator of the singular triangular system has these properties: (i) it is consistent; (ii) the limit distribution is singular in the original coordinates; and (iii) rates of convergence are $O(T)$ in nondegenerate directions and $O(\delta_{KT}T)$ in degenerate directions. In the degenerate direction the limit distribution is the one shown in Proposition 4. Singularity of the limit distribution means that care is needed when undertaking hypothesis testing 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 for some $\phi \in C^1$, functions of dimension q for which the first derivative exists and is continuous,

$$H_0 : \quad \phi(vec(A)) = 0.$$

Suppose $\Omega_{00.x} = RR'$, where R is (m_0, r) matrix with $rank(R) = r < m_0$. So $R'R$ is full rank (r, r) matrix. Then under Assumption K

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

The limiting distribution is mixed normal ($\mathcal{MN}$) and the standard inference methods can be applied. In this case the Wald statistic is written in vectorized form as

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

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

$$rank\left\{\Phi\left(vec(A)\right)\left(\Omega_{00.x}\otimes\left(\int_0^1 B_x B_x'\right)^{-1}\right)\Phi\left(vec(A)\right)'\right\} = q. \tag{1}$$

Under Assumption K, $W \to_d \chi_q^2$. So, under the rank condition (1), the limit distribution of the Wald statistics is invariant to the presence of singularity.

3.1 Violation of the rank condition

We consider the following linear hypothesis

$$H_0: \quad Qvec(A) = R_0, \quad Q \sim q \times m_0 m_x, \quad rankQ = q.$$

Suppose $Q = R_1 \otimes R_2$, with ranks q_1 and q_2 respectively, so that

$$H_0: \quad R_1 A R_2 = R_3, \quad vec\ R_3 = R_0.$$

Then

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

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 does not hold as $\tilde{q}_1 q_2 < q_1 q_2 = q$. This is the case when some of the restrictions isolate directions where FM-OLS is hyperconsistent. The distribution of the Wald test statistics is then nonstandard and depends on nuisance parameters. In general, non-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 similar phenomenon arises in Toda and Phillips (1993), who describe situations where Wald tests of Granger causality do not follow asymptotically chi-squared distributions. For another example, see Phillips (2016), where singularity in the signal matrix gives nonstandard inference.

For an illustration of the consequences of singularity in the present case, consider testing $H_0: \ A = A^0$. Then $Q = I_{m_0 m_x}, R_0 = vec(A^0), R_1 = I_{m_0}, R_2 = I_{m_x}, R_3 = A^0$ 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

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

Note the notational change to W_I to emphasize that the following analysis only considers special restriction structures as above. The rank of $\Omega_{00.x} \otimes \int_0^1 B_x B_x'$ is equal to the rank of the conditional long run variance multiplied by 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 is nonsingular. If the conditional long run variance is nonsingular, the rank condition holds and $W_I \to \chi^2_q$.

In the special case where $\Omega_{00.x} = 0$ and Assumption K holds with $0 < k < 1/2$ under the null hypothesis we have

$$\begin{aligned} W_I &= K^2 \delta_{KT}^{-2} tr\left\{\left(T^{-2}X'X\right)\delta_{KT}T\left(\widehat{A}^+ - A^0\right)'\left(K^2\widehat{\Omega}_{00.x}\right)^{-1}\delta_{KT}T\left(\widehat{A}^+ - A^0\right)\right\} \\ &= O_p\left(K^2\delta_{KT}^{-2}\right) \to_p 0, \end{aligned}$$

because of the rates established in Proposition 3 and because $O_p\left(K^2\delta_{KT}^{-2}\right) = o_p(1)$ for $k < 1/2$.

In the more general case we have the following result.

Theorem 2. *If the conditional long run variance has reduced rank $r < m_0$ and Assumption K holds with $0 < k < 1/2$, then under the null $W_I \to_d \chi^2_{rm_x}$.*

The proof of the above result reveals that the distribution of the Wald test statistic involves the sum of two major components. The first component is the limit in nonsingular directions, which is $\chi^2_{rm_x}$, and the second is the limit in the direction where the conditional long run variance is zero, which is nonstandard, depends on nuisance parameters and decays at the speed $K^2\delta_{KT}^{-2}$. Therefore, for $k < 1/2$ the limit of W has thinner tails than the distribution of $\chi^2_{m_0 m_x}$ and the test is conservative.

4 Finite sample performance

In the following analysis, we run $N = 10000$ simulations for sample sizes $T \in \{50, 100, 1000, 10000\}$. We use R version 3.4.4 and package **cointReg** version 0.2.0 for FM-OLS estimation. The long run variances are estimated using Parzen kernel

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

We have shown that if the rank condition holds, the Wald test is invariant to singularity. If the rank condition fails, the Wald test controls size but is conservative. Using Monte Carlo simulations we study how accurate is the size of the Wald test under singularity and near singularity in finite samples.

We consider the following data generating process (DGP)

$$\begin{aligned} y_t &= Ax_t + u_{0t} \\ x_t &= x_{t-1} + u_{xt}, \quad t = 1, \ldots, T, \end{aligned}$$

where the cointegrating coefficient $A = 2$, and the combined error vector $u_t = (u'_{0t}, u'_{xt})'$ follows the linear process

$$u_t = \eta_t + D_1 \eta_{t-1}, \quad \text{with} \quad \eta_t \sim iidN(0, I_m).$$

We take $m = 2$ and for parameter choices $p \in \{0.0, 0.1, \ldots, 1.0\}$ define

$$D_1 = \begin{bmatrix} -p & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{then} \quad \Omega = \begin{bmatrix} (1-p)^2 & 0 \\ 0 & 1 \end{bmatrix},$$

and $\Omega_{0.xx} = (1-p)^2$. Our theory generalizes results on estimation and testing of cointegrating systems for the case $p = 1$, in which the long run variance is singular and the conditional long run variance is zero. By considering the diagonal D_1, which makes u_{0t} and u_{0xt} independent, we can study the effect of singularity in the long run variance separately from the effect of the long run dependence. When $p = 0$, the long run variance is I_2 and the conditional long run variance is 1, which corresponds to the standard, nonsingular case. We consider estimation of A and testing hypothesis $H_0: \; A = 2$.

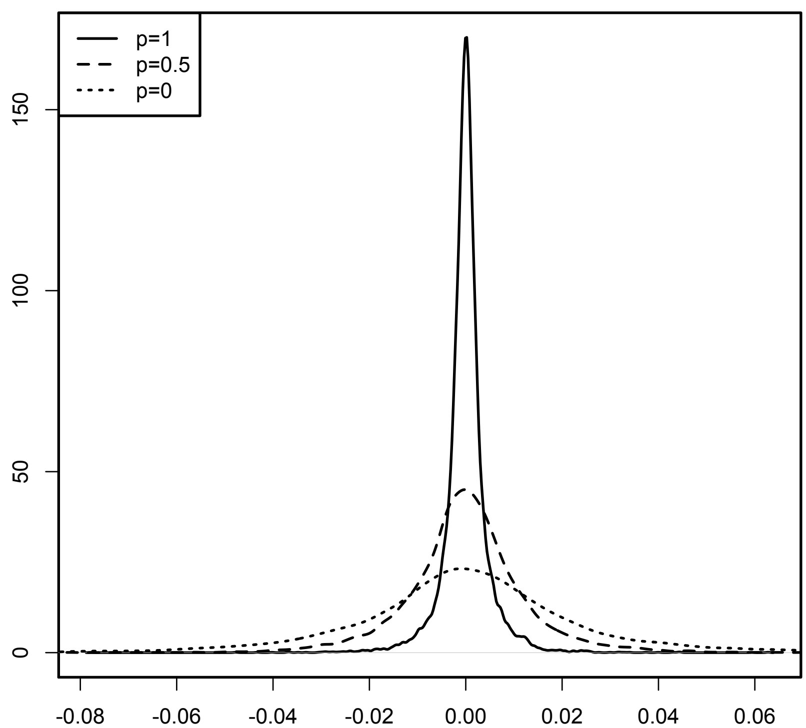

Figure 1: The density of the bias for the sample size $T = 100$ and parameters $p \in \{1.0, 0.5, 0.0\}$

In Figures 1 and 2 the densities of the centred estimator $\widehat{A}^{+} - 2$ are shown for sample sizes $T = 100$ and $T = 1000$. We compare the densities in the singular

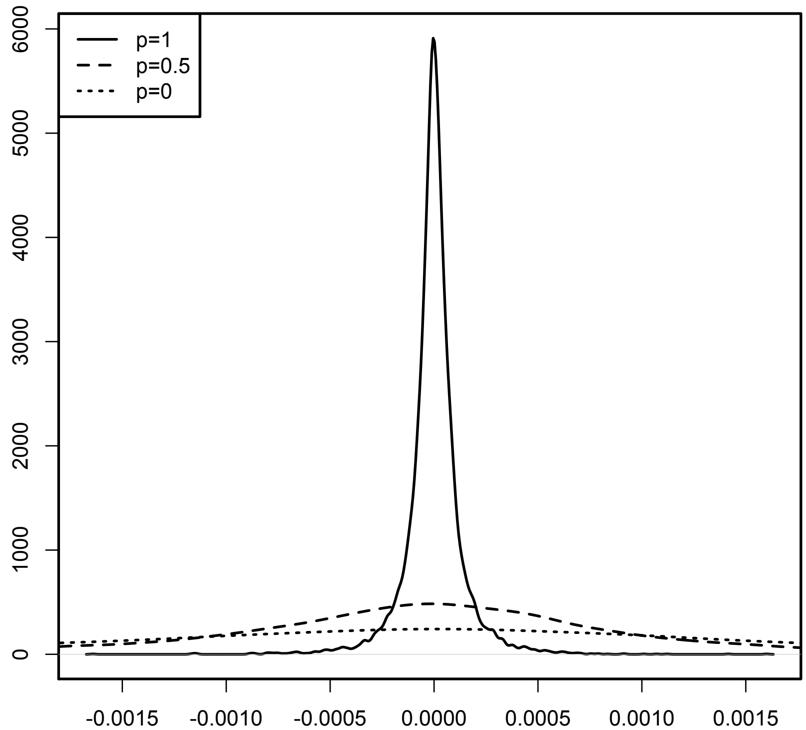

Figure 2: The density of the bias for the sample size $T = 1000$ and parameters $p \in \{1.0, 0.5, 0.0\}$

case ($p = 1$) with two nonsingular cases ($p = 0$ and $p = 0.5$). In both figures the bias in the singular case is much smaller than the bias in the nonsingular cases, with a more pronounced effect for $T = 1000$. Our asymptotic results show higher convergence rates of FM-OLS under singularity, and this effect can be seen already for $T = 100$.

In Figure 3 the densities of the Wald test statistic W for $H_0 : \ A = 2$ are shown for sample size $T = 100$. We compare the densities in the singular case ($p = 1$) with two nonsingular cases ($p = 0$ and $p = 0.5$). In nonsingular cases the test statistics is asymptotically χ^2_1, which density is also plotted. This

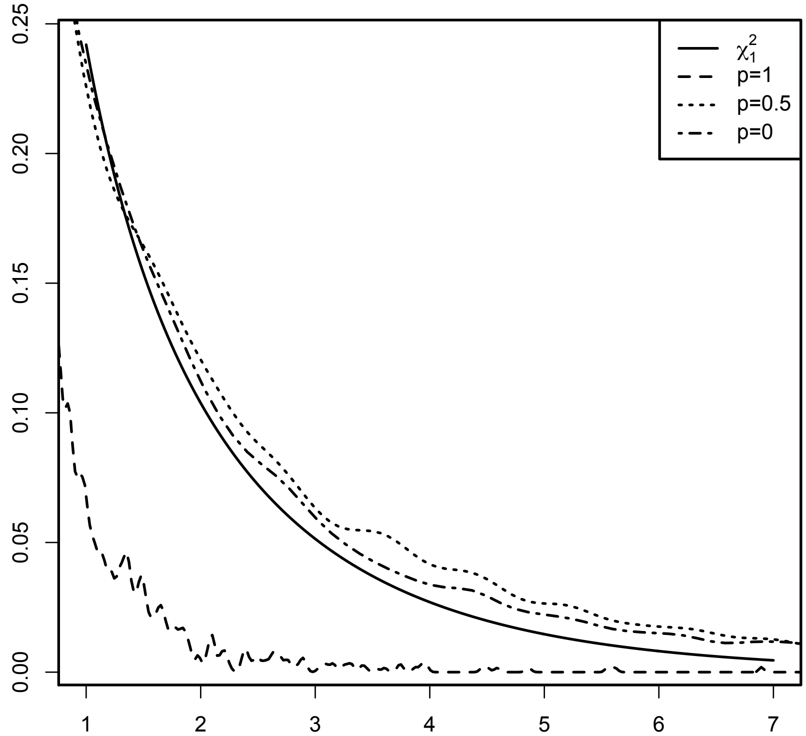

Figure 3: The density of the Wald test statistics for the sample size $T = 100$ and parameters $p \in \{1.0, 0.5, 0.0\}$ together with the χ_1^2 density

approximation is quite accurate for $T = 100$. The density of the test statistic in the singular case has thinner tail, so that the Wald test is conservative.

Further simulation results are in Tables 1 and 2. In the standard case, the bias is 0.003 for $T = 50$ which becomes zero up to the 4th digit for larger samples. The precision of the estimator, measured by the standard deviation of the estimates, increases at rate T. The coverage rates for the test $H_0: \ A = 2$ using the Wald test statistics and χ_1^2 approximation are far above the nominal levels in small samples and become close to the nominal at $T = 1000$. In the singular case, there is no bias even for size $T = 50$ and the precision increases

Table 1: The average bias, standard deviation, and the rejection rates for the nominal 0.01, 0.05, and 0.10 levels of the Wald statistics across 10000 simulations.

	T	p	bias	sd(bias)	0.01	0.05	0.10
1	50	0.0	-0.0003	0.0507	0.086	0.165	0.229
2	100	0.0	0.0000	0.0247	0.047	0.107	0.164
3	1000	0.0	-0.0000	0.0024	0.013	0.059	0.109
4	10000	0.0	-0.0000	0.0002	0.012	0.054	0.098
5	50	0.1	-0.0003	0.0459	0.096	0.178	0.248
6	100	0.1	0.0000	0.0223	0.054	0.120	0.184
7	1000	0.1	-0.0000	0.0022	0.020	0.072	0.129
8	10000	0.1	-0.0000	0.0002	0.018	0.066	0.118
9	50	0.2	-0.0003	0.0412	0.102	0.186	0.258
10	100	0.2	0.0000	0.0199	0.060	0.134	0.199
11	1000	0.2	-0.0000	0.0019	0.027	0.086	0.149
12	10000	0.2	-0.0000	0.0002	0.023	0.078	0.137
13	50	0.3	-0.0002	0.0366	0.103	0.190	0.263
14	100	0.3	0.0000	0.0175	0.063	0.141	0.208
15	1000	0.3	-0.0000	0.0017	0.032	0.096	0.162
16	10000	0.3	-0.0000	0.0002	0.028	0.088	0.150
17	50	0.4	-0.0002	0.0322	0.096	0.181	0.255
18	100	0.4	0.0000	0.0152	0.061	0.141	0.206
19	1000	0.4	-0.0000	0.0014	0.032	0.096	0.163
20	10000	0.4	-0.0000	0.0001	0.028	0.089	0.152
21	50	0.5	-0.0002	0.0279	0.079	0.160	0.230
22	100	0.5	0.0000	0.0129	0.050	0.121	0.188
23	1000	0.5	-0.0000	0.0012	0.026	0.085	0.148
24	10000	0.5	-0.0000	0.0001	0.024	0.078	0.138
25	50	0.6	-0.0001	0.0239	0.055	0.126	0.189
26	100	0.6	0.0000	0.0106	0.031	0.087	0.150
27	1000	0.6	-0.0000	0.0010	0.013	0.055	0.108
28	10000	0.6	-0.0000	0.0001	0.012	0.053	0.098

at a faster rate. The rejection rates, however, are zero even at the 10% level, showing that the test is conservative. Also, note that in the intermediate cases the test is slightly over-sized. For example, for $p = 0.4$ at the 5% level the test rejects 18.1% times for $T = 50$ and 8.9% times for $T = 10000$. More data is required because of autocorrelation in the errors. It is interesting to see that for $p = 0.7$ at the 5% level, the rejection rates are 4.6% at $T = 100$ and 1.8% at $T = 10000$. Here we observe the effect of the two opposite forces on the rejections rates. On the one hand, the test tends to over-reject at small samples; on the other, the test is conservative in near singular case.

Table 2: The average bias, standard deviation, and the rejection rates for the nominal 0.01, 0.05, and 0.10 levels of the Wald statistics across 10000 simulations.

	T	p	bias	sd(bias)	0.01	0.05	0.10
29	50	0.7	-0.0001	0.0204	0.029	0.080	0.132
30	100	0.7	0.0000	0.0085	0.013	0.046	0.087
31	1000	0.7	-0.0000	0.0007	0.003	0.019	0.049
32	10000	0.7	-0.0000	0.0001	0.002	0.018	0.046
33	50	0.8	-0.0001	0.0177	0.011	0.039	0.073
34	100	0.8	0.0000	0.0067	0.003	0.014	0.032
35	1000	0.8	-0.0000	0.0005	0.000	0.001	0.006
36	10000	0.8	-0.0000	0.0000	0.000	0.000	0.005
37	50	0.9	-0.0000	0.0161	0.005	0.018	0.036
38	100	0.9	0.0000	0.0053	0.001	0.003	0.008
39	1000	0.9	-0.0000	0.0003	0.000	0.000	0.000
40	10000	0.9	-0.0000	0.0000	0.000	0.000	0.000
41	50	1.0	-0.0000	0.0160	0.003	0.013	0.026
42	100	1.0	0.0000	0.0049	0.000	0.001	0.004
43	1000	1.0	-0.0000	0.0001	0.000	0.000	0.000
44	10000	1.0	-0.0000	0.0000	0.000	0.000	0.000

5 Evaluating Fiscal Sustainability

Soaring government debt in many countries call for better understanding of fiscal sustainability from both economic and econometric perspective. Econometric analysis of sustainability has a long tradition. To get some insight into sustainability from time series data, Hamilton and Flavin (1986) suggested to test stationarity of the discounted debt. Haikko and Rush (1991), Huag (1991), Trehan and Walsh (1991), Quintos (1995) were among the first to test cointegration between revenues and expenditures. Quintos (1995) calls sustainability "strong" if the revenues and expenditures cointegrate with coefficient $(1, -1)$ and tests the later using FM-OLS based t-statistics. Next we make two remarks regarding the above approach. For a recent discussion of other approaches to evaluate fiscal sustainability, see a chapter in the Handbook of Macroeconomics by D'Erasmo, Mendoza and Zhang (2016).

First, the cointegration between revenues and expenditures is only a sufficient condition for an intertemporal budget constraint (IBC) and that there are many other data generating processes consistent with IBC. It 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 interests 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 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

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

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, Quintos (1995) 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 ΔB_t that can be formulated as $\Delta B_t \sim I(0)$ which corresponds to strong sustainability by Quintos (1995), $G_t - R_t \sim I(0)$, 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 $I(2)$ system, as suggested by Haldrup (1994) and Engsted et al (1997). The results of the present paper show the following: (i) multicointegration can be allowed 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 illustrate these points with the 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: $x_t =$ 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.

The series are plotted in Figure 4. 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., $H_0 : A = 1$. FM-OLS estimation of the full sample gives $\widehat{A}^+ = 0.83$ 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 4 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 the structural breaks we leave for future research. But we do estimate the model for the period from

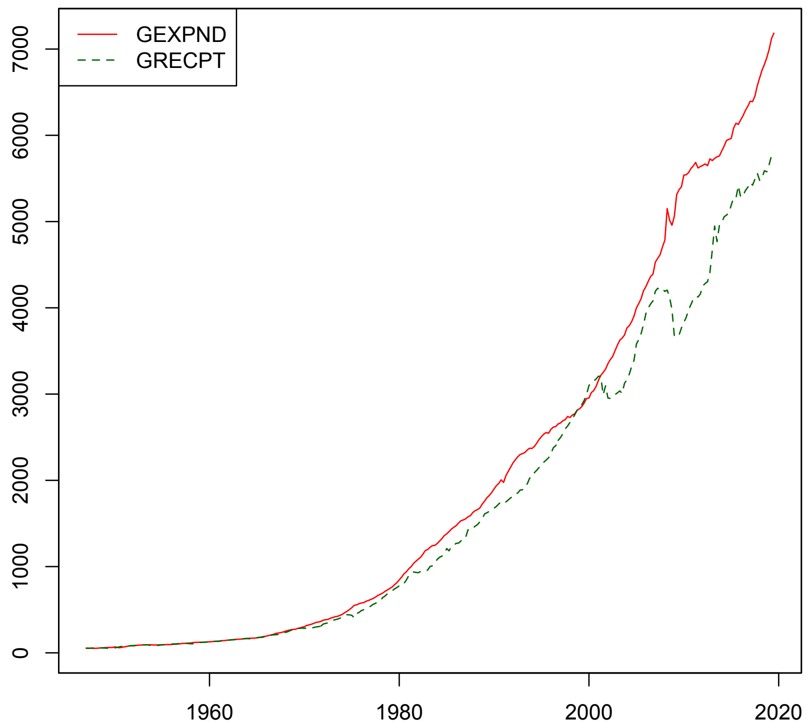

Figure 4: The US Government expenditures and receipts, billions of dollars, seasonally adjusted annual rate, quarterly frequency.

1947:Q1 to 1996:Q4 ($T = 200$) and obtain that $\widehat{A}^+ = 0.87$ 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 $(1, -1)$. We also estimate the cointegration relationship between real revenue and expenditure constructed using the GDP deflator. We take the same data series as in Berenguer-Rico and Carrion-i-Silvestre (2011) (available at the *Journal of Applied Econometrics* Data Archive, http://qed.econ.queensu.ca/jae/2011-v26.2/), 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 $\widehat{A}^+ = 0.92$ 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 multicointegration results in a singular long run error variance matrix. Likewise multicointegration arises when a certain linear combination of the regressor innovations removes the low frequency component in the equation error spectrum. This leads to the long run conditional variance matrix being singular or having a root that is local to zero. The consequence is a higher rate of convergence and non pivotal limit theory in certain directions. We show that the Wald test is invariant to singularity under certain rank conditions. When those conditions fail, the test is conservative. Simulation experiments show that in such situations the test rejection rates are far below the nominal levels under the null hypothesis in singular and near singular cases. We illustrate our results by analyzing the fiscal sustainability of the US government, testing the hypothesis that government revenue and expenditure are cointegrated with coefficient $(1, -1)$, where multicointegration naturally arises if bounds are imposed on deviations of debt from revenue.

The results obtained in this paper motivate the development of new robust approaches to estimating cointegrating relationships that allow for the possible presence of multicointegration, which are pivotal in the presence of singularity. This is an area of current research by the authors.

A Appendix: Proofs

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 linear $I(1)$ process occurs (see Johansen 1992, Engsted and Johansen, 1997) when $z = 1$ is a root, so that $C(1) = \xi\epsilon'$ is reduced rank, and $\xi_\perp' \dot{C}(1)\epsilon_\perp$ is singular. Write $C(L)$ as

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

and its derivative

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

We can take $\xi' = [A, I_{m_x}]$, $\xi_\perp' = [I_{m_x}, -A]$ and $\epsilon' = [D_{21}'(1), D_{22}(1)]$. Then

$$\xi_\perp' \dot{C}(1)\epsilon_\perp = -[D_{11}(1), D_{12}(1)] \begin{bmatrix} D_{21}(1) \\ D_{22}(1) \end{bmatrix}_\perp$$

which is singular if and only if $D(1)$ and equivalently Ω is singular. Indeed, suppose that $D_{22}(1)$ is nonsingular (if not, change the coordinates). Then we can take $\epsilon_\perp' = [I, -D_{21}'(1)D_{22}^{-1}(1)]$, such that $\xi_\perp'\dot{C}(1)\epsilon_\perp = -D_{11}(1) + D_{12}(1)D_{22}^{-1}(1)D_{21}(1)$ is (minus) Shur complement of block $D_{22}(1)$ and its rank by Guttman rank additivity formula is equal to $rank(D(1)) - rank(D_{22}(1))$. □

Proof of Proposition 2. Let $\Omega_{0x.x} = \Omega_{0x} - \Omega_{0x}\Omega_{xx}^{-1}\Omega_{xx}$, $\widehat{\Omega}_{0x.x} = \widehat{\Omega}_{0x} - \Omega_{0x}\Omega_{xx}^{-1}\widehat{\Omega}_{xx}$, $\Delta_{0x.x} = \Delta_{0x} - \Omega_{0x}\Omega_{xx}^{-1}\Delta_{xx}$, and $\widehat{\Delta}_{0x.x} = \widehat{\Delta}_{0x} - \Omega_{0x}\Omega_{xx}^{-1}\widehat{\Delta}_{xx}$. Then

$$\begin{aligned} \widehat{Y}^{+\prime} X = \left(Y' - \widehat{\Omega}_{0x}\widehat{\Omega}_{xx}^{-1}U_x'\right) X &= AX'X + U_0'X - \widehat{\Omega}_{0x}\widehat{\Omega}_{xx}^{-1}U_x'X \\ &= AX'X + U_{0.x}'X - \left(\widehat{\Omega}_{0x}\widehat{\Omega}_{xx}^{-1} - \Omega_{0x}\Omega_{xx}^{-1}\right) U_x'X \end{aligned}$$

and

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

and

$$\widehat{\Omega}_{0x}\widehat{\Omega}_{xx}^{-1} - \Omega_{0x}\Omega_{xx}^{-1} = \widehat{\Omega}_{0x.x}\widehat{\Omega}_{xx}^{-1}.$$

So,

$$\begin{aligned} T\left(\widehat{A}^{+} - A\right) &= \left(T^{-1}\left(\widehat{Y}^{+\prime}X - T\widehat{\Delta}_{0x}^{+} - A\right)\right)\left(T^{-2}X'X\right)^{-1} \\ &= \left(T^{-1}U_{0.x}'X - \widehat{\Delta}_{0x.x}\right)\left(T^{-2}X'X\right)^{-1} \\ &\quad - \widehat{\Omega}_{0x.x}\widehat{\Omega}_{xx}^{-1}\left(T^{-1}U_{x}'X - \widehat{\Delta}_{xx}\right)\left(T^{-2}X'X\right)^{-1}. \end{aligned} \tag{2}$$

From the weak convergence theory for sample covariances developed in Phillips and Durlauf (1986), Phillips (1989)

$$T^{-2}X'X \to_d \int_0^1 B_x B_x', \tag{3}$$

$$T^{-1}U_x'X \to_d \int_0^1 dB_x B_x' + \Delta_{xx}, \tag{4}$$

$$T^{-1}U_0'X \to_d \int_0^1 dB_0 B_x' + \Delta_{0x}. \tag{5}$$

Therefore,

$$T^{-1}U_{0.x}'X \to_d \int_0^1 dB_{0.x} B_x' + \Delta_{0x.x}. \tag{6}$$

By construction, $u_{0.x,t}$ has zero long run covariance with the errors u_{xt} that drive nonstationary component x_t, removing endogeneity of x_t in the long run. Therefore, for any consistent estimator of Ω and Δ, in particular, under Assumption K with $0 < k < 1$,

$$\widehat{\Omega}_{0x.x} \to_p 0_{m_0,m_x} \tag{7}$$

and

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

□

Proof of Proposition 3. Below we show that

$$T^{-1}U_{0.x}'X-\widehat{\Delta}_{0x.x}=T^{-1}\widetilde{\eta}_T x_T'+K^{-2}w''(0)\sum_{h=0}^{\infty}(h+1/2)\,\Gamma_{\tilde{\eta},u_x}(h)$$

$$+O_p\left((KT)^{-1/2}\right)+o_p\left(K^{-2}\right), \tag{8}$$

$$\widehat{\Omega}_{0x.x}=-K^{-2}w''(0)\sum_{h=-\infty}^{\infty}(h+1/2)\,\Gamma_{\tilde{\eta},u_x}(h)$$

$$+O_p\left((KT)^{-1/2}\right)+o_p\left(K^{-2}\right). \tag{9}$$

which together with (3) and (4), expansion (2) and the fact that $T^{-1}\widetilde{\eta}_T x_T'=O_p(T^{-1/2})$ gives $T\left(\widehat{A}^{+}-A\right)=O_p\left(\delta_{KT}^{-1}\right)$. Also,

$$\widehat{\Omega}_{00.x}=-K^{-2}w''(0)\sum_{h=-\infty}^{\infty}\Gamma_{\tilde{\eta},\tilde{\eta}}(h)+O_p\left((KT)^{-1/2}\right)+o_p\left(K^{-2}\right), \tag{10}$$

so $K^2\widehat{\Omega}_{00.x}=O_p(1)$. In our development, we borrow some ideas from the proofs of Lemma 8.1 (a), (b), and (g) in Phillips (1995), although that lemma does not strictly apply to our case. In particular, note that the $I(-1)$ errors appear in Lemma 8.1 in Phillips (1995) from a different source: if the vector x_t is cointegrated, but the cointegrating relationship is unknown, FM-OLS uses the first differences of the whole vector x_t producing linear combination of the first differences of stationary errors which are $I(-1)$. In our case it is assumed that Ω_{xx} is positive definite, i.e. that x_t are full rank nonstationary $I(1)$ and Δx_t are full rank stationary $I(0)$.

We now show (8). Note that

$$T^{-1}U_{0.x}'X-\widehat{\Delta}_{0x.x}=T^{-1}\widetilde{\eta}_T x_T'-T^{-1}\widetilde{\eta}_{-1}'U_x-\widehat{\Delta}_{0x.x} \tag{11}$$

and that $T^{-1}\widetilde{\eta}_T x'_T = O_p(T^{-1/2})$, $T^{-1}\widetilde{\eta}'_{-1}U_x = w(0)\widehat{\Gamma}_{\widetilde{\eta},u_x}(-1)$, while

$$\widehat{\Delta}_{0x.x} + T^{-1}\widetilde{\eta}'_{-1}U_x \tag{12}$$

$$= \sum_{j=0}^{K-1}[w(j/K) - w((j+1)/K)]\widehat{\Gamma}_{\widetilde{\eta},u_x}(j) + w((K-1)/K)\widehat{\Gamma}_{\widetilde{\eta},u_x}(K-1). \tag{13}$$

The first term in (13) is

$$\left(\sum_{j=0}^{K^*} + \sum_{j=K^*+1}^{K-2}\right)[w(j/K) - w((j+1)/K)]\widehat{\Gamma}_{\widetilde{\eta},u_x}(j), \tag{14}$$

for some $K^* = K^b$, with $0 < b < 1$. Applying the second order Taylor expansion of function $w(\cdot)$ at arguments $(j+1)/K$ around j/K,

$$w((j+1)/K) - w((j)/K) = K^{-1}w'(j/K) + 1/2\ K^{-2}w''(j/K)[1 + o(1)], \tag{15}$$

and for $j \le K^*$ we can apply the Taylor expansion of function $w'(\cdot)$ at arguments j/K around 0, where $w'(0) = 0$

$$w'(j/K) = w''(0)(j/K)[1 + o(1)]. \tag{16}$$

Then

$$w((j+1)/K) - w(j/K) = K^{-2}w''(0)(j + 1/2)[1 + o(1)]. \tag{17}$$

Because of expansion (15), the mean of the first term in (14) multiplied by K^2 is

$$K^2\sum_{j=0}^{K^*}[w(j/K) - w((j+1)/K)]\Gamma_{\widetilde{\eta},u_x}(j) \to -w''(0)\sum_{j=0}^{\infty}(j + 1/2)\Gamma_{\widetilde{\eta},u_x}(j) \tag{18}$$

and its variance is $O(1/KT)$. Because of expansion (17), the second term in

(14) is

$$K^{-1}\sum_{j=K^*+1}^{K-2} w'(j/K)\widehat{\Gamma}_{\widetilde{\eta},u_x}(j)[1+O(1/K)], \tag{19}$$

with mean

$$K^{-1}\sum_{j=K^*+1}^{K-2} w'(j/K)(1-j/T)\Gamma_{\widetilde{\eta},u_x}(j)[1+O(1/K)]. \tag{20}$$

The modulus of

$$K^{-1}\sum_{j=K^*+1}^{K-2} w'(j/K)(1-j/T)\Gamma_{\widetilde{\eta},u_x}(j) \tag{21}$$

is dominated by

$$\sup_x |w'(x)| K^{-1}\sum_{j=K^*+1}^{K-2} \|\Gamma_{\widetilde{\eta},u_x}(j)\| \tag{22}$$

$$= const\ K^{-1}\sum_{j>K^*}\sum_{s=0}^{\infty} \|D_s\|\|\widetilde{D}_{s+j}\| \tag{23}$$

$$= const\ K^{-1}{K^*}^{-\nu}\sum_{j>K^*}\sum_{s=0}^{\infty}(s+j)^{\nu}\|D_s\|\|D_{s+j}\| \tag{24}$$

$$= const\ K^{-1}K^{-\nu b}\sum_{s=0}^{\infty}\|D_s\|\sum_{r=0}^{\infty} r^{\nu}\|D_r\| = O(K^{-1-\nu b}) = o(K^{-2}), \tag{25}$$

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

We now bound the second term in (13). By Assumption K (b), $w((K-1)/K) = O(K^{-2})$ when $K\to\infty$. Since $\widetilde{\eta}_t = \sum_{s=0}^{\infty} u_{0.x,t-s}$,

$$\Gamma_{\widetilde{\eta},u_x}(K-1) = \sum_{s=0}^{\infty}\Gamma_{u_{0.x},u_x}(K-1-s) = \sum_{s=-\infty}^{\infty}\Gamma_{u_{0.x},u_x}(s) + o(1) = o(1),\ K\to\infty. \tag{26}$$

Also $\mathrm{Var}(\widehat{\Gamma}_{\widetilde{\eta},u_x}(K-1)) = O(T^{-1})$, therefore $\widehat{\Gamma}_{\widetilde{\eta},u_x}(K-1) = o_p(1)$ and

$$w((K-1)/K)\widehat{\Gamma}_{\widetilde{\eta},u_x}(K-1) = o_p(K^{-2}). \tag{27}$$

Expansion (8) is established.

We now show (9). Note that

$$\begin{aligned}
\widehat{\Omega}_{0x.x} &= \sum_{j=-K+1}^{K-1} w(j/K)\widehat{\Gamma}_{\Delta\widetilde{\eta},u_x}(j) &(28)\\
&= \sum_{j=-K+1}^{K-1} [w(j/K) - w((j+1)/K)]\widehat{\Gamma}_{\widetilde{\eta},u_x}(j) &(29)\\
&+ w((K-1)/K)\widehat{\Gamma}_{\widetilde{\eta},u_x}(K-1) - w((-K+1)/K)\widehat{\Gamma}_{\widetilde{\eta},u_x}(-K). &(30)
\end{aligned}$$

The first term in (30) is

$$\left(\sum_{j=-K^*}^{K^*} + \sum_{|j|>K^*,|j|<K}\right)[w(j/K) - w((j+1)/K)]\widehat{\Gamma}_{\widetilde{\eta},u_x}(j), \tag{31}$$

for some $K^* = K^b$, with $0 < b < 1$. Because of the expansion of $w(\cdot)$ in (15), the mean of the first term in (31) multiplied by K^2 is

$$K^2 \sum_{j=-K^*}^{K^*} [w(j/K) - w((j+1)/K)]\Gamma_{\widetilde{\eta},u_x}(j) \to -w''(0) \sum_{j=-\infty}^{\infty} (j+1/2)\Gamma_{\widetilde{\eta},u_x}(j) \tag{32}$$

and its variance is $O(1/KT)$. The second term in (30) and the second term in (31) are bounded as above. The third term in (30) is also bounded as above because

$$\Gamma_{\widetilde{\eta},u_x}(-K) = \sum_{s=0}^{\infty} \Gamma_{u_{0.x},u_x}(-K-s) = \sum_{s=K}^{\infty} \Gamma_{u_{0.x},u_x}(-s) = o(1),\ K \to \infty. \tag{33}$$

We now show (10). Note that

$$\widehat{\Omega}_{00.x} = \sum_{j=-K+1}^{K-1} w(j/K)\widehat{\Gamma}_{\Delta\widetilde{\eta},\Delta\widetilde{\eta}}(j) \tag{34}$$

$$= \sum_{j=-K+1}^{K-1} [w(j/K) - w((j+1)/K)]\widehat{\Gamma}_{\widetilde{\eta},\Delta\widetilde{\eta}}(j) \tag{35}$$

$$+ w((K-1)/K)\widehat{\Gamma}_{\widetilde{\eta},\Delta\widetilde{\eta}}(K-1) - w((-K+1)/K)\widehat{\Gamma}_{\widetilde{\eta},\Delta\widetilde{\eta}}(-K). \tag{36}$$

The first term in (36) is

$$\left(\sum_{j=-K^*}^{K^*} + \sum_{|j|>K^*,|j|<K}\right)[w(j/K) - w((j+1)/K)]\widehat{\Gamma}_{\widetilde{\eta},\Delta\widetilde{\eta}}(j), \tag{37}$$

for some $K^* = K^b$, with $0 < b < 1$. Because of the expansion of $w(\cdot)$ in (15), the mean of the first term in (37) multiplied by K^2 is

$$K^2 \sum_{j=-K^*}^{K^*} [w(j/K) - w((j+1)/K)]\Gamma_{\widetilde{\eta},\Delta\widetilde{\eta}}(j) \to -w''(0) \sum_{j=-\infty}^{\infty} (j+1/2)\Gamma_{\widetilde{\eta},\Delta\widetilde{\eta}}(j) \tag{38}$$

$$= -w''(0) \sum_{j=-\infty}^{\infty} j\Gamma_{\widetilde{\eta},\Delta\widetilde{\eta}}(j) \tag{39}$$

$$= -w''(0) \sum_{j=-\infty}^{\infty} j\Gamma_{\widetilde{\eta},\widetilde{\eta}}(j) + w''(0) \sum_{j=-\infty}^{\infty} j\Gamma_{\widetilde{\eta},\widetilde{\eta}}(j+1) \tag{40}$$

$$= -w''(0) \sum_{j=-\infty}^{\infty} \Gamma_{\widetilde{\eta},\widetilde{\eta}}(j) = -w''(0)\Omega_{\widetilde{\eta},\widetilde{\eta}}. \tag{41}$$

The second and the third terms in (36) and the second term in (37) are bounded as above. □

Proof of Proposition 4. For $k < 1/4$, (8) and (9) becomes

$$T^{-1}U'_{0.x}X - \widehat{\Delta}_{0x.x} = K^{-2}w''(0)\sum_{h=0}^{\infty}(h+1/2)\,\Gamma_{\tilde{\eta},u_x}(h) + o_p\left(K^{-2}\right),$$

$$\widehat{\Omega}_{0x.x} = -K^{-2}w''(0)\sum_{h=-\infty}^{\infty}(h+1/2)\,\Gamma_{\tilde{\eta},u_x}(h) + o_p\left(K^{-2}\right),$$

which together with (3) and (4) and expansion (2) gives the distribution of $K^2T\left(\widehat{A}^{+} - A\right)$. Also Equation (10) becomes

$$\widehat{\Omega}_{00.x} = -K^{-2}w''(0)\sum_{h=-\infty}^{\infty}\Gamma_{\tilde{\eta},\tilde{\eta}}(h) + o_p\left(K^{-2}\right),$$

from which the distribution of $K^2\widehat{\Omega}_{00.x}$ follows. □

Proof of Theorem 1. We have that

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

where

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

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 $L_R L_\Omega$ keeps nonstationary regressors $x_t = x_{t-1} + u_{xt}$ 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 which have nonsingular long run variance matrix $\Omega_{ff.x}$, $R'y_t^+ = R'Ax_t + u_{f.x,t}$, for which 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 $A_{ii.j} = \left(A_{ii} - A_{ij}A_{jj}^{-1}A_{ji}\right)$. We apply the above formula to the variance-covariance matrix in Wald test statistics

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

where we take into account that

1. $R\widehat{\Omega}_{00.x}R = O_p(1)$, because R isolates nondegenerate direction,
2. $R\widehat{\Omega}_{00.x}R_\perp = O_p(K^{-2})$, can be obtained similar to (9) in the proof of Proposition 3,
3. $R_\perp\widehat{\Omega}_{00.x}R_\perp = O_p(K^{-2})$, can be obtained similar to (10) in the proof of Proposition 3.

Then

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

using that $TR'\left(\widehat{A}^{+}-A^{0}\right) = O_p(1)$ and $\delta_{KT}TR_{\perp}'\left(\widehat{A}^{+}-A^{0}\right) = O_p(1)$ from Theorem 1. For $k < 1/2$, we have $O_p\left(K^{2}\delta_{KT}^{-2}\right) = o_p(1)$, and therefore $W_I \to_d \chi^{2}_{rm_x}$. □

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] D'Erasmo, P.D., Mendoza, E.G. and J. Zhang (2016) What is a Sustainable Public Debt? *Handbook of Macroeconomics*, 2B, 2493–2597.

[4] Engsted, T. and S. Johansen (1999) Grangers 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.

[5] 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.

[6] Hakkio, C.S., and M.Rush (1991) Is the budget deficit too large? *Economic Inquiry* 29, 429–445.

[7] Haldrup, N. (1994) The asymptotics of single-equation comtegration regressions with I(1) and I(2) variables. *Journal of Econometrics* 63, 151–181.

[8] 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.

[9] Haug, A.A. (1991) Cointegration and government borrowing constraints: evidence for the United States. *Journal of Business and Economic Statistics* 9, 97–101.

[10] Johansen, S. (1992) A representation of vector autoregressive processes integrated of order 2. *Econometric Theory* 8, 188–202.

[11] Phillips, P. C. B. (1989) Partially Identified Econometric Models, *Econometric Theory*, 5, 181–240.

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

[13] Phillips, P. C. B (2014) Optimal estimation of cointegrated systems with irrelevant instruments. *Journal of econometrics* 178, 210–224.

[14] Phillips, P. C. B (2016) Inference in near-singular regression. *Advances in Econometrics* 36, 461–486.

[15] Phillips, P. C. B., and S.N. Durlauf (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–495.

[16] 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.

[17] Phillips, P. C. B and I. Kheifets (2019) On Multicointegration. *Working Paper, Yale University.* Presented at the Rimini Time Series Conference, Larnica, Cyprus, June 2019.

[18] Phillips, P. C. B (1991): "Optimal Inference in Cointegrated Systems," Econometrica, 59, 283-306.

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

[20] Toda, H. Y. and P.C.P. Phillips (1993) Vector Autoregressions and Causality *Econometrica*, 61, 1367–1393.

[21] Quintos, C.E. (1995) Sustainability of the deficit process with structural shifts. *Journal of Business and Economic Statistics* 13, 409–417.

[22] 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.