Title

Simple, Robust, and Accurate F and t Tests in Cointegrated Systems

Authors

Hwang, Jungbin
Sun, Yixiao

Publication Date

2016-01-04

Simple, Robust, and Accurate F and t Tests in Cointegrated Systems

Jungbin Hwang and Yixiao Sun*
Department of Economics,
University of California, San Diego

First version: July 2014. This version: December 31, 2015

Abstract

This paper proposes new, simple, and more accurate statistical tests in a cointegrated system that allows for endogenous regressors and serially dependent errors. The approach involves first transforming the time series using some orthonormal basis functions in $L^2[0,1]$, which has energy concentrated at low frequencies, and then running an augmented regression based on the transformed data. The tests are extremely simple to implement as they can be carried out in exactly the same way as if the transformed regression is a classical linear normal regression. In particular, critical values are from the standard F or t distribution. The proposed F and t tests are robust in that they are asymptotically valid regardless of whether the number of basis functions is held fixed or allowed to grow with the sample size. The F and t tests have more accurate size in finite samples than existing tests such as the asymptotic chi-squared and normal tests based on the fully-modified OLS estimator of Phillips and Hansen (1990) and the trend IV estimator of Phillips (2014) and can be made as powerful as the latter tests.

JEL Classification: C12, C13, C32

Keywords: Cointegration, F test, Alternative Asymptotics, Nonparametric Series Method, t test, Transformed and Augmented OLS

1 Introduction

This paper considers a new approach to parameter estimation and inference in a triangular cointegrated regression system. A salient feature of this system is that the I(1) regressors are endogenous. In addition, to maintain generality of the short-run dynamics, we allow the I(0) regression errors to have serial dependence of unknown forms. One of the most popular semi-parametric estimators in this system is the fully modified OLS (FM-OLS) estimator of Phillips and Hansen (1990). The estimator involves using a long run variance and a half long run variance to remove the long run joint dependence and endogeneity bias. Both the long run variance and

*Email: j6hwang@ucsd.edu and yisun@ucsd.edu. Correspondence to: Jungbin Hwang or Yixiao Sun, Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508. We gratefully acknowledge partial research support from NSF under Grant No. SES-1530592.

the half long run variance are estimated nonparametrically. Inference based on the FM-OLS is standard — as in the classical linear regression with stationary or iid data the Wald statistic is asymptotically chi-squared. This is perhaps one of the most elegant and convenient results in time series econometrics. It releases us from having to simulate functionals of Brownian motion.

A drawback of the FM-OLS method is that the asymptotic chi-square test often has large size distortion. The source of the problem is that the estimation errors in the long run variance and half long run variance have been completely ignored in the conventional asymptotic framework adopted in Phillips and Hansen (1990). A new "fixed-b" asymptotic framework has been put forward by Vogelsang and Wagner (2014) but the Wald statistic does not appear to be asymptotically pivotal, making inference difficult and inconvenient. For this reason, Vogelsang and Wagner (2014) proceed to propose a different estimation method called the Integrated-Modified OLS (IM-OLS). They show that the associated test statistics are asymptotically pivotal under the fixed-b asymptotics. However, the inference procedure is quite complicated, and critical values have to be simulated.

In the same spirit of Vogelsang and Wagner (2014), we propose a new estimation method that involves first transforming the data using some orthonormal basis functions and then running an augmented regression based on the transformed data in the second stage. This gives rise to our transformed and augmented (TA) OLS (TAOLS) estimator. Augmentation removes the long run endogeneity problem while transformation eliminates the second order bias that plagues the OLS estimator. A key feature of our asymptotic analysis is that the number of basis functions K is held fixed as the sample size goes to infinity, leading to our fixed-K asymptotic theory. Compared with existing methods such as the FM-OLS of Phillips and Hansen (1990), the Trend Instrument Variable (TIV) of Phillips (2014) and the IM-OLS of Vogelsang and Wagner (2014), our new method enjoys several advantages.

First, under the fixed-K asymptotics, the test statistics based on the TAOLS estimator are asymptotically standard F or t distributed. Since critical values from the F and t distributions are easily available from statistical tables, there's no need to further approximate or simulate a nonstandard limit distribution. In addition, the test statistics can be obtained directly from canned statistical programs that can compute the F and t statistics in a classical linear normal regression. So our method is practically convenient and empirically appealing comparing with the IM-OLS method where both the test statistics and the critical values cannot be easily obtained. In particular, the fixed-b limit of the Wald statistic based on the IM-OLS is highly nonstandard. Critical values have to be simulated.

Second, our TAOLS method is asymptotically equivalent to the TIV method of Phillips (2014). As a by-product, we have established the fixed-K asymptotics of the TIV estimator and the associated test statistics. Under the increasing-K asymptotics where K grows with the sample size at an appropriate rate, Phillips (2014) shows that the Wald statistic and t statistic are asymptotically chi-squared and normal, respectively. While the fixed-K asymptotic distribution is different from the increasing-K asymptotic distribution, we show that the fixed-K asymptotic distribution approaches the increasing-K asymptotic distribution as K increases. As a result, the fixed-K critical values are asymptotically valid regardless of the type of asymptotics we consider. This is a robust property enjoyed by our asymptotic F and t tests.

Third, simulation results show that the asymptotic F and t tests have more accurate size than existing tests such as the asymptotic chi-squared and normal tests based on the FM-OLS or TIV estimators. On the other hand, the asymptotic F and t tests could be made as powerful as the latter tests. This is based on our simulation evidence. It is also consistent with the asymptotic efficiency of our TAOLS estimator under the increasing-K asymptotics. The asymptotic efficiency holds because the TAOLS estimator and the asymptotically efficient FM-OLS estimator have the same asymptotic distribution under the increasing-K asymptotics.

Fourth, taking it literally, the fixed-K asymptotics requires us to use only low-frequency information. Fundamentally, what a cointegrating vector measures is the long run relation among economic time series. For this reason, it is natural to estimate the cointegrating vector using only the long run variation of the underlying time series. Doing so helps us avoid high-frequency contaminations. From this perspective, the fixed K limiting thought experiment not only is an asymptotic device for developing new and more accurate approximations but also has substantive empirical content in economic applications.

Finally, in the presence of a linear trend, we can filter out the trend using a shifted version of standard cosine transforms. Interestingly, regression augmentation, which is necessary to achieve the asymptotic mixed normality for general basis functions, is not needed under the shifted cosine transforms. As a result, we can justify an even simpler OLS estimator — the transformed OLS (TOLS) estimator, which involves only transforming the original regression.

This paper contributes to a large body of literature on semiparametric estimation of cointegrated systems with Phillips and Hansen (1990), Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993) as seminal early contributions. In the FM-OLS setting, partial fixed-b asymptotic theory for cointegration inference has been considered by Bunzel (2006) and Jin, Phillips and Sun (2006) but the fixed-b asymptotics is applied only to the standard error estimator. See Vogelsang and Wagner (2014) for more discussion. Transforming a time series using the basis functions considered in this paper is equivalent to filtering the time series with a particular class of linear filters. The filtering idea has a long history; see, for example, Thomson (1982). For other applications of the idea in cointegration analysis, see Bierens (1997) and Müller and Watson (2013). See also Sun (2006) where series data transformation is used to estimate realized volatility.

The rest of the paper is organized as follows. Section 2 introduces a standard linear cointegration regression and discusses some of the drawbacks of existing methods. Section 3 introduces our TAOLS estimator and establishes the fixed-K asymptotic limits of the TAOLS estimator and the corresponding Wald statistic. Section 4 considers cointegration analysis under cosine or shifted cosine transforms. Section 5 presents simulation evidence. The last section concludes. Proofs are given in the appendix.

2 Model and Existing Literature

Consider the following cointegration model:

$$\begin{aligned} y_t &= \alpha_0 + x_t'\beta_0 + u_{0t} \\ x_t &= x_{t-1} + u_{xt} \end{aligned} \tag{1}$$

for $t = 1, ..., T$, where y_t is a scalar time series and x_t is a $d \times 1$ vector of time series with $x_0 = O_p(1)$. The mean zero error vector $u_t \equiv (u_{0t}, u_{xt}')' \in \mathbb{R}^m$ for $m = d+1$ is jointly stationary with long run variance (LRV) matrix Ω. We partition Ω as follows:

$$\underset{m\times m}{\Omega} = \sum_{j=-\infty}^{\infty} E u_t u_{t-j}' = \begin{pmatrix} \underset{1\times 1}{\sigma_0^2} & \underset{1\times d}{\sigma_{0x}} \\ \underset{d\times 1}{\sigma_{x0}} & \underset{d\times d}{\Omega_{xx}} \end{pmatrix}, \tag{2}$$

and write it as a sum of three conformable components: $\Omega = \Sigma + \Lambda + \Lambda'$ where

$$\Lambda := \sum_{j=1}^{\infty} E u_{t-j} u_t' = \begin{pmatrix} \underset{1\times 1}{\Lambda_{00}} & \underset{1\times d}{\Lambda_{0x}} \\ \underset{d\times 1}{\Lambda_{x0}} & \underset{d\times d}{\Lambda_{xx}} \end{pmatrix} \text{ and } \Sigma := E u_t u_t' = \begin{pmatrix} \underset{1\times 1}{\Sigma_{00}} & \underset{1\times d}{\Sigma_{0x}} \\ \underset{d\times 1}{\Sigma_{x0}} & \underset{d\times d}{\Sigma_{xx}} \end{pmatrix}.$$

The half long run variance Δ is defined to be

$$\Delta = \Sigma + \Lambda = \begin{pmatrix} \Delta_{00} & \Delta_{0x} \\ \Delta_{x0} & \Delta_{xx} \end{pmatrix}. \tag{3}$$

We assume that Ω_{xx} is positive definite so that x_t is a full-rank integrated process.

We shall maintain the Functional Central Limit Theorem (FCLT) below

$$T^{-1/2} \sum_{s=1}^{[T\cdot]} u_s \Rightarrow B(\cdot) = \Omega^{1/2} W(\cdot), \tag{4}$$

where $W(\cdot) := (w_0(\cdot), W_x'(\cdot))'$ is an m-dimensional standard Brownian process. Also, it will be convenient in our asymptotic development to represent the process $B(\cdot)$ using the Cholesky form of $\Omega^{1/2}$:

$$B(\cdot) = \begin{pmatrix} B_0(\cdot) \\ B_x(\cdot) \end{pmatrix} = \begin{pmatrix} \sigma_{0\cdot x} w_0(\cdot) + \sigma_{0x} \Omega_{xx}^{-1/2} W_x(\cdot) \\ \Omega_{xx}^{1/2} W_x(\cdot) \end{pmatrix}, \tag{5}$$

where $\sigma_{0\cdot x}^2 = \sigma_0^2 - \sigma_{0x} \Omega_{xx}^{-1} \sigma_{x0}$ and $\Omega_{xx}^{1/2}$ is a symmetric matrix square root of Ω_{xx}.

To simplify the discussion, we assume that there is no intercept in the regression. Let $X = [x_1', ..., x_T']'$ and $Y = [y_1, ..., y_T]'$. The OLS estimator of β_0 is given by $\hat{\beta}_{OLS} = (X'X)^{-1} X'Y$. It follows from Phillips and Durlauf (1986) and Stock (1987) that

$$T\left(\hat{\beta}_{OLS} - \beta_0\right) = \left(\frac{1}{T^2} \sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} x_t u_{0t}\right) \tag{6}$$

$$\Rightarrow \left(\int_0^1 B_x(r) B_x'(r) dr\right)^{-1} \left(\int_0^1 B_x(r) dB_0(r) + \Delta_{x0}\right), \tag{7}$$

where the presence of Δ_{x0} reflects the second-order endogeneity bias.

Since $B_x(\cdot)$ and $B_0(\cdot)$ are correlated, and Δ and hence Δ_{x0} is unknown, it is not possible to make asymptotically valid inference based on the naive OLS estimator. To overcome these two problems, Phillips and Hansen (1990) suggest the FM-OLS method that involves estimating Ω and Δ in the first step. Typical estimators of Ω and Δ take the following forms:

$$\hat{\Omega} = \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} Q_h\left(\frac{s}{T}, \frac{t}{T}\right) \hat{u}_t \hat{u}_s', \tag{8}$$

$$\hat{\Delta} = \frac{1}{T} \sum_{s=1}^{T} \sum_{t=s}^{T} Q_h\left(\frac{s}{T}, \frac{t}{T}\right) \hat{u}_t \hat{u}_s', \tag{9}$$

where $\hat{u}_t = (\hat{u}_{0t}, u_{xt}')'$ and $\hat{u}_{0t} = y_t - x_t' \hat{\beta}_{OLS}$. In the above definitions of $\hat{\Omega}$ and $\hat{\Delta}$, $Q_h(r, s)$ is a symmetric weighting function that depends on the smoothing parameter h. For conventional kernel LRV estimators, $Q_h(r,s) = k((r-s)/b)$ and we take $h = 1/b$. For the orthonormal series (OS) LRV estimators[1], $Q_h(r,s) = K^{-1}\sum_{j=1}^{K}\phi_j(r)\phi_j(s)$ and we take $h = K$, where $\{\phi_j(r)\}_{j=1}^{K}$ are orthonormal basis functions on $L^2[0,1]$ satisfying $\int_0^1 \phi_j(r)\,dr = 0$. We parametrize h in such a way so that h indicates the amount of smoothing for both types of LRV estimators.

After partitioning $\hat{\Omega}$ and $\hat{\Delta}$ in the same way as Ω and Δ, we define

$$\begin{aligned} y_t^+ &:= y_t - \Delta x_t' \hat{\Omega}_{xx}^{-1}\hat{\sigma}_{x0}, \\ u_t^+ &:= u_t - \Delta x_t' \hat{\Omega}_{xx}^{-1}\hat{\sigma}_{x0}, \\ \mathcal{M} &:= T\left(\hat{\Delta}_{x0} - \hat{\Delta}_{xx}\hat{\Omega}_{xx}^{-1}\hat{\sigma}_{x0}\right). \end{aligned} \tag{10}$$

Then, the FM-OLS estimator is given by

$$\hat{\beta}_{FM} = \left(X'X\right)^{-1}\left(X'Y^+ - \mathcal{M}\right),$$

where $Y^+ = [y_1^+, ..., y_T^+]'$. On the basis of kernel estimators of Ω and Δ, Phillips and Hansen (1990) show that $\hat{\beta}_{FM}$ is asymptotically mixed normal, i.e.,

$$T\left(\hat{\beta}_{FM} - \beta_0\right) \Rightarrow MN\left(0, \sigma_{0\cdot x}^2 \int_0^1 B_x(r)B_x'(r)dr\right). \tag{11}$$

This is in contrast with the limiting distribution of $\hat{\beta}_{OLS}$, which is complicated and has a second order endogeneity bias. Based on a consistent estimator $\hat{\sigma}_{0\cdot x}^2$ of $\sigma_{0\cdot x}^2$, one can obtain t and Wald statistics that are asymptotically normal and chi-square distributed, respectively.

A key step behind Phillips and Hansen's result is that $\hat{\Omega}$, $\hat{\Delta}, \hat{\sigma}_{0\cdot x}^2$ are all approximated by the respective degenerate distributions concentrated at Ω, Δ, and $\sigma_{0\cdot x}^2$. That is, regardless of the kernel function and the bandwidth used in the nonparametric estimators $\hat{\Omega}$, $\hat{\Delta}$, and $\hat{\sigma}_{0\cdot x}^2$, the same asymptotic approximations are used. However, in finite samples, both the kernel function and the bandwidth, especially the latter, do affect the sampling distribution of $\hat{\beta}_{FM}$ and the associated test statistics. For this reason, the normal and chi-squared approximations can be very poor in finite samples. This is because we completely ignore the estimation uncertainty in the nonparametric estimators $\hat{\Omega}$, $\hat{\Delta}$, and $\hat{\sigma}_{0\cdot x}^2$, which can be very high in finite samples. Bunzel (2006) and Jin, Phillips and Sun (2006) develop partial fixed-b asymptotic theory that accounts for the estimation uncertainty in $\hat{\sigma}_{0\cdot x}^2$ but ignore that in $\hat{\Omega}$ and $\hat{\Delta}$.

The degenerate distributional approximations for $\hat{\Omega}$, $\hat{\Delta}$, and $\hat{\sigma}_{0\cdot x}^2$ with consequential normal and chi-squared tests are obtained under the conventional increasing-smoothing asymptotic theory. Instead of this conventional asymptotics, we can use the fixed-smoothing asymptotics to obtain more accurate asymptotic approximations. The fixed-smoothing asymptotics includes the fixed-b asymptotics of Kiefer and Vogelsang (2005) as a special case. For more discussions on these two types of asymptotics, see Sun (2014a, 2014b). There is a growing number of papers on fixed-b asymptotic theory for stationary data starting with Kiefer and Vogelsang (2005). More recently, Vogelsang and Wagner (2014) establish the full-fledged fixed-b asymptotic distribution of the FM-OLS estimator and show that the Wald statistic depends on many nuisance parameters even in the limit. As a result, it is hard to make asymptotically pivotal inference. As an alternative solution, they suggest the Integrated Modified estimator (IM-OLS) which is based on partial sums of the original cointegration regression augmented by the original regressor. They

[1] Sun (2011, 2013) provides more background information on the OS LRV estimators.

invoke the fixed-b asymptotics to approximate the IM-OLS test statistics and show that they are asymptotically pivotal. However, their limiting distributions are quite complicated and highly nonstandard. Critical values have to be simulated for practical implementation.

3 Cointegration Analysis: Augmentation and Transformation

3.1 Model without time trend

To confront several challenges in the literature, we propose a new method to estimate the cointegration model in (1), where no trend is present. We consider the augmented cointegration model:

$$y_t = \alpha_0 + x_t'\beta_0 + \Delta x_t'\delta_0 + u_{0\cdot xt} \tag{12}$$

where $\delta_0 = \Omega_{xx}^{-1}\sigma_{x0}$ is the long run regression coefficient of Δx_t on u_{0t}, and $u_{0\cdot xt} = u_{0t} - u_{xt}'\delta_0$ is the long run regression error of u_{0t} projected onto u_{xt}. The long run variance of $u_{0\cdot xt}$ is $\sigma^2_{0\cdot x}$.

Let $\{\phi_i\}_{i=1}^{\infty}$ be a set of orthonormal basis functions in Hilbert space $L^2[0,1]$. Our new method starts by transforming the original data $\{y_t, x_t', \Delta x_t'\}_{t=1}^{T}$ using the basis functions $\{\phi_i\}_{i=1}^{K}$ for a finite K and then conducts regression analysis based on the transformed data. For each $i = 1, ..., K$, the transformed data $\{\mathbb{W}_i^{\cdot}\}$ are weighted averages of the original data:

$$\begin{aligned}
\mathbb{W}_i^{\alpha} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_i(\frac{t}{T}),\\
\mathbb{W}_i^{y} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}y_t\phi_i(\frac{t}{T}) = \frac{Y'\Phi_i}{\sqrt{T}},\ \mathbb{W}_i^{x} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_t\phi_i(\frac{t}{T}) = \frac{X'\Phi_i}{\sqrt{T}},\\
\mathbb{W}_i^{\Delta x} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta x_t\phi_i(\frac{t}{T}),\ \mathbb{W}_i^{0\cdot x} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{0\cdot xt}\phi_i(\frac{t}{T}),
\end{aligned} \tag{13}$$

where $\Phi_i = [\phi_i(1/T), ..., \phi_i((T-1)/T), \phi_i(1)]'$.

When $\phi_i(r) = \phi_i(1-r)$, which holds for the basis functions we will use, we can write, for example,

$$\mathbb{W}_i^{y} = \frac{1}{\sqrt{T}}\sum_{t=0}^{T-1}y_{T-t}\phi_i(\frac{T-t}{T}) = \frac{1}{\sqrt{T}}\sum_{t=0}^{T-1}y_{T-t}\phi_i(\frac{t}{T}). \tag{14}$$

So $\mathbb{W}_i^{y}$ can be regarded as the output from applying a linear filter to $\{y_t\}_{t=1}^{T}$. The transfer function of this linear filter is

$$H_{Ti}(\omega) = \frac{1}{\sqrt{T}}\sum_{t=0}^{T-1}\phi_i(\frac{t}{T})\exp(\iota t\omega) \text{ for } \iota = \sqrt{-1}. \tag{15}$$

To capture the long run behavior of the processes, we implicitly require that $H_{Ti}(\omega)$ be concentrated around the origin. That is, $H_{Ti}(\omega)$ resembles a band pass filter that passes low frequencies within a certain range and attenuates frequencies outside that range. The requirement can be met by any low-order trigonometric bases such as $\sqrt{2}\sin 2\pi ir$, $\sqrt{2}\cos 2\pi ir$ for a small i. In fact, the transfer functions associated with the first few basis functions in a commonly-used base system in $L^2[0,1]$ are often concentrated around the origin. So the requirement can be met easily.

Based on the augmented regression and the transformed data, we have

$$\mathbb{W}_i^y = \alpha_0 \mathbb{W}_i^\alpha + \mathbb{W}_i^{x\prime}\beta_0 + \mathbb{W}_i^{\Delta x\prime}\delta_0 + \mathbb{W}_i^{0\cdot x} \text{ for } i = 1, ..., K. \tag{16}$$

This can be regarded as a cross sectional regression with K observations. We assume that $K \geq 2d$, which is necessary for consistency. Obviously, there is no point of considering $K > T$, as there is no extra information beyond the first T transforms.

Under the assumption that each function $\phi_i(\cdot)$ is continuously differentiable and satisfies $\int_0^1 \phi_i(r)\, dr = 0$, which we will maintain, we have

$$\mathbb{W}_i^\alpha = \sqrt{T}\int_0^1 \phi_i(r)dr + \sqrt{T}O(1/T) = O\left(1/\sqrt{T}\right) = o(1), \tag{17}$$

and so the effect of the constant term α_0 in (18) is asymptotically negligible for a large T. As a result, our asymptotic theory remains the same regardless of whether an intercept is present or not. To simplify the presentation, we will assume without loss of generality that there is no intercept in the model so that

$$y_t = x_t'\beta_0 + u_{0t}, \; x_t = x_{t-1} + u_{xt} \tag{18}$$

and

$$\mathbb{W}_i^y = \mathbb{W}_i^{x\prime}\beta_0 + \mathbb{W}_i^{\Delta x\prime}\delta_0 + \mathbb{W}_i^{0\cdot x} \text{ for } i = 1, ..., K. \tag{19}$$

Putting (19) in the vector form, we have

$$\mathbb{W}^y = \mathbb{W}^{x\prime}\beta_0 + \mathbb{W}^{\Delta x\prime}\delta_0 + \mathbb{W}^{0\cdot x},$$

where $\mathbb{W}^y = (\mathbb{W}_1^y, ..., \mathbb{W}_K^y)'$ and $\mathbb{W}^x$, $\mathbb{W}^{\Delta x}$ and $\mathbb{W}^{0\cdot x}$ are defined similarly. Running the OLS based on the above equation leads to our Transformed and Augmented OLS (TAOLS) estimator of $\gamma_0 = (\beta_0', \delta_0')'$:

$$\hat{\gamma}_{TAOLS} = (\tilde{\mathbb{W}}'\tilde{\mathbb{W}})^{-1}\tilde{\mathbb{W}}'\mathbb{W}^y$$

where $\tilde{\mathbb{W}} = \left(\mathbb{W}^x, \mathbb{W}^{\Delta x}\right)$.

Let

$$P_x = \mathbb{W}^x(\mathbb{W}^{x\prime}\mathbb{W}^x)^{-1}\mathbb{W}^{x\prime}, \; P_{\Delta x} = \mathbb{W}^{\Delta x}(\mathbb{W}^{\Delta x\prime}\mathbb{W}^{\Delta x})^{-1}\mathbb{W}^{\Delta x\prime},$$

and $M_x = I_K - P_x$, $M_{\Delta x} = I_K - P_{\Delta x}$. Then we can represent $\hat{\gamma}_{TAOLS}$ as

$$\hat{\gamma}_{TAOLS} = \begin{pmatrix} \hat{\beta}_{TAOLS} \\ \hat{\delta}_{TAOLS} \end{pmatrix} = \begin{pmatrix} (\mathbb{W}^{x\prime}M_{\Delta x}\mathbb{W}^x)^{-1}(\mathbb{W}^{x\prime}M_{\Delta x}\mathbb{W}^y) \\ (\mathbb{W}^{\Delta x\prime}M_x\mathbb{W}^{\Delta x})^{-1}(\mathbb{W}^{\Delta x\prime}M_x\mathbb{W}^y) \end{pmatrix}. \tag{20}$$

To establish the asymptotic properties of $\hat{\gamma}_{TAOLS}$, we make the following assumptions.

Assumption 1 *(i) For $i = 1, ..., K$, each function $\phi_i(\cdot)$ is continuously differentiable and satisfies $\int_0^1 \phi_i(x)\, dx = 0$; (ii) The functions $\{\phi_i(\cdot)\}_{i=1}^K$ are orthonormal in $L^2[0,1]$.*

Assumption 2 *The vector process $\{u_t\}_{t=1}^T$ satisfies the FCLT in (4).*

Assumption 1 is mild and is satisfied by many basis functions. For example, $\sqrt{2}\cos(2\pi ir)$ and $\sqrt{2}\sin(2\pi ir)$ satisfy Assumption 1. Assumption 2 is a standard FCLT for time series data.

Under Assumptions 1 and 2, we have

$$\mathbb{W}_i^0 := \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\phi_i\left(\frac{s}{T}\right)u_s \Rightarrow \int_0^1 \phi_i(r)\,dB(r) = \begin{pmatrix} \sigma_{0\cdot x}\nu_i + \sigma_{0x}\Omega_{xx}^{-1/2}\xi_i \\ \Omega_{xx}^{1/2}\xi_i \end{pmatrix} \sim iidN(0,\Omega),$$

where $\nu_i = \int_0^1 \phi_i(r)\,dw_0(r)$ and $\xi_i = \int_0^1 \phi_i(r)\,dW_x(r)$. Since $w_0(\cdot)$ and $W_x(\cdot)$ are independent, we know that $\nu = (\nu_1, ..., \nu_K)'$ and $\xi = (\xi_1, ..., \xi_K)'$ are jointly normal and mutually independent. Also, by the continuous mapping theorem,

$$\frac{1}{T^{3/2}}\sum_{s=1}^{T}\phi_i\left(\frac{s}{T}\right)x_s = \frac{1}{T}\sum_{s=1}^{T}\phi_i\left(\frac{s}{T}\right)\frac{1}{\sqrt{T}}\sum_{\tau=1}^{s}u_{x\tau} \Rightarrow \int_0^1 \phi_i(r)\,B_x(r)dr = \Omega_{xx}^{1/2}\eta_i,$$

where for $\Psi_i(s) = \int_0^s \phi_i(r)dr$,

$$\begin{aligned}\eta_i &= \int_0^1 \phi_i(r)\,W_x(r)dr = \int_0^1 \phi_i(r)\left[\int_0^r dW_x(s)\right]dr = \int_0^1\left(\int_s^1 \phi_i(r)\,dr\right)dW_x(s) \\ &= -\int_0^1\left(\int_0^s \phi_i(r)\,dr\right)dW_x(s) = -\int_0^1 \Psi_i(s)dW_x(s). \end{aligned} \tag{21}$$

For any fixed K, let $\eta :\equiv (\eta_1, \eta_2, ..., \eta_K)' \in \mathbb{R}^{K\times d}$. Then

$$\begin{pmatrix} vec(\eta') \\ vec(\xi') \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} A\otimes I_d & B\otimes I_d \\ B'\otimes I_d & I_K\otimes I_d \end{pmatrix}\right),$$

where $A \in \mathbb{R}^{K\times K}$ and $B \in \mathbb{R}^{K\times K}$ whose (i,j)th components are $[A_{ij}] = \int_0^1\int_0^1 \phi_i(r)\phi_j(s)\min(r,s)drds$ and $[B_{ij}] = -\int_0^1 \Psi_i(r)\phi_j(r)\,dr$, respectively.

Using these properties, the following theorem establishes the asymptotic distribution of the TAOLS estimator.

Theorem 1 *Let Assumptions 1 and 2 hold. Then under the fixed-K asymptotics we have*

$$\Upsilon_T(\hat{\gamma}_{TAOLS} - \gamma_0) \Rightarrow (\tilde{\zeta}'\tilde{\zeta})^{-1}\tilde{\zeta}'\tilde{\nu},$$

where

$$\Upsilon_T = \begin{pmatrix} T\cdot I_d & \underset{d\times d}{0} \\ \underset{d\times d}{0} & I_d \end{pmatrix},\ \tilde{\zeta} = \left(\eta\Omega_{xx}^{1/2}, \xi\Omega_{xx}^{1/2}\right),\ \tilde{\nu} = \sigma_{0\cdot x}\nu,\ \textit{and}\ \tilde{\zeta}\perp\tilde{\nu}.$$

A direct implication of Theorem 1 is that

$$T(\hat{\beta}_{TAOLS} - \beta_0) \Rightarrow \sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\eta' M_\xi\eta\right)^{-1}\eta' M_\xi\nu, \tag{22}$$

$$\hat{\delta}_{TAOLS} - \delta_0 \Rightarrow \sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\xi' M_\eta\xi\right)^{-1}\xi' M_\eta\nu, \tag{23}$$

where $M_\xi = I_K - \xi\left(\xi'\xi\right)^{-1}\xi'$ and $M_\eta = I_K - \eta\left(\eta'\eta\right)^{-1}\eta'$. Conditional (η,ξ), both limiting distributions are normal:

$$\sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\eta' M_\xi\eta\right)^{-1}\eta' M_\xi\nu =^d N\left[0, \sigma_{0\cdot x}^2\Omega_{xx}^{-1/2}(\eta' M_\xi\eta)^{-1}\Omega_{xx}^{-1/2}\right],$$

$$\sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\xi' M_\eta\xi\right)^{-1}\xi' M_\eta\nu =^d N\left[0, \sigma_{0\cdot x}^2\Omega_{xx}^{-1/2}(\xi' M_\eta\xi)^{-1}\Omega_{xx}^{-1/2}\right].$$

So the unconditional limiting distributions are mixed normal. Furthermore, there is no second-order endogeneity bias in the TAOLS estimator. The TAOLS approach has effectively removed the two problems that plague the naive OLS estimator. The first problem, i.e., the asymptotic dependence between the partial sum processes of the regressor and regression error is eliminated because we argument the original regression by the additional regressor Δx_t. The second problem, i.e., the second-order endogeneity bias, is eliminated because we transform the original data and run the regression in the space spanned by the basis functions. In general, both augmentation and transformation are necessary to achieve the asymptotic mixed normality and asymptotic unbiasedness. However, for some special basis functions, augmentation is not necessary for the asymptotic mixed normality. See Section 4 for more detail.

Our TAOLS approach is similar to the Trend Instrumental Variable (TIV) approach of Phillips (2014), which involves solving

$$\begin{aligned}(\hat{\beta}_{TIV}', \hat{\delta}_{TIV}')' &= \underset{(\beta',\delta')'}{\arg\min}(Y - X\beta - \Delta X\delta)'\Phi(\Phi'\Phi)^{-1}\Phi'(Y - X\beta - \Delta X\delta) \\ &= (\tilde{X}'P_\Phi\tilde{X}')^{-1}(\tilde{X}'P_\Phi Y),\ \tilde{X} = [X, \Delta X] \text{ and } P_\Phi = \Phi(\Phi'\Phi)^{-1}\Phi'.\end{aligned} \tag{24}$$

The basis functions $\Phi = [\Phi_1, ..., \Phi_K]$ now act as "irrelevant" and deterministic trend instruments. The TIV approach is closely related to our TAOLS approach. It first projects the data onto the space spanned by the basis functions and then run the regression based on the projected data in the second stage. However, the interpretations are somewhat different. While Phillips (2014) emphasizes the use of the basis functions as irrelevant instruments and how they help reproduce the Karhunen-Loève representation of Brownian motion, we focus more on using the basis functions as low-frequency filters to exact the long run variation and covariation. If trigonometric bases are used, our approach is closer to the frequency domain approach of Phillips (1991a), although no frequency domain technique is needed here.

The following proposition shows that $\hat{\beta}_{TIV}$ and $\hat{\beta}_{TAOLS}$ are asymptotically equivalent under the fixed-K asymptotics.

Proposition 2 *Let Assumptions 1 and 2 hold.*

(i) Under the fixed-K asymptotics, $T(\hat{\beta}_{TIV} - \beta_0) = T(\hat{\beta}_{TAOLS} - \beta_0) + o_p(1)$.

(ii) Let V_K *be a random variable with distribution* $MN\left[0, \sigma_{0\cdot x}^2\Omega_{xx}^{-1/2}(\eta' M_\xi \eta)^{-1}\Omega_{xx}^{-1/2}\right]$. *Assume that* $\{\phi_i(\cdot)\}_{i=1}^{\infty}$ *is a complete orthonormal system in*

$$L_0^2[0,1] = \left\{ f(\cdot) \in L^2[0,1] : \int_0^1 f(r)\, dr = 0 \right\}.$$

Then as $K \to \infty$,

$$V_K \Rightarrow MN\left[0, \sigma_{0\cdot x}^2 \Omega_{xx}^{-1/2} \left(\int_0^1 \tilde{W}_x(r)\tilde{W}_x(r)' dr \right)^{-1} \Omega_{xx}^{-1/2}\right]$$

where $\tilde{W}_x(r) = W_x(r) - \int_0^1 W_x(s)\, ds$ *is the demeaned version of* $W_x(r)$.

Given the asymptotic equivalence in Proposition 2(i), our fixed-K asymptotic theory applies to the TIV estimator. This can be regarded as a by-product of our paper. For the TIV estimator, Phillips (2014) considers only the increasing-K asymptotics under which T and K go to infinity and $K/T \to 0$ at an appropriate rate. Phillips and Liao (2014, Lemma 5.1) considers the fixed-K limit of $\hat{\beta}_{TIV}$ in the case with a scalar regressor. In contrast to their claim on the existence of asymptotic bias, our mixed normal representation in Theorem 1 shows that there is no second-order asymptotic bias in both TAOLS and TIV estimators.

The conditional variance in Proposition 2(ii) is the semiparametric efficiency bound in the sense of Phillips (1991b). Here we do not aim at achieving the bound *per se*. Instead, our goal is to come up with a more accurate approximation for the given K value in a finite sample situation. Proposition 2(ii) indicates that the TAOLS estimator could become more efficient for a larger K and ultimately reach the semiparametric efficiency bound under the increasing-K asymptotics. So from this alternative asymptotic point of view, there is no loss of efficiency in our TAOLS approach.

The asymptotics in Proposition 2(ii) is obtained for a fixed K as $T \to \infty$ and then letting $K \to \infty$. This is a type of sequential asymptotics. The sequential asymptotics provides a smooth transition from our fixed-K asymptotics to the increasing-K asymptotics in Phillips (2005, 2014). There is no discontinuity between the fixed-K approximation for a large K value and the increasing-K approximation.

The asymptotic mixed normality and unbiasedness facilitate hypothesis testing. Suppose that we are interested in testing

$$H_0 : R\beta_0 = r \text{ vs. } H_1 : \ R\beta_0 \neq r, \tag{25}$$

where R is a $p \times d$ matrix. If $\sigma^2_{0\cdot x}$ is known, then we would construct the following Wald statistic:

$$\tilde{F}(\hat{\beta}_{TAOLS}) = \frac{1}{\sigma^2_{0\cdot x}}(R\hat{\beta}_{TAOLS} - r)' \left[R(\mathbb{W}^{x\prime} M_{\Delta x}\mathbb{W}^{x})^{-1}R'\right]^{-1} (R\hat{\beta}_{TAOLS} - r)/p.$$

When $p = 1$ and for one-sided alternative hypothesis, we would construct the following t statistic:

$$\tilde{t}(\hat{\beta}_{TAOLS}) = \frac{R\hat{\beta}_{TAOLS} - r}{\sqrt{\sigma^2_{0\cdot x}R(\mathbb{W}^{x\prime} M_{\Delta x}\mathbb{W}^{x})^{-1}R'}}.$$

Under the null hypothesis in (25), we can invoke Theorem 1 to obtain

$$\tilde{F}(\hat{\beta}_{TAOLS}) \Rightarrow Q'[\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\tilde{R}']^{-1}Q/p, \tag{26}$$

where

$$\tilde{R} = R\Omega_{xx}^{-1/2} \text{ and } Q = \tilde{R}(\eta' M_\xi \eta)^{-1}\eta' M_\xi \nu. \tag{27}$$

By construction, Q follows the mixed normal distribution $MN\left[0, \tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\tilde{R}'\right]$. Conditional on $\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\tilde{R}'$,

$$Q'\left(\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\tilde{R}'\right)^{-1} Q/p \sim \chi^2_p/p.$$

The conditional distribution does not depend on the conditioning variable $\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\tilde{R}'$. So χ^2_p/p is also the unconditional distribution. That is, the infeasible test statistic $\tilde{F}(\hat{\beta}_{TAOLS})$ converges in distribution to χ^2_p/p. Similarly, $\tilde{t}(\hat{\beta}_{TAOLS})$ converges to the standard normal distribution.

The presence of the unknown long run variance $\sigma^2_{0\cdot x}$ in $\tilde{F}(\hat{\beta}_{TAOLS})$ and $\tilde{t}(\hat{\beta}_{TAOLS})$ hinders their practical applications. In practice, we have to estimate $\sigma^2_{0\cdot x}$ in order to construct the test statistics. Given that $\sigma^2_{0\cdot x}$ is the approximate variance of the error term in the TAOLS regression, it is natural to estimate $\sigma^2_{0\cdot x}$ by

$$\hat{\sigma}^2_{0\cdot x} = \frac{1}{K}\sum_{i=1}^{K}\left(\hat{\mathbb{W}}_i^{0\cdot x}\right)^2 = \frac{1}{K}\mathbb{W}^{0\cdot x\prime}\left[I_K - \tilde{\mathbb{W}}(\tilde{\mathbb{W}}'\tilde{\mathbb{W}})^{-1}\tilde{\mathbb{W}}'\right]\mathbb{W}^{0\cdot x},$$

where $\hat{\mathbb{W}}_i^{0\cdot x} = \mathbb{W}_i^y - \mathbb{W}_i^{x\prime}\hat{\beta}_{TAOLS} - \mathbb{W}_i^{\Delta x\prime}\hat{\delta}_{TAOLS}$. With the estimator $\hat{\sigma}^2_{0\cdot x}$, we can construct the feasible $F(\hat{\beta}_{TAOLS})$ and $t(\hat{\beta}_{TAOLS})$ as follows:

$$F(\hat{\beta}_{TAOLS}) = \frac{1}{\hat{\sigma}^2_{0\cdot x}}(R\hat{\beta}_{TAOLS} - r)'\left[R(\mathbb{W}^{x\prime}M_{\Delta x}\mathbb{W}^x)^{-1}R'\right]^{-1}(R\hat{\beta}_{TAOLS} - r)/p, \tag{28}$$

$$t(\hat{\beta}_{TAOLS}) = \frac{R\hat{\beta}_{TAOLS} - r}{\sqrt{\hat{\sigma}^2_{0\cdot x}R(\mathbb{W}^{x\prime}M_{\Delta x}\mathbb{W}^x)^{-1}R'}}.$$

The theorem below establishes the limiting null distributions of $F(\hat{\beta}_{TAOLS})$ and $t(\hat{\beta}_{TAOLS})$ under the fixed-K asymptotics.

Theorem 3 *Let Assumptions 1 and 2 hold. Under the fixed-K asymptotics, we have*

$$F(\hat{\beta}_{TAOLS}) \Rightarrow \frac{K}{K-2d}\cdot F_{p,K-2d} \text{ and}$$
$$t(\hat{\beta}_{TAOLS}) \Rightarrow \sqrt{\frac{K}{K-2d}}\cdot t(K-2d).$$

Theorem 3 shows that both $F(\hat{\beta}_{TAOLS})$ and $t(\hat{\beta}_{TAOLS})$ are asymptotically pivotal and have standard limiting distributions. This is in contrast with the IM-OLS approach of Vogelsang and Wagner (2014) where the corresponding limiting distributions are nonstandard. A great advantage of our approach is that critical values can be obtained from statistical tables and software packages. There is no need to simulate nonstandard critical values.

Our asymptotic distributions are also in sharp contrast with the chi-squared (χ^2_p/p) and standard normal distributions. The latter distributions are the limits for the infeasible test statistics. In fact, under the increasing-K asymptotics as developed in Phillips (2014), the latter distributions are also the limits of the feasible statistics $F(\hat{\beta}_{TAOLS})$ and $t(\hat{\beta}_{TAOLS})$. So the increasing-K asymptotics effectively assumes that $\sigma^2_{0\cdot x}$ is known in large samples, and hence completely ignores the estimation uncertainty in $\hat{\sigma}^2_{0\cdot x}$.

Let $F^\alpha_{p,K-2d}$ and χ^α_p be the $(1-\alpha)$ quantiles from the standard $F_{p,K-2d}$ and χ^2_p distributions, respectively. Then we can use the modified F critical value $K/(K-2d)F^\alpha_{p,K-2d}$ to carry out our F test. This critical value is larger than the scaled chi-squared critical value χ^α_p/p for two reasons. First, $F^\alpha_{p,K-2d} > \chi^\alpha_p$ because the F distribution $F_{p,K-2d}$ has a random denominator as compared to the corresponding chi-square distribution. Second, the multiplicative factor $K/(K-2d)$ is greater than 1. The difference between the two critical values depends on the value of K. It can be quite large when K is small. However, as K increases, $K/(K-2d)F^\alpha_{p,K-2d}$ approaches χ^α_p/p. There is a smooth transition from a fixed-K critical value to the corresponding increasing-K critical value. So the critical value $K/(K-2d)F^\alpha_{p,K-2d}$ is asymptotically valid regardless of whether K is held fixed or allowed to grow with the sample size. In this sense, $K/(K-2d)F^\alpha_{p,K-2d}$ is a robust critical value.

3.2 Model with a linear trend

In this subsection, we consider a more general version of (1) by including a time trend in the cointegration model. The model is now given by

$$\begin{aligned} y_t &= x_t'\beta_0 + \mu_0 t + u_{0t}, \\ x_t &= x_{t-1} + u_{xt}. \end{aligned} \tag{29}$$

Define $\mathbb{W}_i^{tr} = T^{-1/2}\sum_{t=1}^{T} \phi_i(t/T)t$ for $i = 1, ..., K$ and $\mathbb{W}^{tr} = (\mathbb{W}_1^{tr}, ..., \mathbb{W}_K^{tr})'$. Then, the transformed regression in (18) is naturally generalized to

$$\mathbb{W}_i^y = \mathbb{W}_i^{x'}\beta_0 + \mathbb{W}_i^{\Delta x\prime}\delta_0 + \mathbb{W}_i^{tr}\mu_0 + \mathbb{W}_i^{0\cdot x} \text{ for } i = 1, ..., K. \tag{30}$$

As we discussed before, an intercept can be included in (29) and (30) but our approach is asymptotically invariant to location shifts. The TAOLS estimator for β_0, δ_0 and μ_0 is now given by

$$\left(\hat{\beta}_{TAOLS}', \hat{\delta}_{TAOLS}', \hat{\mu}_{OLS}'\right)' = (\tilde{\mathbb{W}}_{tr}'\tilde{\mathbb{W}}_{tr})^{-1}\tilde{\mathbb{W}}_{tr}'\mathbb{W}^y, \tag{31}$$

where $\tilde{\mathbb{W}}_{tr} = \left(\mathbb{W}^x,\ \mathbb{W}^{\Delta x}, \mathbb{W}^{tr}\right)$.

Let $\hat{\mathbb{W}}_{i,tr}^{0\cdot x} = \mathbb{W}_i^y - \mathbb{W}_i^{x\prime}\hat{\beta}_{TAOLS} - \mathbb{W}_i^{\Delta x\prime}\hat{\delta}_{TAOLS} - \mathbb{W}_i^{tr}\hat{\mu}_{OLS}$ and $(\hat{\sigma}_{0\cdot x}^{tr})^2 = K^{-1}\sum_{i=1}^{K}(\hat{\mathbb{W}}_{i,tr}^{0\cdot x})^2$. Then we can construct the Wald statistic and t statistic as follows:

$$\begin{aligned} F_{tr}(\hat{\beta}_{TAOLS}) &= \frac{1}{(\hat{\sigma}_{0\cdot x}^{tr})^2}(R\hat{\beta}_{TAOLS} - r)\left[R(\mathbb{W}^{x\prime}M_{\Delta x,tr}\mathbb{W}^x)^{-1}R'\right]^{-1}(R\hat{\beta}_{TAOLS} - r)/p, \\ t_{tr}(\hat{\beta}_{TAOLS}) &= \frac{R\hat{\beta}_{TAOLS} - r}{\sqrt{(\hat{\sigma}_{0\cdot x}^{tr})^2 R(\mathbb{W}^{x\prime}M_{\Delta x,tr}\mathbb{W}^x)^{-1}R'}}, \end{aligned}$$

where $M_{\Delta x,tr} = I_K - \mathbb{W}_{\Delta x,tr}\left(\mathbb{W}_{\Delta x,tr}'\mathbb{W}_{\Delta x,tr}\right)^{-1}\mathbb{W}_{\Delta x,tr}'$ and $\mathbb{W}_{\Delta x,tr} = \left(\mathbb{W}^{\Delta x}, \mathbb{W}^{tr}\right)$.

Theorem 4 *Let Assumptions 1 and 2 hold. Assume that* $a := \left(\int_0^1 \phi_1(r)\,rdr, ..., \int_0^1 \phi_K(r)\,rdr\right)' \neq 0$. *Under the fixed-K asymptotics, we have (i)*

$$\Upsilon_{T,tr}\begin{pmatrix} \hat{\beta}_{TAOLS} - \beta_0 \\ \hat{\delta}_{TAOLS} - \delta_0 \\ \hat{\mu}_{TAOLS} - \mu_0 \end{pmatrix} \Rightarrow \begin{pmatrix} \sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\eta' M_{\xi,a}\eta\right)^{-1}\eta' M_{\xi,a}\nu \\ \sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\xi' M_{\eta,a}\xi\right)^{-1}\xi' M_{\eta,a}\nu \\ \sigma_{0\cdot x}\left(a' M_{\eta,\xi}a\right)^{-1}a' M_{\eta,\xi}\nu \end{pmatrix}, \tag{32}$$

where

$$\Upsilon_{T,tr} = \begin{pmatrix} \Upsilon_T & 0 \\ 0 & T^{3/2} \end{pmatrix},$$

and M_υ is the project matrix projecting onto the orthogonal complement of the column space of υ.

(ii)

$$F_{tr}(\hat{\beta}_{TAOLS}) \Rightarrow \frac{K}{K-2d-1}F_{p,K-2d-1} \text{ and } t_{tr}(\hat{\beta}_{TAOLS}) \Rightarrow \sqrt{\frac{K}{K-2d-1}}t(K-2d-1). \tag{33}$$

Theorem 4(ii) is entirely analogous to Theorem 3. The effect of having an additional trend regressor $\mathbb{W}_i^{Tr}$ is reflected by the adjustment in the multiplicative correction factor and the degrees of freedom in the F and t distributions.

The asymptotic F and t limit theory resembles those in the classical linear normal regressions (CLNR) with K iid observations. The multiplicative correction is a type of degrees-of-freedom correction. Had we followed the standard practice in the CLNR and define

$$(\hat{\sigma}_{0\cdot x})^2 = \frac{1}{K-2d}\sum_{i=1}^{K}(\hat{\mathbb{W}}_i^{0\cdot x})^2 \text{ and } (\hat{\sigma}_{0\cdot x}^{tr})^2 = \frac{1}{K-2d-1}\sum_{i=1}^{K}(\hat{\mathbb{W}}_{i,tr}^{0\cdot x})^2, \tag{34}$$

we would not have to make the multiplicative correction. That is, the Wald statistic will be asymptotically F distributed, and the t statistic will be asymptotically t distributed.

Observing that we compute the standard error of the TAOLS estimator as if the errors in the transformed regression are homoskedastic, which does hold in large samples, our Wald statistic $F_{tr}(\hat{\beta}_{TAOLS})$ with (34) as the error variance estimator is numerically identical to the F statistic based on the residual sum of squares under the restricted and unrestricted models. So we can obtain $F_{tr}(\hat{\beta}_{TAOLS})$ (and $t_{tr}(\hat{\beta}_{TAOLS})$) from the output of any simple and very basic regression program as long as it works at least for the CLNR with homoskedastic errors. The only step that we have to take is to get the data into the transformed form. A cautionary note is that we do not include the intercept in the transformed and augmented regression.

If instead of a linear trend we have the polynomial trends $(t, t^2, ..., t^g)$, then the same proof of Theorem 4 can be invoked to show that

$$F_{tr}(\hat{\beta}_{TAOLS}) \Rightarrow \frac{K}{K-2d-g}F_{p,K-2d-g} \text{ and } t_{tr}(\hat{\beta}_{TAOLS}) \Rightarrow \sqrt{\frac{K}{K-2d-g}}t(K-2d-g), \tag{35}$$

where $2d+g$ is now the number of parameters to be estimated.

As a by-product, we can perform the endogeneity test, i.e., a test of whether $\delta_0 = 0$, in exactly the same way as if the transformed regression is a CLNR. This can be justified asymptotically using the same argument for Theorem 3 or 4.

3.3 Selecting the number of basis functions

In principle, we can use any finite number of orthonormal basis functions satisfying Assumption 1 in our fixed-K framework. However, Proposition 2 indicates that a larger K leads to a more efficient estimator. On the other hand, when K is too large, the TAOLS estimator will suffer from the asymptotic bias that is not captured by the fixed-K asymptotics. For example, if we set K equal to the sample size, which is the upper bound for K, the TAOLS estimator will be the same as the augmented OLS estimator which suffers from the second order asymptotic bias. So there is an opportunity to select K to trade-off the variance effect with the bias effect.

A direct approach to data-driven choice of K is to first develop a high order expansion of $\hat{\beta}_{TAOLS}$ from which we obtain the approximate mean squared error (AMSE) of $\hat{\beta}_{TAOLS}$ and then select K to minimize AMSE($\hat{\beta}_{TAOLS}$). For hypothesis testing, a direct approach is to derive the optimal choice of K that minimizes the Type II error of our proposed Wald test or t test subject to a control of the Type I error. The direct approaches are ambitious. Phillips (2014) discusses some of the technical challenges behind the direct approaches. We leave them for future research.

An indirect approach that appears to work well is based on the bias and variance of the LRV estimator. Following a large literature on LRV estimation, Phillips (2005) proposes to select K

by minimizing the AMSE of $\hat{\Omega}$ defined in (8). In the present setting, we have

$$\hat{\Omega} = \frac{1}{K}\sum_{i=1}^{K}\left(\hat{\mathbb{W}}_i^u\right)\left(\hat{\mathbb{W}}_i^u\right)' \text{ for } \hat{\mathbb{W}}_i^u = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{u}_t\phi_i(\frac{t}{T})$$

where $\hat{u}_t = (y_t - x_t\hat{\beta}_{OLS}, \Delta x_t')'$ or $\hat{u}_t = (y_t - x_t\hat{\beta}_{OLS} - \hat{\mu}_{OLS}t, \Delta x_t')'$ depending on whether a linear trend is present or not. Suppose that we use the cosine and sine basis functions:

$$\left\{\sqrt{2}\cos 2j\pi r, \sqrt{2}\sin 2j\pi r, j = 1, ..., K/2\right\}. \tag{36}$$

Then the AMSE-optimal K^* is given by

$$\begin{aligned} K^*_{MSE} &= \left\lceil \left(\frac{tr(I_{m^2} + \mathbb{K}_{mm})(\Omega \otimes \Omega)}{4vec\,(B)'\,vec\,(B)}\right)^{1/5} T^{4/5} \right\rceil \\ \text{for } B &= -\frac{\pi^2}{6}\sum_{h=-\infty}^{\infty} h^2\Gamma_u(h),\ \Gamma_u(h) = Eu_tu_{t-h}' \end{aligned} \tag{37}$$

where $\mathbb{K}_{mm}$ is the $m^2 \times m^2$ commutation matrix and I_{m^2} is the $m^2 \times m^2$ identity matrix.

Recall that K has to be large enough to ensure the consistency of the TAOLS estimator and the associated tests. Suppose that we are interested in testing the significance of all regressors in the TA regression without a trend. Then the limiting distribution of the Wald statistic is the F distribution with the denominator degrees of freedom $K - 2d$. For this F distribution to have a finite variance, we require $K - 2d \geq 5$, i.e., $K \geq 2d + 5$. So in finite samples, it is reasonable to set K equal to $K^*_{MSE,c}$ with

$$K^*_{MSE,c} = \max(2d + 5, K^*_{MSE}). \tag{38}$$

When a linear trend is included, we make an obvious adjustment and set K equal to the following $K^*_{MSE,c}$:

$$K^*_{MSE,c} = \max(2d + 6, K^*_{MSE}).$$

There is another reason to avoid a large K. Cointegration is fundamentally a long run relationship. To estimate the cointegrating vector, we should employ a regression that uses only the long run variation of the underlying variables. The short run variation can help only when the short run relationship coincides with the long run relationship. If the two types of relationships differ from each other, then going beyond a reasonable value of K runs the risk of being struck by short run contaminations. A trade-off between the asymptotic efficiency and robustness with respect to short run contaminations leads us to consider a moderate K value.

Under the cosine and sine basis functions given in (36), the transformed data consist of the real and imaginary parts of the Discrete Fourier Transforms (DFT) of the original data. In this case, a useful rule of thumb choice is provided in Müller (2014) and Müller and Watson (2013). These papers propose to select a K value to reflect business cycle frequencies or below. For example, with $T = 64$ years of post-World-War-II macro data, the choice of $K = 16$ value captures the long run movements of macro data lower than the commonly accepted business cycle period of $T/(K/2) = 8$ years.

4 Cointegration Analysis with Cosine Bases

4.1 Model without time trend

We go back to the model without an intercept and time trend, i.e., the model in (18), but we drop the augmented term in (19) and consider the following equation

$$\mathbb{W}_i^y = \mathbb{W}_i^{x\prime}\beta_0 + \mathbb{W}_i^0 \tag{39}$$

where by definition

$$\mathbb{W}_i^0 = \mathbb{W}_i^{\Delta x\prime}\delta_0 + \mathbb{W}_i^{0\cdot x} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{0t}\phi_i\left(\frac{t}{T}\right).$$

Define the transformed OLS (TOLS) estimator $\hat{\beta}_{TOLS}$ to be

$$\hat{\beta}_{TOLS} = \left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1}\mathbb{W}^{x\prime}\mathbb{W}^y. \tag{40}$$

In general, dropping $\mathbb{W}_i^{\Delta x}$ in (19) will lead to an omitted variable bias unless the correlation between $\mathbb{W}^{\Delta x}$ and $\mathbb{W}^x$ is zero. The zero correlation is ensured by the following assumption.

Assumption 3 *The basis functions satisfy* $\int_0^1 \Psi_i(r)\phi_j(r)\,dr = 0$ *with* $\Psi_i(r) = \int_0^r \phi_i(s)\,ds$ *for* $i, j = 1, ..., K$.

Recall that

$$\begin{aligned}\frac{\mathbb{W}_i^x}{T} &\Rightarrow \Omega_{xx}^{1/2}\int_0^1 \phi_i(r)W_x(r)dr = -\Omega_{xx}^{1/2}\int_0^1 \Psi_i(r)dW_x(r),\\ \mathbb{W}_j^0 &\Rightarrow \sigma_{0\cdot x}\int_0^1 \phi_j(r)\,dw_0(r) + \sigma_{0x}\Omega_{xx}^{-1/2}\int_0^1 \phi_j(r)dW_x(r),\end{aligned} \tag{41}$$

for $i, j = 1, ..., K$ where $w_0(r)$ and $W_x(r)$ are independent Brownian motion processes. The asymptotic distribution of $(\mathbb{W}_i^x/T, \mathbb{W}_j^0)$ is jointly normal with covariance

$$\text{cov}\left(\int_0^1 \phi_i(r)B_x(r)dr, \int_0^1 \phi_j(r)dB_0(r)dr\right) = -\Omega_{xx}^{1/2}\left(\int_0^1 \Psi_i(r)\phi_j(r)\,dr\right)\Omega_{xx}^{-1/2}\sigma_{x0}.$$

Thus, $T^{-1}\mathbb{W}_i^x$ and $\mathbb{W}_j^{\Delta x}$ are asymptotically independent if the basis functions satisfy Assumption 3.

Lemma 5 *The cosine functions*

$$\phi_j^c(r) = \sqrt{2}\cos(2j\pi r) \ \ \textit{for } j = 1, ..., K \tag{42}$$

satisfy Assumptions 1 and 3.

The lemma not only shows that Assumption 3 can hold but also gives the set of simple and commonly-used cosine functions as an example. Although there may be other functions that satisfy Assumption 3, we have the cosine functions in mind when developing the asymptotic results in this section. We are not aware of other commonly-used basis functions that also satisfy Assumption 3.

Theorem 6 *Consider the model in (18). Let Assumptions 1–3 hold. Under the fixed-K asymptotics, we have*

$$T(\hat{\beta}_{TOLS} - \beta_0) \Rightarrow MN\left(0, \sigma_0^2 \Omega_{xx}^{-1/2} (\eta'\eta)^{-1} \Omega_{xx}^{-1/2}\right).$$

It is interesting to see that the transformed OLS estimator is asymptotically unbiased and mixed normal. To some extent, the use of the special basis functions such as the cosine functions kills two birds with one stone. There is no need to augment the original regression in order to achieve the asymptotic mixed normality.

Given the mixed normality of the limiting distribution, it is reasonable to make inference based on $\hat{\beta}_{TOLS}$. The Wald statistic and t statistic are

$$F_c(\hat{\beta}_{TOLS}) = \frac{1}{(\hat{\sigma}_0^c)^2} (R\hat{\beta}_{TOLS} - r)' \left[R\left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1} R'\right]^{-1} (R\hat{\beta}_{TOLS} - r)/p, \tag{43}$$

$$t_c(\hat{\beta}_{TOLS}) = \frac{R\hat{\beta}_{TOLS} - r}{\hat{\sigma}_0^c \sqrt{R\left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1} R'}}, \tag{44}$$

where

$$(\hat{\sigma}_0^c)^2 = \frac{1}{K}\sum_{i=1}^{K} \left(\hat{\mathbb{W}}_{ci}^0\right)^2 \text{ where } \hat{\mathbb{W}}_{ci}^0 = \mathbb{W}_i^y - \mathbb{W}_i^{x\prime}\hat{\beta}_{TOLS}.$$

Following a proof similar to that of Theorem 3, we can show that

$$F_c(\hat{\beta}_{TOLS}) \Rightarrow \frac{K}{K-d} \cdot F_{p,K-d} \text{ and } t_c(\hat{\beta}_{TOLS}) \Rightarrow \sqrt{\frac{K}{K-d}} \cdot t(K-d).$$

The above results are clearly analogous to the well-known results in a CLNR with K iid observations and d regressors.

4.2 Model with a linear trend

We consider the cointegration system with a linear trend as given in (29). Dropping the regressors $\mathbb{W}_i^{\Delta x}$ and $\mathbb{W}_i^{tr}$ in (30), we obtain

$$\mathbb{W}_i^y = \mathbb{W}_i^{x\prime}\beta_0 + \left(\mathbb{W}_i^{tr}\mu_0 + \mathbb{W}_i^0\right) \text{ for } i = 1, ..., K \tag{45}$$

where $\mathbb{W}_i^{tr}\mu_0 + \mathbb{W}_i^0 = \mathbb{W}_i^{tr}\mu_0 + \mathbb{W}_i^{\Delta x\prime}\delta_0 + \mathbb{W}_i^{0\cdot x}$ is the composite error. In general, the transformed OLS estimator obtained by regressing $\mathbb{W}_i^y$ on $\mathbb{W}_i^x$ is not consistent even if cosine transforms are used. The reason is that the composite error is not mean zero and is correlated with the included regressor. In fact,

$$\begin{aligned}
\mathbb{W}_i^{tr} &= \frac{\sqrt{2}}{\sqrt{T}} \sum_{t=1}^{T} t\cos\left(\frac{2\pi i t}{T}\right) = T\sqrt{2T}\left[\frac{1}{T}\sum_{t=1}^{T} \frac{t}{T}\cos\left(\frac{2\pi i t}{T}\right)\right] \\
&= T\sqrt{2T}\left[\int_0^1 r\cos(2\pi i r)\,dr + O\left(\frac{1}{T}\right)\right] = O(\sqrt{T})
\end{aligned} \tag{46}$$

using $\int_0^1 r\cos(2\pi i r)\,dr = 0$. So the composite error grows with the sample size at the rate of $\sqrt{T}$, and as a result the transformed OLS estimator obtained in the absence of the trend term is not consistent.

A simple way to fix this problem is to use shifted cosine transforms. Let

$$\phi^c_{Ti}(r) = \phi^c_i(r - \frac{1}{2T}) = \sqrt{2}\cos\left(2\pi i(r - \frac{1}{2T})\right) \text{ for } i = 1, ..., K \tag{47}$$

be the finite sample shifted version of $\{\phi^c_i(r)\}_{i=1}^K$[2]. We define

$$\breve{\mathbb{W}}^v_i = \frac{1}{\sqrt{T}}\sum_{t=1}^T v_t\phi^c_{Ti}\left(\frac{t}{T}\right) \text{ for } v = y,\ x,\ \Delta x \text{ and} \tag{48}$$

$$\breve{\mathbb{W}}^{tr}_i = \frac{1}{\sqrt{T}}\sum_{t=1}^T t\phi^c_{Ti}\left(\frac{t}{T}\right),\ \breve{\mathbb{W}}^0_i = \frac{1}{\sqrt{T}}\sum_{t=1}^T u_{0t}\phi^c_{Ti}\left(\frac{t}{T}\right). \tag{49}$$

It follows from Lemma 8 in Bierens (1997) that $\breve{\mathbb{W}}^{tr}_i = 0$ for any $i = 1, ..., K$. Also, it is easy to show that $T^{-1/2}\sum_{t=1}^T \phi^c_{Ti}(t/T) = 0$ for all $i = 1, ..., K$. So utilizing $\{\phi^c_{Ti}(r)\}_{i=1}^K$ as the basis functions filters out both the intercept and linear trend in the original equation (29)[3]. As a result, we have

$$\breve{\mathbb{W}}^y_i = \breve{\mathbb{W}}^{x\prime}_i\beta_0 + \breve{\mathbb{W}}^0_i \text{ for } i = 1, ..., K. \tag{50}$$

On the basis of this equation, the transformed OLS estimator of β_0 is given by

$$\hat{\beta}_{TOLS} = \left(\breve{\mathbb{W}}^{x\prime}\breve{\mathbb{W}}^x\right)^{-1}\breve{\mathbb{W}}^{x\prime}\breve{\mathbb{W}}^y.$$

In view of $\phi^c_{Ti}(r) = \phi^c_i(r) + O(1/T)$ uniformly for $r \in [0, 1]$, we have

$$\begin{aligned}\breve{\mathbb{W}}^x_i &\Rightarrow \Omega^{1/2}_{xx}\eta_i \overset{d}{=} \Omega^{1/2}_{xx}\int_0^1 \phi^c_i(r)\, W_x(r)dr,\\ \breve{\mathbb{W}}^{\Delta x}_j &\Rightarrow \Omega^{1/2}_{xx}\xi_j \overset{d}{=} \Omega^{1/2}_{xx}\int_0^1 \phi^c_j(r)\, dW_x(r),\end{aligned} \tag{51}$$

for $i, j = 1, ..., K$, and $\breve{\mathbb{W}}^x_i$ and $\breve{\mathbb{W}}^{\Delta x}_j$ are asymptotically independent. Using these and the same proof for Theorem 6, we can prove the theorem below.

Theorem 7 *Consider the model in (29). Let Assumptions 1 and 2 hold. Suppose that the shifted cosine transforms are used. Then under the fixed-K asymptotics,*

$$T(\hat{\beta}_{TOLS} - \beta_0) \Rightarrow MN\left(0, \sigma^2_0\Omega^{-1/2}_{xx}(\eta'\eta)^{-1}\Omega^{-1/2}_{xx}\right).$$

It follows from the theorem and the arguments similar to the proof of Theorem 3 that

$$F_c(\hat{\beta}_{TOLS}) \Rightarrow \frac{K}{K-d}\cdot F_{p,K-d} \text{ and } t_c(\hat{\beta}_{TOLS}) \Rightarrow \sqrt{\frac{K}{K-d}}\cdot t(K-d), \tag{52}$$

where $F_c(\hat{\beta}_{TOLS})$ and $t_c(\hat{\beta}_{TOLS})$ are defined in the same way as in (43) and (44).

[2] The cosine weight functions $\phi^c_{Ti}(t/T)$ are known as Chebishev time polynomials of even orders. See Hamming (1973) for details. Bierens (1997) shows that the cosine basis functions enjoy a certain optimality property for hypothesis testing.

[3] Sun (2011) also uses the cosine basis functions in OS LRV estimation in order to achieve invariance with respect to the intercept and linear trend.

For a cointegration model without a trend, it is not hard to show that Theorem 6 and the asymptotic results for the test statistics thereafter remain the same if we use the shifted cosine transforms in place of the original cosine transforms. That is, for a cointegration model without a trend, it does not matter asymptotically whether the shifted cosine transforms or the original cosine transforms are employed. However, the shifted cosine transforms lead to the TOLS estimator that is invariant to the presence of a linear trend. This is a nice property that is not enjoyed by the original cosine transforms. For this reason, the shifted cosine transforms are preferred over the original cosine transforms.

4.3 Augment or not: asymptotic efficiency comparison

Suppose that we use the shifted cosine transforms. Regardless of whether there is a linear time trend, we have two different estimators of β_0, both of which are asymptotically mixed normal. The first one is the TAOLS estimator and the second one is the TOLS estimator. The difference is whether the underlying regression is augmented or not. In this subsection, we address the relative efficiency of the two estimators.

For the model without a time trend, it follows from (22) and Theorem 6 that the asymptotic variances of $\hat{\beta}_{TAOLS}$ and $\hat{\beta}_{TOLS}$ conditioning on (η, ξ) are

$$V_{TAOLS} = \sigma_{0\cdot x}^2 \Omega_{xx}^{-1/2} (\eta' M_\xi \eta)^{-1} \Omega_{xx}^{-1/2}, \tag{53}$$

$$V_{TOLS} = \sigma_0^2 \Omega_{xx}^{-1/2} (\eta' \eta)^{-1} \Omega_{xx}^{-1/2}, \tag{54}$$

where we call that $\eta = (\eta_1, ..., \eta_K)'$, $\xi = (\xi_1, ..., \xi_K)'$, $\eta_i = \int \phi_i^c(r) W_x(r) dr$ and $\xi_i = \int \phi_i^c(r) dW_x(r)$.

For the model with a linear time trend, we know that $a = 0$ in Theorem 4. So no transformed time trend can be included in the transformed and augmented regression. In this case, we can follow the same proof of Theorem 4 and show that the asymptotic variance of $\hat{\beta}_{TAOLS}$ is $\sigma_{0\cdot x}^2 \Omega_{xx}^{-1/2} (\eta' M_\xi \eta)^{-1} \Omega_{xx}^{-1/2}$. On the other hand, the asymptotic variance of $\hat{\beta}_{TOLS}$ is $\sigma_0^2 \Omega_{xx}^{-1/2} (\eta' \eta)^{-1} \Omega_{xx}^{-1/2}$ as indicated by Theorem 7. That is, the asymptotic variance formulae in (53) and (54) hold regardless of whether a linear trend is included in the model or not.

For any conforming vector $c \in \mathbb{R}^d$, we have

$$\begin{aligned} & c'(V_{TAOLS}^{-1} - V_{TOLS}^{-1})c \\ &= \frac{c'\Omega_{xx}^{1/2}}{\sigma_{0\cdot x}} (\eta' M_\xi \eta - \frac{\sigma_{0\cdot x}^2}{\sigma_0^2} \eta'\eta) \frac{\Omega_{xx}^{1/2} c}{\sigma_{0\cdot x}} \\ &= \frac{c'\Omega_{xx}^{1/2}}{\sigma_{0\cdot x}} \left[\eta' (I_K - P_\xi) \eta - \left(\frac{\sigma_0^2 - \sigma_{0x}\Omega_{xx}^{-1}\sigma_{x0}}{\sigma_0^2} \right) \eta'\eta \right] \frac{\Omega_{xx}^{1/2} c}{\sigma_{0\cdot x}} \\ &= \frac{c'\Omega_{xx}^{1/2}\eta'}{\sigma_{0\cdot x}} \left[I_K \cdot \left(\frac{\sigma_{0x}\Omega_{xx}^{-1}\sigma_{x0}}{\sigma_0^2} \right) - P_\xi \right] \frac{\eta \Omega_{xx}^{1/2} c}{\sigma_{0\cdot x}} \\ &= \tilde{c}' \left[I_K \cdot \varrho^2 - P_\xi \right] \tilde{c} \end{aligned} \tag{55}$$

where $\tilde{c} = \eta \Omega_{xx}^{1/2} c / \sigma_{0\cdot x}$, $P_\xi = \xi(\xi'\xi)^{-1}\xi'$, and

$$\varrho^2 = \frac{\sigma_{0x}\Omega_{xx}^{-1}\sigma_{x0}}{\sigma_0^2} = \arg\max_\ell \left(\frac{\ell'\sigma_{x0}}{\sqrt{\ell'\Omega_{xx}\ell}\sigma_0} \right)^2 \in [0, 1]. \tag{56}$$

By definition, ϱ^2 is the squared long run canonical correlation coefficient between u_{0t} and u_{xt}. If $\varrho^2 = 0$, then $c'(V_{TAOLS}^{-1} - V_{TOLS}^{-1})c = -\tilde{c}' P_\xi \tilde{c} \leq 0$ almost surely. In this case, the asymptotic variance of $\hat{\beta}_{TAOLS}$ is always larger than the asymptotic variance of $\hat{\beta}_{TOLS}$. Intuitively, when the long run canonical correlation between u_{0t} and u_{xt} is zero, including the additional regressor $\mathbb{W}^{\Delta x}$ will not help reduce the size of the error term in the transformed regression. However, the presence of $\mathbb{W}^{\Delta x}$ reduces the strength of the signal in $\mathbb{W}^{x}$ even though they are asymptotically independent. That is why $\hat{\beta}_{TAOLS}$ is asymptotically less efficient. On the other hand, when $\varrho^2 = 1$, which holds if the long run variation of u_{0t} can be perfectly predicted by u_{xt}, we have $c'(V_{TAOLS}^{-1} - V_{TOLS}^{-1})c = \tilde{c}'(I_K - P_\xi)\tilde{c} \geq 0$ almost surely. In this case, the benefit of including the additional regressor $\mathbb{W}^{\Delta x}$ outweighs the cost, and it is worthwhile to include $\mathbb{W}^{\Delta x}$ to get the asymptotically more efficient estimator $\hat{\beta}_{TAOLS}$.

There are many scenarios between these two extreme cases. Whether the asymptotic distribution of $\hat{\beta}_{TAOLS}$ has a larger variance than that of $\hat{\beta}_{TOLS}$ depends on the value of ϱ^2.

Proposition 8 *If $\varrho^2 \geq d/K$, then $\hat{\beta}_{TAOLS}$ has a smaller asymptotic variance than $\hat{\beta}_{TOLS}$, i.e., asymvar$(\hat{\beta}_{TAOLS}) - \text{asymvar}(\hat{\beta}_{TAOLS})$ is negative semidefinite. Otherwise, $\hat{\beta}_{TAOLS}$ has a larger asymptotic variance than $\hat{\beta}_{TOLS}$.*

4.4 AMSE Rule

For the cosine basis function $\{\sqrt{2}\cos 2i\pi r\}_{i=1}^{K}$ we can follow Phillips (2005) and Sun (2011) and show that the AMSE-optimal K^* is given by

$$\begin{aligned} K_{MSE}^{c*} &= \left\lceil \left(\frac{1}{16} \frac{tr(I_{m^2} + \mathbb{K}_{mm})(\Omega \otimes \Omega)}{4vec(B)' \, vec(B)} \right)^{1/5} T^{4/5} \right\rceil \\ &\simeq \left(\frac{1}{16} \right)^{1/5} K_{MSE}^* = K_{MSE}^*(0.57) \end{aligned} \quad (57)$$

where K_{MSE}^* is the AMSE-optimal K given in (37) for the basis functions given in (36). Following the same argument for (38), we recommend making an adjustment in finite samples and set K equal to $\max(K_{MSE}^{c*}, d+5)$.

Given the smaller choice of K, the use of cosine basis functions rather than the complete cosine and sine basis functions may lead to a less efficient estimator of β_0. However, the cosine basis functions enjoy two advantages that the complete basis functions do not. First, it automatically filters out the time trend regressor so that we do not have to worry about the estimation error in trend extraction. Second, the use of cosine basis function renders it unnecessary in some scenarios to include the first difference regressor in the regression and thus saves some degrees of freedom. These two advantages may offset the efficiency loss from having to select a smaller K.

5 Simulation Study

5.1 DGP without time trend

We compare the finite sample performance of our method with several existing methods in the literature. Our first DGP is a cointegration regression model without a time trend. We follow Phillips (2014) and consider:

$$\begin{matrix} y_t = \alpha_0 + x_t'\beta_0 + u_{0t} \\ x_t = x_{t-1} + u_{xt} \end{matrix}, \; u_t = \begin{pmatrix} u_{0t} \\ u_{xt} \end{pmatrix} = \Theta u_{t-1} + \epsilon_t \quad (58)$$

where

$$\epsilon_t = \begin{pmatrix} \epsilon_{0t} \\ \epsilon_{xt} \end{pmatrix} \sim \text{i.i.d } N(0, \Sigma), \ \Theta = \rho \cdot I_{d+1}, \ \Sigma = J_{d+1,d+1} \cdot \varphi + I_{d+1} \cdot (1 - \varphi)$$

and $J_{p,q}$ is the $p \times q$ matrix of ones. The parameter ρ controls the persistence of individual components in $u_t = (u_{0t}, u_{xt}')' \in \mathbb{R}^{d+1}$ and the second parameter φ characterizes comovements among the components of u_t. The dimension d of x_t is set to be 2, and the true coefficients are set to be $\alpha_0 = 3$ and $\beta_0 = (1, 1)'$.

We are interested in testing $H_0 : \beta_0 = (1, 1)'$ vs $H_1 : \beta_0 \neq (1, 1)'$. We consider the Wald type of tests based on four different estimators: the FM-OLS estimator of Phillips and Hansen (1990), the TIV estimator of Phillips (2014), the IM-OLS estimator by Vogelsang and Wagner (2014), and the TAOLS estimator proposed in this paper. The first two tests employ the increasing-smoothing asymptotic approximation and use chi-square critical values. The IM-OLS test employs the fixed-b asymptotic approximation with simulated critical values[4]. The TAOLS test employs the fixed-K asymptotic approximation and scaled standard F critical values.

For the FM-OLS and IM-OLS methods, we consider the Bartlett, Parzen and Quadratic Spectral kernels with the smoothing parameter b selected by the data-driven method given in Andrews (1991). The plug-in model used is the VAR(1). For the TIV and TAOLS estimators, we consider the cosine and sine basis functions given in (36) with K selected based on the formula in (37). The results reported here are obtained without making the adjustment given in (38). However, the lower bound of $K \geq 2d + 1$ is imposed.

Figures 1 and 2 and Table 1 report the empirical size of different tests for

$$\rho \in \{0.05, 0.20, 0.35, 0.50, 0.70, 0.90\} \text{ and } \varphi = 0.75.$$

The empirical size is computed using 10,000 simulation replications. The nominal size of all tests is 5%. Table 2 reports the average of the data-driven smoothing parameters. It is clear that, for all values for ρ and sample sizes $T \in \{100, 200\}$, the TAOLS test with F critical values outperforms all other tests by a large margin. For example, when $\rho = 0.9$ and $T = 200$, the empirical size of the tests based on the FM-OLS (Bartlett), IM-OLS (Bartlett) and TIV estimators is reported to be as high as 74%, 35% and 72%, respectively. There is some reduction in size distortion when other kernels are employed for the FM-OLS and IM-OLS methods: 45% for FM-OLS (QS) and 17% for IM-OLS (QS), but the size distortion is still substantial. In contrast, our proposed F test has either no size distortion or small size distortion. Simulation results not reported here show that using the F critical values can also dramatically reduce the size distortion of the TIV test. Our findings are consistent with the literature on heteroskedasticity and autocorrelation robust (HAR) inference such as Sun (2013, 2014a), Sun, Phillips, and Jin (2008), and Kiefer and Vogelsang (2005) which provide theoretical justifications and simulation evidence on the accuracy of the fixed-smoothing approximations.

Next, we investigate the finite sample power of each procedure. The power is size-adjusted so that the comparison is meaningful. The DGP's are the same except that the parameters of interest are from the local alternative hypothesis $\beta = \beta_0 + \theta/T$. The choice rules for K and b are also the same as before. Each power curve is drawn against $\|\theta\|$, which measures the magnitude of the local departure. Figures 3–6 present the size-adjusted power curve of each procedure for

[4]In our simulation, we use the simulated critical values reported by Vogelsang and Wagner (2014) in their supplementary materials.

$\rho \in \{0.35, 0.50, 0.70, 0.90\}$, $\varphi = 0.75$ and $T = 100, 200$. The results are briefly summarized as follows.

First, the FM-OLS test with second-order kernels such as Parzen and QS kernels yields the highest power under all DGP's we consider. The TAOLS method outperforms the IM-OLS method when $\rho = 0.35$ and 0.50, but it starts to under-perform the IM-OLS when the dependence becomes strong, i.e., $\rho = 0.75$ and 0.90. It is not surprising that the FM-OLS method achieves higher power because it effectively uses both low and high frequency components to estimate the cointegrating vector with modification only in the second stage. However, the FM-OLS method can be fragile if there are high frequency contaminations. In addition, the FM-OLS test has very large size distortion. For example, for $T = 200$, the empirical size of the FM-OLS test with the QS kernel is 26% when $\rho = 0.75$. It increases to 45% when $\rho = 0.90$.

Second, the power of the TAOLS test is lower especially when the dependence is strong such as $\rho = 0.75$ and 0.90. This can be explained by the small K values selected by the AMSE rule. According to Table 2, the average values of K's are 7.12 ($\rho = 0.75$) and 6.00 ($\rho = 0.90$) which are very close to the lower bound of the admissible values for K[5]. So, even though using a small K gives us very successful size control in finite samples, there is a power loss.

Third, Figures 7 and 8 show that the power of the TAOLS test increases, as K increases. The power starts to dominate that of the FM-OLS method as K crosses some threshold value. For example, when $\rho = 0.75$, with $K = 24$ the power of the TAOLS test is slightly higher than that of the FM-OLS (QS) test. When $\rho = 0.90$, the TAOLS test becomes more powerful than the FM-OLS test when K increases to 10, which is close to the lower bound of $2d + 5$ given in (38). There is always a trade-off between power improvement and size distortion. Simulation results not reported here show that the empirical size of the TAOLS test under $\rho = 0.90$ increases from 6% to 10% when K increases from 6 to 10. That is, had we used the adjusted formula in (38), we would have obtained a test that is as nearly powerful as the FM-OLS test. The cost of doing so is the increase of size distortion ranging from 1 percentage point to 5 percentage points. This is a relatively small cost comparing with the size distortion of 40% for the FM-OLS(QS) test.

To sum up, when we use the data-driven K given in (37), the TAOLS-based F test is remarkably accurate. It is much more accurate than the FM-OLS and TIV tests that use the chi-square approximation. It is more accurate than the IM-OLS test, which also uses a fixed-smoothing approximation. However, the size accuracy is achieved at the cost of some power loss, especially when the process is highly autocorrelated. When we use the adjusted K given in (38), the TAOLS-based F test becomes as powerful as the FM-OLS test, but there is some sacrifice in size accuracy. However, the size distortion is still much lower than that of the FM-OLS chi-square test. Depending on our tolerance towards size distortion, we may use either (37) or (38) to select K.

5.2 DGP with a time trend

The second data generating process generalizes (58) by including a linear time trend:

$$\begin{aligned} y_t &= \alpha_0 + \mu_0 t + x_t' \beta_0 + u_{0t} \\ x_t &= x_{t-1} + u_{xt} \end{aligned}$$

[5] It follows from Theorem 3 that K has to be larger than $2d + 1$, which is 5 in the present setting.

where the true parameters are $(\alpha_0, \mu_0, \beta_0') = (3, 0.05, 1, 1)$. We fix ρ at 0.75 and consider $\varrho^2 \in \{0.05, 0.20, 0.35, 0.50, 0.75, 0.90\}$. It is straightforward to obtain the LRV Ω of u_t as

$$\Omega = (I - \Theta)^{-1} \Sigma (I - \Theta)^{-1\prime} = \left(\frac{1}{1-\rho}\right)^2 \cdot \Sigma$$
$$= \left(\frac{1}{1-\rho}\right)^2 \cdot \begin{pmatrix} 1 & \varphi \cdot J_{1,d} \\ \varphi \cdot J_{d,1} & J_{d,d} \cdot \varphi + I_d \cdot (1-\varphi) \end{pmatrix}. \quad (59)$$

It then follows that

$$\varrho^2 = \frac{d\varphi^2}{1 + \varphi (d-1)}.$$

From the above formula, we can back out the value of φ that produces the desired value of ϱ^2.

As before, we compare the performances of various Wald type tests. For the FM-OLS, IM-OLS and TAOLS methods, we consider the same tests as in the previous subsection except that we take the linear trend into account. We also consider the shifted cosine transforms, leading to the TAOLS-C and TOLS-C estimators and the corresponding tests where the underlying transformed regressions do not include a trend term.

Table 3 and Figures 9 and 10 report the empirical size of each test. We observe the dominating finite sample performances of TAOLS, TOLS-C and TAOLS-C compared to other testing methods such as FM-OLS and IM-OLS. The decent performances of TAOLS-C and TOLS-C methods indicate that the shifted cosine transforms successfully filter out the linear trend and remove the endogeneity bias.

Figure 11 compares the finite sample power performances of TAOLS-C and TOLS-C when $T = 200$. The simulation evidence is consistent with our theoretical results in Section 4.3: as ϱ^2 decreases, the efficiency of TAOLS-C relative to TOLS-C decreases. So, the power curves of TOLS-C and TAOLS-C cross each other at a certain value $\varrho^2 \in (0, 1)$.

6 Conclusion

This paper provides a simple, robust and more accurate approach to parameter estimation and inference in a triangular cointegrating system. Cointegration is fundamentally a long run relationship. Our approach echoes this key observation by focusing only on data transformations that capture the long run variation and covariation of the underlying time series. In this respect, our approach resembles the frequency domain approach that uses only low-frequency information, but it avoids the complications of frequency domain techniques. From a practical point of view, our approach enjoys two major advantages. First, the more accurate approximations we derived under the so-call fixed-K asymptotics are the standard F and t distributions. Second, test statistics can be obtained from the usual regression output. So our asymptotic F and t tests are just as easy to implement as the F and t tests in a classical linear normal regression. A simulation study shows that our tests are much more accurate than the chi-square tests.

A key open question is how to select the number of basis functions. While we have suggested a data-driven approach, it does not directly target at the problem under consideration. It will be interesting to select the number of basis functions to minimize the approximate mean squared error of the point estimator of the cointegration vector. If we are interested in interval estimation or hypothesis testing, then the number of basis functions should be oriented at optimizing the underlying objects such as the coverage probability error, the interval length, and the type I and type II errors. There may also be room to select optimal basis functions. We hope to address some of these questions in future research.

7 Appendix of Proofs

Proof of Theorem 1. By the definition of $\hat{\gamma}_{TAOLS}$ and Υ_T, we have

$$\Upsilon_T\left(\hat{\gamma}_{TAOLS}-\gamma_0\right)=(\Upsilon_T^{-1}\tilde{\mathbb{W}}'\tilde{\mathbb{W}}\Upsilon_T^{-1})^{-1}\Upsilon_T^{-1}\tilde{\mathbb{W}}'\mathbb{W}^{0\cdot x}. \quad (60)$$

Note that $\tilde{\mathbb{W}}\Upsilon_T^{-1}=\left(\mathbb{W}^x/T,\,\mathbb{W}^{\Delta x}\right)$ where

$$\mathbb{W}^x/T=(\mathbb{W}_1^x/T,...,\mathbb{W}_K^x/T)'=\left(\frac{1}{T^{3/2}}\sum_{t=1}^T\phi_1(\frac{t}{T})x_t,...,\frac{1}{T^{3/2}}\sum_{t=1}^T\phi_K(\frac{t}{T})x_t\right)'$$

and

$$\mathbb{W}^{\Delta x}=\left(\mathbb{W}_1^{\Delta x},...,\mathbb{W}_K^{\Delta x}\right)'=\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T\phi_1(\frac{t}{T})u_{xt},...,\frac{1}{\sqrt{T}}\sum_{t=1}^T\phi_K(\frac{t}{T})u_{xt}\right)'.$$

By Assumption 1 and the continuous mapping theorem,

$$\frac{1}{T^{3/2}}\sum_{t=1}^T\phi_i(\frac{t}{T})x_t\Rightarrow\Omega_{xx}^{1/2}\left(\int_0^1\phi_i(r)W_x(r)dr\right)=\Omega_{xx}^{1/2}\eta_i, \quad (61)$$

$$\text{and } \frac{1}{\sqrt{T}}\sum_{t=1}^T\phi_j(\frac{t}{T})u_{xt}\Rightarrow\Omega_{xx}^{1/2}\left(\int_0^1\phi_j(r)dW_x(r)\right):=\Omega_{xx}^{1/2}\xi_j, \quad (62)$$

hold jointly over $i,j=1,...,K$. So

$$\mathbb{W}^x/T\Rightarrow(\Omega_{xx}^{1/2}\eta')'\text{ and }\mathbb{W}^{\Delta x}\Rightarrow(\Omega_{xx}^{1/2}\xi')',$$

and

$$\tilde{\mathbb{W}}\Upsilon_T^{-1}\Rightarrow\left((\Omega_{xx}^{1/2}\eta')',(\Omega_{xx}^{1/2}\xi')'\right)=\left(\eta\Omega_{xx}^{1/2},\xi\Omega_{xx}^{1/2}\right). \quad (63)$$

Similarly, we have

$$\begin{aligned}\mathbb{W}^{0\cdot x}&=\left(\frac{1}{T^{1/2}}\sum_{t=1}^T\phi_1(\frac{t}{T})u_{0\cdot xt},...,\frac{1}{T^{1/2}}\sum_{t=1}^T\phi_K(\frac{t}{T})u_{0\cdot xt}\right)'\\&\Rightarrow(\sigma_{0\cdot x}\nu_1,\sigma_{0\cdot x}\nu_2,...,\sigma_{0\cdot x}\nu_K)'=\sigma_{0\cdot x}\nu \quad (64)\end{aligned}$$

where $\nu=[\nu_1,...,\nu_K]'\sim N(0,I_K)$. The above convergence holds jointly with (63), i.e.,

$$\left(\tilde{\mathbb{W}}\Upsilon_T^{-1},\mathbb{W}^{0\cdot x}\right)\Rightarrow\left(\tilde{\zeta},\tilde{\nu}\right)\text{ where }\tilde{\zeta}=\left(\eta\Omega_{xx}^{1/2},\xi\Omega_{xx}^{1/2}\right),\ \tilde{\nu}:=\sigma_{0\cdot x}\nu\text{ and }\tilde{\zeta}\perp\tilde{\nu}. \quad (65)$$

Using this result, we have

$$\begin{aligned}\Upsilon_T\left(\hat{\gamma}_{TAOLS}-\gamma_0\right)&=(\Upsilon_T^{-1}\tilde{\mathbb{W}}'\tilde{\mathbb{W}}\Upsilon_T^{-1})^{-1}\Upsilon_T^{-1}\tilde{\mathbb{W}}'\mathbb{W}^{0\cdot x}+o_p\left(1\right)\\&\Rightarrow(\tilde{\zeta}'\tilde{\zeta})^{-1}\tilde{\zeta}'\tilde{\nu}\overset{d}{=}MN\left[0,\sigma_{0\cdot x}^2(\tilde{\zeta}'\tilde{\zeta})^{-1}\right].\end{aligned}$$

The weak limit can be written more explicitly as

$$
\begin{aligned}
(\tilde{\zeta}'\tilde{\zeta})^{-1}\tilde{\zeta}'\tilde{\nu} &= \sigma_{0\cdot x}\begin{pmatrix} \Omega_{xx}^{1/2}\eta'\eta\Omega_{xx}^{1/2} & \Omega_{xx}^{1/2}\eta'\xi\Omega_{xx}^{1/2} \\ \Omega_{xx}^{1/2}\xi'\eta\Omega_{xx}^{1/2} & \Omega_{xx}^{1/2}\xi'\xi\Omega_{xx}^{1/2} \end{pmatrix}^{-1}\begin{pmatrix} \Omega_{xx}^{1/2}\eta' \\ \Omega_{xx}^{1/2}\xi' \end{pmatrix}\nu \\
&= \sigma_{0\cdot x}\begin{pmatrix} \Omega_{xx}^{-1/2} & 0 \\ 0 & \Omega_{xx}^{-1/2} \end{pmatrix}\begin{pmatrix} \eta'\eta & \eta'\xi \\ \xi'\eta & \xi'\xi \end{pmatrix}^{-1}\begin{pmatrix} \eta' \\ \xi' \end{pmatrix}\nu \\
&= \begin{pmatrix} \sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\eta' M_{\xi}\eta\right)^{-1}\eta' M_{\xi}\nu \\ \sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\xi' M_{\eta}\xi\right)^{-1}\xi' M_{\eta}\nu \end{pmatrix}. \qquad (66)
\end{aligned}
$$

So the representations in (22) and (23) hold. ■

Proof of of Proposition 2. Part (i). By Lemma A in Section 6 of Phillips (2005), we have

$$\Phi'\Phi = I_K + O(\frac{1}{T}) \text{ and } (\Phi'\Phi)^{-1} = I_K + O(\frac{1}{T})$$

for any fixed K. Then, it is straightforward to show that

$$P_{\Phi} = \Phi\left(\Phi'\Phi\right)^{-1}\Phi' = \Phi\left(I_K + O(T^{-1})\right)\Phi' = \Phi\Phi' + O(T^{-1}).$$

Now

$$T(\hat{\beta}_{TIV} - \beta_0) = (X'S_{\Phi}X)^{-1}(X'S_{\Phi}U^{0\cdot x}) = \left(\frac{X'}{T^{3/2}}S_{\Phi}\frac{X}{T^{3/2}}\right)^{-1}\left(\frac{X'}{T^{3/2}}S_{\Phi}\frac{U^{0\cdot x}}{T^{1/2}}\right) \qquad (67)$$

where $U^{0\cdot x} = (u_{0\cdot x,1}, ..., u_{0\cdot x,T})'$ and $S_{\Phi} = P_{\Phi} - P_{\Phi}\Delta X(\Delta X' P_{\Phi}\Delta X)^{-1}\Delta X' P_{\Phi}$. Note that

$$
\begin{aligned}
S_{\Phi} &= \Phi\left\{\left(\Phi\Phi'\right)^{-1} - \left(\Phi\Phi'\right)^{-1}\frac{\Phi'\Delta X}{\sqrt{T}}\left[\left(\frac{\Delta X'\Phi}{\sqrt{T}}\right)\left(\Phi\Phi'\right)^{-1}\left(\frac{\Phi'\Delta X}{\sqrt{T}}\right)\right]^{-1}\frac{\Delta X'\Phi}{\sqrt{T}}\left(\Phi\Phi'\right)^{-1}\right\}\Phi' \\
&= \Phi\left\{I_K + O\left(T^{-1}\right) - \left[I_K + O\left(T^{-1}\right)\right]\mathbb{W}^{\Delta x}\left\{\mathbb{W}^{\Delta x\prime}\left[I_K + O\left(T^{-1}\right)\right]\mathbb{W}^{\Delta x}\right\}^{-1}\mathbb{W}^{\Delta x\prime}\left[I_K + O\left(T^{-1}\right)\right]\right\}\Phi' \\
&= \Phi\left(I_K - P_{\Delta x}\right)\Phi' + o_p(1),
\end{aligned}
$$

we have

$$
\begin{aligned}
\frac{X'}{T^{3/2}}S_{\Phi}\frac{X}{T^{3/2}} &= \frac{X'\Phi}{T^{3/2}}\left(I_K - P_{\Delta x} + o_p(1)\right)\frac{\Phi'X}{T^{3/2}} \\
&= \frac{\mathbb{W}^{x\prime}}{T}\left(I_K - P_{\Delta x}\right)\frac{\mathbb{W}^{x}}{T} + o_p(1),
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{X'}{T^{3/2}}S_{\Phi}\frac{U^{0\cdot x}}{T^{1/2}} &= \left(\frac{X'\Phi}{T^{3/2}}\right)'\left(I_K - P_{\Delta x} + o_p(1)\right)\frac{\Phi'U^{0\cdot x}}{\sqrt{T}} \\
&= \frac{\mathbb{W}^{x\prime}}{T}\left(I_K - P_{\Delta x}\right)\mathbb{W}^{0\cdot x} + o_p(1). \qquad (68)
\end{aligned}
$$

Combining these two representations and using Theorem 1, we have

$$
\begin{aligned}
T(\hat{\beta}_{TIV} - \beta_0) &= \left(\frac{\mathbb{W}^{x\prime}}{T}\left(I_K - P_{\Delta x}\right)\frac{\mathbb{W}^{x}}{T}\right)^{-1}\left(\frac{\mathbb{W}^{x\prime}}{T}\left(I_K - P_{\Delta x}\right)\right)\mathbb{W}^{0\cdot x} + o_p(1) \\
&= T(\hat{\beta}_{TAOLS} - \beta_0) + o_p(1).
\end{aligned}
$$

Part (ii): For any conformable vector c and $\tau \in \mathbb{R}$, we have

$$\lim_{K\to\infty} P\left(c'V_K < \tau\right) = \lim_{K\to\infty} EG\left(\frac{\tau}{\sigma_{0\cdot x}\sqrt{c'\left[\Omega_{xx}^{-1/2}(\eta' M_\xi \eta)^{-1}\Omega_{xx}^{-1/2}\right]c}}\right),$$

where $G(\cdot)$ is the cdf of the standard normal distribution. Note that $\int_0^1 W_x(r)W_x(r)'dr$ is positive definite with probability one. By the definition of weak convergence and the continuous mapping theorem, it suffices to show that

$$\eta' M_\xi \eta \Rightarrow \int_0^1 \tilde{W}_x(r)\tilde{W}_x(r)'dr \text{ as } K \to \infty.$$

We first show that $\eta'\eta \Rightarrow \int_0^1 \tilde{W}_x(r)\tilde{W}_x(r)'dr$. Given that $\{\phi_i(\cdot)\}_{i=1}^{\infty}$ is a complete orthonormal system in $L_0^2[0,1]$, we know that $\sum_{i=1}^K \phi_i(r)\phi_i(s)$ converges to the Dirac delta function in that

$$\left\| \int_0^1 \left(\sum_{i=1}^K \phi_i(r)\phi_i(s)\right) f(r)dr - f(s)\right\|_{L^2} \to 0 \tag{69}$$

for any $f \in L_0^2[0,1]$ as $K \to \infty$ where $\|\cdot\|_{L^2}$ is the L^2 norm. But

$$\begin{aligned}\eta'\eta &= \sum_{i=1}^K \eta_i\eta_i' = \sum_{i=1}^K \left(\int_0^1 \phi_i(r)W_x(r)dr\right)\left(\int_0^1 \phi_i(s)W_x(s)ds\right)' \\ &= \int_0^1\int_0^1 \left(\sum_{i=1}^K \phi_i(r)\phi_i(s)\right)\tilde{W}_x(r)\tilde{W}_x'(s)drds\end{aligned}$$

where $\tilde{W}_x(s) \in L_0^2[0,1]$ almost surely, and so

$$\eta'\eta \to \int_0^1 \tilde{W}_x(r)\tilde{W}_x'(r)dr \tag{70}$$

almost surely. This implies that $\eta'\eta \Rightarrow \int_0^1 \tilde{W}_x(r)\tilde{W}_x(r)'dr$.

Next, we prove $\eta' P_\xi \eta = o_p(1)$. We have

$$\begin{aligned}\eta' P_\xi \eta &= \eta'\xi\left(\xi'\xi\right)^{-1}\xi'\eta = \frac{1}{K}\cdot \eta'\xi\left(\frac{\xi'\xi}{K}\right)^{-1}\xi'\eta \\ &= \frac{1}{K}\cdot\left(\sum_{i=1}^K \eta_i\xi_i'\right)\left(\frac{1}{K}\sum_{i=1}^K \xi_i\xi_i'\right)^{-1}\left(\sum_{i=1}^K \xi_i\eta_i'\right) \\ &= \frac{1}{K}\cdot\left(\sum_{i=1}^K \eta_i\xi_i'\right)\left(\sum_{i=1}^K \xi_i\eta_i'\right) + o_p(1)\end{aligned} \tag{71}$$

where the last equality follows from the result that $\xi_i \overset{i.i.d}{\sim} N(0, I_d)$ for $i = 1, ..., K$. For the term $\sum_{i=1}^{K} \eta_i \xi_i'$, we have

$$\begin{aligned}
E\sum_{i=1}^{K}\eta_i\xi_i' &= -\sum_{i=1}^{K} E\left[\int_0^1 \Psi_i(s)\, dW_x(s)\right]\left[\int_0^1 \phi_i(r)\, dW_x'(r)\right] \\
&= -I_d \sum_{i=1}^{K}\int_0^1 \Psi_i(r)\,\phi_i(r)\,dr = -I_d\int_0^1 \sum_{i=1}^{K}\Psi_i(r)\,\phi_i(r)\,dr \\
&= -I_d\int_0^1 \left(\int_0^1 \sum_{i=1}^{K}\phi_i(s)\left\{s\le r\right\}ds\right)\phi_i(r)dr \\
&= -I_d\int_0^1 \left[\int_0^1 \sum_{i=1}^{K}\phi_i(s)\phi_i(r)\left\{s\le r\right\}dr\right]ds \to -I_d\int_0^1 ds = -I_d
\end{aligned}$$

and

$$\begin{aligned}
& var\left[vec\left(\sum_{i=1}^{K}\eta_i\xi_i'\right)\right] \\
&= var\left[\sum_{i=1}^{K}\int_0^1\int_0^1 \Psi_i(r)\phi_i(s)vec\left(dW_x(r)\,dW_x'(s)\right)\right] \\
&= var\left[\sum_{i=1}^{K}\int_0^1\int_0^1 \Psi_i(r)\phi_i(s)\left(dW_x(s)\otimes dW_x(r)\right)\right] \\
&= \sum_{i=1}^{K}\sum_{j=1}^{K} E\left[\int_0^1\int_0^1\int_0^1\int_0^1 \Psi_i(r)\phi_i(s)\Psi_j(p)\phi_j(q)\left[dW_x(s)\otimes dW_x(r)\right]\left[dW_x'(q)\otimes dW_x'(p)\right]\right] \\
&= \sum_{i=1}^{K}\sum_{j=1}^{K}\left[\int_0^1\int_0^1 \Psi_i(r)\phi_i(r)\Psi_j(p)\phi_j(p)vec(I_d)vec(I_d)'drdp\right] \\
&+ \sum_{i=1}^{K}\sum_{j=1}^{K}\int_0^1\int_0^1 \Psi_i(r)\phi_i(s)\Psi_j(r)\phi_j(s)drds\left[I_d\otimes I_d\right]\left(I_{d^2}+\mathbb{K}_{d,d}\right)
\end{aligned}$$

where $\mathbb{K}_{d,d}$ is the $d^2 \times d^2$ commutation matrix. Now

$$\sum_{i=1}^{K}\sum_{j=1}^{K}\left[\int_0^1\int_0^1 \Psi_i(r)\phi_i(r)dr\int_0^1 \Psi_j(p)\phi_j(p)dp\right]vec(I_d)vec(I_d)' = 0$$

because $\int_0^1 \Psi_i(r)\phi_i(r)dr = \frac{1}{2}\left[\Psi_i(r)\right]^2\Big|_0^1 = 0$. Also

$$\begin{aligned}
&\int_0^1\int_0^1 \Psi_i(r)\phi_i(s)\Psi_j(r)\phi_j(s)drds \\
&= \int_0^1 \Psi_i(r)\Psi_j(r)dr\int_0^1 \phi_i(s)\phi_j(s)ds = 1_{\{i=j\}}\int_0^1\left[\Psi_i(r)\right]^2dr \\
&= 1\left\{i=j\right\}\int_0^1\left[\int_0^r \phi_i(s)ds\right]^2dr. \qquad (72)
\end{aligned}$$

As a result, we have

$$
\begin{aligned}
& var\left[vec\left(\sum_{i=1}^{K}\eta_i\xi_i'\right)\right] \\
&= \sum_{i=1}^{K}\int_0^1\left[\int_0^r \phi_i(s)ds\right]^2 dr\times\left[I_d\otimes I_d\right]\left(I_{d^2}+\mathbb{K}_{d,d}\right) \\
&= \int_0^1\int_0^1\int_0^1\sum_{i=1}^{K}\phi_i(p)\phi_i(q)1\left\{p\leq r\right\}1\left\{q\leq r\right\}dpdqdr\times\left[I_d\otimes I_d\right]\left(I_{d^2}+\mathbb{K}_{d,d}\right) \\
&\to \int_0^1\int_0^1 1\left\{p\leq r\right\}dpdr\times\left[I_d\otimes I_d\right]\left(I_{d^2}+\mathbb{K}_{d,d}\right) \\
&= \int_0^1 rdr\times\left[I_d\otimes I_d\right]\left(I_{d^2}+\mathbb{K}_{d,d}\right)=\frac{1}{2}\left[I_d\otimes I_d\right]\left(I_{d^2}+\mathbb{K}_{d,d}\right) \qquad (73)
\end{aligned}
$$

as $K\to\infty$. In view of the mean and variance orders, we have $\sum_{i=1}^{K}\eta_i\xi_i'=O_p(1)$. It then follows that

$$
\begin{aligned}
\eta' P_\xi \eta &= \frac{1}{K}\cdot\left(\sum_{k=1}^{K}\eta_i\xi_i'\right)\left(\sum_{k=1}^{K}\xi_i\eta_i'\right)+o_p(1) \\
&= O_p(\frac{1}{K})+o_p(1)=o_p(1). \qquad (74)
\end{aligned}
$$

Combining (70) and (74) yields

$$
\eta' M_\xi \eta=\eta'\eta-\eta' P_\xi\eta=\eta'\eta+o_p(1)\Rightarrow\int_0^1\tilde{W}_x(r)\tilde{W}_x(r)'dr
$$

as desired. ■

Proof of of Theorem 3. We prove only the result for the Wald statistic as the proof goes through for the t statistic with obvious modifications. Using (65), we have

$$
\begin{aligned}
\hat{\sigma}^2_{0\cdot x} &= \frac{1}{K}\mathbb{W}^{0\cdot x\prime}\left[I_K-\tilde{\mathbb{W}}\left(\tilde{\mathbb{W}}'\tilde{\mathbb{W}}\right)^{-1}\tilde{\mathbb{W}}'\right]\mathbb{W}^{0\cdot x} \\
&= \frac{1}{K}\mathbb{W}^{0\cdot x\prime}\left\{I_K-\tilde{\mathbb{W}}\Upsilon_T^{-1}\left[\left(\tilde{\mathbb{W}}\Upsilon_T^{-1}\right)'\tilde{\mathbb{W}}\Upsilon_T^{-1}\right]^{-1}\left(\tilde{\mathbb{W}}\Upsilon_T^{-1}\right)'\right\}\mathbb{W}^{0\cdot x} \\
&\Rightarrow \sigma^2_{0\cdot x}\frac{1}{K}\nu'\left[I_K-\tilde{\zeta}\left(\tilde{\zeta}'\tilde{\zeta}\right)^{-1}\tilde{\zeta}'\right]\nu.
\end{aligned}
$$

But

$$
\begin{aligned}
&\tilde{\zeta}\left(\tilde{\zeta}'\tilde{\zeta}\right)^{-1}\tilde{\zeta}' \\
&=\left(\eta\Omega_{xx}^{1/2},\xi\Omega_{xx}^{1/2}\right)\left(\begin{array}{cc}\Omega_{xx}^{1/2}\eta'\eta\Omega_{xx}^{1/2} & \Omega_{xx}^{1/2}\eta'\xi\Omega_{xx}^{1/2}\\ \Omega_{xx}^{1/2}\xi'\eta\Omega_{xx}^{1/2} & \Omega_{xx}^{1/2}\xi'\xi\Omega_{xx}^{1/2}\end{array}\right)^{-1}\left(\begin{array}{c}\Omega_{xx}^{1/2}\eta'\\ \Omega_{xx}^{1/2}\xi'\end{array}\right) \\
&=\left(\eta\Omega_{xx}^{1/2},\xi\Omega_{xx}^{1/2}\right)\left[\left(\begin{array}{cc}\Omega_{xx}^{1/2} & 0\\ 0 & \Omega_{xx}^{1/2}\end{array}\right)\left(\begin{array}{cc}\eta'\eta & \eta'\xi\\ \xi'\eta & \xi'\xi\end{array}\right)\left(\begin{array}{cc}\Omega_{xx}^{1/2} & 0\\ 0 & \Omega_{xx}^{1/2}\end{array}\right)\right]^{-1}\left(\begin{array}{c}\Omega_{xx}^{1/2}\eta'\\ \Omega_{xx}^{1/2}\xi'\end{array}\right) \\
&=\left(\eta\Omega_{xx}^{1/2},\xi\Omega_{xx}^{1/2}\right)\left(\begin{array}{cc}\Omega_{xx}^{-1/2} & 0\\ 0 & \Omega_{xx}^{-1/2}\end{array}\right)\left(\begin{array}{cc}\eta'\eta & \eta'\xi\\ \xi'\eta & \xi'\xi\end{array}\right)^{-1}\left(\begin{array}{cc}\Omega_{xx}^{-1/2} & 0\\ 0 & \Omega_{xx}^{-1/2}\end{array}\right)\left(\begin{array}{c}\Omega_{xx}^{1/2}\eta'\\ \Omega_{xx}^{1/2}\xi'\end{array}\right) \\
&=(\eta,\xi)\left(\begin{array}{cc}\eta'\eta & \eta'\xi\\ \xi'\eta & \xi'\xi\end{array}\right)^{-1}\left(\begin{array}{c}\eta'\\ \xi'\end{array}\right):=P_{\zeta},
\end{aligned}
$$

where $\zeta=(\eta,\xi)\in\mathbb{R}^{K\times 2d}$, so $\hat{\sigma}_{0\cdot x}^2\Rightarrow\sigma_{0\cdot x}^2\frac{1}{K}\nu' M_{\zeta}\nu$ for $M_{\zeta}=I_K-P_{\zeta}$. Combining this with

$$T\left(R\hat{\beta}_{TAOLS}-r\right)\Rightarrow\sigma_{0\cdot x}R\Omega_{xx}^{-1/2}\left(\eta' M_{\xi}\eta\right)^{-1}\eta' M_{\xi}\nu$$

and

$$
\begin{aligned}
&R(\frac{\mathbb{W}^{x\,\prime}}{T}M_{\Delta x}\frac{\mathbb{W}^{x}}{T})^{-1}R' \\
&=R\left\{\frac{\mathbb{W}^{x\,\prime}}{T}\left[I_K-\mathbb{W}^{\Delta x}(\mathbb{W}^{\Delta x\prime}\mathbb{W}^{\Delta x})^{-1}\mathbb{W}^{\Delta x\prime}\right]\frac{\mathbb{W}^{x}}{T}\right\}^{-1}R' \\
&\Rightarrow R\left\{\Omega_{xx}^{1/2}\eta'\left[I_K-\xi\Omega_{xx}^{1/2}\left(\Omega_{xx}^{1/2}\xi'\xi\Omega_{xx}^{1/2}\right)^{-1}\Omega_{xx}^{1/2}\xi'\right]\eta\Omega_{xx}^{1/2}\right\}^{-1}R' \\
&=R\left\{\Omega_{xx}^{1/2}\eta'\left(I_K-\xi\left(\xi'\xi\right)^{-1}\xi'\right)\eta\Omega_{xx}^{1/2}\right\}^{-1}R' \\
&=R\Omega_{xx}^{-1/2}\eta' M_{\xi}\eta\Omega_{xx}^{-1/2}R', \qquad (75)
\end{aligned}
$$

we have

$$
\begin{aligned}
F(\hat{\beta}_{TAOLS})&\Rightarrow\frac{K}{p}\frac{\left[\tilde{R}\left(\eta' M_{\xi}\eta\right)^{-1}\eta' M_{\xi}\nu\right]'\left(\tilde{R}\eta' M_{\xi}\eta\tilde{R}'\right)^{-1}\left[\tilde{R}\left(\eta' M_{\xi}\eta\right)^{-1}\eta' M_{\xi}\nu\right]}{\nu' M_{\zeta}\nu} \\
&=\frac{K}{p}\frac{Q'\left(\tilde{R}\eta' M_{\xi}\eta\tilde{R}'\right)^{-1}Q}{\nu' M_{\zeta}\nu}, \qquad (76)
\end{aligned}
$$

where

$$Q'\left(\tilde{R}\left(\eta' M_{\xi}\eta\right)^{-1}\tilde{R}'\right)^{-1}Q\overset{d}{=}\chi_p^2\text{ and }\nu' M_{\zeta}\nu\overset{d}{=}\chi_{K-2d}^2.$$

Note that conditional on $\zeta=(\eta,\xi)$, $M_{\zeta}\nu=\left[I_K-\zeta\left(\zeta'\zeta\right)^{-1}\zeta'\right]\nu$ and $\eta' M_{\xi}\nu$ are independent, as both $M_{\zeta}\nu$ and $\eta' M_{\xi}\nu$ are normal and the conditional covariance is

$$cov\left(M_{\zeta}\nu,\eta' M_{\xi}\nu\right)=\left[I_K-\zeta\left(\zeta'\zeta\right)^{-1}\zeta'\right]M_{\xi}\eta=0.$$

So conditional on ζ, the numerator and the denominator in (76) are independent chi-square variates. This implies that

$$\frac{K}{p}\frac{Q'\left(\tilde{R}\eta' M_\xi \eta \tilde{R}'\right)^{-1} Q}{\nu' M_\zeta \nu} = \frac{K}{K-2d}\frac{Q'\left(\tilde{R}\eta' M_\xi \eta \tilde{R}'\right)^{-1} Q/p}{\nu' M_\zeta \nu/(K-2d)} \overset{d}{=} \frac{K}{K-2d}F_{p,K-2d}$$

conditional on ζ. But the conditional distribution does not depend on the conditioning variable ζ, so it is also the unconditional distribution. We have therefore proved that

$$F(\hat{\beta}_{TAOLS}) \Rightarrow \frac{K}{K-2d}F_{p,K-2d}. \tag{77}$$

■

Proof of Theorem 4. We follow the same step as in the proof of Theorem 1. We consider only $F_{tr}(\hat{\beta}_{TAOLS})$. The proof for $t_{tr}(\hat{\beta}_{TAOLS})$ is similar.

Let

$$\Upsilon_{T,tr} = \begin{pmatrix} \Upsilon_T & 0 \\ 0 & T^{3/2} \end{pmatrix}$$

Then

$$\Upsilon_{T,tr}\begin{pmatrix} \hat{\beta}_{TAOLS}-\beta_0 \\ \hat{\delta}_{TAOLS}-\delta_0 \\ \hat{\mu}_{TAOLS}-\mu_0 \end{pmatrix} = (\Upsilon_{T,tr}^{-1}\tilde{\mathbb{W}}_{tr}'\tilde{\mathbb{W}}_{tr}\Upsilon_{T,tr}^{-1})^{-1}\Upsilon_{T,tr}\tilde{\mathbb{W}}_{tr}'\mathbb{W}^y \Rightarrow \left(\tilde{\zeta}_{tr}'\tilde{\zeta}_{tr}\right)^{-1}\tilde{\zeta}_{tr}'\tilde{\nu},$$

where

$$\tilde{\zeta}_{tr} = \left(\eta\Omega_{xx}^{1/2}, \xi\Omega_{xx}^{1/2}, a\right).$$

Some simple calculations show that

$$\left(\tilde{\zeta}_{tr}'\tilde{\zeta}_{tr}\right)^{-1}\tilde{\zeta}_{tr}'\tilde{\nu} = \begin{pmatrix} \sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\eta' M_{\xi,a}\eta\right)^{-1}\eta' M_{\xi,a}\nu \\ \sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\xi' M_{\eta,a}\xi\right)^{-1}\xi' M_{\eta,a}\nu \\ \sigma_{0\cdot x}\left(a' M_{\eta,\xi}a\right)^{-1}a' M_{\eta,\xi}\nu \end{pmatrix}.$$

So part (i) of the theorem holds. In particular,

$$T(\hat{\beta}_{TAOLS}-\beta_0) \Rightarrow \sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\eta' M_{\xi,a}\eta\right)^{-1}\eta' M_{\xi,a}\nu. \tag{78}$$

Following the same steps in the proof of Theorem 3, we have $(\hat{\sigma}_{0\cdot x}^{tr})^2 \Rightarrow \sigma_{0\cdot x}^2\frac{1}{K}\nu' M_{\zeta,a}\nu$. Combining this with (78), we have

$$\begin{aligned} F_{tr}(\hat{\beta}_{TAOLS}) &\Rightarrow \frac{K}{p}\frac{\left[\tilde{R}\left(\eta' M_{\xi,a}\eta\right)^{-1}\eta' M_{\xi,a}\nu\right]'\left(\tilde{R}\eta' M_{\xi,a}\eta\tilde{R}'\right)^{-1}\left[\tilde{R}\left(\eta' M_{\xi,a}\eta\right)^{-1}\eta' M_{\xi,a}\nu\right]}{\nu' M_{\zeta,a}\nu} \\ &=^d \frac{K}{K-2d-1}F_{p,K-2d-1}. \end{aligned} \tag{79}$$

■

Proof of Lemma 5. The cosine functions clearly satisfy Assumption 1. Note that

$$\begin{aligned}\Psi_i^c(r) &:= \int_0^r \phi_i^c(s)\, ds = \sqrt{2}\int_0^r \cos(2i\pi s) ds \\ &= \sqrt{2}\, \frac{\sin(2i\pi s)}{2i\pi}\bigg|_0^r = \frac{\sin(2i\pi r)}{\sqrt{2}i\pi},\end{aligned}$$

we have, for $i \neq j$,

$$\begin{aligned}\int_0^1 \Psi_i^c(r)\phi_j^c(r)\, dr &= \int_0^1 \frac{\sin(2i\pi r)}{\sqrt{2}i\pi}\cdot\sqrt{2}\cos(2j\pi r)\, dr \\ &= \frac{1}{i\pi}\int_0^1 \sin(2i\pi r)\cos(2j\pi r) dr \\ &= \frac{1}{2i\pi}\left(\int_0^1 \sin(2(i+j)\pi r)\, dr + \int_0^1 \sin(2(i-j)\pi r)\, dr\right) \\ &= \frac{1}{2i\pi}\left(-\frac{\cos(2(i+j)\pi r)}{2(i+j)\pi}\bigg|_0^1\right) + \frac{1}{2i\pi}\left(-\frac{\cos(2(i-j)\pi r)}{2(i-j)\pi}\bigg|_0^1\right) = 0.\end{aligned}$$

For $i = j$, we have

$$\int_0^1 \Psi_i^c(r)\phi_j^c(r)\, dr = \frac{1}{2i\pi}\left(\int_0^1 \sin(2(i+j)\pi r)\, dr\right) = \frac{1}{2i\pi}\left(-\frac{\cos(2(i+j)\pi r)}{2(i+j)\pi}\bigg|_0^1\right) = 0.$$

Therefore $\int_0^1 \Psi_i^c(r)\phi_j^c(r)\, dr = 0$ for any given $i, j = 1, ..., K$. That is, the cosine functions also satisfy Assumption 3. ■

Proof of Theorem 6. We have

$$T(\hat{\beta}_{TOLS} - \beta_0) \Longrightarrow \Omega_{xx}^{-1/2}\left(\sum_{i=1}^{K}\eta_i\eta_i'\right)^{-1}\left(\sum_{i=1}^{K}\eta_i\psi_i\right),$$

where

$$\eta_i := \int_0^1 \phi_i(r)W_x(r)dr \text{ and } \psi_i := \sigma_{0\cdot x}\nu_i + \sigma_{x0}\Omega_{xx}^{-1/2}\xi_i$$

for

$$\xi_i := \int_0^1 \phi_i(r)dW_x(r) \text{ and } \nu_i = \int_0^1 \phi_i(r)dw_0(r).$$

Since $\psi := (\psi_1, ..., \psi_K)' \sim N(0, \sigma_0^2 I_K)$ and $\eta \perp \psi$, we can represent the limiting distribution as zero mean mixed normal distribution

$$\Omega_{xx}^{-1/2}\left(\sum_{i=1}^{K}\eta_i\eta_i'\right)^{-1}\left(\sum_{i=1}^{K}\eta_i\psi_i\right) = \Omega_{xx}^{-1/2}\left(\eta'\eta\right)^{-1}\eta'\psi \overset{d}{=} MN\left(0, \sigma_0^2\Omega_{xx}^{-1/2}(\eta'\eta)^{-1}\Omega_{xx}^{-1/2}\right)$$

as desired. ■

Proof of Proposition 8. Note that $Ec'(V_{TAOLS}^{-1} - V_{TOLS}^{-1})c = c'E\left[E(V_{TAOLS}^{-1} - V_{TOLS}^{-1})|\eta\right]c$ for any conforming vector $c \in \mathbb{R}^d$. From equation (55) and the independence between η and ξ, we have

$$c'\left[E(V_{TAOLS}^{-1} - V_{TOLS}^{-1})|\eta\right]c = \frac{c'\Omega_{xx}^{1/2}\eta'}{\sigma_{0\cdot x}}\left(\varrho^2 I_K - E[P_\xi]\right)\frac{\eta\Omega_{xx}^{1/2}c}{\sigma_{0\cdot x}}. \tag{80}$$

where the (i,j)th element of $P_\xi = \xi(\xi'\xi)^{-1}\xi'$ is

$$p_{i,j} = \xi_i' \left(\sum_{s=1}^{K} \xi_s \xi_s' \right)^{-1} \xi_j.$$

We want to show that $E[P_\xi] = d/K \cdot I_K$. That is, $E[p_{ii}] = d/K$ for $i = 1, ..., K$ and $E[p_{ij}] = E[p_{ji}] = 0$ for all $i \neq j$. Since $\xi_i \overset{i.i.d}{\sim} N(0, I_d)$, it is easy to show that

$$\xi_i' \left(\sum_{s=1}^{K} \xi_s \xi_s' \right)^{-1} \xi_i \overset{d}{=} \xi_j' \left(\sum_{s=1}^{K} \xi_s \xi_s' \right)^{-1} \xi_j \text{ for all } i, j = 1, ..., K. \tag{81}$$

So, we have $E[p_{ii}] = E[p_{jj}]$, i.e., all the diagonal elements of $E[P_\xi]$ are same. In other words, $E[p_{11}] = E[p_{22}] = ... = E[p_{KK}] = \lambda$ for some λ. This gives us

$$\lambda K = tr[EP_\xi] = E[tr\left[\xi(\xi'\xi)^{-1}\xi'\right] = d$$

and so $\lambda = E[p_{ii}] = d/K$ for $i = 1, ..., K$. For the off-diagonal elements, we note that the distribution of $p_{i,j}$ is symmetric around zero, which implies that $E[p_{ij}] = E[p_{ji}] = 0$ for all $i \neq j$. Therefore,

$$Ec'(V_{TAOLS}^{-1} - V_{TOLS}^{-1})c = \left(\varrho^2 - \frac{d}{K} \right) E \left\| \frac{\eta \Omega_{xx}^{1/2} c}{\sigma_{0 \cdot x}} \right\|^2, \tag{82}$$

and this immediately leads to the desired result. ■

References

[1] Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–854.

[2] Bierens, H. J. (1997). Nonparametric cointegration analysis. *Journal of Econometrics*, 77(2), 379–404.

[3] Bunzel, H. (2006). Fixed-b asymptotics in single-equation cointegration models with endogenous regressors. *Econometric Theory* 22, 743–755.

[4] Hamming, R.W. (1973). Numerical Methods for Scientists and Engineers. Dover Publications, New York.

[5] Jin, S., Phillips, P.C.B., Sun, Y. (2006). A new approach to robust inference in cointegration. *Economics Letters* 91, 300–306.

[6] Kiefer, N.M., Vogelsang, T.J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* 21, 1130–1164.

[7] Müller, U. K. (2014). HAC corrections for strongly autocorrelated time series. *Journal of Business & Economic Statistics* 32(3), 311–322.

[8] Müller, U. K., Watson, M.W. (2013). Low-frequency robust cointegration testing. *Journal of Econometrics* 174(2), 66–81.

[9] Phillips, P.C.B., (1991a). Spectral regression for cointegrated time series. In: Barnett, W., Powell, J., Tauchen, G. (Eds.), *Nonparametric and Semiparametric Methods in Economics and Statistics.* Cambridge University Press, New York, pp. 413–435.

[10] Phillips, P.C.B. (1991b). Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.

[11] Phillips, P.C.B. (2005). HAC estimation by automated regression. *Econometric Theory* 21(1), 116–142.

[12] Phillips, P.C.B. (2014). Optimal estimation of cointegrated systems with irrelevant instruments, *Journal of Econometrics* 178(2), 210–224.

[13] Phillips, P.C.B., Durlauf, S.N. (1986). Multiple regression with integrated processes. *Review of Economic Studies* 53, 473–496.

[14] Phillips, P.C.B., Hansen, B.E. (1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99–125.

[15] Phillips, P.C.B., Liao, Z. (2014). Series estimation of stochastic processes: Recent Developments and Econometric Applications, in Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics, Edited by J.S. Racine, L. Su, and A. Ullah.

[16] Phillips, P.C.B., Loretan, M. (1991). Estimating long run economic equilibria. *Review of Economic Studies* 58, 407–436.

[17] Saikkonen, P. (1991). Asymptotically efficient estimation of cointegrating regressions. *Econometric Theory* 7, 1–21.

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

[19] Stock, J.H., Watson, M.W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.

[20] Sun, Y. (2011). Robust trend inference with series variance estimator and testing-optimal smoothing parameter. *Journal of Econometrics* 164(2), 345–66.

[21] Sun, Y. (2013). A heteroskedasticity and autocorrelation robust F test using orthonormal series variance estimator. *Econometrics Journal* 16(1), 1–26.

[22] Sun, Y. (2014a). Let's Fix It: fixed-b asymptotics versus small-b asymptotics in heteroscedasticity and autocorrelation robust inference. *Journal of Econometrics* 178(3), 659–677.

[23] Sun, Y. (2014b). Fixed-smoothing asymptotics in a two-step GMM framework. *Econometrica* 82(6), 2327–2370.

[24] Sun, Y., Phillips, P.C.B., Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica* 76(1), 175–94.

[25] Thomson, D.J. (1982). Spectrum estimation and harmonic analysis. *IEEE Proceedings*, 70, 1055–1096.

[26] Vogelsang, T.J., Wagner, M. (2014). Integrated modified OLS estimation and fixed-b inference for cointegrating regressions. *Journal of Econometrics*, 178(2), 741–760.

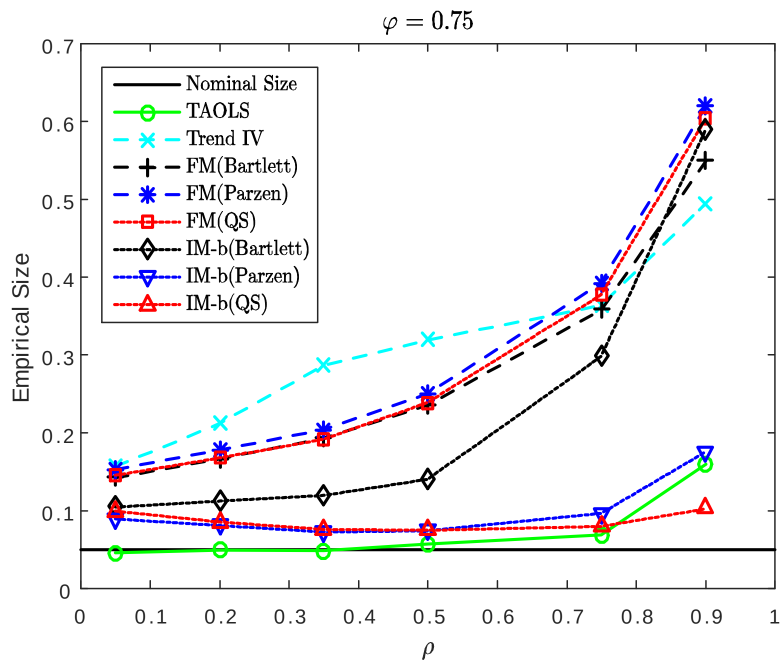

Figure 1: Empirical size of various tests for $\varphi = 0.75$ and $T = 100$.

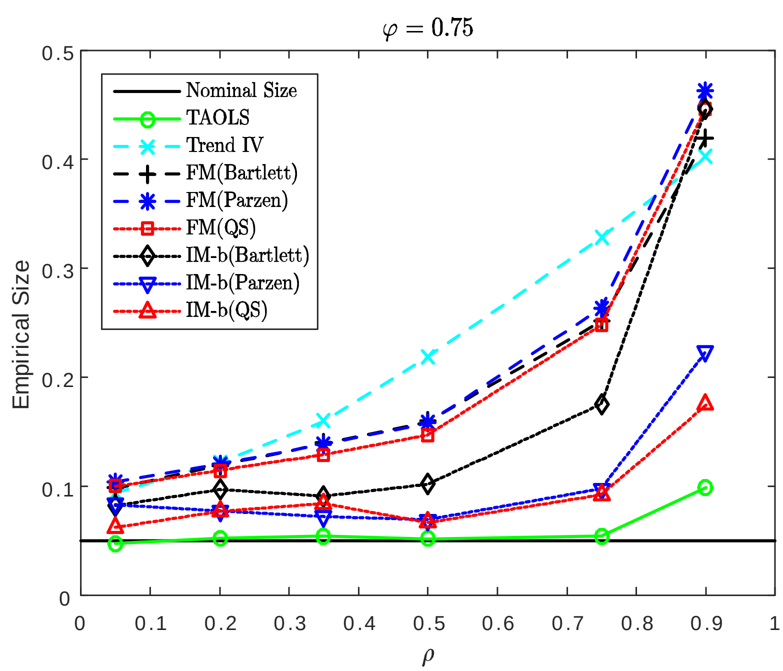

Figure 2: Empirical size of various tests for $\varphi = 0.75$ and $T = 200$.

			T = 100					
	TAOLS	**TIV**	**FM-OLS**			**IM-OLS**		
ρ	**DFT**		**Bartlett**	**Parzen**	**QS**	**Bartlett**	**Parzen**	**QS**
0.05	0.0473	0.1629	0.1448	0.1572	0.1515	0.1051	0.0879	0.0988
0.20	0.0498	0.2161	0.1876	0.1776	0.1705	0.0986	0.0790	0.0813
0.35	0.0504	0.3017	0.2557	0.2044	0.1919	0.0821	0.0725	0.0749
0.50	0.0542	0.4809	0.3843	0.2451	0.2329	0.0758	0.0684	0.0669
0.75	0.0478	0.7153	0.6901	0.3955	0.3777	0.1389	0.0953	0.0756
0.90	0.0712	0.7520	0.7923	0.6220	0.6114	0.4349	0.1757	0.1017
			T = 200					
	TAOLS	**TIV**	**FM-OLS**			**IM-OLS**		
ρ	**DFT**		**Bartlett**	**Parzen**	**QS**	**Bartlett**	**Parzen**	**QS**
0.05	0.0472	0.0961	0.0981	0.1045	0.1028	0.0681	0.0779	0.0600
0.20	0.0532	0.1269	0.1299	0.1219	0.1162	0.0944	0.0813	0.0822
0.35	0.0529	0.1622	0.1704	0.1399	0.1332	0.0826	0.0721	0.0815
0.50	0.0534	0.2248	0.2629	0.1670	0.1594	0.0841	0.0738	0.0705
0.75	0.0560	0.5754	0.6183	0.2693	0.2571	0.1192	0.1012	0.0960
0.90	0.0603	0.7237	0.7371	0.4631	0.4494	0.3460	0.2226	0.1674

Table 1: Empirical size of various tests for $\varphi = 0.75$.

		T = 100					
	TAOLS/TIV	**FM-OLS**			**IM-OLS**		
ρ	**K**	**Bartlett**	**Parzen**	**QS**	**Bartlett**	**Parzen**	**QS**
0.05	19.0040	0.0292	0.0556	0.0276	0.0289	0.0548	0.0272
0.20	14.2524	0.0543	0.0749	0.0372	0.0547	0.0747	0.0371
0.35	10.7156	0.1113	0.1015	0.0504	0.1123	0.1016	0.0505
0.50	8.0284	0.2461	0.1403	0.0697	0.2443	0.1396	0.0693
0.75	6.0310	0.8882	0.2796	0.1389	0.8775	0.2711	0.1347
0.90	6.0012	0.9983	0.5234	0.2611	0.9946	0.4966	0.2480
		T = 200					
	TAOLS/TIV	**FM-OLS**			**IM-OLS**		
ρ	**K**	**Bartlett**	**Parzen**	**QS**	**Bartlett**	**Parzen**	**QS**
0.05	36.1384	0.0139	0.0283	0.0142	0.0138	0.0280	0.0140
0.20	25.6522	0.0250	0.0397	0.0197	0.0252	0.0398	0.0198
0.35	18.6720	0.0541	0.0550	0.0273	0.0548	0.0552	0.0274
0.50	13.5976	0.1257	0.0771	0.0383	0.1264	0.0772	0.0383
0.75	7.1238	0.7303	0.1603	0.0796	0.7267	0.1583	0.0786
0.90	6.0078	0.9997	0.3351	0.1664	0.9990	0.3219	0.1599

Table 2: Averages of the data-driven smoothing parameters K and b for $\varphi = 0.75$.

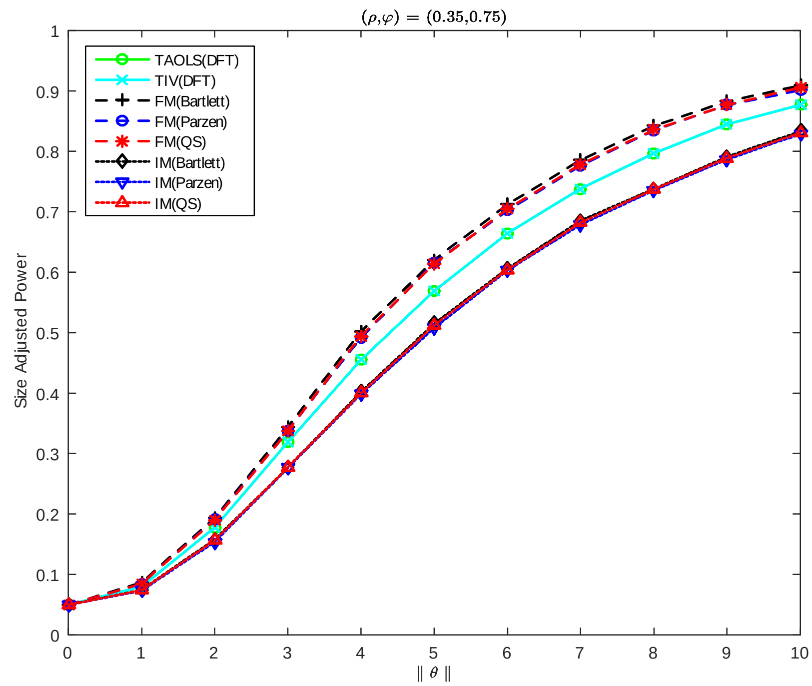

Figure 3: Size-adjusted power curves at $\rho = 0.35$, $\varphi = 0.75$ and $T = 200$.

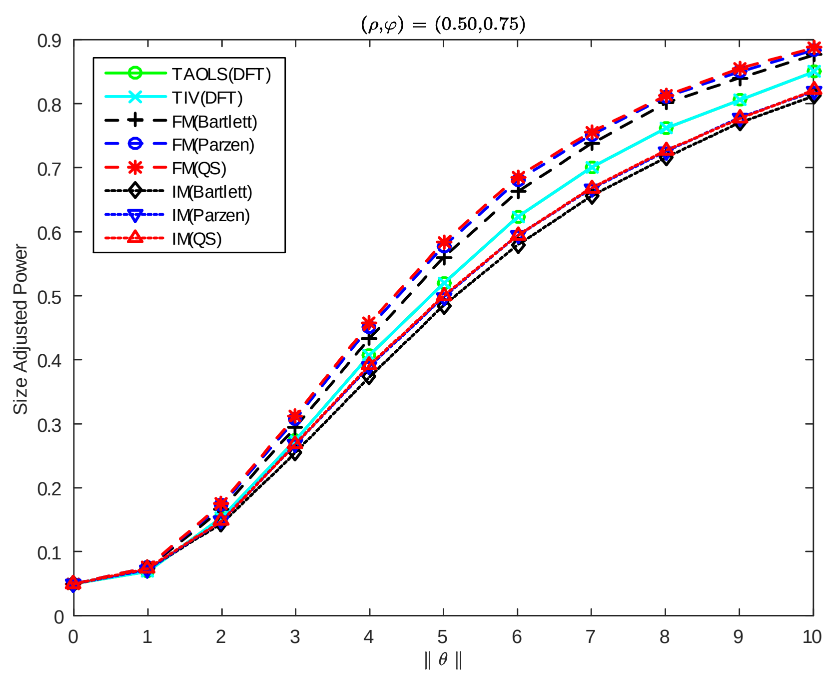

Figure 4: Size-adjusted power curves at $\rho = 0.50$, $\varphi = 0.75$ and $T = 200$.

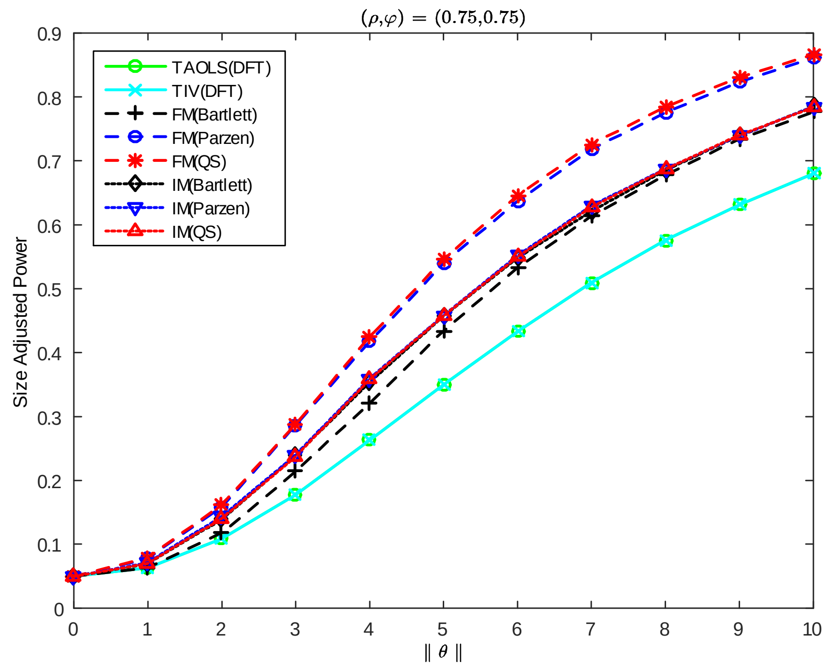

Figure 5: Size-adjusted power curves at $\rho = 0.75$, $\varphi = 0.75$ and $T = 200$.

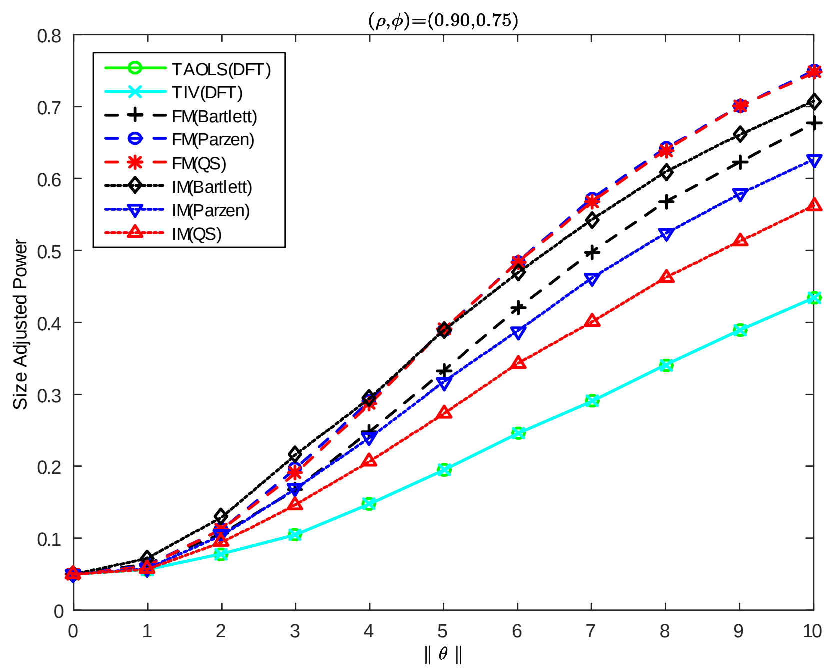

Figure 6: Size-adjusted power curves at $\rho = 0.90$, $\varphi = 0.75$ and $T = 200$.

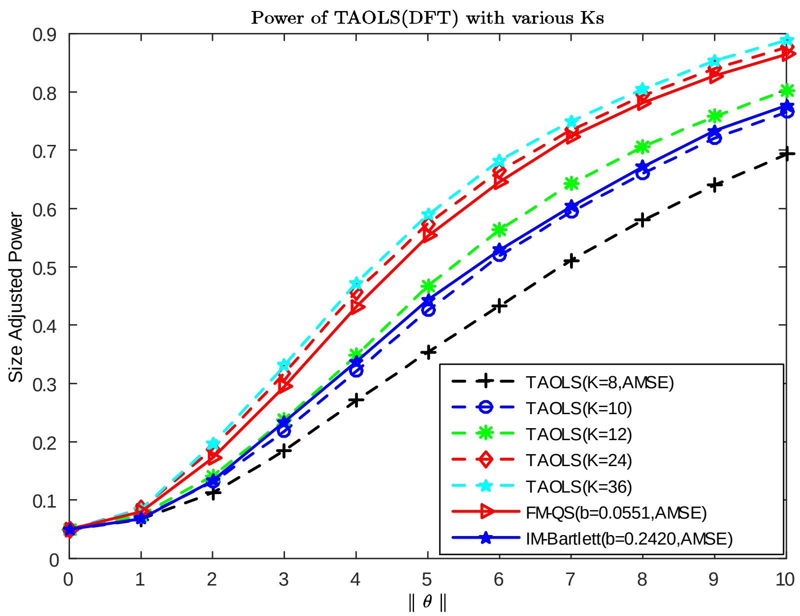

Figure 7: Size-adjusted power curves at $\rho = 0.75$, $\varphi = 0.75$ and $T = 200$ with K values between $8 \sim 36$.

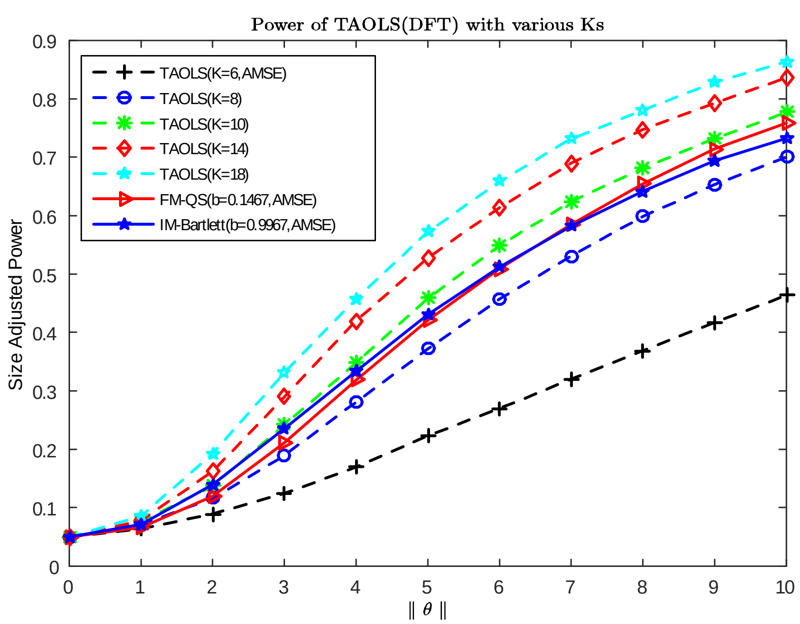

Figure 8: Size-adjusted power curves at $\rho = 0.90$, $\varphi = 0.75$ and $T = 200$ with K values between $6 \sim 18$.

$T = 100$							
	TAOLS		**TOLS**	**FM-OLS**		**IM-OLS**	
ϱ^2	**DFT**	**Cosine**	**Cosine**	**Bartlett**	**QS**	**Bartlett**	**QS**
0.05	0.0554	0.0586	0.0795	0.8011	0.4374	0.2816	0.1182
0.20	0.0507	0.0564	0.0822	0.7992	0.4189	0.2733	0.1182
0.35	0.0518	0.0580	0.0763	0.7911	0.4143	0.2481	0.0941
0.50	0.0499	0.0633	0.0846	0.7902	0.4071	0.2283	0.0845
0.75	0.0521	0.0662	0.0839	0.7551	0.3473	0.1749	0.0709
0.90	0.0514	0.0632	0.0830	0.6894	0.2795	0.0899	0.0668
$T = 200$							
	TAOLS		**TOLS**	**FM-OLS**		**IM-OLS**	
ϱ^2	**DFT**	**Cosine**	**Cosine**	**Bartlett**	**QS**	**Bartlett**	**QS**
0.05	0.0556	0.0527	0.0597	0.6760	0.2524	0.1839	0.1133
0.20	0.0549	0.0545	0.0577	0.6772	0.2657	0.1846	0.1122
0.35	0.0524	0.0522	0.0623	0.6862	0.2617	0.1667	0.1056
0.50	0.0502	0.0524	0.0587	0.6846	0.2582	0.1565	0.1021
0.75	0.0502	0.0517	0.0546	0.6800	0.2356	0.1106	0.0892
0.90	0.0491	0.0514	0.0632	0.6469	0.1911	0.0711	0.0927

Table 3: Empirical size of various tests in the cointegration model with a linear time trend and $\rho = 0.75$.

T = 100						
	DFT	**Cosine**	**FM-OLS**		**IM-OLS**	
ϱ^2	**K**	**K**	**Bartlett**	**QS**	**Bartlett**	**QS**
0.05	6.0916	6.0000	0.7713	0.1232	0.8092	0.1282
0.20	6.0606	6.0002	0.8014	0.1273	0.8317	0.1313
0.35	6.0414	6.0000	0.8255	0.1314	0.8483	0.1333
0.50	6.0336	6.0000	0.8567	0.1358	0.8622	0.1346
0.75	6.0534	6.0000	0.8801	0.1373	0.8342	0.1277
0.90	6.2974	6.0092	0.8177	0.1260	0.7084	0.1105
T = 200						
	DFT	**Cosine**	**FM-OLS**		**IM-OLS**	
ϱ^2	**K**	**K**	**Bartlett**	**QS**	**Bartlett**	**QS**
0.05	8.2486	6.0372	0.4947	0.0670	0.5292	0.0684
0.20	7.9748	6.0184	0.5403	0.0697	0.5711	0.0710
0.35	7.6838	6.0136	0.5928	0.0726	0.6173	0.0736
0.50	7.4500	6.0100	0.6451	0.0754	0.6571	0.0756
0.75	7.0746	6.0080	0.7548	0.0808	0.7233	0.0778
0.90	7.4216	6.0984	0.7596	0.0786	0.6650	0.0717

Table 4: The averages of K and b selected by the AMSE rule in the cointegration model with a linear time trend and $\rho = 0.75$.

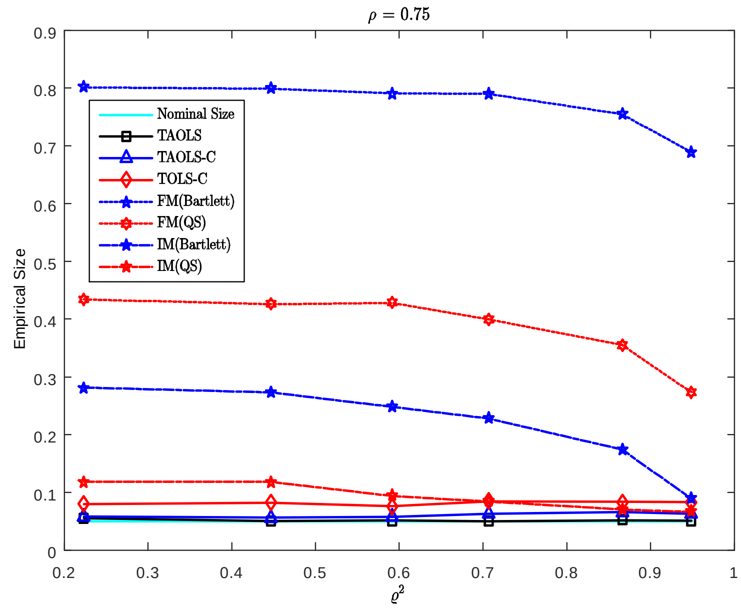

Figure 9: Empirical size of various tests for the cointegration model with a linear trend for $\rho = 0.75$ and $T = 100$.

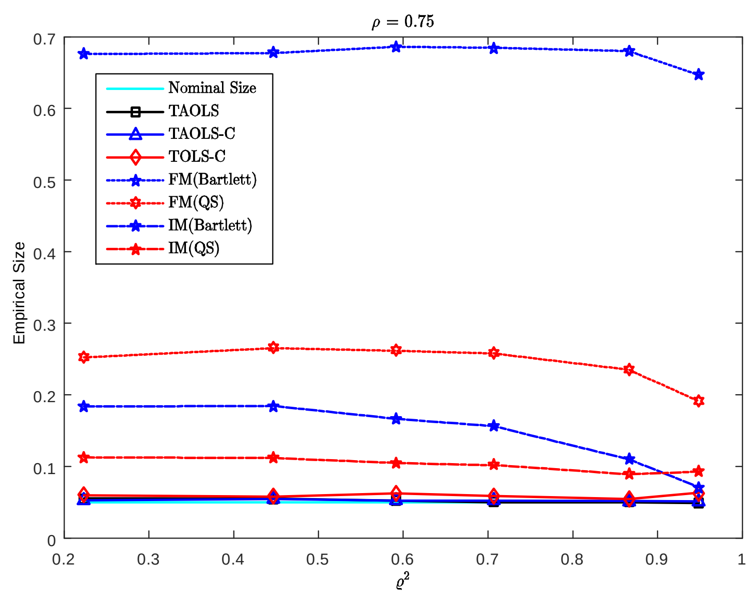

Figure 10: Empirical size of various tests for the cointegration model with a linear trend for $\rho = 0.75$ and $T = 200$.

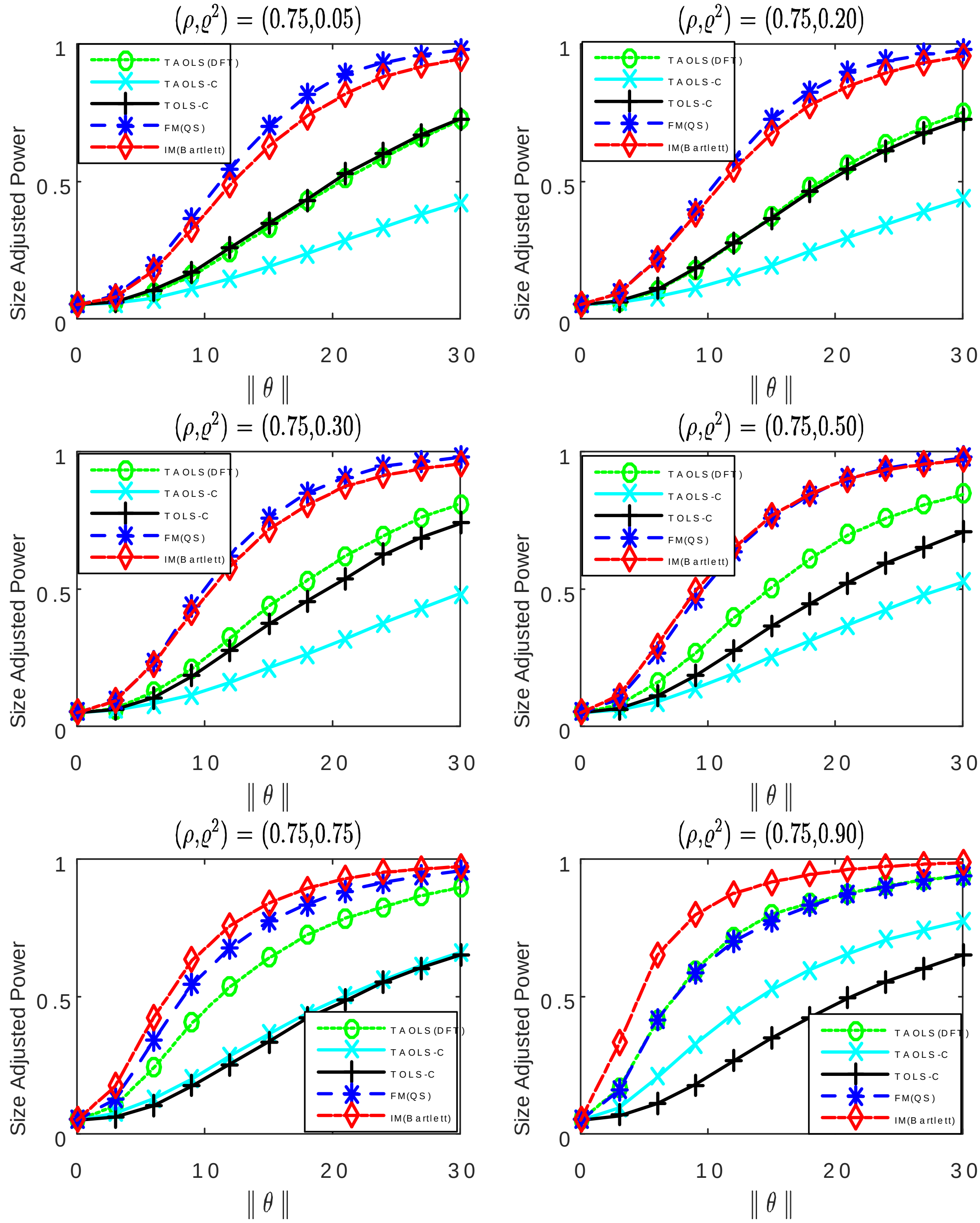

Figure 11: Size-adjusted power curves of the TAOLS-C and TOLS-C F tests and other tests with $\varrho^2 \in \{0.05, 0.20, 0.35, 0.50, 0.75, 0.90\}$, $\rho = 0.75$ and $T = 200$.