Some Extensions of Asymptotic F and t Theory in Nonstationary Regressions

Yixiao Sun*
Department of Economics,
University of California, San Diego

October 30, 2021

Abstract

We develop and extend the asymptotic F and t test theory in linear regression models where the regressors could be deterministic trends, unit-root processes, and near-unit-root processes, among others. We consider both the exogenous case where the regressors and the regression error are independent and the endogenous case where they are correlated. In the former case, we design a new set of basis functions that are invariant to the parameter estimation uncertainty and use them to construct a new series long run variance estimator. We show that the F-test version of the Wald statistic and the t statistic are asymptotically F and t distributed, respectively. In the latter case, we show that the asymptotic F and t theory is still possible, but we have to develop it in a pseudo-frequency domain. The F and t approximations are more accurate than the more commonly used chi-squared and normal approximations. The resulting F and t tests are also easy to implement — they can be implemented in exactly the same way as the F and t tests in a classical normal linear regression.

JEL Classification: C12, C13, C32

Keywords: F distribution, Fixed-smoothing Asymptotics, Heteroscedasticity and Autocorrelation, Series Long Run Variance Estimator, Nonstationary Process, t Distribution, Transformed and Augmented OLS, Unit Root.

*Email: yisun@ucsd.edu. Correspondence to: Yixiao Sun, Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508. For helpful comments, the author thanks Yoosoon Chang, and two anonymous referees.

1 Introduction

The paper considers time series regressions in a nonstationary framework. The regressors can be deterministic trends, unit-root processes, or near-unit-root processes while the regression error is stationary but with an unknown autocorrelation function. The regressions we consider include trend regressions, cointegration regressions, and predictive regressions as special cases. For all these regressions, a main challenge for statistical inference is to account for the nonparametric autocorrelation in the regression error and the possible nonparametric correlation between the innovation of the regressor process and the regressor error. Standard practice is to estimate the unknown autocorrelation and correlation using a nonparametric kernel method but ignore the nonparametric estimation error for the convenience of statistical inference. However, by its nonparametric nature, the nonparametric estimation error can be large in finite samples. As a result, ignoring the estimation error can lead to highly unreliable inferences.

Recent literature has developed the fixed-smoothing asymptotics, an alternative type of asymptotics, to capture the nonparametric estimation error. The fixed-smoothing asymptotic theory, which includes, as a special case, the fixed-b asymptotic theory of Kiefer and Vogelsang (2002a,b, 2005), was originally developed in the time series setting. It has been extended to accommodate spatial data, spatial and temporal data, panel data, and clustered data. See Sun (2018) for a recent discussion. However, the fixed-smoothing asymptotic distributions are often nonstandard and thus not very convenient to use.

To obtain more accurate but at the same time more convenient fixed-smoothing approximations, we can use the series approach to correct the potential endogenous bias and robustify the inference. The underlying series variance estimator has a long history. A classical example is the average periodograms estimator, which involves taking a simple average of the first few periodograms. An appealing feature of the series approach is that we have the freedom to choose and design the basis functions so that the fixed-smoothing asymptotic distributions become standard F and t distributions. See, for example, Sun (2011) for trend regression, Müller (2007), Sun (2013), Sun (2014a), Lazarus et al. (2021) in the first-step GMM/OLS setting, and Sun (2014b), Hwang and Sun (2017), Martínez-Iriarte et al. (2020) in the two-step GMM setting, Sun and Kim (2015) and Liu and Sun (2019) for spatial and panel data settings (difference-in-differences regressions). More recently, adopting the framework of Chang et al. (2018), Pellatt and Sun (2020) develops the asymptotic F theory in a continuous-time setting. Earlier papers along this line of research include Phillips (2005) and Sun (2006).

The aim of the present paper is to review the asymptotic F and t theory in the nonstationary framework and extend it to cover the nonstationary cases that the theory is currently lacking. Section 2 considers the exogeneous case where the regressors follow either a deterministic trend function or a stochastic trend. It introduces a new idea to design the bases for series variance estimation. This approach involves projecting any given candidate bases onto the orthogonal complement of the column space spanned by the regressors. The projection ensures that the new series variance estimator is invariant to the parameter estimation error. After proper normalization, the projected bases are orthonormal. This enables us to establish that the associated Wald and t statistics are asymptotically F and t distributed, respectively. Section 3 examines the case with endogenous stochastic trends. The regressors can be unit-root or near-unit-root processes but are correlated with the regression error. Here we follow Phillips (2014) and Hwang and Sun (2018) and cast the regression as a low-frequency instrumental variable regression. Effectively, we convert a highly nonstandard inference problem into a standard inference problem in a classical normal linear regression. The asymptotic F and t theory then follows naturally. Section 4 presents a simulation study that demonstrates the higher size-accuracy of the proposed F test in a cointegration regression with exogenous regressors similar to that considered in Phillips and Park (1988). The last section concludes.

2 Deterministic and exogenous cases

Consider the regression model:

$$Y_t = X_t'\beta_0 + u_t,\ t = 1, 2, \ldots, T,$$

where $X_t \in \mathbb{R}^d$ for $d \geq 1$ is either a deterministic trend process, a unit-root process, or a near-unit-root process, and $\{u_t\}$ is a stationary zero-mean process that is independent of $\{X_t\}$. An intercept can be included as the first element of X_t.

For some $p \times d$ matrix $R = (R(i,j))$ and $p \times 1$ vector r, we are interested in testing

$$H_0 : R\beta_0 = r \text{ against } H_1 : R\beta_0 \neq r.$$

When $p = 1$, we may be interested in testing the null against a one-sided alternative, say, testing $H_0 : R\beta_0 = r$ against $H_1 : R\beta_0 > r$.

Based on the observations $\{X_t, Y_t\}_{t=1}^T$, we estimate β_0 by the OLS estimator:

$$\hat{\beta} = \left(\sum_{t=1}^T X_t X_t'\right)^{-1} \sum_{t=1}^T X_t Y_t.$$

For many regression models we consider here, the OLS estimator is asymptotically as efficient as the GLS estimator. These include polynomial trend regressions and cointegration regressions. For the former case, see Grenander and Rosenblatt (1957), and for the latter case, see Phillips and Park (1988), which also provides general discussions on the reason for the asymptotic equivalence between OLS and GLS.

Assumption 1 *(i) For some diagonal matrix D_T, we have*

$$\begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} u_t \\ D_T^{-1}X_{[Tr]} \end{pmatrix} \Rightarrow \begin{pmatrix} B_u(r) \\ B_x(r) \end{pmatrix} := \begin{pmatrix} \omega_u W_u(r) \\ B_x(r) \end{pmatrix},$$

where $B_x(\cdot)$ is either a deterministic function or a stochastic process and $W_u(\cdot)$ is a standard Brownian motion that is independent of $B_x(\cdot)$. (ii) $\int_0^1 B_x(r) B_x(r)' dr$ is of full rank d almost surely.

Assumption 1 requires that $B_x(r)$ and $W_u(r)$ be independent. When $B_x(r)$ is not random, this holds trivially. When $B_x(r)$ is random, this assumption amounts to assuming that the long run correlation between $\{u_t\}$ and $\{\sqrt{T}D_T^{-1}\Delta X_t\}$ converges to zero. The latter condition holds if X_t is strictly exogenous in the sense that $\{X_t : t = 1, \ldots, T\}$ is independent of $\{u_t : t = 1, \ldots, T\}$. We consider the strictly exogenous case in this section and defer the endogenous case to the next section.

We now provide a few examples where Assumption 1 holds. These examples also show that our framework accommodates different types of regressions.

Example 1 ***Deterministic trend.*** *Let $X_t = (1, t, t^2)'$, in which case $D_T = diag(1, T, T^2)$ and $B_x(r) = \left(1, r, r^2\right)'$. One can also consider other types of trend functions.*

Example 2 ***Unit-root process.*** *Let $X_t = (1, \tilde{X}_t')'$ and $\tilde{X}_t$ be a unit-root process:*

$$\tilde{X}_t = \tilde{X}_{t-1} + u_{x,t} \text{ for } t = 1, 2, \ldots, T,$$

where $\tilde{X}_0 = o_p(\sqrt{T})$ and $u_{x,t}$ is a stationary zero-mean process. In this case,

$$D_T = diag(1, \sqrt{T}, \ldots, \sqrt{T}) \text{ and } B_x(r) = \begin{pmatrix} 1 \\ \Omega_{xx}^{1/2} W_x(r) \end{pmatrix},$$

where $W_x(\cdot)$ is a standard Brownian motion process, Ω_{xx} is the long run variance of $\{u_{x,t} := \Delta\tilde{X}_t\}$ and $\Omega_{xx}^{1/2}$ is the unique and symmetric matrix square root of Ω_{xx} such that $\Omega_{xx}^{1/2}(\Omega_{xx}^{1/2})' = \Omega_{xx}$.

Example 3 ***Near-unit-root process.*** *Let* $X_t = (1, \tilde{X}_t')'$ *and* $\tilde{X}_t$ *be a near-unit-root process:*

$$\tilde{X}_t = \left(1 - \frac{c}{T}\right)\tilde{X}_{t-1} + u_{x,t} \text{ for } t = 1, 2, \ldots, T,$$

where $c > 0$, $\tilde{X}_0 = o_p(\sqrt{T})$, *and* $u_{x,t}$ *is a stationary zero-mean process. In this case,*

$$D_T = diag(1, \sqrt{T}, \ldots, \sqrt{T}) \text{ and } B_x(r) = \begin{pmatrix} 1 \\ \Omega_{xx}^{1/2} J_{c,x}(r) \end{pmatrix},$$

where $J_{c,x}(\cdot)$ *is the Ornstein-Uhlenbeck (OU) process defined by*

$$dJ_{c,x}(r) = -cJ_{c,x}(r)\,dr + dW_x(r)$$

with $J_{c,x}(0) = 0$*. That is,* $J_{c,x}(r) = \int_0^r e^{-c(r-s)} dW_x(s)$.

Example 4 ***Structural break.*** *Let* $X_t' = [X_t^{\circ\prime} 1(t \leq \lambda T), X_t^{\circ\prime} 1(t > \lambda T)]$ *and* $\beta_0 = (\beta_{10}', \beta_{20}')'$ *so that*

$$Y = X_t'\beta_0 + u_t = X_t^{\circ\prime} 1(t \leq \lambda T)\beta_{10} + X_t^{\circ\prime} 1(t > \lambda T)\beta_{20} + u_t.$$

This model allows for a structural break in the linear relationship between Y_t *and* X_t°*. The possible break takes place at time* $t = \lambda T$*, where, for convenience,* λT *is assumed to be an integer. Assuming that*

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t \\ D_T^{-1} X_{[Tr]}^{\circ} \end{pmatrix} \Rightarrow \begin{pmatrix} \omega_u W_u(r) \\ B_x^{\circ}(r) \end{pmatrix}$$

for some D_T *and stochastic processes* $W_u(\cdot)$ *and* $B_x^{\circ}(\cdot)$*, we have*

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t \\ D_T^{-1} X_{[Tr]} \end{pmatrix} \Rightarrow \begin{pmatrix} \omega_u W_u(r) \\ B_x(r) \end{pmatrix},$$

where

$$B_x(r) = \begin{pmatrix} B_x^{\circ}(r) 1\{r \leq \lambda\} \\ B_x^{\circ}(r) 1\{r > \lambda\} \end{pmatrix}.$$

Example 5 ***High-order integrated process.*** *Let* $X_t = (1, \tilde{X}_t')'$ *and* $\tilde{X}_t$ *be an* $I(2)$ *process:*

$$(1 - L)^2 \tilde{X}_t = u_{x,t} \text{ for } t = 1, 2, \ldots, T,$$

where L *is the lag operator,* $\tilde{X}_0 = o_p(\sqrt{T})$*, and* $u_{x,t}$ *is a stationary zero-mean process. In this case,*

$$D_T = diag(1, T^{3/2}, \ldots, T^{3/2}) \text{ and } B_x(r) = \begin{pmatrix} 1 \\ \Omega_{xx}^{1/2} \int_0^r W_x(s)\,ds \end{pmatrix},$$

where $W_x(\cdot)$ *is a standard Brownian motion process.*

In the above examples, the nonstationary regressors are of the same type. In principle, our setting can also accommodate regressors of different types given in these examples. The asymptotic F and t theory holds as long as Assumption 1 is satisfied.

Under Assumption 1, we have

$$\sqrt{T}D_T(\hat{\beta}-\beta) = \left(\frac{1}{T}\sum_{t=1}^{T}\left(D_T^{-1}X_t\right)\left(D_T^{-1}X_t\right)'\right)^{-1}\left(\sum_{t=1}^{T}\left(D_T^{-1}X_t\right)\frac{u_t}{\sqrt{T}}\right)$$
$$\Rightarrow \omega_u\left[\int_0^1 B_x(r)\,B_x(r)'\,dr\right]^{-1}\int_0^1 B_x(r)\,dW_u(r).$$

Different elements of $\hat{\beta}$ may have different rates of convergence. To find the asymptotic distribution of $R(\hat{\beta}-\beta_0)$, we need to find the slowest rate of convergence among the elements of $\hat{\beta}$ that are involved in each restriction. More specifically, for the i-th restriction $R(i,\cdot)\beta = r_i$ where $R(i,\cdot)$ is the i-th row of R and r_i is the i-th element of r, we define the set

$$\mathcal{S}_i := \{j: \text{ for } j\in\{1,2,\ldots,d\} \text{ such that } R(i,j)\neq 0\},$$

which consists of the indices of the coefficients that appear in the i-th restriction. The rate of convergence of $R(i,\cdot)\hat{\beta}$ is given by $\sqrt{T}\min_{j\in S_i} D_T(j,j)$. Let

$$\tilde{D}_T = \operatorname{diag}\left(\min_{j\in\mathcal{S}_1} D_T(j,j),\ldots,\min_{j\in\mathcal{S}_p} D_T(j,j)\right),$$

which is a $p\times p$ diagonal matrix. Then $\lim_{T\to\infty}\tilde{D}_T R D_T^{-1} = A$ for a matrix $A\in\mathbb{R}^{p\times d}$ whose (i,j)-th element $A(i,j)$ is equal to

$$A(i,j) = \lim_{T\to\infty}\tilde{D}_T(i,i)\,R(i,j)\,/D_T(j,j) = R(i,j)\,1\{j\in\mathcal{S}_i\}.$$

That is, A is the same as R after we zero out the elements in each row for which the corresponding coefficients can be estimated at a faster rate. So, under the null H_0,

$$\begin{aligned}
\tilde{D}_T\sqrt{T}\left(R\hat{\beta}-r\right) &= \tilde{D}_T\sqrt{T}R\left(\hat{\beta}-\beta_0\right)\\
&= \left(\tilde{D}_T R D_T^{-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\left(D_T^{-1}X_t\right)\left(D_T^{-1}X_t\right)'\right)^{-1}\left(\sum_{t=1}^{T}\left(D_T^{-1}X_t\right)\frac{u_t}{\sqrt{T}}\right)\\
&\Rightarrow \omega_u A\left[\int_0^1 B_x(r)\,B_x(r)'\,dr\right]^{-1}\int_0^1 B_x(r)\,dW_u(r)\\
&:= \omega_u\int_0^1 B_x^*(r)\,dW_u(r), \qquad (1)
\end{aligned}$$

where

$$B_x^*(r) = A\left[\int_0^1 B_x(r)\,B_x(r)'\,dr\right]^{-1} B_x(r).$$

The above asymptotic theory forms the basis for testing H_0 against H_1, but we still have to estimate the long run variance ω_u^2. Here we employ the series approach. Let $\{\phi_i(\cdot), i = 1, 2, \ldots, K\}$ be some basis functions on $L^2[0,1]$. The series estimator of ω_u^2 is

$$\hat{\omega}_u^2 = \frac{1}{K} \sum_{i=1}^{K} \left[\frac{1}{\sqrt{T}} \sum_{i=1}^{T} \phi_i \left(\frac{t}{T} \right) \hat{u}_t \right]^2 \tag{2}$$

where $\hat{u}_t = Y_t - X_t' \hat{\beta}$. As a rule of thumb, we can use the formula developed by Phillips (2005) to choose K. See Section 4 for more details.

Based on $\hat{\omega}_u^2$, we construct the Wald statistic

$$F_T = \frac{\left(R\hat{\beta} - r\right)' \left[R \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1} R \right]^{-1} \left(R\hat{\beta} - r \right)}{p \hat{\omega}_u^2}. \tag{3}$$

When $p = 1$, we construct the t statistic

$$t_T = \frac{R\hat{\beta} - r}{\hat{\omega}_u \sqrt{R \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1} R}}.$$

Let

$$\tilde{\phi}_i(r) = \phi_i(r) - \left[\int_0^1 \phi_i(s) B_x(s)' ds \right] \left[\int_0^1 B_x(\tau) B_x(\tau)' d\tau \right]^{-1} B_x(r), \tag{4}$$

which is the projection of $\phi_i(r)$ onto the orthogonal complement of the space spanned by $B_x(r)$. By construction, $\int_0^1 \tilde{\phi}_i(r) B_x'(r) dr = 0$.[1]

Denote $\tilde{\boldsymbol{\phi}}(r) = (\tilde{\phi}_1(r), \ldots, \tilde{\phi}_K(r))'$. To obtain the weak limits of $\hat{\omega}_u^2$, F_T, and t_T, we make the following assumption on the basis functions.

Assumption 2 *(i) For each $i = 1, \ldots, K$, $\phi_i(\cdot)$ is continuously differentiable. (ii) $\int_0^1 \tilde{\boldsymbol{\phi}}(r) \tilde{\boldsymbol{\phi}}(r)' dr$ is of rank K almost surely.*

Using Assumption 1, we have

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \hat{u}_t \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \left(Y_t - X_t' \hat{\beta} \right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \left(Y_t - X_t' \beta_0 - X_t' (\hat{\beta} - \beta_0) \right) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t - \frac{1}{T} \sum_{t=1}^{[Tr]} D_T^{-1} X_t' \left(\frac{1}{T} \sum_{t=1}^{T} \left(D_T^{-1} X_t \right) \left(D_T^{-1} X_t \right)' \right)^{-1} \left(\sum_{t=1}^{T} \left(D_T^{-1} X_t \right) \frac{u_t}{\sqrt{T}} \right) \\
&\Rightarrow B_u(r) - \left(\int_0^r B_x(s)' ds \right) \left[\int_0^1 B_x(\tau) B_x(\tau)' d\tau \right]^{-1} \int_0^1 B_x(\tau) dB_u(\tau).
\end{aligned}$$

[1] Here '0' stands for a vector of zeros, and its dimension may be different at different occurrences.

Combining this with Assumption 2(i), we have

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_i\left(\frac{t}{T}\right)\hat{u}_t \\
&\Rightarrow \int_0^1 \phi_i(r)\,dB_u(r) - \left[\int_0^1 \phi_i(s)\,B_x(s)'ds\right]\left[\int_0^1 B_x(\tau)B_x(\tau)'d\tau\right]^{-1}\int_0^1 B_x(\tau)dB_u(\tau) \\
&= \int_0^1 \left\{\phi_i(r) - \left[\int_0^1 \phi_i(s)\,B_x(s)'ds\right]\left[\int_0^1 B_x(\tau)B_x(\tau)'d\tau\right]^{-1} B_x(r)\right\} dB_u(r) \\
&= \omega_u \int_0^1 \tilde{\phi}_i(r)\,dW_u(r).
\end{aligned}
$$

Hence, for a fixed K,

$$
\hat{\omega}_u^2 \Rightarrow \omega_u^2 \frac{1}{K}\sum_{i=1}^{K}\left[\int_0^1 \tilde{\phi}_i(r)\,dW_u(r)\right]^2.
$$

Under the fixed-K asymptotics, $\hat{\omega}_u^2$ converges weakly to a random variable that is proportional to ω_u^2. This is sufficient for asymptotically pivotal inference.

For the variance term in the test statistic F_T, we have

$$
\begin{aligned}
&\tilde{D}_T R\left(\sum_{t=1}^{T} X_t X_t'\right)^{-1} R\tilde{D}_T \\
&= \tilde{D}_T R D_T^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\left(D_T^{-1}X_t\right)\left(D_T^{-1}X_t\right)'\right)^{-1}\left(\tilde{D}_T R D_T^{-1}\right)' \\
&\Rightarrow A\left[\int_0^1 B_x(r)\,B_x(r)'\right]^{-1} A' = \int_0^1 B_x^*(r)\,B_x^*(r)'\,dr.
\end{aligned}
$$

Using the above weak convergence results and that for $R\hat{\beta} - r$ in (1), we obtain, for a fixed K,

$$
\begin{aligned}
F_T &= \frac{1}{p\hat{\omega}_u^2}\left\{\tilde{D}_T\sqrt{T}R\left(\hat{\beta}-\beta_0\right)\right\}'\left[\tilde{D}_T R\left(\sum_{t=1}^{T} X_t X_t'\right)^{-1} R\tilde{D}_T\right]^{-1}\left\{\tilde{D}_T\sqrt{T}R\left(\hat{\beta}-\beta_0\right)\right\} \\
&\Rightarrow \frac{\left\{\int_0^1 B_x^*(r)\,dW_u(r)\right\}'\left\{\int_0^1 B_x^*(r)\,B_x^*(r)'\,dr\right\}^{-1}\left\{\int_0^1 B_x^*(r)\,dW_u(r)\right\}/p}{\sum_{i=1}^{K}\left[\int_0^1 \tilde{\phi}_i(r)\,dW_u(r)\right]^2/K} \\
&:= \frac{\eta_0'\eta_0/p}{\sum_{i=1}^{K}\eta_i^2/K} := F_\infty, \qquad (5)
\end{aligned}
$$

where

$$
\begin{aligned}
\eta_0 &= \left\{\int_0^1 B_x^*(r)\,B_x^*(r)'\,dr\right\}^{-1/2}\int_0^1 B_x^*(r)\,dW_u(r), \\
\eta_i &= \int_0^1 \tilde{\phi}_i(r)\,dW_u(r) \text{ for } i = 1, 2, \ldots, K.
\end{aligned}
$$

In general, $\int_0^1 B_x^*(r) B_x^*(r)' dr$ may not be invertible. Here we assume that $A = \lim_{T\to\infty} \tilde{D}_T R D_T^{-1}$ is of full (row) rank p so that $\int_0^1 B_x^*(r) B_x^*(r)' dr$ is invertible almost surely. The above asymptotic theory allows us to make asymptotically valid inferences, but the limiting distribution F_∞ is nonstandard, and so critical values have to be simulated.

We proceed to study how we may obtain a standard fixed-K asymptotic distribution. Note that B_x^* is a function of B_x. Conditional on B_x, η_0 is a standard normal vector. So $\eta_0'\eta_0 \sim \chi_p^2$ conditional on B_x. Moreover, conditional on B_x, both η_0 and η_i are normal, and their conditional covariance given B_x is

$$cov(\eta_0, \eta_i) = \left\{ \int_0^1 B_x^*(r) B_x^*(r)' dr \right\}^{-1/2} \left(\int_0^1 B_x^*(r) \tilde{\phi}_i(r) dr \right) = 0 \text{ for } i = 1, \ldots, K.$$

Here, the second equality holds because

$$\int_0^1 B_x^*(r) \tilde{\phi}_i(r) dr = A \left[\int_0^1 B_x(r) B_x(r)' dr \right]^{-1} \int_0^1 B_x(r) \tilde{\phi}_i(r) dr = 0,$$

where we have used $\int_0^1 B_x(r) \tilde{\phi}_i(r) dr = 0$. Therefore, conditional on B_x, η_0 and $\{\eta_i, i = 1, \ldots, K\}$ are independent.

To reduce the asymptotic distribution F_∞ to a standard F distribution, we hope that η_i is i.i.d. $N(0,1)$ conditional on B_x. For this, we require $\tilde{\phi}_i(r)$ to be orthonormal (conditional on B_x). But

$$\begin{aligned}
&\int_0^1 \tilde{\phi}_{i_1}(r) \tilde{\phi}_{i_2}(r) dr \\
&= \int_0^1 \phi_{i_1}(r) \phi_{i_2}(r) dr - \left[\int_0^1 \phi_{i_1}(r) B_x(r)' dr \right] \left[\int_0^1 B_x(\tau) B_x(\tau)' d\tau \right]^{-1} \int_0^1 B_x(s) \phi_{i_2}(s) ds \\
&= \int_0^1 \int_0^1 \phi_{i_1}(r) \left\{ \delta(r-s) - B_x(r)' \left[\int_0^1 B_x(\tau) B_x(\tau)' d\tau \right]^{-1} B_x(s) \right\} \phi_{i_2}(s) dr ds \\
&:= \int_0^1 \int_0^1 \phi_{i_1}(r) C(r,s) \phi_{i_2}(s) dr ds,
\end{aligned}$$

where

$$C(r,s) = \delta(r-s) - B_x(s)' \left[\int_0^1 B_x(\tau) B_x(\tau)' d\tau \right]^{-1} B_x(r),$$

and $\delta(\cdot)$ is the Dirac delta function such that

$$\int_0^1 \int_0^1 \phi_{i_1}(s) \delta(r-s) \phi_{i_2}(r) dr ds = \int_0^1 \phi_{i_1}(r) \phi_{i_2}(r) dr.$$

To design the basis functions $\{\phi_i(r)\}$ such that $\{\tilde{\phi}_i(r)\}$ are orthonormal on $L^2[0,1]$, we require that $\{\phi_i(r)\}$ be orthonormal with respect to the weighting function $C(r,s)$, that is,

$$\int_0^1 \int_0^1 C(r,s) \phi_{i_1}(r) \phi_{i_2}(s) dr ds = 1\{i_1 = i_2\} \tag{6}$$

almost surely. Because of the general form of $C(r,s)$ and its randomness, commonly-used basis functions do not satisfy the above condition.

Instead of searching for the basis functions that satisfy (6), we search for their discrete versions: the basis vectors. For each basis function $\phi_i(r)$, the corresponding basis vector is defined as

$$\boldsymbol{\phi}_i = \left(\phi_i\left(\frac{1}{T}\right), \phi_i\left(\frac{2}{T}\right), \ldots, \phi_i\left(\frac{T}{T}\right)\right)'.$$

We note that it is the basis vectors that are used in the variance estimator. Basis functions only appear in the limiting distribution when $T \to \infty$.

Let

$$\mathbf{C}_T = T \cdot M_X \text{ for } M_X = I_T - X\left(X'X\right)^{-1} X',$$

where $X = (X_1, ..., X_T)' \in \mathbb{R}^{T\times d}$. By definition, $\mathbf{C}_T$ is symmetric and positive semidefinite. It is the discrete version of $C(r,s)$. For any two vectors $r_1, r_2 \in \mathbb{R}^T$, we define their inner product as

$$\langle r_1, r_2\rangle = r_1'\mathbf{C}_T r_2/T^2. \tag{7}$$

The discrete analogue of (6) is

$$\left\langle \boldsymbol{\phi}_{i_1}, \boldsymbol{\phi}_{i_2}\right\rangle = 1\{i_1 = i_2\} \text{ for } i_1, i_2 = 1, \ldots, K. \tag{8}$$

Given any set of basis vectors $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K$, let $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K)$ be the $T\times K$ matrix of these basis vectors. Define

$$\boldsymbol{\phi}^M = \sqrt{T}\left(M_X\boldsymbol{\phi}\right)\left[\left(M_X\boldsymbol{\phi}\right)' M_X\boldsymbol{\phi}\right]^{-1/2},$$

where the superscript 'M' signifies that the basis vectors in $\boldsymbol{\phi}^M$ are obtained via a transformation involving M_X. We have

$$\begin{aligned} & T^{-2}(\boldsymbol{\phi}^M)'\mathbf{C}_T\boldsymbol{\phi}^M \\ & = TT^{-2}\left[\left(M_X\boldsymbol{\phi}\right)' M_X\boldsymbol{\phi}\right]^{-1/2}\boldsymbol{\phi}' M_X \cdot \mathbf{C}_T \cdot M_X\boldsymbol{\phi}\left[\left(M_X\boldsymbol{\phi}\right)' M_X\boldsymbol{\phi}\right]^{-1/2} \\ & = \left[\boldsymbol{\phi}' M_X\boldsymbol{\phi}\right]^{-1/2}\left(\boldsymbol{\phi}' M_X\boldsymbol{\phi}\right)\left[\boldsymbol{\phi}' M_X\boldsymbol{\phi}\right]^{-1/2} = I_K. \end{aligned}$$

That is, the columns of the matrix $\boldsymbol{\phi}^M$ satisfy the conditions in (8).

If we use $\{\boldsymbol{\phi}_i^M = (\phi_{i,1}^M, \ldots, \phi_{i,T}^M)', i = 1, ..., M\}$, the columns of $\boldsymbol{\phi}^M$, in constructing the variance estimator, that is, we estimate ω_u^2 by

$$\hat{\omega}_{u,M}^2 = \frac{1}{K}\sum_{i=1}^{K}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_{i,t}^M\hat{u}_t\right]^2,$$

then

$$\begin{aligned}\hat{\omega}_{u,M}^2 &= \frac{1}{TK}\hat{u}'\boldsymbol{\phi}^M\left(\boldsymbol{\phi}^M\right)'\hat{u} \\ &= \frac{1}{K}\left(\hat{u}'M_X\boldsymbol{\phi}\right)\left(\boldsymbol{\phi}'M_X\boldsymbol{\phi}\right)^{-1}\left(\boldsymbol{\phi}'M_X\hat{u}\right) = \frac{1}{K}\left(\hat{u}'\boldsymbol{\phi}\right)\left(\boldsymbol{\phi}'M_X\boldsymbol{\phi}\right)^{-1}\left(\boldsymbol{\phi}'\hat{u}\right) \\ &\Rightarrow \omega_u^2\frac{1}{K}\left[\int_0^1 \tilde{\boldsymbol{\phi}}(r)'\,dW_u(r)\right]\left[\int_0^1 \tilde{\boldsymbol{\phi}}(r)\,\tilde{\boldsymbol{\phi}}(r)'\,dr\right]^{-1}\left[\int_0^1 \tilde{\boldsymbol{\phi}}(r)\,dW_u(r)\right]. \qquad (9)\end{aligned}$$

In the above, the third equality holds because $\hat{u} = M_X u$ and so $\hat{u}'M_X\boldsymbol{\phi} = u'M_X'M_X\boldsymbol{\phi} = u'M_XM_X\boldsymbol{\phi} = u'M_X\boldsymbol{\phi} = \hat{u}'\boldsymbol{\phi}$. As a result,

$$F_T^* := \frac{\left(R\hat{\beta}-r\right)'\left[R\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}-r\right)/p}{\left(\hat{u}'M_X\boldsymbol{\phi}\right)\left(\boldsymbol{\phi}'M_X\boldsymbol{\phi}\right)^{-1}\left(\boldsymbol{\phi}'M_X\hat{u}\right)/K} \Rightarrow \frac{\eta_0'\eta_0/p}{\eta'\eta/K},$$

where, as before

$$\eta_0 = \left\{\int_0^1 B_x^*(r)\,B_x^*(r)'\,dr\right\}^{-1/2}\int_0^1 B_x^*(r)\,dW_u(r),$$

but now

$$\eta := (\eta_1, \eta_2, \ldots, \eta_K)' = \left[\int_0^1 \tilde{\boldsymbol{\phi}}(r)\,\tilde{\boldsymbol{\phi}}(r)'\,dr\right]^{-1/2}\left[\int_0^1 \tilde{\boldsymbol{\phi}}(r)\,dW_u(r)\right].$$

Under Assumption 2(ii), $\eta \sim N(0, I_K)$ conditional on B_x. Also, as we have shown before, η is independent of η_0 conditional on B_x. So, conditional on B_x,

$$\frac{\eta_0'\eta_0/p}{\eta'\eta/K} \sim F_{p,K}.$$

Given that the conditional distribution $F_{p,K}$ does not depend on the conditioning variable (i.e., B_x), $\frac{\eta_0'\eta_0/p}{\eta'\eta/K} \sim F_{p,K}$ unconditionally. Therefore, $F_T^* \Rightarrow F_{p,K}$. We formalize this result and a similar result for a t statistic in the theorem below.

Theorem 1 *Let Assumptions 1 and 2 hold. Assume further that* $\lim_{T\to\infty}\tilde{D}_TRD_T^{-1}$ *is of full row rank p. Then, for a fixed K as* $T\to\infty$,

$$\begin{aligned}F_T^* &:= \frac{\left(R\hat{\beta}-r\right)'\left[R\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}-r\right)/p}{\left(\hat{u}'M_X\boldsymbol{\phi}\right)\left(\boldsymbol{\phi}'M_X\boldsymbol{\phi}\right)^{-1}\left(\boldsymbol{\phi}'M_X\hat{u}\right)/K} \to F_{p,K} \\ t_T^* &:= \frac{R\hat{\beta}-r}{\sqrt{\left(\hat{u}'M_X\boldsymbol{\phi}\right)\left(\boldsymbol{\phi}'M_X\boldsymbol{\phi}\right)^{-1}\left(\boldsymbol{\phi}'M_X\hat{u}\right)/K}\sqrt{R\left(\sum_{t=1}^T X_tX_t'\right)^{-1}R'}} \to t_K,\end{aligned}$$

where $M_X = I_T - X\left(X'X\right)^{-1}X'$, $F_{p,K}$ *is the standard F distribution with degrees of freedom* (p, K), *and* t_K *is the standard t distribution with degrees of freedom K.*

Remark 1 *In order to develop the asymptotic F and t theory, we use the novel series variance estimator* $(TK)^{-1}\hat{u}'\boldsymbol{\phi}^M\left(\boldsymbol{\phi}^M\right)'\hat{u}$ *instead of the usual series variance estimator* $(TK)^{-1}\left(\hat{u}'\boldsymbol{\phi}\right)\left(\boldsymbol{\phi}'\hat{u}\right)$. *To obtain* $\boldsymbol{\phi}^M$, *we first project* $\boldsymbol{\phi}$ *on the orthogonal complement of the column space of* X *to obtain* $M_X\boldsymbol{\phi}$ *and then orthonormalize it into* $\boldsymbol{\phi}^M = \sqrt{T}\left(M_X\boldsymbol{\phi}\right)\left[\left(M_X\boldsymbol{\phi}\right)' M_X\boldsymbol{\phi}\right]^{-1/2}$. *Note that*

$$(TK)^{-1}\hat{u}'\boldsymbol{\phi}^M\left(\boldsymbol{\phi}^M\right)'\hat{u} = (TK)^{-1}u'\boldsymbol{\phi}^M\left(\boldsymbol{\phi}^M\right)'u.$$

The series variance estimator is the same regardless of whether $\hat{u}$ *or the true* u *is used. The projection, therefore, ensures that the series variance estimator* $(TK)^{-1}\hat{u}'\boldsymbol{\phi}^M\left(\boldsymbol{\phi}^M\right)'\hat{u}$ *is invariant to the parameter estimation error. When* X *is random, the projection is onto a random subspace. For series variance estimation, the idea of using data-dependent and randomly orthonormalized basis functions is new in the literature.*

Remark 2 *From a theoretical point of view, the implied basis functions we use in the series variance estimation are the elements of the following vector:*

$$\tilde{\boldsymbol{\phi}}^*(r) = \left[\int_0^1 \tilde{\boldsymbol{\phi}}(\tau)\,\tilde{\boldsymbol{\phi}}(\tau)'\,d\tau\right]^{-1/2}\tilde{\boldsymbol{\phi}}(r).$$

From any given vector of basis functions $\boldsymbol{\phi}(r) = (\phi_1(r), \ldots, \phi_K(r))'$, *we first use the projection operation given in (4) to obtain* $\tilde{\boldsymbol{\phi}}(r) = (\tilde{\phi}_1(r), \ldots, \tilde{\phi}_K(r))'$ *and then use the orthonormalization (i.e., premultiplying* $\tilde{\boldsymbol{\phi}}(r)$ *by* $[\int_0^1 \tilde{\boldsymbol{\phi}}(\tau)\,\tilde{\boldsymbol{\phi}}(\tau)'\,d\tau]^{-1/2}$*) to obtain* $\tilde{\boldsymbol{\phi}}^*(r) := [\int_0^1 \tilde{\boldsymbol{\phi}}(\tau)\,\tilde{\boldsymbol{\phi}}(\tau)'\,d\tau]^{-1/2}\tilde{\boldsymbol{\phi}}(r)$. *For practical implementation, we do not need to find the implied basis functions, as we only need to use the corresponding basis vectors* $\left(\boldsymbol{\phi}' M_X\boldsymbol{\phi}\right)^{-1/2}\left(\boldsymbol{\phi}' M_X\right)$ *to compute* $\hat{\omega}_u^2$. *The basis functions in* $\tilde{\boldsymbol{\phi}}^*(r)$ *appear only in the asymptotic distributions of the long run variance estimator* $\hat{\omega}_u^2$ *and the test statistics* F_T^* *and* t_T^*. *More specifically, we can rewrite (9) as*

$$\hat{\omega}_{u,M}^2 \Rightarrow \omega_u^2\frac{1}{K}\left[\int_0^1 \tilde{\boldsymbol{\phi}}^*(r)\,dW_u(r)\right]'\left[\int_0^1 \tilde{\boldsymbol{\phi}}^*(r)\,dW_u(r)\right].$$

The basis functions in $\tilde{\boldsymbol{\phi}}^*(r)$ *appear in our asymptotic distributions only via the above weak limit of* $\hat{\omega}_{u,M}^2$.

Remark 3 *Theorem 1 extends Sun (2011) to allow for more general trend functions. For trend regressions, Sun (2011) considers a linear trend with* $X_t = (1,t)'$ *so that* $B_x(r) = (1,r)'$ *and employs the cosine basis functions* $\phi_i(r) = \sqrt{2}\cos 2\pi i r$ *for* $i = 1, 2, \ldots, K$. *These functions are special in that they are orthonormal on* $L^2[0,1]$ *and satisfy*

$$\int_0^1 \phi_i(s)\,B_x(s)\,ds = \begin{pmatrix} \int_0^1 \sqrt{2}\cos(2\pi s)\,ds \\ \int_0^1 s\sqrt{2}\cos(2\pi s)\,ds \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

As a result, $\tilde{\phi}_i(r) = \phi_i(r) - \left[\int_0^1 \phi_i(s) B_x(s)' ds\right] \left[\int_0^1 B_x(\tau) B_x(\tau)' d\tau\right]^{-1} B_x(r) = \phi_i(r)$ *for all* i*. That is, the projection does not change the original basis functions. Hence,* $\eta_i := \int_0^1 \tilde{\phi}_i(r) dW_u(r) = \int_0^1 \phi_i(r) dW_u(r)$ *is i.i.d.* $N(0,1)$*. The asymptotic* F *and* t *theory can then be established directly with the cosine basis functions. However, the requirements that* $\tilde{\phi}_i(r) = \phi_i(r)$ *for all* i *and* $\{\phi_i(r)\}$ *are orthonormal on* $L^2[0,1]$ *severely limit the set of basis functions we can use. For example, Sun (2011) has to rule out the sine functions* $\{\sqrt{2}\sin 2\pi i r\}$ *with consequential adverse effect on the power of the resulting test. In contrast, Theorem 1 allows us to use any basis functions that satisfy Assumption 2. We can do so because the projection step preemptively purges the effect of the parameter estimation uncertainty and the orthonormalization step ensures the orthonormality of the implied basis functions* $\{\tilde{\phi}_i^*(r)\}$*.*

Remark 4 *For regressions with strictly exogenous integrated regressors, Theorem 1 can be regarded as an F test version of Park and Phillips (1988) (Theorem 5.4) where the asymptotic chi-square test theory was developed. We note that an asymptotic F theory can not be established for the usual test statistic constructed based on a kernel (long-run) variance estimator. A series variance estimator with carefully crafted basis functions/vectors appears to be indispensable for the asymptotic F theory.*

Remark 5 *For cointegration regressions,Bunzel (2006) and Jin et al. (2006) develop the fixed-b asymptotic theory for studentized test statistics. The asymptotic distributions in these two papers are nonstandard. In contrast, the asymptotic distributions in Theorem 1 are standard F and t distributions and are thus more convenient for practical use. In the presence of endogeneity, the case considered in the next section, we may employ the series long run and half long run variance estimators to obtain the fully-modified OLS estimator of Phillips and Hansen (1990). Suppose we ignore the estimation error in the half long run variance estimator, which effectively reduces the problem back to the exogenous case, then we can use the F and t approximations for inference. This may not be completely satisfactory, because the F and t distributions are not the exact asymptotic distributions of the F and t statistics under the fixed-K asymptotics. We use the F and t approximations only because they are expected to be more accurate than the chi-square and normal approximations. For the endogenous case, exact and asymptotically valid F and t tests are developed in the next section.*

Remark 6 *Theorem 1 allows for near-unit-root processes. It appears to be the first time that an asymptotic F and t theory is established in this setting. However, see Sun (2014c) for the F* *and t limit theory in a different setting where the regressor error is a near-unit-root process and Guo et al. (2018) for the asymptotic t theory in an autoregression where the process is moderately explosive. Theorem 1 also allows for I(2) processes. To the best of our knowledge, this has not been considered in the literature before.*

Remark 7 *The asymptotic F and t theory in the structural break setting given above appears to be new. Sun and Wang (2021) establish the asymptotic F and t theory in a structural break model, but they consider only the case when X_t° is stationary. Here we allow X_t° to be nonstationary.*

Remark 8 *Theorem 1 provides a unified framework that accommodates various nonstationary regressors. The idea may be extended under suitable conditions to allow for fractionally integrated processes, slowing-varying trend regressors, and nonlinear trends. See Phillips (2007) for slowing-varying trend regressions.*

To conclude this section, we outline the steps in conducting the asymptotic F and t tests:

(i) Estimate β_0 by the OLS estimator $\hat{\beta}$ and calculate the residual $\hat{u} = Y - X\hat{\beta}$.

(ii) Construct the $T \times K$ matrix $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K)$ of the basis vectors. We recommend using the following Fourier series as the basis vectors:

$$\begin{aligned} \boldsymbol{\phi}_{2i-1} &= (\sqrt{2}\cos 2\pi i\frac{t}{T})_{t=1}^{T} = \sqrt{2}(\cos 2\pi i\frac{1}{T}, \cos 2\pi i\frac{2}{T}, ..., \cos 2\pi i\frac{T}{T})' \\ \boldsymbol{\phi}_{2i} &= (\sqrt{2}\sin 2\pi i\frac{t}{T})_{t=1}^{T} = \sqrt{2}(\sin 2\pi i\frac{1}{T}, \sin 2\pi i\frac{2}{T}, ..., \sin 2\pi i\frac{T}{T})' \end{aligned} \tag{10}$$

for $i = 1, 2, ..., K/2$ assuming that K is even.

(iii) For $M_X = I_T - X\left(X'X\right)^{-1}X'$, compute the test statistics F_T^* and t_T^* defined in Theorem 1.

(iv) For the F test, compare F_T^* with critical values from the standard F distribution $F_{p,K}$. For the t test, compare t_T^* with critical values from the standard t distribution t_K.

3 Endogenous case

3.1 Cointegration regression

We consider the model

$$\begin{aligned} Y_t &= \alpha_0 + X_t'\beta_0 + u_{0t} \\ X_t &= \left(1 - \frac{c}{T}\right)X_{t-1} + u_{xt} \end{aligned} \tag{11}$$

for $t = 1, \ldots, T$, where Y_t is a scalar time series and X_t is a $d_x \times 1$ vector of time series with $X_0 = o_p(\sqrt{T})$. Here we single out the intercept because we allow the rest of the regressors X_t to be endogenous: $\{u_{xt}\}$ and $\{u_{0t}\}$ can be arbitrarily correlated. We assume that $c \geq 0$ so that both unit-root and near-unit-root processes can be accommodated.

We are interested in constructing a confidence interval for $R\beta_0$. The confidence interval can then be used in testing whether $R\beta_0 = r$ for some $r \in \mathbb{R}^p$.

We maintain the following assumption on $u_t = (u_{0t}, u_{xt}')'$. Note that the definition of u_t is different from that in the previous section.

Assumption 3 *The functional central limit theorem holds:*

$$T^{-1/2} \sum_{t=1}^{[T\cdot]} u_t \Rightarrow B(\cdot) = \Omega^{1/2} W(\cdot), \tag{12}$$

where $W(\cdot) := (W_0(\cdot), W_x'(\cdot))'$ *is a* $(d_x + 1)$*-dimensional standard Brownian process,*

$$\Omega = \sum_{j=-\infty}^{\infty} E u_t u_{t-j}' = \begin{pmatrix} \underset{1\times 1}{\omega_0^2} & \underset{1\times d_x}{\omega_{0x}} \\ \underset{d_x \times 1}{\omega_{x0}} & \underset{d_x \times d_x}{\Omega_{xx}} \end{pmatrix}, \tag{13}$$

and Ω *is positive definite.*

Using the Cholesky form of $\Omega^{1/2}$, we can write $B(\cdot)$ as

$$B(\cdot) = \begin{pmatrix} B_0(\cdot) \\ B_x(\cdot) \end{pmatrix} = \begin{pmatrix} \omega_{0\cdot x} W_0(\cdot) + \omega_{0x} \Omega_{xx}^{-1/2} W_x(\cdot) \\ \Omega_{xx}^{1/2} W_x(\cdot) \end{pmatrix}, \tag{14}$$

where $\omega_{0\cdot x}^2 = \omega_0^2 - \omega_{0x}\Omega_{xx}^{-1}\omega_{x0}$ and $\Omega_{xx}^{1/2}$ is the symmetric and positive-definite matrix square root of Ω_{xx}.

In the presence of endogeneity, $B_0(\cdot)$ and $B_x(\cdot)$ will be dependent, and the OLS estimator $\hat{\beta}$ of β_0 will have a second-order endogeneity bias and a complicated asymptotic distribution. To remove the endogeneity bias and restore the asymptotic (mixed) normality of $\hat{\beta}$, we may use the fully modified OLS estimator of Phillips and Hansen (1990). This estimator involves using a long run variance and a half long run variance to remove the dependence between $B_0(\cdot)$ and $B_x(\cdot)$ and the endogeneity bias. Both the long run variance and the half long run variance are estimated nonparametrically. However, the estimation uncertainty, which is potentially very high, is ignored in the asymptotic chi-square and normal approximations. For this reason, the chi-square and normal tests often have large size distortion; see, for example, Vogelsang and Wagner (2014).

To confront the size distortion problem, we follow Hwang and Sun (2018)[2] and consider a different estimation approach. We assume that c is known. If X_t is a unit-root process, we know that $c = 0$. So the assumption is not restrictive. However, when X_t is a near-unit-root process, we do not know c in general, and the assumption becomes restrictive. In this case, we can first construct a confidence interval for c and then use Bonferroni's method to construct the confidence bound for $R\beta_0$. In this case, our asymptotic theory below forms the basis for the Bonferroni's method. For more details, see, for example, Phillips (2015) for such a practice in predictive regressions. The same method can be used here.

Define the quasi-differenced process $\Delta_c X_t$ as

$$\Delta_c X_t = X_t - \left(1 - \frac{c}{T}\right) X_{t-1}.$$

Obviously, $\Delta_c X_t = u_{x,t}$. Let $\delta_0 = \Omega_{xx}^{-1}\omega_{x0}$ be the long run regression coefficient when u_{0t} is regressed on u_{xt} and

$$u_{0\cdot x,t} = u_{0t} - (\Delta_c X_t)'\delta_0 = u_{0t} - u_{x,t}'\delta_0$$

be the corresponding long run regression error. Then, we obtain the augmented regression

$$Y_t = \alpha_0 + X_t'\beta_0 + (\Delta_c X_t)'\delta_0 + u_{0\cdot x,t}, \tag{15}$$

where, by definition, the long run correlation between $u_{0\cdot x,t}$ and $u_{x,t}$ is zero. The augmentation is designed to purge the dependence between $B_0(\cdot)$ and $B_x(\cdot)$.

Note that the zero long run correlation between $u_{0\cdot x,t}$ and $u_{x,t}$ does not rule out that $u_{0\cdot x,t}$ may be still correlated with X_t and $\Delta_c X_t$. Hence, the (augmented) OLS estimator (denoted by $\hat{\beta}_{AOLS}$) of β_0 based on the augmented regression can still have a second-order endogeneity bias. More precisely, the mean of the asymptotic distribution of $T(\hat{\beta}_{AOLS} - \beta_0)$ may not be zero. Ignoring the nonzero mean leads to invalid and unreliable statistical inferences.

To remove the endogeneity bias, we follow Phillips (2014) and Hwang and Sun (2018) and run the regression in a different domain, which resembles the frequency domain, but any set of orthonormal basis functions in $L^2[0,1]$ can be used. For convenience, we refer to this domain as the pseudo-frequency domain. Let $\{\phi_i\}_{i=1}^K$ be a set of K such basis functions on $L^2[0,1]$. For

[2] Hwang and Sun (2018) tackle the size distortion problem in a cointegration regression where $\{X_t\}$ are unit-root processes. Here we generalize their asymptotic theory to allow $\{X_t\}$ to be near-unit-root processes. We employ the same argument as in Hwang and Sun (2018). To make the paper self-contained, we outline the main steps of the argument here

each $i=1,\ldots,K$, we transform all variables in the augmented regression into

$$\begin{aligned}\mathbb{W}_i^{\alpha} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_i(\frac{t}{T}),\\ \mathbb{W}_i^{y} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_t\phi_i(\frac{t}{T}),\ \ \mathbb{W}_i^{x} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t\phi_i(\frac{t}{T}),\\ \mathbb{W}_i^{\Delta_c x} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\Delta_c X_t)\phi_i(\frac{t}{T}),\ \ \mathbb{W}_i^{0\cdot x} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{0\cdot x,t}\phi_i(\frac{t}{T}).\end{aligned} \tag{16}$$

The basis transformation is designed to extract the long run component in the time series data.

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

$$\mathbb{W}_i^{y} = \alpha_0\mathbb{W}_i^{\alpha} + \mathbb{W}_i^{x\prime}\beta_0 + \mathbb{W}_i^{\Delta_c x\prime}\delta_0 + \mathbb{W}_i^{0\cdot x} \text{ for } i=1,\ldots,K. \tag{17}$$

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

$$\mathbb{W}_i^{\alpha} = \sqrt{T}\int_0^1\phi_i(r)dr + \sqrt{T}O(1/T) = O(1/\sqrt{T}) = o(1). \tag{18}$$

So

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

where

$$\mathbb{W}_{\alpha,i}^{0\cdot x} = \mathbb{W}_i^{0\cdot x} + \mathbb{W}_i^{\alpha} = \mathbb{W}_i^{0\cdot x} + o(1).$$

Our estimation and inference will be based on equation (19), which can be regarded as a low-frequency regression.

Putting (19) in a vector form, we have

$$\mathbb{W}^{y} = \mathbb{W}^{x}\beta_0 + \mathbb{W}^{\Delta_c x}\delta_0 + \mathbb{W}_{\alpha}^{0\cdot x}, \tag{20}$$

where $\mathbb{W}^y = (\mathbb{W}_1^y,\ldots,\mathbb{W}_K^y)'$ and $\mathbb{W}^x$, $\mathbb{W}^{\Delta x}$, and $\mathbb{W}_{\alpha}^{0\cdot x}$ are defined similarly. Running OLS based on the above equation leads to the 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 $\tilde{\mathbb{W}} = \left(\mathbb{W}^x, \mathbb{W}^{\Delta x}\right)$. See Hwang and Sun (2018) for discussions on the efficiency and robustness of this estimator. The TAOLS estimator can be regarded as an IV estimator based on the augmented equation in (15) using the basis vectors $\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_K$ as the instruments. See Phillips (2014) for more details.

Let

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

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

$$\hat{\gamma}_{TAOLS} = \begin{pmatrix} \hat{\beta}_{TAOLS} \\ \hat{\delta}_{TAOLS} \end{pmatrix} = \begin{pmatrix} (\mathbb{W}^{x\prime}M_{\Delta_c x}\mathbb{W}^x)^{-1}(\mathbb{W}^{x\prime}M_{\Delta_c x}\mathbb{W}^y) \\ (\mathbb{W}^{\Delta_c x\prime}M_x\mathbb{W}^{\Delta_c x})^{-1}(\mathbb{W}^{\Delta_c x\prime}M_x\mathbb{W}^y) \end{pmatrix}. \tag{21}$$

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

Assumption 4 *(i) For every $i = 1, \ldots, K$, $\phi_i(\cdot)$ is continuously differentiable; (ii) for every $i = 1, \ldots, K$, $\phi_i(\cdot)$ satisfies $\int_0^1 \phi_i(r)\,dr = 0$; (iii) the functions $\{\phi_i(\cdot)\}_{i=1}^K$ are orthonormal in $L^2[0,1]$.*

Under Assumptions 3 and 4(i&ii), we can use summation by parts, the continuous mapping theorem, and integration by parts to obtain

$$\begin{aligned}
\mathbb{W}^{0\cdot x}_{\alpha,i} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_i(\frac{t}{T})u_{0\cdot xt} + \mathbb{W}^{\alpha}_i \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_i(\frac{t}{T})\left(u_{0t} - u_{xt}'\delta_0\right) + o(1) \\
&\Rightarrow \int_0^1 \phi_i(r)\,d\left[B_0(r) - B_x(r)'\delta_0\right] \\
&= \omega_{0\cdot x}\int_0^1 \phi_i(r)\,dW_0(r) := \omega_{0\cdot x}\nu_i.
\end{aligned}$$

Similarly,

$$\frac{\mathbb{W}^x_i}{T} = \frac{1}{T^{3/2}}\sum_{t=1}^{T}\phi_i\left(\frac{t}{T}\right)X_t \Rightarrow \Omega_{xx}^{1/2}\int_0^1 \phi_i(r)\,J_{c,x}(r)dr := \xi_i,$$

and

$$\mathbb{W}^{\Delta_c x}_i \Rightarrow \int_0^1 \phi_i(r)\,dB_x(r) = \Omega_{xx}^{1/2}\int_0^1 \phi_i(r)\,dW_x(r) := \eta_i.$$

Let

$$\begin{aligned}
\nu &\equiv (\nu_1, \nu_2, \ldots, \nu_K)' \in \mathbb{R}^{K\times 1}, \\
\xi &\equiv (\xi_1, \xi_2, \ldots, \xi_K)' \in \mathbb{R}^{K\times d_x}, \\
\eta &\equiv (\eta_1, \eta_2, \ldots, \eta_K)' \in \mathbb{R}^{K\times d_x},
\end{aligned}$$

and $\zeta = (\xi, \eta)$. Then

$$\left(\mathbb{W}^x/T, \mathbb{W}^{\Delta x}, \mathbb{W}^{0\cdot x}_{\alpha}\right) \Rightarrow (\xi, \eta, \omega_{0\cdot x}\nu), \tag{22}$$

where $\zeta \perp \nu$. Also, it follows from Assumption 4(iii) that $\nu \sim N(0, I_K)$. In particular, for

$$\Upsilon_T = \begin{pmatrix} T \cdot I_{d_x} & \underset{d_x \times d_x}{O} \\ \underset{d_x \times d_x}{O} & I_{d_x} \end{pmatrix},$$

where O is a matrix of zeros whose dimension may be different at different occurrences, we have $\tilde{\mathbb{W}}\Upsilon_T^{-1} \Rightarrow \zeta$. It then follows that

$$\begin{aligned} \Upsilon_T \left(\hat{\gamma}_{TAOLS} - \gamma_0\right) &= (\Upsilon_T^{-1}\tilde{\mathbb{W}}'\tilde{\mathbb{W}}\Upsilon_T^{-1})^{-1}(\tilde{\mathbb{W}}\Upsilon_T^{-1})'\mathbb{W}_\alpha^{0\cdot x} \\ &\Rightarrow \omega_{0\cdot x}(\zeta'\zeta)^{-1}\zeta'\nu = \omega_{0\cdot x} \begin{pmatrix} \left(\xi' M_\eta \xi\right)^{-1} \xi' M_\eta \nu \\ \left(\eta' M_\xi \eta\right)^{-1} \eta' M_\xi \nu \end{pmatrix}, \end{aligned}$$

where

$$M_\eta = I_{d_x} - \eta \left(\eta'\eta\right)^{-1} \eta' \text{ and } M_\xi = I_{d_x} - \xi \left(\xi'\xi\right)^{-1} \xi'.$$

We formalize the above asymptotic result in the theorem below.

Theorem 2 *Let Assumptions 3 and 4 hold. Then under the fixed-K asymptotics where K is held fixed as $T \to \infty$, we have*

$$T(\hat{\beta}_{TAOLS} - \beta_0) \Rightarrow \omega_{0\cdot x} \left(\xi' M_\eta \xi\right)^{-1} \xi' M_\eta \nu, \tag{23}$$

$$\hat{\delta}_{TAOLS} - \delta_0 \Rightarrow \omega_{0\cdot x} \left(\eta' M_\xi \eta\right)^{-1} \eta' M_\xi \nu, \tag{24}$$

jointly.

Except for the difference in the definitions and the distributions of $\zeta = (\xi, \eta)$, Theorem 2 is identical to Theorem 1 of Hwang and Sun (2018).

Conditional on (ξ, η), both limiting distributions in Theorem 2 are normal with mean zero. There is no second-order endogeneity bias in the TAOLS estimator. As in Hwang and Sun (2018), the TAOLS approach successfully removes the two problems that plague the usual OLS estimator. It paves the way for developing standard inference procedures.

To make inferences on $R\beta_0$, we estimate $\omega_{0\cdot x}^2$ by

$$\begin{aligned} \hat{\omega}_{0\cdot x}^2 &= \frac{1}{K}\sum_{i=1}^{K} \left(\mathbb{W}_i^y - \mathbb{W}_i^{x\prime}\hat{\beta}_{TAOLS} - \mathbb{W}_i^{\Delta_c x\prime}\hat{\delta}_{TAOLS}\right)^2 \\ &= \frac{1}{K}\mathbb{W}_\alpha^{0\cdot x\prime}\left[I_K - \tilde{\mathbb{W}}(\tilde{\mathbb{W}}'\tilde{\mathbb{W}})^{-1}\tilde{\mathbb{W}}'\right]\mathbb{W}_\alpha^{0\cdot x}. \end{aligned}$$

We can then construct the test statistic

$$F(\hat{\beta}_{TAOLS}) = \frac{1}{\hat{\omega}_{0\cdot x}^2}\left[R\left(\hat{\beta}_{TAOLS} - \beta_0\right)\right]'\left[R(\mathbb{W}^{x\prime}M_{\Delta_c x}\mathbb{W}^x)^{-1}R'\right]^{-1}\left[R\left(\hat{\beta}_{TAOLS} - \beta_0\right)\right]/p, \tag{25}$$

and for $p=1$,

$$t(\hat{\beta}_{TAOLS}) = \frac{R\left(\hat{\beta}_{TAOLS}-\beta_0\right)}{\sqrt{\hat{\omega}_{0\cdot x}^2 R(\mathbb{W}^{x\prime}M_{\Delta_c x}\mathbb{W}^x)^{-1}R'}}.$$

Using (22), we have

$$\begin{aligned}
F(\hat{\beta}_{TAOLS}) &\Rightarrow \frac{\left[R\left(\xi' M_\eta \xi\right)^{-1}\xi' M_\eta \nu\right]' \left[R\left(\xi' M_\eta \xi\right)^{-1}R\right]^{-1}\left[R\left(\xi' M_\eta \xi\right)^{-1}\xi' M_\eta \nu\right]/p}{K^{-1}\nu'\left[I_K-\zeta(\zeta'\zeta)^{-1}\zeta\right]\nu} \\
&= \frac{\left\|\left[R\left(\xi' M_\eta \xi\right)^{-1}R\right]^{-1/2}R\left(\xi' M_\eta \xi\right)^{-1}\xi' M_\eta \nu\right\|^2/p}{\left\|\left[I_K-\zeta(\zeta'\zeta)^{-1}\zeta\right]\nu\right\|^2/K}.
\end{aligned}$$

Conditional on $\zeta=(\xi,\eta)$, we have

$$\begin{aligned}
\left\|\left[R\left(\xi' M_\eta \xi\right)^{-1}R\right]^{-1/2}R\left(\xi' M_\eta \xi\right)^{-1}\xi' M_\eta \nu\right\|^2 &\sim \chi_p^2 \\
\left\|\left[I_K-\zeta(\zeta'\zeta)^{-1}\zeta\right]\nu\right\|^2 &\sim \chi_{K-2d_x}^2
\end{aligned}$$

and conditional on ζ,

$$\begin{aligned}
&cov\left(R\left(\xi' M_\eta \xi\right)^{-1}\xi' M_\eta \nu, \left[I_K-\zeta(\zeta'\zeta)^{-1}\zeta\right]\nu\right) \\
&= R\left(\xi' M_\eta \xi\right)^{-1}\xi' M_\eta\left[I_K-\zeta(\zeta'\zeta)^{-1}\zeta\right] \\
&= R\left(\xi' M_\eta \xi\right)^{-1}\xi' M_\eta - R\left(\xi' M_\eta \xi\right)^{-1}\xi'\left[M_\eta\xi, O\right](\zeta'\zeta)^{-1}\zeta \\
&= R\left(\xi' M_\eta \xi\right)^{-1}\xi' M_\eta - [R,O]\begin{pmatrix} \left(\xi' M_\eta \xi\right)^{-1}\xi' M_\eta \\ \left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \end{pmatrix} = O.
\end{aligned}$$

So, conditional on ζ, the numerator and the denominator of the limiting distribution of $F(\hat{\beta}_{TAOLS})$ follow independent chi-square distributions. Hence,

$$\frac{\left\|\left[R\left(\xi' M_\eta \xi\right)^{-1}R\right]^{-1/2}R\left(\xi' M_\eta \xi\right)^{-1}\xi' M_\eta \nu\right\|^2/p}{\left\|\left[I_K-\zeta(\zeta'\zeta)^{-1}\zeta\right]\nu\right\|^2/\left(K-2d_x\right)} \sim F_{p,K-2d_x}.$$

But $F_{p,K-2d_x}$ does not depend on the conditioning variable ζ, thus, it is also the unconditional distribution. We have, therefore, shown that $F(\hat{\beta}_{TAOLS}) \Rightarrow \frac{K}{K-2d_x}\cdot F_{p,K-2d_x}$. We collect this and the result on the t statistic in the theorem below.

Theorem 3 *Let Assumptions 3 and 4 hold. Assume that $K > 2d_x$. Under the fixed-K asymptotics, we have*

$$F^*(\hat{\beta}_{TAOLS}) := \frac{K - 2d_x}{K} F(\hat{\beta}_{TAOLS}) \Rightarrow F_{p,K-2d_x} \text{ and}$$
$$t^*(\hat{\beta}_{TAOLS}) := \sqrt{\frac{K - 2d_x}{K}} t(\hat{\beta}_{TAOLS}) \Rightarrow t_{K-2d_x} \text{ for } p = 1,$$

where $F_{p,K-2d_x}$ is the standard F distribution with degrees of freedom p and $K - 2d_x$, and t_{K-2d_x} is the standard t distribution with degrees of freedom $K - 2d_x$.

Remark 9 *If we pretend that all variables in the regression (20) are distributed exactly as their respective asymptotic normal distributions, then we obtain a classical normal linear regression model (CNLRM). The asymptotic F and t theory in Theorem 3 is the same as the exact F and t theory in a CNLRM. The test statistic $F^*(\hat{\beta}_{TAOLS})$ can be equivalently computed by the classical formula that compares the sums of squared residuals for restricted and unrestricted regressions.*

Remark 10 *To establish the asymptotic F and t theory, we employ a conditioning argument by conditioning on ζ. The exact form of the distribution of ζ is not essential. The asymptotic F and t theory holds regardless of the distribution of ζ. While the distribution of ζ in Hwang and Sun (2018) is different from what we have here, the asymptotic F and t distributions in Theorem 3 are the same as those in Theorem 3 of Hwang and Sun (2018). There is an opportunity to extend the asymptotic F and t theory further to allow for other distributions of ζ; see, for example, Pellatt and Sun (2020).*

3.2 Predictive regression

As a variant of the model in the previous subsection, we consider the predictive regression:

$$\begin{aligned} Y_t &= \alpha_0 + X_{t-1}'\beta_0 + u_{0t}, \\ X_t &= \left(1 - \frac{c}{T}\right) X_{t-1} + u_{xt}. \end{aligned} \tag{26}$$

To describe the information filtration, we let $(u_{0t}', \varepsilon_{xt}')'$ be a martingale difference sequence. We assume that $u_{xt} = g(\varepsilon_{xt}, \varepsilon_{x,t-1}, \ldots)$ for some measurable function g.

There is a large econometric and finance literature on this type of regression; see Phillips (2015) for a recent review. Here we do not restrict $\{u_{xt}\}$ to be a martingale difference sequence, but we still maintain Assumption 3. Our assumption is less restrictive than most of the existing literature where the martingale difference assumption is maintained. However, see Hjalmarsson (2007), which allows u_{xt} to be a linear process driven by $\{\varepsilon_{xt}\}$.

Define $\mathbb{W}_i^{\alpha}, \mathbb{W}_i^{y}, \mathbb{W}_i^{\Delta_c x}, \mathbb{W}_i^{0\cdot x}$ in the same way as before, but define $\mathbb{W}_{i,-1}^{x}$ as

$$\mathbb{W}_{i,-1}^{x} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_{t-1} \phi_i(\frac{t}{T}).$$

The transformed and augmented regression model becomes

$$\mathbb{W}_i^{y} = \mathbb{W}_{i,-1}^{x\prime} \beta_0 + \mathbb{W}_i^{\Delta_c x\prime} \delta_0 + \mathbb{W}_{\alpha,i}^{0\cdot x} \text{ for } i = 1, \ldots, K, \tag{27}$$

which is the same as (19) but with $\mathbb{W}_{i,-1}^{x}$ replacing $\mathbb{W}_i^{x}$. Noting that

$$\frac{\mathbb{W}_{i,-1}^{x}}{T} = \frac{1}{T^{3/2}} \sum_{s=1}^{T} \phi_i\left(\frac{s}{T}\right) X_{s-1} \Rightarrow \Omega_{xx}^{1/2} \int_0^1 \phi_i(r) J_{c,x}(r) dr = \xi_i,$$

the asymptotic distribution of $\mathbb{W}_{i,-1}^{x}/T$ is the same as that of $\mathbb{W}_i^{x}/T$. Therefore, all variables in (27) have the same asymptotic distributions as those in (19). It then follows that the asymptotic F and t theory in Theorem 3 continues to hold for the predictive regression.

To conclude this section, we outline the steps in conducting the asymptotic F and t tests[3]:

(i) Choose a sequence of orthonormal basis functions on $L^2[0,1]$. For example, we can use $\left\{\sqrt{2}\cos 2\pi r i, \sqrt{2}\sin 2\pi r i\right\}$ for $i = 1, 2, \ldots, K/2$.

(ii) Project the time series data Y_t, X_t, and $\Delta_c X_t$ onto the space spanned by the basis vectors to obtain $\mathbb{W}_i^{y}, \mathbb{W}_i^{x}$, and $\mathbb{W}_i^{\Delta_c x}$.

(iii) Estimate

$$\mathbb{W}_i^{y} = \mathbb{W}_i^{x\prime} \beta_0 + \mathbb{W}_i^{\Delta_c x\prime} \delta_0 + \mathbb{W}_{\alpha,i}^{0\cdot x} \text{ for } i = 1, \ldots, K. \tag{28}$$

by OLS with $\mathbb{W}_{\alpha,i}^{0\cdot x}$ as the regression error.

(iv) Conduct inferences in the usual way, treating the above as the CNLRM. More specifically, use the F and t distributions to construct confidence intervals and perform hypothesis testing.

4 A Simulation Study

In this section, we compare the finite sample performances of the proposed F test with those of the conventional chi-squared tests. There has already been some simulation evidence that the F test (or the t test) has more accurate size than the chi-squared tests (or the normal tests) in some

[3]We consider the cointegration regression here. For the predictive regression, we only need to change X_t into X_{t-1}.

nonstationary regressions, such as the linear trend regression in Sun (2011) and the cointegration regression with endogeneity in Hwang and Sun (2018). We have covered a few other nonstationary regressions in this paper. In light of the space constraint and given that this paper is a tribute to Joon Park's seminal contributions to econometrics, we consider only the cointegration regression of Phillips and Park (1988) where the regressors are strictly exogeneous. More specifically, we consider the following data generating process with three regressors $X_t = (1, \tilde{X}_t')' \in \mathbb{R}^3$ and $\tilde{X}_t = \left(\tilde{X}_{t1}, \tilde{X}_{t2}\right)' \in \mathbb{R}^2$:

$$\begin{aligned} Y_t &= X_t'\beta_0 + u_t, \\ \tilde{X}_t &= \tilde{X}_{t-1} + u_{x,t}, \\ u_t &= \rho u_{t-1} + \epsilon_{yt}, \\ u_{x,t} &= \rho u_{x,t-1} + \epsilon_{xt}, \end{aligned}$$

where $(\epsilon_{yt}, \epsilon_{xt}')'$ is i.i.d. $N(0, I_3)$.[4] Since $\{\epsilon_{yt}\}_{t=1}^T$ and $\{\epsilon_{xt}\}_{t=1}^T$ are independent, $\tilde{X}_t$ is strictly exogeneous. The parameter ρ controls the persistence of u_t and each component of u_{xt}. We consider the following values of ρ :

$$\rho \in \{0.05, 0.20, 0.35, 0.50, 0.75, 0.90\}.$$

Without loss of generality, we set the true coefficient vector to be $\beta_0 = (\beta_{10}, \beta_{20}, \beta_{30})' = (1, 1, 1)'$.

We test $H_0 : R\beta_0 = r$ against $H_1 : R\beta_0 \neq r$ where

$$R = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } r = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

so that $p = 2$. That is, we perform a test on whether the coefficients β_{20} and β_{30} on the nonconstant regressors are jointly one. We consider two sample sizes $T = 100$ and 200. The number of simulation replications is 10000.

In Phillips and Park (1988), $\{u_t\}$ is known to follow an AR process with a known AR order, and hence the long run variance of $\{u_t\}$ can be estimated by running an AR regression based on the OLS residuals. In contrast, here we assume that we do not know the true data generating process for $\{u_t\}$, and we estimate the long run variance of $\{u_t\}$ nonparametrically.

Depending on the long run variance estimator and the critical value used, we consider three groups of 5% tests. The first group consists of two tests that are based on the series long run variance estimator. Using the Fourier basis functions given in (10), we compute the test statistic

$$F_{T,Fourier} = \left(R\hat{\beta} - r\right)' \left[R\left(X'X\right)^{-1} R\right]^{-1} \left(R\hat{\beta} - r\right) / \left(p\hat{\omega}_u^2\right),$$

[4]Other distributions have been considered, but the simulation results are qualitatively close to what we report here.

where $\hat{\omega}_u^2 = (KT)^{-1} \left(\boldsymbol{\phi}'\hat{u}\right)' \left(\boldsymbol{\phi}'\hat{u}\right)$. See equations (2) and (3). Both tests in this group are based on the test statistic $F_{T,Fourier}$. The first test uses the simulated critical value from the nonstandard distribution F_∞ given in (5) and is referred to as "Fourier-F_∞." The second test uses the 5% critical value from $\chi_2^2/2$, a (normalized) chi-squared distribution[5] and is referred to as "Fourier-χ_2^2".

The second group of tests is similar to the first group but is based on the transformed Fourier series. The first test in this group is the F test detailed at the end of Section 2. The second test in this group uses the same test statistic but employs the 5% critical value from $\chi_2^2/2$. We refer to these two tests as "Transformed-Fourier-$F_{2,K}$" and "Transformed-Fourier-χ_2^2", respectively.

The third group consists of three chi-squared tests that use kernel estimators of the long run variance ω_u^2 of $\{u_t\}$. We include the Bartlett, Parzen, or Quadratic Spectral (QS) kernels in our simulation study, but here we focus on the case with the QS kernel, as the results for the other two cases are qualitatively similar. Given the QS kernel $k_{QS}(\cdot)$, we first construct

$$\tilde{\omega}_u^2 = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{u}_t \hat{u}_s k_{QS}\left(\frac{t-s}{bT}\right),$$

where b is a smoothing parameter, and then compare the test statistic

$$F_{T,QS} = \left(R\hat{\beta} - r\right)' \left[R\left(X'X\right)^{-1} R\right]^{-1} \left(R\hat{\beta} - r\right) / \left(p\tilde{\omega}_u^2\right)$$

with the critical value from $\chi_2^2/2$. We refer to the resulting test as "QS-χ_2^2".

For the series-based tests, we need to choose K, and for the kernel-based tests, we need to choose b. We consider both pre-specified values and data-driven choices. In the former case, we set $K = 8$ and 16. Hwang and Sun (2018) consider these two values of K in their simulation study and provide some justifications. Nevertheless, these two values should be regarded as rule-of-thumb choices. For each $K = 8$ or 16, we obtain the following comparable value of b :

$$b = \left(\int_{-\infty}^{\infty} k_{QS}^2(x)\, dx \cdot K\right)^{-1} = 1/K.$$

Under the above relationship between K and b, the Fourier-series and QS-kernel LRV estimators have the same asymptotic variance under the conventional asymptotics.

For the data-driven choices of K and b, we use the MSE-based rules developed by Phillips (2005) and Andrews (1991), respectively. We employ the AR(1) plug-in implementation. After fitting an AR(1) model to the residual process $\{\hat{u}_t\}$ by OLS, we compute

$$\hat{K} = 0.7134\, [\hat{\alpha}(2)]^{-1/5}\, T^{4/5} \text{ and } \hat{b}_{QS} = 1.3221\, [\hat{\alpha}(2)\, T]^{1/5} / T,$$

[5]More precisely, $\chi_2^2/2$ stands for the distribution of $\mathcal{Z}^2/2$ where $\mathcal{Z}^2 \sim \chi_2^2$.

where $\hat{\alpha}(2) = 4\hat{\rho}^2 (1-\hat{\rho})^{-4}$ and $\hat{\rho}$ is the estimated AR coefficient. These data-driven choices of K and b can be justified in the asymptotic framework of Andrews (1991), but they are not necessarily most suitable for testing problems. It is beyond the scope of this paper to derive a testing-optimal choice of K or b along the lines of Sun et al. (2008).

Figure 1 reports the empirical type I errors of the five tests when $K = 8$ and $b = 1/8$ for $T = 100$ and $T = 200$. The results for the case with $K = 16$ are qualitatively similar and are omitted here. The figure shows that the empirical type I errors of the chi-squared tests are substantially larger than the nominal significance level. Ignoring the estimation error in the long run variance estimator, each chi-squared test over-rejects the null hypothesis. In contrast, the proposed F test and the nonstandard F_∞ test have very accurate size. The empirical null rejection probability for both tests is very close to the nominal significance level, except when ρ is large (i.e., when ρ is larger than 0.75). We note that the F test and the nonstandard F_∞ test achieve similar size accuracy. Given that there is no need to simulate critical values, we recommend the more convenient F test.

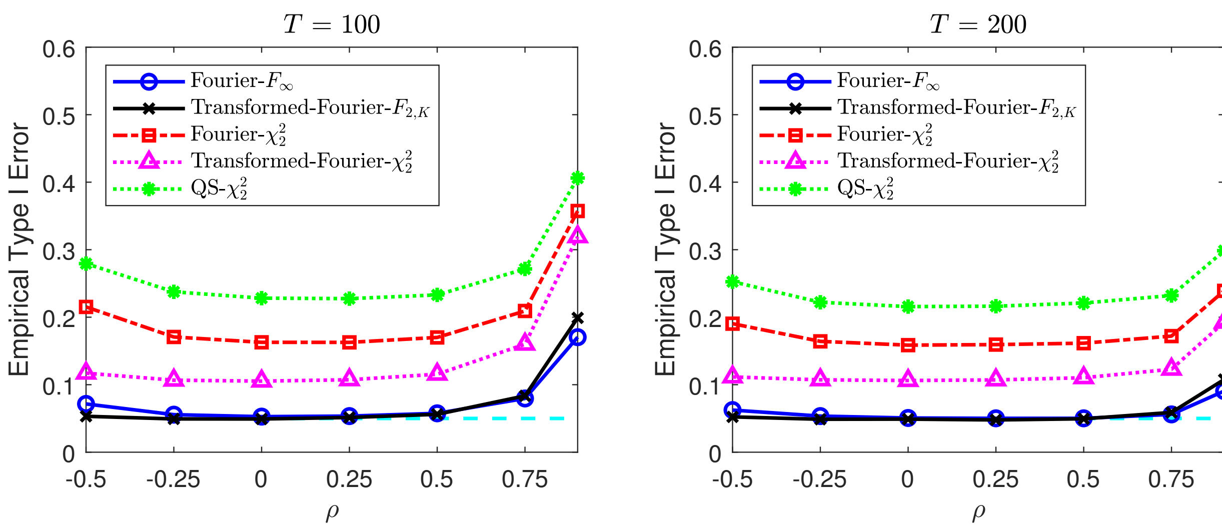

Figure 1: Empirical Type I error of different 5% tests with smoothing parameters $K = 8$ and $b = 1/K$.

In Figure 1, the size distortion of the chi-squared tests does not decrease significantly with the sample size. The reason is that the smoothing parameter K is fixed at $K = 8$ and the smoothing parameter b is fixed at 1/8. Hence, the estimation uncertainty in the LRV estimator remains more or less the same across the two sample sizes $T = 100$ and $T = 200$. As a result, the size performance of the chi-squared tests does not improve much as the sample size increases.

Figure 2 reports the empirical type I errors when K and b are data-driven. The F test and the nonstandard F_∞ test still have more accurate size than any of the chi-squared tests, and they achieve more or less the same size accuracy. However, the F test is more convenient to use and hence is recommended.

Another observation from Figure 2 is that the size distortion of all tests decreases as the sample size increases from 100 to 200. This is a feature of data-driven choices of the smoothing parameters. Unreported results show that as the sample size increases, the average value of the data-driven K's increases and the average value of the data-driven b's decreases. Hence, a larger sample size leads to LRV estimators with smaller variability. As a result, all tests become less size-distorted as the sample size increases.

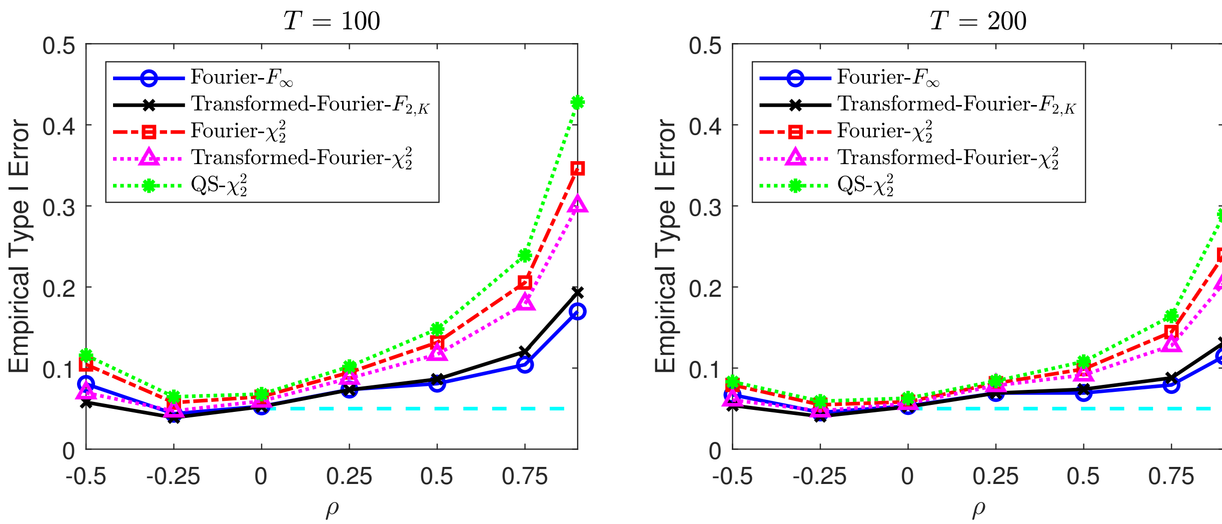

Figure 2: Empirical Type I error of different 5% tests with data-driven smoothing parameters

We have simulated the size-adjusted power curves. The two tests in the first group "Fourier-F_∞" and "Fourier-χ_2^2" have the same size-adjusted power, as they are based on the same test statistic. Similarly, the two tests in the second group "Transformed-Fourier-$F_{2,K}$" and "Transformed-Fourier-χ_2^2" have the same size-adjusted power. It suffices to consider only three tests for the size-adjusted power comparison, and we denote the three tests by "Fourier", "Transformed-Fourier", and "QS" in our power figure.

We consider the same data generating processes as before but now $\beta_1 = \beta_{10}$ and

$$(\beta_2, \beta_3)' = (\beta_{20}, \beta_{30})' + \theta/T$$

for some θ. For each simulation replication, we draw a different value of θ uniformly over a circle. We plot the size-adjusted power as a function of the radius $\|\theta\|$ in Figure 3 when $\rho = 0.5$ and

$K = 8$. The figure shows that "Transformed Fourier" is slightly more powerful than "Fourier" and "QS". For other parameter configurations (i.e., different values of ρ) and smoothing parameter choices, "Transformed Fourier" is as powerful as "Fourier" and "QS".

To sum up, we have found that for the cointegration regression considered here, the proposed F test has more accurate size than the commonly used chi-squared tests, and it is as powerful as and sometimes more powerful than the chi-squared tests.

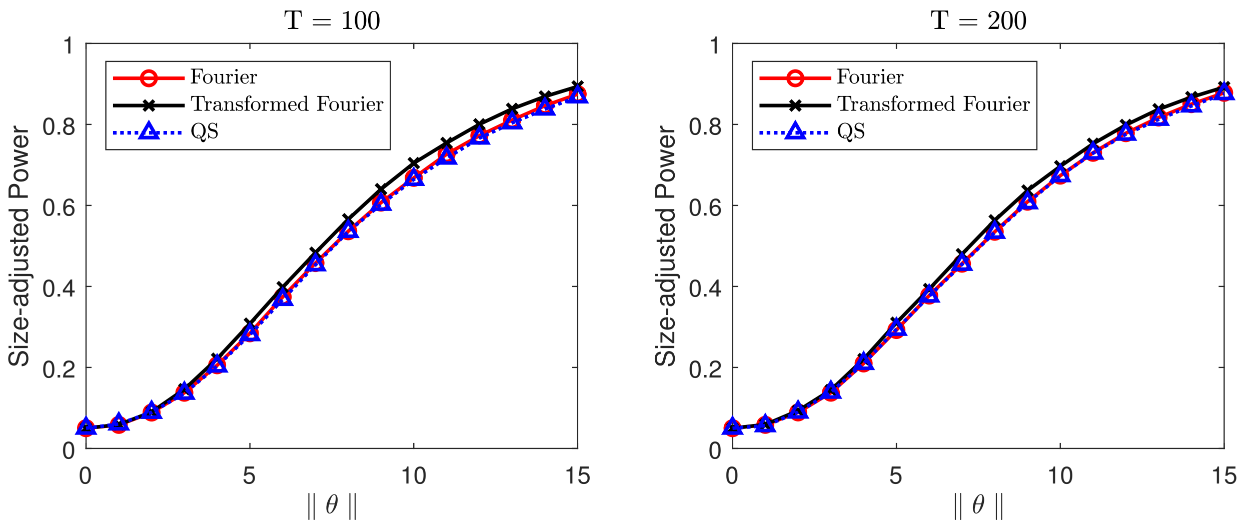

Figure 3: Size-adjusted power of different 5% tests with smoothing parameters $K = 8$ and $b = 1/K$.

5 Conclusion

This paper has developed the asymptotic F and t theory for regressions with nonstationary regressors of general forms. Linear and nonlinear trend functions, unit-root processes, near-unit-root processes with an additional complication from a structural break are all accommodated. While our asymptotic theory covers some existing results, it also covers many new cases that the asymptotic F and t theory is currently lacking in the literature. Depending on whether the regressors are endogenous or not, we develop the asymptotic F and t theory in different domains: the time domain or the pseudo-frequency domain. In both cases, statistical inferences are very easy to implement. In particular, in the latter case, we only need to transform our data using real matrix multiplications and then conduct the F and t tests as if we have a classical normal linear regression model. There is no need to use complex exponentials or explicitly estimate the long run variance and half long run variance.

As discussed before, it will be interesting to extend the theory further to allow the regressors to be fractionally integrated or follow a slow-varying trend. It will also be interesting to extend the theory to cover regressions with both nonstationary regressors, exogenous or not, and stationary regressors, such as the regressions considered by Park and Phillips (1989). The idea of using data-dependent and random basis functions in series variance estimation may be extended to functional data analysis.

References

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

Bunzel, H. (2006). Fixed-b asymptotics in single-equation cointegration models with endogenous regressors. *Econometric Theory*, 22(4):743–755.

Chang, Y., Lu, Y., and Park, J. Y. (2018). Understanding regressions with observations collected at high frequency over long span. Working Paper 2018-10, University of Sydney, School of Economics.

Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley publication in mathematical statistics. Wiley.

Guo, G., Sun, Y., and Wang, S. (2018). Testing for moderate explosiveness. *The Econometrics Journal*, 22(1):73–95.

Hjalmarsson, E. (2007). Fully modified estimation with nearly integrated regressors. *Finance Research Letters*, 4(2):92–94.

Hwang, J. and Sun, Y. (2017). Asymptotic F and t tests in an efficient GMM setting. *Journal of Econometrics*, 198:277–295.

Hwang, J. and Sun, Y. (2018). Simple, robust, and accurate F and t tests in cointegrated systems. *Econometric Theory*, 34(5):949–984.

Jin, S., Phillips, P. C. B., and Sun, Y. (2006). A new approach to robust inference in cointegration. *Economics Letters*, 91(2):300–306.

Kiefer, N. M. and Vogelsang, T. J. (2002a). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory*, 18:1350–1366.

Kiefer, N. M. and Vogelsang, T. J. (2002b). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica*, 70:2093–2095.

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

Lazarus, E., Lewis, D. J., and Stock, J. H. (2021). The size-power tradeoff in har inference. *Econometrica*, 89(5):2497–2516.

Liu, C. and Sun, Y. (2019). A simple and trustworthy asymptotic t test in difference-in-differences regressions. *Journal of Econometrics*, 210(2):327–362.

Martínez-Iriarte, J., Sun, Y., and Wang, X. (2020). Asymptotic F tests under possibly weak identification. *Journal of Econometrics*, 218(1):140–177.

Müller, U. K. (2007). A theory of robust long-run variance estimation. *Journal of Econometrics*, 141:1331–1352.

Park, J. Y. and Phillips, P. C. B. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, 4(3):468–497.

Park, J. Y. and Phillips, P. C. B. (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory*, 5(1):95–131.

Pellatt, D. and Sun, Y. (2020). Asymptotic F test in regressions with observations collected at high frequency over long span. Working Paper, Department of Economics, UC San Diego.

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

Phillips, P. C. B. (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory*, 23(4):557–614.

Phillips, P. C. B. (2014). Optimal estimation of cointegrated systems with irrelevant instruments. *Journal of Econometrics*, 178:210–224. Recent Advances in Time Series Econometrics.

Phillips, P. C. B. (2015). Halbert White Jr. Memorial JFEC Lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics*, 13:521–555.

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

Phillips, P. C. B. and Park, J. Y. (1988). Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors. *Journal of the American Statistical Association*, 83(401):111–115.

Sun, Y. (2006). Best quadratic unbiased estimators of integrated variances in the presence of market microstructure noise. Working Paper, Department of Economics, UC San Diego.

Sun, Y. (2011). Robust trend inference with series variance estimator and testing-optimal smoothing parameter. *Journal of Econometrics*, 164:345–366.

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

Sun, Y. (2014a). Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. *Journal of Econometrics*, 178:659–677.

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

Sun, Y. (2014c). Fixed-smoothing asymptotics and asymptotic F and t tests in the presence of strong autocorrelation. *Advances in Econometrics*, 33:23–63.

Sun, Y. (2018). Comment on "HAR inference: Recommendations for practice" by Lazarus, Lewis, Stock, and Watson. *Journal of Business & Economic Statistics*, 36(4):565–568.

Sun, Y. and Kim, M. S. (2015). Asymptotic F test in a GMM framework with cross sectional dependence. *Review of Economics and Statistics*, 97:210–223.

Sun, Y., Philips, P. C. B., and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica*, 76:175–194.

Sun, Y. and Wang, X. (2021). An asymptotically F-distributed Chow test in the presence of heteroscedasticity and autocorrelation. *Econometric Reviews*, 0(0):1–35.

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