Title

Asymptotic F test in Regressions with Observations Collected at High Frequency over Long Span

Authors
Pellatt, Daniel
Sun, Yixiao

Publication Date
2020-10-29

Asymptotic F test in Regressions with Observations Collected at High Frequency over Long Span

Daniel Pellatt and Yixiao Sun*
Department of Economics
UC San Diego

May 2020

Abstract

This paper proposes tests of linear hypotheses when the variables may be continuous-time processes with observations collected at a high sampling frequency over a long span. Utilizing series long run variance (LRV) estimation in place of the traditional kernel LRV estimation, we develop easy-to-implement and more accurate F tests in both stationary and nonstationary environments. The nonstationary environment accommodates endogenous regressors that are general semimartinglales. The F tests can be implemented in exactly the same way as in the usual discrete-time setting. The F tests are, therefore, robust to the continuous-time or discrete-time nature of the data. Simulations demonstrate the improved size accuracy and competitive power of the F tests relative to existing continuous-time testing procedures and their improved versions. The F tests are of practical interest as recent work by Chang et al. (2018) demonstrates that traditional inference methods can become invalid and produce spurious results when continuous-time processes are observed on finer grids over a long span.

JEL Classification: C12, C13, C22

Keywords: continuous time model, F distribution, high frequency regression, long run variance estimation

1 Introduction

The advent of high-frequency data poses challenges for classical inference and modeling procedures. For linear regression analysis with observations collected over time, as the grid of observed times becomes finer, continuous-time properties of the underlying processes may conflict with traditional assumptions framed in a discrete-time setting. An immediate concern is the validity of inference procedures when the data generating processes may be continuous-time in nature. Another concern is automating procedures so that inference depends on fewer technical and theoretical modeling decisions. At what sampling frequency should a researcher consider moving to an explicitly continuous-time framework? If continuous-time modeling requires accounting for the sampling frequency, what measurement constitutes a single unit of time? A month or a year? Designing trustworthy inference procedures in realistic sample sizes is also a concern.

*Email: dpellatt@ucsd.edu; yisun@ucsd.edu. Address correspondence to Department of Economics, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA

In this paper we propose tests of linear hypotheses that aim to address the above concerns. Recently Chang et al. (2018) considers statistical inference in this setting, highlighting how traditional hypothesis tests can become spurious when observations are collected at high frequency over a long time span. They show that it is essential to use an autocorrelation-robust variance or long run variance to construct test statistics and make valid inferences. They utilize the continuous-time kernel LRV estimator developed in Lu and Park (2019). Adopting the traditional asymptotic specification that ensures the consistency of the kernel LRV estimator, they show that the test statistics are asymptotically chi-squared. One takeaway from Chang et al. (2018) is that not all kernel-based LRV estimation procedures can be applied without explicitly accounting for the continuous-time environment. A "high-frequency-compatible" bandwidth is desired. It is interesting that the parametric plug-in bandwidth choice of Andrews (1991) is high-frequency-compatible while the nonparametric analogue of Newey and West (1994) is not.

In this paper we build on Chang et al. (2018) and propose convenient and trustworthy tests in regressions with high-frequency data collected over a long span. We consider both common regressions with stationary regressors and cointegrating regressions with nonstationary regressors. Due to self-normalization, our tests yield valid inference in the continuous-time setting and would also be valid if the observations were generated from a discrete-time process satisfying standard linear regression assumptions. A practitioner does not have to make any difficult decisions — they can simply use all the observed data, and they can compute the test statistic and perform hypothesis testing in exactly the same way in both the discrete-time and continuous-time settings.

We make several contributions along different dimensions. First, we adopt the more recent fixed-smoothing asymptotic framework. In the discrete-time setting, it is well known that randomness in LRV estimators can lead to significant size distortion of the associated chi-squared tests in finite samples. The same problem is present in the continuous-time setting. By employing the fixed-smoothing asymptotic framework as in Sun (2011, 2013), we show that our test statistics are asymptotically F distributed in both stationary and nonstationary settings. The F approximations capture the randomness of the LRV estimators and are more accurate than the chi-squared approximations.

Second, the asymptotic F theory is based on the series LRV estimator, and we characterize its asymptotic bias and variance in the high-frequency setting. The series LRV estimator involves projecting the discretized data onto a sequence of orthonormal basis functions and then taking an average of the outer-products of the projection coefficients. The number of orthonormal basis functions, denoted by K, is the smoothing parameter in this type of nonparametric variance estimator. Based on the asymptotic bias and variance, we develop a data-driven and automated choice of K in the high-frequency setting. Our rule of selecting of K extends that of Phillips (2005), which considers the series LRV estimator in the low-frequency discrete-time setting[1]. Furthermore, we allow for a general class of orthonormal basis functions while Phillips (2005) focuses on sine and cosine functions. See Lazarus et al. (2018) for some practical guidance on using the series LRV estimator with low-frequency discrete-time data.

Third, in a discrete-time cointegrating model, it is common to accommodate endogenous regressors. Following this practice, we allow the regressors to be endogenous in the continuous-time nonstationary setting. This constitutes another departure from Chang et al. (2018) which considers only the case with exogenous regressors. To deal with the endogeneity, we follow Hwang

[1]Typical examples of low-frequency discrete-time data inlcude daily, weekly, monthly, and yearly data. The frequency here refers to the sampling frequency, namely the number of times we can draw observations per unit of time. It does not refer to the speed that a process completes a cycle.

and Sun (2018), but we have to introduce some modifications to facilitate the asymptotic analysis. However, the continuous-time test statistic is computationally identical to the discrete-time statistic in Hwang and Sun (2018), and they are shown to have the same limiting F distribution.

Fourth, in the nonstationary setting, we establish the asymptotic F distribution for a wider class of regressor processes. The scaled regressor process may converge to a general stochastic process that includes the Brownian motion as a special case. To a great extent, our asymptotic F theory goes beyond its counterpart in the low-frequency discrete-time setting where the nonstationary process is a unit root process and thus converges to a Brownian motion after appropriate normalization.

Finally, we show that in both stationary and nonstationary settings, our F test remains asymptotically valid when the regression error contains additional low-frequency discrete-time measurement noise. In both settings, the discrete measurement noise is dominated by the continuous-time error component and hence does not affect our asymptotic theory.

The class of series LRV estimators is closely related to the class of kernel LRV estimators; see, for example, the discussion in Sun (2011). In essence, a series LRV estimator can be regarded as a kernel LRV estimator with a generalized kernel function. The fixed-K approach adopted here is analogous to the "fixed-b" approach employed in Kiefer and Vogelsang (2005). Fixed-b asymptotics can be developed for the kernel-based test statistics in Chang et al. (2018). However, the limiting distributions are nonstandard and hard to use. They can also be nonpivotal in the nonstationary setting (see Vogelsang and Wagner (2014) for the possible nonpivotality). This provides a further justification on the use of series LRV estimation in designing convenient and accurate inference procedures in finite samples.

The outline of the paper is as follows. In Section 2 we introduce the basic setup of the regression problem at hand. In Section 3 we consider the case where the regressors are stationary and consider a data-driven approach to selecting K. In Section 4 we consider the nonstationary case with cointegration. Section 5 evaluates the finite sample performances of the proposed F tests, Section 6 considers the impact of an additive error component of discrete nature, and Section 7 concludes. Proofs are given in the appendix.

2 Basic setting and Assumptions

Consider a continuous-time regression of the form

$$Y_t = X_t'\beta_0 + U_t,$$

where each of $Y_t \in \mathbb{R}, X_t \in \mathbb{R}^{d\times 1}$ and $U_t \in \mathbb{R}$ is a continuous-time process for $t \in [0, T]$ with sample paths that are right continuous with left limits (cadlag). We will assume that U_t is stationary and $E(U_t|X_s, s \in [0, T]) = 0$.

We do not observe the processes continuously. Instead, for some small sampling interval δ, we observe $\{x_i, y_i\}_{i=1}^n$ where[2]

$$x_i = X_{i\delta}; y_i = Y_{i\delta}$$

for $i = 1, ..., n$ and $n = T/\delta$. $\{(x_i, y_i)\}$ satisfies

$$y_i = x_i'\beta_0 + u_i,$$

[2]For notational simplicity, we assume that T/δ is an integer.

where $u_i = U_{i\delta}$ is unobserved. We are interested in testing $H_0 : R\beta_0 = r$ versus $H_1 : R\beta_0 \neq r$ for some $p \times d$ matrix R with a full row rank.

Given the discrete data $\{x_i, y_i\}_{i=1}^n$, we estimate β_0 by

$$\hat{\beta}_D = \left(\sum_{i=1}^n x_i x_i'\right)^{-1} \sum_{i=1}^n x_i y_i.$$

Define

$$\hat{\beta}_C = \left[\int_0^T X_t X_t' dt\right]^{-1} \left[\int_0^T X_t Y_t dt\right],$$

which is the least-square analogue of $\hat{\beta}_D$ in the space $L^2[0,T]$ using the continuous data $\{(X_t, Y_t), t \in [0,T]\}$. $\hat{\beta}_C$ is not feasible, and we use it only as a benchmark for comparison.

3 The Stationary Case

In this section, we consider the case that X_t is a stationary process and defer the case with a nonstationary X_t to Section 4. An intercept can be included in X_t.

3.1 The test statistic

To make inferences on $R\beta_0$, we often first find the rate of convergence of $\hat{\beta}_D - \beta_0$, establish the asymptotic distribution of a rescaled version of $\hat{\beta}_D - \beta_0$ and then construct the test statistic based on an estimated asymptotic variance. Instead of following these conventional steps, we use heuristic arguments and construct the test statistic directly. The *approximate* variance of $\hat{\beta}_D - \beta_0$ is

$$\left(\sum_{i=1}^n x_i x_i'\right)^{-1} \text{var}\left(\sum_{i=1}^n x_i u_i\right)\left(\sum_{i=1}^n x_i x_i'\right)^{-1}.$$

To test the null of $R\beta_0 = r$, we construct the test statistic

$$F_T = \left(R\hat{\beta}_D - r\right)' \left[R\left(\sum_{i=1}^n x_i x_i'\right)^{-1} \widehat{\text{var}}\left(\sum_{i=1}^n x_i \hat{u}_i\right)\left(\sum_{i=1}^n x_i x_i'\right)^{-1} R'\right]^{-1} \left(R\hat{\beta}_D - r\right)/p,$$

where $\hat{u}_i = y_i - x_i'\hat{\beta}_D$ and $\widehat{\text{var}}\left(\sum_{i=1}^n x_i \hat{u}_i\right)$ is an estimator of the *approximate* variance of $\sum_{i=1}^n x_i u_i$.

We use the series estimator for the approximate variance. Let $\{\phi_j(\cdot)\}$ be some basis functions on $L^2[0,1]$. The series variance estimator is given by

$$\widehat{\text{var}}\left(\sum_{i=1}^n x_i \hat{u}_i\right) = \frac{1}{K}\sum_{j=1}^K \left[\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right)(x_i \hat{u}_i)\right]^{\otimes 2},$$

where $a^{\otimes 2} = aa'$ for any vector a and K is a tuning parameter. Note that the basis functions are evaluated at i/n instead of i/T. We have, therefore, effectively ignored the high-frequency nature of the time series observations. The test statistic is then

$$F_T = \left(R\hat{\beta}_D - r\right)' \left[R\left(\sum_{i=1}^n x_i x_i'\right)^{-1} \frac{1}{K}\sum_{j=1}^K \left[\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right)(x_i \hat{u}_i)\right]^{\otimes 2} \left(\sum_{i=1}^n x_i x_i'\right)^{-1} R'\right]^{-1} \left(R\hat{\beta}_D - r\right)/p.$$

The form of the test statistic is exactly the same as what we would use for a standard time series regression. To construct the test statistic, we can ignore the fact that our observations come from sampling continuous-time processes.

The test statistic F_T takes a self-normalized form. This will become more transparent if we consider the special case that $d = p = 1$ and $K = 1$. In this case, we take $R = 1$ without loss of generality, and the test statistic becomes

$$F_T = \left(\frac{\sum_{i=1}^{n} (x_i u_i)}{\sum_{i=1}^{n} \phi\left(\frac{i}{n}\right)(x_i \hat{u}_i)} \right)^2 := (t_T)^2 .$$

The numerator in the t statistic t_T is a simple sum of $x_i u_i$ while the denominator is a weighted sum of $x_i \hat{u}_i$ with non-diminishing and bounded weights. We expect the numerator and denominator to be of the same order of magnitude no matter what δ is. As a result, t_T and F_T will be stochastically bounded for any sampling interval δ. In this sense, the denominator normalizes the numerator, and thus no additional normalization is needed. This form of self-normalization leads to the invariance of our testing procedure to the sampling interval, which we will develop in greater detail.

We consider the asymptotics along the limiting sequence $\delta \to 0$ and $T \to \infty$. The asymptotics would best reflect the finite sample situation where the observations are collected at high frequency ($\delta \to 0$) over a long span ($T \to \infty$). To develop the more accurate fixed-smoothing asymptotic approximations, we hold K fixed as $\delta \to 0$ and $T \to \infty$. For the stationary case in this section, we maintain the assumptions below.

Assumption 3.1 *As $\delta \to 0$ and $T \to \infty$,*

$$\frac{1}{n} \sum_{i=1}^{[nr]} x_i x_i' = \frac{1}{T} \int_0^{Tr} X_t X_t' dt + o_p(1)$$

and

$$\frac{1}{T} \int_0^{Tr} X_t X_t' dt \xrightarrow{p} rQ$$

uniformly over $r \in [0, 1]$ for some positive definite matrix Q.

Assumption 3.2 *As $\delta \to 0$ and $T \to \infty$,*

$$\frac{\delta}{\sqrt{T}} \sum_{i=1}^{[nr]} x_i u_i = \frac{1}{\sqrt{T}} \int_0^{Tr} X_t U_t dt + o_p(1)$$

uniformly over $r \in [0, 1]$, and

$$\frac{1}{\sqrt{T}} \int_0^{Tr} X_t U_t dt \xrightarrow{d} \Omega^{1/2} W_d(r),$$

where $W_d(r)$ is the $d \times 1$ standard Brownian motion process,

$$\Omega = \lim_{T \to \infty} \operatorname{var}\left(\frac{1}{\sqrt{T}} \int_0^T X_t U_t dt \right) = \int_{-\infty}^{\infty} \Gamma_{XU}(\tau) d\tau,$$

$\Gamma_{XU}(\tau) = E\left[X_t U_t U_{t-\tau} X_{t-\tau}'\right]$*, and $\Omega^{1/2}$ is the matrix square root of Ω so that $\Omega^{1/2}\left(\Omega^{1/2}\right)' = \Omega$.*

Assumption 3.3 *For $j = 1, \ldots, K$, each function $\phi_j(\cdot)$ is piecewise monotonic, twice continuously differentiable, and $\int_0^1 \phi_j(t)dt = 0$. Also, $\{\phi_j(\cdot)\}_{j=1}^K$ form an orthonormal set in $L^2[0,1]$.*

In Assumption 3.2, we view $T^{-1/2}\int_0^{Tr} X_t U_t dt$ for $r \in [0,1]$ as a random element of $\mathbb{D}^d[0,1]$, the space of cadlag functions from $[0,1]$ to $\mathbb{R}^{d\times 1}$ endowed with the Skorokhod topology. The weak convergence in Assumption 3.2 is defined on this space. The assumption constrains the magnitude of cumulative growth in the jump components of the process $X_t U_t$ and stipulates that the continuous-time integral can be recovered from discrete observations as $\delta \to 0$ and $T \to \infty$.

Sufficient conditions for Assumptions 3.1 and 3.2 can be found in Chang et al. (2018) (Assumptions A and C1). Note that

$$\frac{\delta}{\sqrt{T}} \sum_{i=1}^{[nr]} x_i u_i = \frac{\delta}{\sqrt{n\delta}} \sum_{i=1}^{[nr]} x_i u_i = \frac{1}{\sqrt{n/\delta}} \sum_{i=1}^{[nr]} x_i u_i.$$

Because $\{x_i u_i\}$ are highly correlated for a small δ, in order to obtain a well-defined weak limit, we need to normalize the partial sum by $\sqrt{n/\delta}$, which is larger than the usual normalization factor $\sqrt{n}$ by an order of magnitude. Assumption 3.3 is standard in the literature on orthonormal series variance estimation. See, for example, Assumption 1(b) in Sun (2014a).

Let $\Lambda(n,\delta) = \sqrt{n/\delta}$, which is the effective scale of $\sum_{i=1}^n x_i u_i$. Then

$$\begin{aligned}\sqrt{T}\left[\hat{\beta}_D - \beta_0\right] &= \delta\Lambda(n,\delta)\left[\hat{\beta}_D - \beta_0\right] \\ &= \left(\frac{1}{\delta\Lambda(n,\delta)^2}\sum_{i=1}^n x_i x_i'\right)^{-1} \frac{1}{\Lambda(n,\delta)}\sum_{i=1}^n x_i u_i.\end{aligned}$$

Using Assumption 3.1, we have

$$\frac{1}{\delta\Lambda(n,\delta)^2}\sum_{i=1}^n x_i x_i' = \frac{1}{n}\sum_{i=1}^n X_{i\delta}X_{i\delta}' = \frac{1}{T}\int_0^T X_t X_t' dt + o_p(1).$$

Using Assumption 3.2, we have

$$\frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{[nr]} x_i u_i = \frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{[nr]} (X_{i\delta}U_{i\delta}) = \frac{1}{\sqrt{T}}\int_0^{Tr} X_t U_t dt + o_p(1).$$

Therefore,

$$\begin{aligned}\sqrt{T}(\hat{\beta}_D - \beta_0) &= \left[\frac{1}{T}\int_0^T X_t X_t' dt + o_p(1)\right]^{-1}\left[\frac{1}{\sqrt{T}}\int_0^T X_t U_t dt + o_p(1)\right] \\ &= \left[\frac{1}{T}\int_0^T X_t X_t' dt\right]^{-1}\left[\frac{1}{\sqrt{T}}\int_0^T X_t U_t dt\right] + o_p(1) \\ &= \sqrt{T}(\hat{\beta}_C - \beta_0) + o_p(1).\end{aligned}$$

Assumptions 3.1 and 3.2 ensure that $\sqrt{T}(\hat{\beta}_D - \beta_0)$ and $\sqrt{T}(\hat{\beta}_C - \beta_0)$ are asymptotically equivalent. Invoking these two assumptions again, we obtain the asymptotic distribution of $\sqrt{T}(\hat{\beta}_D - \beta)$ and another key result in the lemma below.

Lemma 3.1 *Let Assumptions 3.1 and 3.2 hold. Then*

$$\sqrt{T}(\hat{\beta}_D - \beta_0) \xrightarrow{d} Q^{-1}\Omega^{1/2} W_d\left(1\right).$$

In addition, let Assumption 3.3 hold. Then

$$\frac{1}{\Lambda\left(n,\delta\right)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_i\hat{u}_i \xrightarrow{d} \Omega^{1/2}\int_0^1 \phi_j\left(r\right)dW_d(r)$$

jointly for $j = 1, 2, ..., K$.

Lemma 3.1 shows that $\hat{\beta}_D$ converges to β_0 at the rate of $\sqrt{T}$. We do not obtain the rate of $\sqrt{n}$, which is the rate for the discrete-time data with n observations. For high-frequency data sampled from a continuous-time process, the effective sample size is the time span T rather than the number of observations n.

Using Lemma 3.1, we have, under the null hypothesis:

$$\begin{aligned}
F_T &= \delta\Lambda\left(n,\delta\right)\left(R\hat{\beta}_D - r\right)' \\
&\times \left[R\left(\frac{1}{\delta\Lambda\left(n,\delta\right)^2}\sum_{i=1}^{n}x_i x_i'\right)^{-1}\frac{1}{K}\sum_{j=1}^{K}\left[\frac{1}{\Lambda\left(n,\delta\right)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)\left(x_i\hat{u}_i\right)\right]^{\otimes 2}\left(\frac{1}{\delta\Lambda\left(n,\delta\right)^2}\sum_{i=1}^{n}x_i x_i'\right)^{-1}R'\right]^{-1} \\
&\times \delta\Lambda\left(n,\delta\right)\left(R\hat{\beta}_D - r\right)/p \\
&\xrightarrow{d} \left[RQ^{-1}\Omega^{1/2}W_d\left(1\right)\right]'\left\{RQ^{-1}\Omega^{1/2}\frac{1}{K}\sum_{j=1}^{K}\left[\int_0^1\phi_j\left(r\right)dW_d(r)\right]^{\otimes 2}\Omega^{1/2}Q^{-1}R'\right\}^{-1}\left[RQ^{-1}\Omega^{1/2}W_d\left(1\right)\right]/p.
\end{aligned}$$

Note that $RQ^{-1}\Omega^{1/2}W_d\left(r\right) \overset{d}{=} \left[RQ^{-1}\Omega Q^{-1}R'\right]^{1/2}W_p\left(r\right)$ for a $p \times 1$ standard Brownian motion process $W_p\left(\cdot\right)$ and that $RQ^{-1}\Omega Q^{-1}R'$ is of a full rank. We have

$$F_T \xrightarrow{d} \left[W_p\left(1\right)\right]'\left\{\frac{1}{K}\sum_{j=1}^{K}\left[\int_0^1\phi_j\left(r\right)dW_p(r)\right]^{\otimes 2}\right\}^{-1}W_p\left(1\right)/p.$$

Under Assumption 3.3, $\left[\int_0^1\phi_j\left(r\right)dW_p(r)\right]^{\otimes 2}$ is iid Wishart distributed. The above limiting distribution is equal to Hotelling's T^2 distribution. In view of the relationship between the T^2 and F distributions (e.g., Bilodeau and Brenner (2010)), we have the following theorem.

Theorem 3.1 *Let Assumptions 3.1 – 3.3 hold. Then, for a fixed* $K \geq p$,

$$F_T \xrightarrow{d} \frac{K}{K-p+1}F_{p,K-p+1},$$

where $F_{p,K-p+1}$ *is the* F *distribution with degrees of freedom* p *and* $K-p+1$.

If we use the OLS variance estimator that ignores the autocorrelation, we would construct the test statistic as follows

$$F_{T,OLS} = \left(R\hat{\beta}_D - r\right)' \times \left[R\hat{\sigma}_u^2\left(\sum_{i=1}^{n}x_i x_i'\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_D - r\right)/p$$

where $\hat{\sigma}_u^2 = n^{-1} \sum_{i=1}^n \hat{u}_i^2$ is an estimator of the variance σ_u^2 of U_t. Then

$$\delta F_{T,OLS} = \sqrt{T} \left(R\hat{\beta}_D - r \right)' \times \left[R\hat{\sigma}_u^2 \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} R' \right]^{-1} \sqrt{T} \left(R\hat{\beta}_D - r \right) /p$$
$$\to^d \left[RQ^{-1}\Omega^{1/2} W_d (1) \right]' \times \left[\sigma_u^2 RQ^{-1} R' \right]^{-1} \left[RQ^{-1}\Omega^{1/2} W_d (1) \right] /p.$$

So, as $\delta \to 0$, $F_{T,OLS} \to \infty$ with probability approaching one. Consequently, using $F_{T,OLS}$ for inference can lead to the spurious finding of a significant relationship that does not actually exist. See Chang et al. (2018) for more details.

To illustrate the key difference between the variance estimators underlying F_T and $F_{T,OLS}$, consider the special case with $K = d = p = 1$. Then the ratio of the autocorrelation robust variance estimator to the OLS variance estimator is

$$\frac{\left[\sum_{i=1}^n \phi_j \left(\frac{i}{n} \right) (x_i \hat{u}_i) \right]^2}{\hat{\sigma}_u^2 \sum_{i=1}^n x_i^2} = \frac{\Lambda (n, \delta)^2}{n} \frac{\left[\frac{1}{\Lambda(n,\delta)} \sum_{i=1}^n \phi_j \left(\frac{i}{n} \right) x_i \hat{u}_i \right]^2}{\hat{\sigma}_u^2 \frac{1}{n} \sum_{i=1}^n x_i^2}$$
$$= \frac{1}{\delta} \cdot \frac{\left[\frac{1}{\Lambda(n,\delta)} \sum_{i=1}^n \phi_j \left(\frac{i}{n} \right) x_i \hat{u}_i \right]^2}{\hat{\sigma}_u^2 \frac{1}{n} \sum_{i=1}^n x_i^2}.$$

Note that the second factor converges to a nondegenerate distribution. So the ratio will diverge at the rate of $1/\delta$. That is, by ignoring the high-autocorrelation of $x_i u_i$, especially when δ is small, the OLS variance estimator under-estimates the true variation of the OLS estimator by a factor of $1/\delta$. This explains why F_T is stochastically bounded while $F_{T,OLS}$ explodes as $\delta \to 0$ and $T \to \infty$.

3.2 Optimal choice of K

In this subsection, we establish the MSE-optimal choice of K. Our theoretical results are the high-frequency continuous-time counterparts of Phillips (2005), which develops the MSE-optimal choice of K in LRV estimation for a fully observed discrete-time process. We allow for more general basis functions while Phillips (2005) considers only sine and cosine basis functions. Thus, even for usual discrete-time processes, our theoretical development goes beyond Phillips (2005).

As it is well known in the fixed-smoothing literature, the fixed-K framework does not allow us to develop an optimal choice of K, as the (squared) asymptotic bias and variance are not of the same order of magnitude. Here we follow the large HAR literature and resort to the more traditional increasing K asymptotics under which $K \to \infty$ in order to find an optimal K to balance the (squared) asymptotic bias and variance.

To abstract away the technical issues that will not affect the practical implementation, we define the infeasible variance estimator:

$$\hat{\Omega}^* = \frac{1}{K} \sum_{j=1}^K \left[\frac{1}{\Lambda (n, \delta)} \sum_{i=1}^n \phi_j \left(\frac{i}{n} \right) (x_i u_i) \right]^{\otimes 2}.$$

$\hat{\Omega}^*$ is infeasible because u_i is not observed. We choose K to minimize the asymptotic MSE of $\hat{\Omega}^*$. We could alternatively follow Andrews (1991) to find the approximate and truncated MSE of the feasible estimator $\hat{\Omega}$ and use it to guide the choice of K. These two approaches will lead to the same formula for the MSE-optimal K. Here we opt for the simpler approach.

Assumption 3.4 *The following holds as $K \to \infty, \delta \to 0$ and $T \to \infty$:*

(i) $var\left[\text{vec}(\hat{\Omega}^*)\right] = var\left[\text{vec}\left(\Omega^{1/2}\frac{1}{K}\sum_{j=1}^{K}\left[\int_0^1 \phi_j(r)\,dW_d(r)\right]^{\otimes 2}\Omega^{1/2}\right)\right](1+o(1))$;

(ii) For $\Gamma_{XU}(\tau) = E\left(X_t U_t U_{t-\tau} X'_{t-\tau}\right)$,

$$\delta \sum_{k=-\infty}^{\infty} (k\delta)^2 \|\Gamma_{XU}(k\delta)\| \leq \infty \text{ and } \int_{-\infty}^{\infty} \tau^2 \|\Gamma_{XU}(\tau)\|\, d\tau < \infty;$$

(iii) $\delta \sum_{k=-n+1}^{n-1} \Gamma_{XU}(k\delta) - \int_{-T}^{T} \Gamma_{XU}(\tau)\, d\tau = O(\delta)$;

(iv) For some constant C, $\sup_{r\in[0,1]} |\phi_j(r)| \leq C$, $\sup_{r\in[0,1]} \left|\dot{\phi}_j(r)\right| \leq jC$, $\phi_j^2(1) + \phi_j^2(0) \leq C$, *and for some constant* $c_\phi \neq 0$,

$$\lim_{K\to\infty}\left[-\frac{1}{K^3}\sum_{j=1}^{K}\frac{1}{2}\int_0^1 \phi_j(r)\,\ddot{\phi}_j(r)\,dr\right] = c_\phi$$

where $\dot{\phi}_j$ and $\ddot{\phi}_j$ are the first and second order derivatives of ϕ_j.

Assumption 3.4(i) is a high-level assumption. We assume that the limit of the exact finite sample variance of $vec(\hat{\Omega}^*)$ is equal to the variance of its limiting distribution, namely the asymptotic variance. From a theoretical point of view, this is plausible if we have enough moment conditions. Alternatively, we simply use the asymptotic variance in place of the exact finite sample variance to obtain an approximate MSE. This is, in fact, a typical approach for smoothing parameter choice in a nonparametric setting when the exact finite sample variance is difficult, if not impossible, to obtain.

Assumption 3.4(ii) imposes some conditions on the integrability of the covariance function $\Gamma_{XU}(\tau)$. Note that

$$\begin{aligned}
&\left|\delta \sum_{k=-\infty}^{\infty} (k\delta)^2 \|\Gamma_{XU}(k\delta)\| - \int_{-\infty}^{\infty} \tau^2 \|\Gamma_{XU}(\tau)\|\, d\tau\right| \\
&\leq \sum_{k=-\infty}^{\infty}\left[\int_{k\delta}^{(k+1)\delta}\left|\left[(k\delta)^2\|\Gamma_{XU}(k\delta)\| - \tau^2\|\Gamma_{XU}(\tau)\|\right]\right| d\tau\right] \\
&\leq \sum_{k=-\infty}^{\infty}\int_{k\delta}^{(k+1)\delta}\max_{t\in[k\delta,(k+1)\delta]}\frac{\partial\left[t^2\|\Gamma_{XU}(t)\|\right]}{\partial t}\delta d\tau \\
&= \sum_{k=-\infty}^{\infty}\max_{t\in[k\delta,(k+1)\delta]}\frac{\partial\left[t^2\|\Gamma_{XU}(t)\|\right]}{\partial t}\cdot\delta^2.
\end{aligned}$$

If the above sum is finite and $\int_{-\infty}^{\infty}\tau^2\|\Gamma_{XU}(\tau)\|\,d\tau < \infty$, then $\delta\sum_{k=-\infty}^{\infty}(k\delta)^2\|\Gamma_{XU}(k\delta)\| < \infty$.

Similarly,

$$\begin{aligned}
&\delta \sum_{k=-n+1}^{n-1} \Gamma_{XU}(k\delta) - \int_{-T}^{T} \Gamma_{XU}(\tau)\, d\tau \\
&= \sum_{k=-n+1}^{n-1} \left[\int_{k\delta}^{(k+1)\delta} \left[\Gamma_{XU}(k\delta) - \Gamma_{XU}(\tau)\right] d\tau \right] + O(\delta) \\
&= \left[\sum_{k=-n+1}^{n-1} \max_{t\in[k\delta,(k+1)\delta]} \frac{\partial \Gamma_{XU}(t)}{\partial t} \delta + O(1) \right] \delta.
\end{aligned}$$

Therefore, Assumption 3.4(iii) holds if $\sum_{k=-n+1}^{n-1} \max_{t\in[k\delta,(k+1)\delta]} \frac{\partial \Gamma_{XU}(t)}{\partial t} \delta < \infty$.

Assumption 3.4(iv) contains some additional mild conditions on the basis functions. When the derivatives do not exist on a set of measure zero, we can let the derivatives take any finite value on this set. The assumption is satisfied for the sine and cosine basis functions such as

$$\phi_{2j-1}(r) = \sqrt{2}\cos(2\pi j r) \text{ and } \phi_{2j}(r) = \sqrt{2}\sin(2\pi j r) \text{ for } j = 1, \ldots, K/2. \quad (1)$$

For the above set of Fourier bases, we have

$$\ddot{\phi}_{2j-1}(r) = -\sqrt{2}\,(2\pi j)^2 \cos(2\pi j r) \text{ and } \ddot{\phi}_{2j}(r) = -\sqrt{2}\,(2\pi j)^2 \sin(2\pi j r) \text{ for } j = 1, \ldots, K/2,$$

and hence

$$\begin{aligned}
c_\phi &= -\lim_{K\to\infty} \frac{1}{K^3} \sum_{j=1}^{K} \frac{1}{2} \int_0^1 \phi_j(r)\, \ddot{\phi}_j(r)\, dr \\
&= \lim_{K\to\infty} \frac{1}{K^3} \sum_{j=1}^{K/2} \frac{4\pi^2 j^2}{2} \left[\int_0^1 2\sin(2\pi j r)^2\, dr + \int_0^1 2\cos(2\pi j r)^2\, dr \right] \\
&= \lim_{K\to\infty} \frac{1}{K^3} \sum_{j=1}^{K/2} 4\pi^2 j^2 = \int_0^{1/2} 4\pi^2 x^2 dx = \frac{\pi^2}{6}.
\end{aligned}$$

We will use the Fourier bases in our simulation study.

For a kernel function $k(\cdot)$ with Parzen exponent q, the asymptotic bias of the kernel LRV estimator depends on the "Parzen parameter" c_k defined by

$$c_k = \lim_{x\to 0} \frac{1 - k(x)}{x^q}.$$

The parameter c_ϕ in Assumption 3.4(iv) plays the same role in series LRV estimation as c_k does in kernel LRV estimation. Here, the assumptions imposed on the basis functions ensure that the resulting series LRV estimator is analogous to a kernel LRV estimator with a second-order kernel (i.e., its Parzen exponent q is equal to 2). There are other sets of basis functions such as Legendre polynomials that deliver series LRV estimators with asymptotic properties similar to the kernel LRV estimators based on a first-order kernel (e.g., the Bartlett kernel). See Lazarus et al. (2018) for more discussion. Hwang and Sun (2018) discuss why the set of Legendre polynomials may not be a good choice. We focus on second-order series LRV estimators in this paper.

Theorem 3.2 *Let Assumption 3.4 hold.*
(i) The variance of $\hat{\Omega}^*$ *satisfies*

$$\text{var}\left[\text{vec}(\hat{\Omega}^*)\right] = \frac{1}{K}\left(\Omega\otimes\Omega\right)\left(\mathbb{I}_{d^2}+\mathbb{K}_{dd}\right)\left(1+o\left(1\right)\right),$$

where $\mathbb{I}_{d^2}$ *is the* $d^2\times d^2$ *identity matrix and* $\mathbb{K}_{dd}$ *is the* $d^2\times d^2$ *commutation matrix.*
(ii) The bias of $\hat{\Omega}^*$ *satisfies*

$$E\left(\hat{\Omega}^*-\Omega\right) = -c_\phi\frac{K^2}{T^2}B\left(1+o\left(1\right)\right)+O\left(\delta+\frac{\left(\log n\right)^2}{T^2}+\frac{1}{T}+o\left(\frac{K^2}{T^2}\right)\right),$$

where

$$B=\int_{-\infty}^{\infty}\tau^2\Gamma_{XU}\left(\tau\right)d\tau.$$

The variance and bias expressions are similar to those in the case with low-frequency discrete-time data. Their interpretations are also similar. For example, when X_tU_t is positively autocorrelated such that $\Gamma_{XU}\left(\tau\right)>0$ for all τ, then $B>0$ and $\hat{\Omega}^*$ is biased downward. This is entirely analogous to the discrete-time case. Note that the dominating bias is equal to $-c_\phi K^2T^{-2}B$ instead of $-c_\phi K^2n^{-2}B$. The latter can be shown to be the dominating bias in the discrete-time case with n observations. A takeaway from this comparison is that the effective sample size for high-frequency data is the time span T instead of the number of observations n over this time span. When we use the effective sample size T in the bias expression, the asymptotic bias depends only on B, which is an intrinsic feature of the continuous-time process. In particular, the asymptotic bias does not depend on δ. This may appear counter-intuitive. We may argue that the process becomes more persistent for a smaller δ, and so we expect a larger absolute bias for a smaller δ. Such an argument is valid if we represent the asymptotic bias in terms of n, namely $-c_\phi\left(K^2n^{-2}\right)\left(B\delta^{-2}\right)$. Smaller δ indeed leads to a larger bias for a given n, but n becomes larger for a smaller δ. The net effect is that the asymptotic bias depends on the effective sample size T but not n or δ separately.

Define[3]

$$\text{MSE}(\hat{\Omega}^*) = E\text{vec}\left(\hat{\Omega}^*-\Omega\right)'\text{vec}\left(\hat{\Omega}^*-\Omega\right),$$

which is the mean square error of vec$(\hat{\Omega}^*)$. It follows from Theorem 3.2 that

$$\begin{aligned}&\text{MSE}(\hat{\Omega}^*)\\ &=\text{tr}\left[\left\{\Omega\otimes\Omega\right\}\left(\mathbb{I}_{d^2}+\mathbb{K}_{dd}\right)\right]\frac{1}{K}+c_\phi^2\text{vec}\left(B\right)'\text{vec}\left(B\right)\frac{K^4}{T^4}\\ &+o\left(\frac{1}{K}+\frac{K^4}{T^4}\right)+O\left(\delta^2+\frac{\left(\log n\right)^4}{T^4}+\frac{1}{T^2}\right).\end{aligned}$$

[3]It is possible to weigh different elements of vec $\left(\hat{\Omega}^*-\Omega\right)$ differentlly by defining

$$\text{MSE}(\hat{\Omega}^*) = E\text{vec}\left(\hat{\Omega}^*-\Omega\right)'\mathcal{W}\text{vec}\left(\hat{\Omega}^*-\Omega\right)$$

for some matrix $\mathcal{W}$. Here we have implicitly chosen $\mathcal{W}$ to be an identity matrix.

Ignoring the terms that will be shown to be of smaller order and optimizing MSE($\hat{\Omega}^*$) for K, we obtain the formula[4]

$$K = \kappa\left(\Omega, B\right)^{1/5} T^{4/5}, \tag{2}$$

where

$$\kappa\left(\Omega, B\right) := \left(\frac{tr\left[\left\{\Omega \otimes \Omega\right\}\left(\mathbb{I}_{d^2} + \mathbb{K}_{dd}\right)\right]}{4c_\phi^2 \text{vec}\left[B\right]' \text{vec}\left[B\right]}\right).$$

When $K = \kappa\left(\Omega, B\right)^{1/5} T^{4/5}$, the first two terms in MSE($\hat{\Omega}^*$) are of order $T^{-4/5}$. To ensure the terms that we ignore are indeed of smaller order, we require that

$$\delta^2 + \frac{(\log n)^4}{T^4} + \frac{1}{T^2} = o\left(T^{-4/5}\right).$$

If we set $\delta = O(T^{-\tau})$, then we require $\tau > 2/5$.

In the case of low-frequency discrete-time data with sample size n, the optimal choice of K is given by

$$K_D = \kappa\left(\Omega_D, B_D\right)^{1/5} n^{4/5}, \tag{3}$$

where

$$\kappa\left(\Omega_D, B_D\right) := \frac{tr\left[\left\{\Omega_D \otimes \Omega_D\right\}\left(\mathbb{I}_{d^2} + \mathbb{K}_{dd}\right)\right]}{4c_\phi^2 \text{vec}\left[B_D\right]' \text{vec}\left[B_D\right]}.$$

See Phillips (2005). In the above, Ω_D and B_D are the discrete analogues of Ω and B. If we use the low-frequency formula for K and set $K = cn^{4/5}$ for some constant $c > 0$, then we obtain a sub-optimal rate of K for the high-frequency data. The choice of $K = cn^{4/5}$ is too large for high-frequency data. For this type of data, the neighboring observations are highly correlated, and a smaller K is desired.

Now suppose we pretend that $\{z_i = x_i u_i\}_{i=1}^n$ is a low-frequency discrete-time process with n observations, and we use a parametric AR(1) plug-in approach to implement (3). We fit an AR(1) model to each component of z_i :

$$z_{i,j} = \rho_j z_{i-1,j} + e_j \text{ for } j = 1, 2, ..., d$$

with the AR parameter and error variance estimated by

$$\hat{\rho}_j = \frac{\sum_{i=2}^n z_{i,j} z_{i-1,j}}{\sum_{i=2}^n z_{i-1,j}^2} \text{ and } \hat{\sigma}_j^2 = \frac{1}{n}\sum_{i=2}^n \left(z_{i,j} - \hat{\rho}_j z_{i-1,j}\right)^2.$$

On the basis of the above plug-in estimates, we compute

$$\hat{\kappa}_D = \frac{1}{8c_\phi^2}\left(\sum_{j=1}^d \frac{\hat{\rho}_j^2 \hat{\sigma}_j^4}{\left(1 - \hat{\rho}_j\right)^8}\right)^{-1}\left(\sum_{j=1}^d \frac{\hat{\sigma}_j^4}{\left(1 - \hat{\rho}_j\right)^4}\right)$$

and let

$$\hat{K}_D = \hat{\kappa}_D^{1/5} n^{4/5}. \tag{4}$$

[4]Given that K is an integer, we should round $\kappa\left(\Omega, B\right)^{1/5} T^{4/5}$ up to the next integer and use it as K. We ignore this for the theoretical analysis but implement it in the simulation study.

The question is whether the so obtained $\hat{K}_D$ is of the optimal order $T^{4/5}$ with probability approaching one. On the surface, the answer is no, as $\hat{K}_D$ is apparently of order $n^{4/5}$. However, under the AR(1) plug-in implementation, $\hat{\kappa}_D$ is not a fixed constant. In fact, following Chang et al. (2018) (Lemma 4.2), we can show that as $\delta \to 0$ and $T \to \infty$,

$$\hat{\rho}_j = 1 - c_{1j}\delta + o_p(\delta) \text{ and } \hat{\sigma}_j^2 = c_{2j}\delta + o_p(\delta)$$

for some constants $c_{1j} > 0$ and $c_{2j} > 0$. Essentially, $\{z_{i,j}\}$ is a highly persistent process with the autocorrelation approaching unity at the rate of δ. The smaller δ is, the higher the autocorrelation is. As $\delta \to 0$, $\{z_{i,j}\}$ is effectively a near unit root process with the innovation variance proportional to the sampling interval δ. Plugging the above results into $\hat{\kappa}_D$ yields

$$\begin{aligned} \hat{\kappa}_D &= \frac{1}{8c_\phi^2}\left(\sum_{j=1}^{d}\frac{(c_{2j})^2\delta^2}{(c_{1j}\delta)^8}\right)^{-1}\left(\sum_{j=1}^{d}\frac{(c_{2j})^2\delta^2}{(c_{1j}\delta)^4}\right)(1+o_p(1)) \\ &= \frac{1}{8c_\phi^2}\left(\sum_{j=1}^{d}\frac{c_{2j}^2}{c_{1j}^8}\right)^{-1}\left(\sum_{j=1}^{d}\frac{c_{2j}^2}{c_{1j}^4}\right)\delta^4(1+o_p(1)). \end{aligned}$$

As a result,

$$\begin{aligned} \hat{K}_D = \hat{\kappa}_D^{1/5}n^{4/5} &= \left[\frac{1}{8c_\phi^2}\left(\sum_{j=1}^{d}\frac{c_{2j}^2}{c_{1j}^8}\right)^{-1}\left(\sum_{j=1}^{d}\frac{c_{2j}^2}{c_{1j}^4}\right)\right]^{1/5}\delta^{4/5}n^{4/5}(1+o_p(1)) \\ &= \left[\frac{1}{8c_\phi^2}\left(\sum_{j=1}^{d}\frac{c_{2j}^2}{c_{1j}^8}\right)^{-1}\left(\sum_{j=1}^{d}\frac{c_{2j}^2}{c_{1j}^4}\right)\right]^{1/5}T^{4/5}(1+o_p(1)). \end{aligned}$$

With probability approaching one, the rate of $\hat{K}_D$ is the same as the optimal rate of $T^{4/5}$. So the AR(1) plug-in implementation leads to a rate-optimal choice of K. Chang et al. (2018) call this feature of the AR(1) plug-in high-frequency compatible.

It should be noted that in the discrete-time setting it is typical to truncate the AR estimator at 0.97. See footnote 8 of Andrews (1991). Here, we should not follow this practice, as we rely on the convergence of $1 - \hat{\rho}_j$ to zero at the rate of δ to achieve the high-frequency compatibility. Had we truncated the initial AR estimator at 0.97 or any fixed number less than 1, $\hat{\kappa}_D$ would be bounded away from zero with probability approaching one. As a result, $\hat{K}_D$ would be of order $n^{4/5}$ and we would lose the high-frequency compatibility. Computationally, without truncating the initial AR estimator, we may have $1 - \hat{\rho}_j = 0$ and encounter the "divided by zero" problem. To avoid this, we can truncate the AR estimator so that $1 - \hat{\rho}_j$ is larger than the machine epsilon.

Note that the high-frequency compatible rate of K is $O\left(T^{4/5}\right)$, which is smaller than $n^{4/5}$ by an order of magnitude. So, when T is small, K may be small too, and the fixed-K asymptotics may be more accurate.

To conclude this section, we have shown that, in the stationary case, we do not need to change our estimation and inference methods to account for the high-frequency nature of the data. We can use exactly the same approach as we would do in the case with low-frequency data: the test statistic is constructed in the same way, and the smoothing parameter is chosen in the same way. The only caveat is that we should use a parametric AR(1) plug-in to obtain the data-driven smoothing parameter. Using the nonparametric approach of Newey and West (1994) will lead to a sub-optimal rate for the smoothing parameter. See Chang et al. (2018) for the details.

4 The Nonstationary Case

In this section, we consider linear hypothesis testing for cointegrating regressions in the continuous-time setting. The model is

$$Y_t = \alpha_0 + X_t'\beta_0 + U_{0t},$$

where each of $X_t \in \mathbb{R}^{d\times 1}$ is a nonstationary process and U_{0t} is a stationary zero-mean process. All processes are assumed to be right continuous with left limits almost surely. As in the stationary case, only a discrete set of points $\{x_i, y_i\}_{i=1}^n$ are observed,

$$x_i = X_{i\delta}; y_i = Y_{i\delta},$$

for $i = 1, ..., n$ and $n = T/\delta$. The discrete-time model is

$$y_i = \alpha_0 + x_i'\beta_0 + u_{0i}.$$

The object of interest is the slope parameter β_0, and we aim at testing $H_0 : R\beta_0 = r$ against $H_0 : R\beta_0 \neq r$ where $R \in \mathbb{R}^{p\times d}$ is of rank p. Note that here we single the intercept out of the slope parameter, and the hypothesis of interest involves only the slope parameter.

Following the discrete-time framework of Hwang and Sun (2018), we allow dependence between the regressor processes and the regression error process. We consider the same limiting experiment where $\delta \to 0$ and $T \to \infty$ for a fixed K and maintain the following assumptions.

Assumption 4.1 *For a sequence of $d \times d$ diagonal matrices (Λ_T),*

$$\begin{pmatrix} \Lambda_T^{-1} X_{Tr} \\ \frac{1}{\sqrt{T}} \int_0^{Tr} U_{0s} ds \end{pmatrix} \Rightarrow \begin{pmatrix} X^\circ(r) \\ B(r) \end{pmatrix} \text{ for } r \in [0, 1]$$

as $T \to \infty$, where $X^\circ(\cdot)$ is a continuous (a.s.) semimartingale and

$$B(r) = [X^\circ(r)]' \theta_0 + \sigma_0 W_0(r), \quad r \in [0, 1], \tag{5}$$

for $\sigma_0 \neq 0$ and a standard Brownian motion $W_0(\cdot)$ that is independent of $X^\circ(\cdot)$.

Assumption 4.2 *(i) $\Lambda_T^{-1} X_0 = o_p(1)$. (ii) As $\delta \to 0$ and $T \to \infty$,*

$$\frac{\delta}{\sqrt{T}} \sum_{i=1}^{[nr]} u_{0i} = \frac{1}{\sqrt{T}} \int_0^{Tr} U_{0t} dt + o_p(1),$$

uniformly over $r \in [0, 1]$.

Assumption 4.3 *Let*

$$\xi_j = \int_0^1 \phi_j(r) dX^\circ(r), \quad \eta_j = \int_0^1 \phi_j(r) X^\circ(r) dr, \quad j = 1, \ldots, K,$$

and

$$\xi = (\xi_1, \ldots, \xi_K)' \in \mathbb{R}^{K\times d}, \quad \eta = (\eta_1, \ldots, \eta_K)' \in \mathbb{R}^{K\times d}, \; \zeta = (\eta, \xi) \in \mathbb{R}^{K\times 2d}.$$

With probability one, $\zeta'\zeta$ and $\xi'\xi$ are of (full) ranks $2d$ and d, respectively.

Similar to Assumption 3.2, the weak convergence in Assumption 4.1 is defined on $\mathbb{D}^{d+1}[0,1]$, the space of cadlag functions from $[0,1]$ to $\mathbb{R}^{(d+1)\times 1}$ endowed with the Skorokhod topology.

Assumption 4.1 is the continuous-time analogue of the traditional invariance principles. For example, in the discrete-time setting, Vogelsang and Wagner (2014) model x_i as $x_i = \sum_{s=1}^{i} u_{xs} + x_0$ and assume that $\{u_{xi}\}$ is correlated with $\{u_{0i}\}$ and that $T^{-1/2}\sum_{i=1}^{[nr]}(u_{0i}, u_{xi}')'$ converges weakly to a Brownian motion. For intuition, Assumption 4.1 permits the form $X_t = \int_0^t U_{xs}ds + X_0$ where (U_{xt}) may be correlated with (U_{0t}) and X° is Brownian motion. However, Assumption 4.1 also admits potentially desirable properties of (X_t) such as non-differentiability. Lu and Park (2019) discusses sufficient conditions under which invariance principles hold in the discrete-time and continuous-time settings. Additional references can be found in Chang et al. (2018) where it is noted that wide classes of continuous-time processes, such as general null recurrent diffusions and jump diffusions, satisfy the requirement that $\Lambda_T^{-1}X_{T(\cdot)}$ converges weakly in $\mathbb{D}^d[0,1]$ for some sequence Λ_T. Our assumption that X° takes continuous sample paths (a.s.) is utilized only to simplify the exposition. Analogs of Assumptions A and D2 in Chang et al. (2018) could be used in its place.

As in the stationary case, Assumption 4.2 constrains the jump intensity and magnitude in the error process (U_{0t}) and assumes that the gap between the (scaled) average of the discrete observations and the scaled integral of the continuous process is asymptotically negligible. Mild primitive conditions under which Assumption 4.2 is satisfied are discussed in greater detail in Chang et al. (2018). Technical assumption 4.3 specifies that if we reduce attention from the typically infinite dimensional object X° to a finite dimensional set of (random) vectors, there are no inadvertent further reductions in dimension. It essentially requires that, with probability one, the $L^2[0,1]$ projection coefficients of components of X° in the directions ϕ_j and $\dot{\phi}_j, j = 1, \ldots, K$, form $2K$ linearly independent vectors. For a given choice of $\{\phi_j\}_{j=1}^K$, such as the first K Fourier basis functions, this is satisfied by virtually all continuous-time processes used in practice.

Now we detail the testing procedure. Let

$$\Delta_{T,\delta}x_i = \frac{\sqrt{T}\Lambda_T^{-1}}{\delta}\Delta x_i \text{ and } \Delta x_i = x_i - x_{i-1}.$$

To alleviate the notational burden, we write $\Delta_{T,\delta}x_i$ as $\tilde{\Delta}x_i$.

Augmenting the discrete-time model by $\tilde{\Delta}x_i$, we obtain

$$y_i = \alpha_0 + x_i'\beta_0 + \tilde{\Delta}x_i'\theta_0 + u_{0\cdot xi},$$

where

$$u_{0\cdot xi} = u_{0i} - \tilde{\Delta}x_i'\theta_0.$$

Assume that $K \geq 2d+1$. The testing steps are as follows:

1. Create the transformed data $\{\mathbb{W}_j^y, \mathbb{W}_j^x, \mathbb{W}_j^{\tilde{\Delta}x}\}_{j=1}^K$ where the K weighted observations are described for $j \in \{1, \ldots, K\}$ by

$$\mathbb{W}_j^y = \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right)y_i, \quad \mathbb{W}_j^x = \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right)x_i,$$
$$\mathbb{W}_j^{\tilde{\Delta}x} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right)\tilde{\Delta}x_i. \tag{6}$$

Denote the matrix forms of transformed data by

$$\underset{K\times 1}{\mathbb{W}^y = (\mathbb{W}_1^y, \ldots, \mathbb{W}_K^y)'}, \quad \underset{K\times d}{\mathbb{W}^x = (\mathbb{W}_1^x, \ldots, \mathbb{W}_K^x)'}, \quad \underset{K\times d}{\mathbb{W}^{\tilde{\Delta}x} = (\mathbb{W}_1^{\tilde{\Delta}x}, \ldots, \mathbb{W}_K^{\tilde{\Delta}x})'}.$$

2. Regress $\mathbb{W}^y$ on $\mathbb{W}^x$ and $\mathbb{W}^{\tilde{\Delta}x}$ by OLS (the transformed and augmented OLS (TAOLS) regression). Do not include an intercept. Denote the coefficients associated with $\mathbb{W}^x$ by $\hat{\beta}_{TAOLS}$, the coefficients associated with $\mathbb{W}^{\tilde{\Delta}x}$ by $\hat{\theta}_{TAOLS}$, and let $\hat{\mathbb{W}}^{0\cdot x}$ be the residual vector from this regression. Combining the matrices $\mathbb{W}^x$ and $\mathbb{W}^{\tilde{\Delta}x}$ into $\widetilde{\mathbb{W}} = (\mathbb{W}^x, \mathbb{W}^{\tilde{\Delta}x})$, we can write these objects as

$$\underset{2d\times 1}{\hat{\gamma}} \equiv \begin{pmatrix} \hat{\beta}_{TAOLS} \\ \hat{\theta}_{TAOLS} \end{pmatrix} = \left(\widetilde{\mathbb{W}}'\widetilde{\mathbb{W}}\right)^{-1} \widetilde{\mathbb{W}}'\mathbb{W}^y, \quad \hat{\mathbb{W}}^{0\cdot x} = \mathbb{W}^y - \widetilde{\mathbb{W}}\hat{\gamma}. \tag{7}$$

3. To test $H_0 : R\beta_0 = r$, we calculate the following test statistic

$$F_{TAOLS} = \frac{1}{\hat{\sigma}_{0\cdot x}^2} \left(R\hat{\beta}_{TAOLS} - r\right)' \left[R \left(\mathbb{W}^{x\prime} M_{\tilde{\Delta}x} \mathbb{W}^x\right)^{-1} R'\right]^{-1} \left(R\hat{\beta}_{TAOLS} - r\right) /p, \tag{8}$$

where $M_{\tilde{\Delta}x} = I_K - \mathbb{W}^{\tilde{\Delta}x}(\mathbb{W}^{\tilde{\Delta}x\prime}\mathbb{W}^{\tilde{\Delta}x})^{-1}\mathbb{W}^{\tilde{\Delta}x\prime}$ and

$$\hat{\sigma}_{0\cdot x}^2 = \frac{1}{K} \sum_{j=1}^K \left(\hat{\mathbb{W}}_j^{0\cdot x}\right)^2 = \frac{1}{K} \hat{\mathbb{W}}^{0\cdot x\prime} \hat{\mathbb{W}}^{0\cdot x}. \tag{9}$$

These three steps are identical to the procedure in Hwang and Sun (2018) except that $\tilde{\Delta}x_i$, instead of Δx_i, is used in the augmented regression. Such a modification serves to facilitate theoretical developments only. Since $\tilde{\Delta}x_i$ is proportional to Δx_i, the modification has no effect on the test statistic F_{TAOLS}. For practical implementation, we can follow exactly the same procedure as in Hwang and Sun (2018), utilizing Δx_i in place of $\tilde{\Delta}x_i$ in steps 1-3.

Define

$$\mathbb{W}_j^{0\cdot x} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_j\left(\frac{i}{n}\right)[u_{0i} - \theta_0' \tilde{\Delta}x_i], \quad \underset{K\times 1}{\mathbb{W}^{0\cdot x} = (\mathbb{W}_1^{0\cdot x}, \ldots, \mathbb{W}_K^{0\cdot x})'},$$

Similarly to the notation in Assumption 4.3, for $j = 1, \ldots, K$, let

$$\nu_j = \sigma_0 \int_0^1 \phi_j(r) dW_0(r),$$

and

$$\nu = (\nu_1, \ldots, \nu_K)' \in \mathbb{R}^{K\times 1}.$$

The following lemma establishes the weak limits of $\mathbb{W}^x, \mathbb{W}^{\tilde{\Delta}x}$, and $\mathbb{W}^{0\cdot x}$.

Lemma 4.1 *Let Assumptions 3.3, 4.1, and 4.2 hold. Then, as $\delta \to 0$ and $T \to \infty$,*

$$\left(\frac{1}{\sqrt{n}} \mathbb{W}^x \Lambda_T^{-1}, \sqrt{\delta}\mathbb{W}^{\tilde{\Delta}x}, \sqrt{\delta}\mathbb{W}^{0\cdot x}\right) \xrightarrow{d} (\eta, \xi, \nu).$$

Let $R_{\ell,\cdot}$ be the ℓ-th row of R. Since we do not require that all elements of (X_{Tr}) converge at the same rate, the rate of convergence of $R_{\ell,\cdot}\hat{\beta}_{TAOLS}$ depends on the element of $\hat{\beta}_{TAOLS}$ that has the slowest rate of convergence. To capture this, we define the $p \times p$ diagonal matrix $\tilde{\Lambda}_T$ with the (ℓ, ℓ)-th element given by

$$\tilde{\Lambda}_T(\ell,\ell) = \left[\max_{1\le j\le d}\left\{a_{\ell j} : a_{\ell j} = 1_{R_{\ell,j}\neq 0}/\Lambda_T(j,j)\right\}\right]^{-1} \text{ for each } \ell = 1, ..., p. \tag{10}$$

Then $\lim_{T\to\infty} \tilde{\Lambda}_T R \Lambda_T^{-1} = A$ for some matrix $A \in \mathbb{R}^{p\times d}$. We will require that A is of rank p, a condition that is clearly satisfied when there is no heterogeneity in rates of convergence, for example, $A = R$ when $\Lambda_T = \sqrt{T} I_d$.

The following theorem is analogous to Theorems 1 and 3 of Hwang and Sun (2018).

Theorem 4.1 *Let Assumptions 3.3, 4.1–4.3 hold. Denote* $\gamma_0 = (\beta_0', \theta_0')'$ *and*

$$\Upsilon_T = \begin{pmatrix} \sqrt{T}\Lambda_T & \underset{d\times d}{0} \\ \underset{d\times d}{0} & \mathbb{I}_d \end{pmatrix}.$$

Then, as $\delta \to 0$ *and* $T \to \infty$,

$$\Upsilon_T(\hat{\gamma} - \gamma_0) \xrightarrow{d} (\zeta'\zeta)^{-1}\zeta'\nu \overset{d}{=} MN\left[0, \sigma_0^2(\zeta'\zeta)^{-1}\right].$$

In particular, denoting $M_\xi = I_K - \xi(\xi'^{-1}\xi'$,

$$\sqrt{T}\Lambda_T\left(\hat{\beta}_{TAOLS} - \beta_0\right) \xrightarrow{d} (\eta' M_\xi \eta)^{-1}\eta' M_\xi \nu \overset{d}{=} MN\left[0, \sigma_0^2(\eta' M_\xi\eta)^{-1}\right].$$

Additionally, provided that $\lim_{T\to\infty}\tilde{\Lambda}_T R\Lambda_T^{-1}$ *is of rank* p *and* $K \ge 2d+1$,

$$F_{TAOLS} \xrightarrow{d} \frac{K}{K-2d}\cdot F_{p,K-2d},$$

where $F_{p,K-2d}$ *is the* F *distribution with degrees of freedom* p *and* $K - 2d$.

The theorem shows that the testing procedure of Hwang and Sun (2018) adapts to the continuous-time setting without any modification: The asymptotic F test is, therefore, robust to the sampling frequency of the data. From an applied point of view, we do not have to be concerned about whether we have high-frequency data or low-frequency data. This gives us much practical convenience.

Note that the asymptotic F theory does not depend on the specific form of the limiting process $X^\circ(\cdot)$. In the proof of the theorem, we show that the asymptotic distribution conditional on $X^\circ(\cdot)$ is an F distribution, which does not depend on the conditioning process $X^\circ(\cdot)$. Hence, the asymptotic distribution is also the F distribution unconditionally. We note that the conditioning argument works only for the series-based TA regression. Had we used the kernel-based approach such as that in Vogelsang and Wagner (2014), we may not obtain a pivotal distribution if the parameters governing $X^\circ(\cdot)$ can not be scaled out.

To implement the F test, we still need to choose K. Ideally we want to select K to tradeoff the type I and type II errors of the F test, but this is well beyond the scope of this paper. Note that the variance estimator in (9) takes a similar form to that in the stationary case; the infeasible estimator is now

$$\hat{\sigma}_{0\cdot x}^2 = \frac{1}{K}\sum_{j=1}^{K}\left[\frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)(u_i - \theta_0'\tilde{\Delta}x_i)\right]^2.$$

As a practical rule of thumb, we can adapt the data-driven procedure in the stationary case, taking $\theta_0 = 0$ as it would be in the exogenous case. Alternatively, we could perform the data driven approach of the stationary case on the sample version of the series $\{u_i, \tilde{\Delta}x_i\}$. This is the approach taken in Hwang and Sun (2018). However, the latter option requires a stance taken on δ, which may not be desirable and does not yield meaningful differences in our simulations. As a rule of thumb, one can proceed as follows:

1. Estimate the model $y_i = \alpha_0 + x_i'\beta_0 + u_{0i}$ by OLS to obtain the residual

$$\hat{u}_{0i} = y_i - \hat{\alpha}_{OLS} - x_i'\hat{\beta}_{OLS}.$$

2. On the basis of $\{\hat{u}_{0i}\}$, use the series method to estimate the long run variance of $\{u_i\}$, computing the AR(1) data-driven $\hat{K}_D$ using the formula in (4).

3. Let $\hat{K}^* = \max(\hat{K}_D, 2d+3)$ and use $\hat{K}^*$ to construct the TA regression.

4. Compute the F test statistic in the TAOLS regression. Perform the asymptotic F Test using $\frac{\hat{K}^*}{\hat{K}^*-2d}\cdot F_{p,\hat{K}^*-2d}$ as the reference distribution.

If there is a trend in the model so that

$$Y_t = \alpha_0 + X_t'\beta_0 + \Gamma_t\lambda_0 + U_{0t},$$

where $\Gamma_t \in R^{d_\Gamma}$ consists of deterministic trend functions, then we can show that the asymptotic F test for $H_0 : R\beta_0 = r$ based on the transformed and augmented OLS regression is still asymptotically valid. We only need to adjust the degrees of freedom:

$$F_{TAOLS} \to^d \frac{K}{K-2d-d_\Gamma}\cdot F_{p,K-2d-d_\Gamma}.$$

See Hwang and Sun (2018) for more details.

5 Simulation Evidence

In this section we conduct simulations to evaluate the finite-sample size and power properties of the proposed F tests. For the stationary setting, we consider the model

$$Y_t = \beta_{01} + X_t\beta_{02} + U_t, \quad 0 \le t \le T,$$

with $\beta_{01} = 0$ and $\beta_{02} = 1$. We test $H_0 : (\beta_{01}, \beta_{02})' = (0,1)'$ versus $H_1 : (\beta_{01}, \beta_{02})' \neq (0,1)'$. (X_t) and (U_t) are chosen as stationary Ornstein-Uhlenbeck (OU) processes described by

$$dX_t = -\kappa_x X_t dt + \sigma_x dV_t \;\text{ and }\; dU_t = -\kappa_u U_t dt + \sigma_u dW_t,$$

where $(\kappa_x, \sigma_x) = (0.1020, 1.5514)$, $(\kappa_u, \sigma_u) = (6.9011, 2.7566)$, and (V_t) and (W_t) are independent standard Brownian motions. The parameter values of the OU processes are obtained from Chang et al. (2018), who estimate (κ_x, σ_x) by fitting an OU process to 3-month T-bill rates from 1971 to 2016 and estimate (κ_u, σ_u) by fitting an OU process to the residuals obtained by regressing 3-month eurodollar rates on these T-bill rates.

In the nonstationary setting, we consider the model

$$Y_t = \alpha_0 + X_{1,t}\beta_{01} + X_{2,t}\beta_{02} + U_t, \quad 0 \le t \le T,$$

with $\alpha_0 = 0, \beta_{01} = 1, \beta_{02} = 1$. We test $H_0 : (\beta_{01}, \beta_{02})' = (1,1)'$ versus $H_1 : (\beta_{01}, \beta_{02})' \neq (1,1)'$. In this setting we model $(X_{j,t})$, $j \in \{1,2\}$, as Brownian motions and (U_t) as a stationary OU process. In particular, for $j \in \{1,2\}$, we have

$$dX_{j,t} = \sigma_j dZ_{j,t} \text{ and } dU_t = -\kappa_u U_t dt + \sigma_u dZ_{3,t},$$

where $\sigma_1 = \sigma_2 = 0.0998$, $(\kappa_u, \sigma_u) = (1.5717, 0.0097)$, and

$$\begin{pmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \varphi & \sqrt{1-\varphi^2} & 0 \\ \varphi & \frac{\varphi-\varphi^2}{\sqrt{1-\varphi^2}} & \sqrt{1-\left(\varphi^2 + \frac{(\varphi-\varphi^2)^2}{1-\varphi^2}\right)} \end{pmatrix} \begin{pmatrix} W_{1,t} \\ W_{2,t} \\ W_{3,t} \end{pmatrix}.$$

Here $W_{1,t}$, $W_{2,t}$, and $W_{3,t}$ are independent standard Brownian motions and $\varphi \ge 0$. In this setup, each $(Z_{j,t}), j \in \{1,2,3\}$, is a standard Brownian motion and $Corr(Z_{k,t}, Z_{\ell,t}) = \varphi$ when $k \neq \ell$. The parameter values here also originate from Chang et al. (2018); (σ_1) comes from fitting a Brownian motion process to log US/UK exchange rate spot price data over 1979 to 2017. (κ_u, σ_u) are estimated by fitting an OU process to the residuals from regressing log US/UK exchange rate forward prices on the log US/UK exchange rate spot prices. We consider both $\varphi = 0$ (the exogeneous case) and $\varphi = 0.75$ (the endogenous case).

In addition to the baseline values of κ_x and κ_u, we also multiply κ_x and κ_u by 4, 1/2, and 1/4, allowing some variation in the mean reversion parameters of the stationary elements of the simulations. As the mean reversion parameter gets closer to zero, the stationary OU process becomes more persistent and behaves more like a nonstationary Brownian motion.

In both the stationary and nonstationary settings, we consider $T = 30$ and $T = 60$. The stochastic processes are generated using the transition densities of Brownian motion and the OU process except in the case when $\varphi = 0.75$. In this case, transition densities are used to generate all processes except that U_t is constructed via Euler's method once $Z_{3,t}$ is generated. Discrete samples are collected at various frequencies between $\delta = 1/252$ and $\delta = 1/4$. In each scenario, we replicate the simulation 5000 times.

To implement the testing procedures described in the earlier sections, we utilize the sine and cosine basis functions given in (1) and choose K via the data driven procedures described in Sections 3 and 4. In our figures described below, results corresponding to these tests are denoted "Series F", and there are different figures for the stationary and nonstationary settings. As K increases, in both the stationary and nonstationary settings, the limiting distributions of the test statistics approach the scaled chi-squared distribution $\chi^2(p)/p$. The scaled chi-squared approximation can also be obtained by letting $K \to \infty, \delta \to 0$ and $T \to \infty$ jointly[5]. Utilizing the critical values from this distribution with our test statistics, we denote the related results by "Series Chi2."

[5]The scaling factor of of $1/p$ arises because the test statistics are scaled by p.

To compare the F tests with some existing tests, we carry out the kernel-based tests of Chang et al. (2018). For their tests, we employ the quadratic spectral kernel and utilize Andrews (1991)'s bandwidth selection procedure, which is among the best performers in the simulations in Chang et al. (2018). In our figures, the results corresponding to the QS kernel are denoted "Kernel Chi2." To include the fixed-b version of their tests, we note that the test statistics of Chang et al. (2018) in the stationary setting and in the nonstationary setting with exogeneous regressors, without any change in form, have fixed-b counterparts in the discrete-time settings of Kiefer and Vogelsang (2005) and Jin et al. (2006), respectively. Utilizing arguments similar to what we present here and in Vogelsang and Wagner (2014), it is not difficult to ascertain that the limiting distributions identified in these papers are also applicable in our simulation set up with exogenous regressors. In the cointegrating regression with endogenous regressors, the fixed-b asymptotics of Jin et al. (2006) is not applicable to the test statistic of Chang et al. (2018), as it does not account for endogeneity. To use the fixed-b asymptotics of Vogelsang and Wagner (2014), which accounts for endogeneity, we have to run a different set of regressions and alter the test statistic. This would require further theoretical development and is not considered in our simulations. The tests utilizing the fixed-b approximations of Kiefer and Vogelsang (2005) and Jin et al. (2006) for the test statistics in Chang et al. (2018) are denoted by "Kernel fixed-b" in our figures.

5.1 Size study

Figures 1 − 4 display the empirical sizes in the different simulation scenarios.

Figures 1 and 2 show that in the stationary setting, the series-based F test exhibits less size distortion than all chi-squared tests under consideration. The improvement in the size accuracy of the F test over the chi-square tests is more visible when the underlying OU processes have smaller mean reversion parameters κ_x and κ_u and thus become more persistent. This is consistent with the literature on HAR inference in the discrete-time setting. See, for example, Sun (2013), Sun (2014b), Sun et al. (2008), and Kiefer and Vogelsang (2005) for simulation evidence and theoretical developments. The F test performs similarly to the fixed-b version of the test in Chang et al. (2018) adapted from Kiefer and Vogelsang (2005). This is expected, because both types of tests utilize nonparametric LRV estimators, and both are based on fixed-smoothing asymptotic approximations. The advantage of the series-based F test is that it is more convenient to use, as critical values are readily available from statistical tables and standard programming environments. There is no need to simulate a nonstandard fixed-smoothing asymptotic distribution, an unavoidable and formidable task if we use a kernel-based fixed-smoothing test. We note in passing that all chi-squared tests have similar performances, regardless of whether series-based or kernel-based LRV estimators are used. This provides further simulation evidence that the type of LRV estimators used does not matter much. What matters more are the reference distributions used in the testing procedures.

In the nonstationary setting with exogenous regressors, the performance of the F test relative to the fixed-b version of the test in Chang et al. (2018) adapted from Jin et al. (2006) and the chi-squared tests is qualitatively similar to that in the stationary setting. In particular, the F test and the fixed-b test achieve more or less the same size control. However, the fixed-b tests in this setting aren't developed fully for the continuous-time setting. The validity of the fixed-b test relies not only on the exogeneity of the regressors but also crucially on the premise that the limiting process (X°) is a Brownian motion process. While this does not cause problems in our simulation setting where the premise holds, the fixed-b asymptotic distribution is, in general, a functional of (X°), which may contain additional nuisance-parameters beyond its scale. A benefit of our approach here is that the conditioning argument in the proof of Theorem 4.1 bypasses reliance on the distributional form of (X°). Such a conditioning argument does not go through if we use a kernel LRV estimator.

In the nonstationary setting with endogenous regressors, to the best our knowledge, the F test in Section 4 appears to be the only asymptotically valid test in the literature. Unsurprisingly, it exhibits better size properties than all other tests, including the fixed-b version of the test in Chang et al. (2018). We note that the presence of the endogeneity bias can lead to large size distortion, especially when the chi-square approximation is used. For example, when $\varphi = 0.75$, $T = 30$, and κ_u is 1/4 of the baseline value, the null rejection probability of the 5% chi-squared test of Chang et al. (2018) can be as high as 60%.

Figures 1 – 4 further show that the size properties of all tests are not sensitive to the sampling interval δ, and all tests become more accurate when T increases. This is consistent with our theoretical results that the effective sample size is T and is unrelated to δ. Intuitively, for a given time span T, as δ decreases, the number of sampled observations n increases, but at the same time, the sampled observations become more persistent. These two effects offset each other, leading to an effective sample size of T.

5.2 Power study

Figures 5 and 6 investigate the empirical power properties of the test procedures in finite samples; the power is size adjusted. To evaluate the power of the tests, we use the baseline designs. When generating the data, each of the parameters being tested is multiplied by $1 - \psi$ for a range of $\psi \in [0, 1]$. To keep the visualization simple, we focus only on the frequencies $\delta = 1/252$ and $\delta = 1/4$. As the power is size adjusted and the tests employ two sets of test statistics, ours and those in Chang et al. (2018), the figures only display the comparison for the series-based approach in Sections 3 and 4 and the kernel-based approach in Chang et al. (2018). In the figures, the higher frequency $\delta = 1/252$ is denoted "h", and the lower frequency $\delta = 1/4$ is denoted "l."

Figure 5 shows that, in the stationary setting, all tests have almost indistinguishable power curves. In the nonstationary setting with exogenous regressors, the series-based tests have competitive power relative to the kernel-based tests, although when $T = 30$ the former are slightly less powerful. This could be explained by the MSE-optimality of the QS kernel among the second-order positive-definite kernels. In the nonstationary setting with endogenous regressors, the comparison is not as meaningful, as the tests of Chang et al. (2018) have significant size distortion. Nevertheless, the series-based tests still have competitive power, especially when $T = 60$. When $T = 30$, the series-based tests are somewhat less powerful. This is not unexpected, as the series-based method corrects for the endogenous bias, and as a result introduces additional noise to the point estimator and renders the associated test less powerful.

Figures 5 and 6 also show that the power properties of all tests are not sensitive to the choice of δ. In each scenario, the power curves for $\delta = 1/252$ and $\delta = 1/4$ are virtually identical. This echoes the finding that the size properties are not sensitive to δ. In each scenario, all tests become more powerful when T is larger, reflecting that it is the time span T, not the number of observations n, that is the effective sample size.

6 Robustness to Additive Low-frequency Noise

Here we consider the implications of including additional low-frequency noises in the regression error. We show that, under reasonable assumptions, the additive noises do not affect the testing procedures of previous sections. This is done to address two potential concerns. First, there may be covariates relevant to (Y_t) that are not continuous-time in nature and must be absorbed by the error term in the regression model. Depending on the observation frequency, it may be reasonable to expect this sort of error in some or all discretized observations. Second, this noise could alternatively be interpreted as microstructure noise. Particularly when working with financial data at high frequencies, market frictions and transcription errors may add noise to asset return data beyond the theoretical objects of interest such as returns satisfying a no-arbitrage condition. Among several possible references, see Hansen and Lunde (2006) and Barndorff-Nielsen et al. (2008) for discussions addressing microstructure noise in an alternative setting where the objective is ex post volatility measurement in asset returns.

One way to model the low-frequency noise is to assume that at the observation times $i\delta$, $i = 1, \ldots, n$, the true $Y_{i\delta}$ is not observed. Instead, we observe $Y_{i\delta}$ up to an additive noise term ϵ_i. That is, we now observe $y_i = Y_{i\delta} + \epsilon_i$. For $i = 1, \ldots, n$ and $n = T/\delta$, the *observed* discretized model in the stationary case is then

$$y_i = x_i'\beta_0 + u_i + \epsilon_i. \tag{11}$$

In the nonstationary setting, the *observed* discretized model is

$$y_i = \alpha_0 + x_i'\beta_0 + u_{0i} + \epsilon_i. \tag{12}$$

We will work with the following assumption.

Assumption 6.1 *(i) The process $\{\epsilon_i\}_{i=1}^n$ is independent of the continuous-time processes (X_t) and (U_t). (ii) As a discrete-time process, $\{\epsilon_i\}_{i=1}^n$ is stationary and strongly mixing with mixing coefficients $\{\varphi_n\}_{n=1}^\infty$ that satisfy $\sum_{n=1}^\infty \varphi_n^{1/2} < \infty$. (iii) $E\epsilon_1 = 0$ and $E\epsilon_1^2 < \infty$.*

Assumption 6.1 allows for a weakly dependent noise process where the dependence is tied to the distance in terms of sampling frequency units (i.e., $i - k$) rather than the distance in terms of the units of T. In the limiting experiment of this setting, as δ becomes small, noises at nearby sampling points exhibit dependence, but as the number of observations between any fixed time points $t_1, t_2 \in [0, T]$ gets large, the dependence between ϵ_{t_1} and ϵ_{t_2} becomes small. This form of dependence is consistent with microstructure noise assumptions in the literature on ex post variation measurement with high-frequency data; see, for example, Barndorff-Nielsen et al. (2008) and Aït-Sahalia et al. (2008).

Lemma 6.1 *Let Assumption 6.1 hold. Assume that (X_t) is stationary and $\Gamma_X(\tau) = E(X_t X_{t-\tau}')$, $\tau \geq 0$, is bounded. Then, as $\delta \to 0$ with T fixed, or as $T \to \infty$ with δ fixed, or as $\delta \to 0$ and $T \to \infty$,*

$$\sum_{i=1}^n x_i \epsilon_i = O_p(\sqrt{n}).$$

Whereas high serial correlation in $\{x_i u_i\}_{i=1}^n$ for small δ leads to $\sum_{i=1}^n x_i u_i = O_p\left(\sqrt{n/\delta}\right)$, the tie between the noise dependence structure and the sampling frequency yields that $\sum_{i=1}^n x_i \epsilon_i$ is an order of magnitude smaller (in probability) than $\sum_{i=1}^n x_i u_i$ despite the persistence in $\{x_i\}_{i=1}^n$. Consequently, we will see that the results from Section 3 continue to hold.

In the stationary setting, when the observed $\{y_i\}_{i=1}^n$ take the form in (11), we have

$$\sqrt{T}\left(\hat{\beta}_D-\beta_0\right)=\left(\frac{1}{\delta\Lambda\left(n,\delta\right)^2}\sum_{i=1}^n x_i x_i'\right)^{-1}\frac{1}{\Lambda\left(n,\delta\right)}\sum_{i=1}^n x_i(u_i+\epsilon_i).$$

In the series LRV estimator of Section 3, a key object now becomes

$$\frac{1}{\Lambda\left(n,\delta\right)}\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right)x_i\hat{u}_i=\frac{1}{\Lambda\left(n,\delta\right)}\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right)x_i\left[u_i+\epsilon_i-x_i'\left(\hat{\beta}_D-\beta\right)\right].$$

It follows from Lemma 6.1 that

$$\frac{1}{\Lambda\left(n,\delta\right)}\sum_{i=1}^n x_i\epsilon_i=\frac{\sqrt{\delta}}{\sqrt{n}}\sum_{i=1}^n x_i\varepsilon_i=\sqrt{\delta}O_p(1)=o_p(1),$$

and similarly, as each ϕ_j is bounded under Assumption 3.3,

$$\frac{1}{\Lambda\left(n,\delta\right)}\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right)x_i\epsilon_i=o_p(1),\quad j=1,\ldots,K.$$

Therefore, under the conditions of Lemma 6.1, the objects in Lemma 3.1 differ now only by additive $o_p(1)$ terms, and both objects there still jointly converge in distribution to the same limits. It follows that under the conditions of Lemma 6.1, Theorem 3.1 remains valid, and this is summarized in the following corollary to Lemma 6.1.

Corollary 6.1 *Consider the stationary setting and let the conditions of Lemma 6.1 hold. Theorem 3.1 remains valid when the observed discretized model contains additional low-frequency noises as in* (11)*.*

In the nonstationary setting, it is easy to see that similar arguments to those in the proof of Lemma 6.1, without any conditions on $E(X_tX_{t-\tau}')$, yield the following lemma.

Lemma 6.2 *Let Assumptions 6.1 and 3.3 hold. Then for $j=1,\ldots,K$,*

$$\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right)\epsilon_i=O_p\left(\sqrt{n}\right).$$

Now, the only change to the objects in Lemma 4.1 is that

$$\sqrt{\delta}\mathbb{W}_j^{0\cdot x}=\sqrt{\delta}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right)\left[u_{0i}+\epsilon_i-\theta_0'\tilde{\Delta}x_i\right]\right\}.$$

As

$$\sqrt{\delta}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right)\epsilon_i\right\}=\sqrt{\delta}O_p(1)=o_p(1),$$

the only difference is an additive $o_p(1)$ term, and Lemma 4.1 remains valid provided that Assumption 6.1 holds. Consequently, we have the following corollary to Lemma 6.2.

Corollary 6.2 *Consider the nonstationary setting and let Assumption 6.1 hold. Then Theorem 4.1 remains valid when the observed discretized model contains additional low-frequency noises as in* (12)*.*

Corollaries 6.1 and 6.2 show that our asymptotic theory is robust to the presence of additive low-frequency noises. The estimation and inference procedures of Sections 3 and 4 can be carried out without any modification.

7 Conclusion

This paper provides a simple approach to linear hypothesis testing that is robust to potential continuity of the underlying data generating processes. The test procedures demonstrate reduced size distortion in finite samples relative to existing approaches and can accommodate endogeneity in cointegration-type regressions. From a practical point of view, the tests have several desirable characteristics. Their direct correspondence to analogous discrete-time procedures clears the practitioner from modeling choices that could influence test results. Additionally, the limiting distributions do not need any complicated simulations to derive critical values as some discrete-time fixed-b approaches require; the tests rely only on standard F-distributions. In the cointegrating regression setting, more accurate tests are delivered while maintaining greater generality with regard to the limiting behavior of the regressor process. Lastly, our asymptotic F theory remains valid in the presence of additive low-frequency noises in the regressor error.

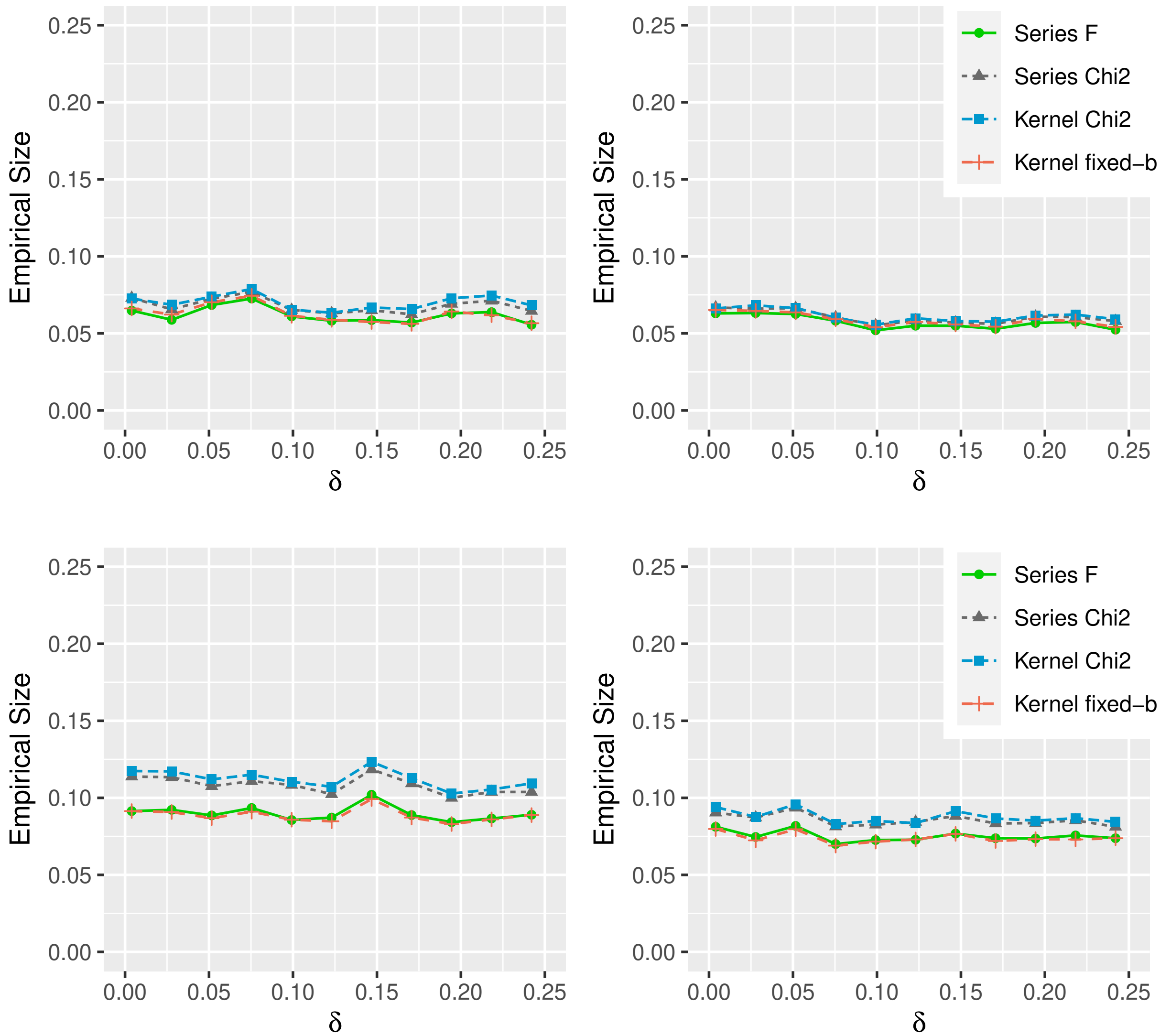

Figure 1: Empirical sizes for the stationary setting. The left hand side corresponds to $T = 30$ and the right to $T = 60$. In the bottom row, the processes are as described in 5. In the top row κ_u and κ_x are multiplied by 4.

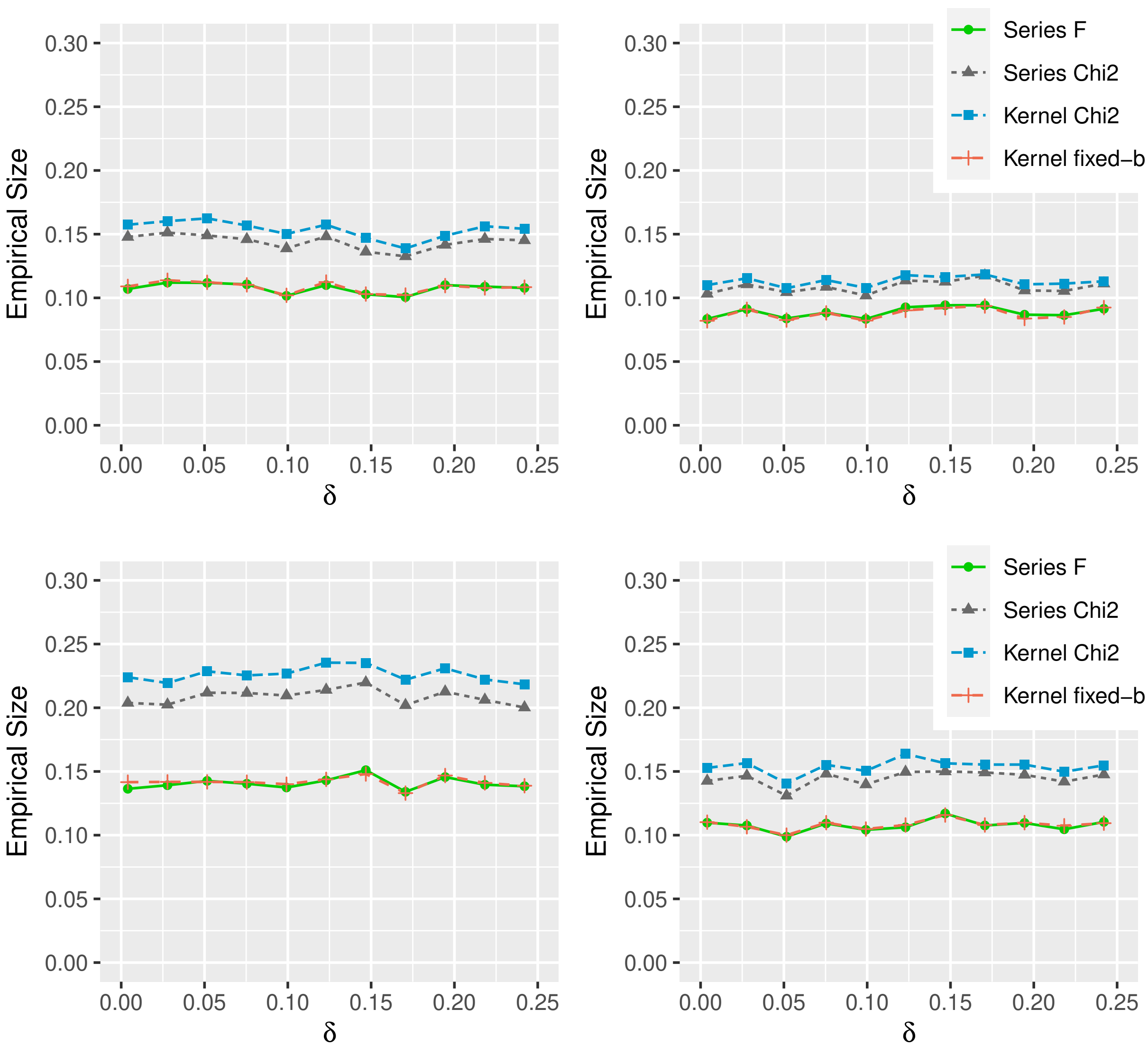

Figure 2: Empirical sizes for the stationary setting. The left hand side corresponds to $T = 30$ and the right to $T = 60$. In the top panel, κ_u and κ_x are multiplied by $1/2$ and in the bottom panels they are multiplied by $1/4$.

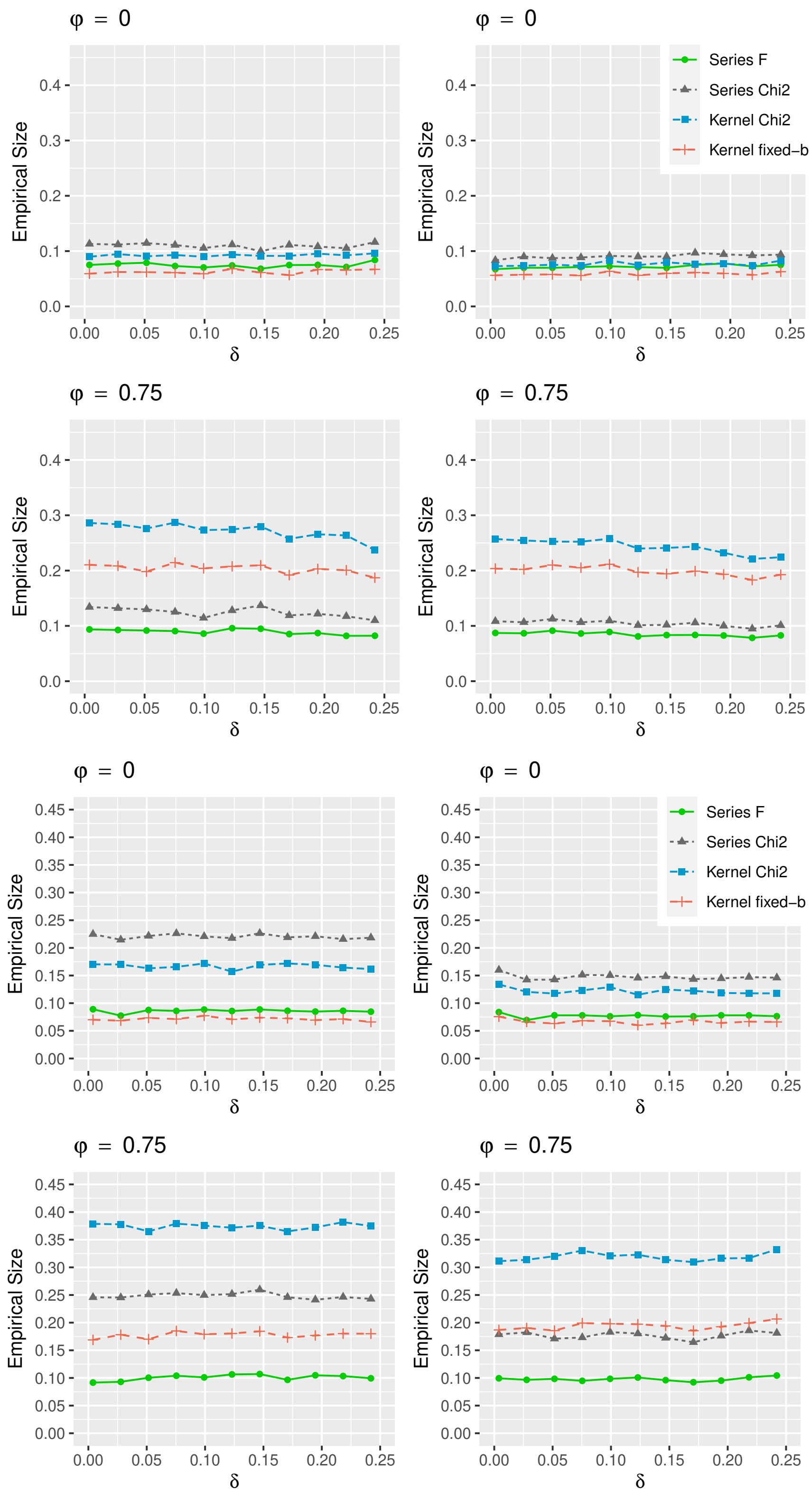

Figure 3: Empirical sizes, nonstationary setting. The left is $T = 30$ and the right is $T = 60$. In the top two rows, κ_u is multiplied by 4, and in the bottom two rows, κ_u is multiplied by 1.

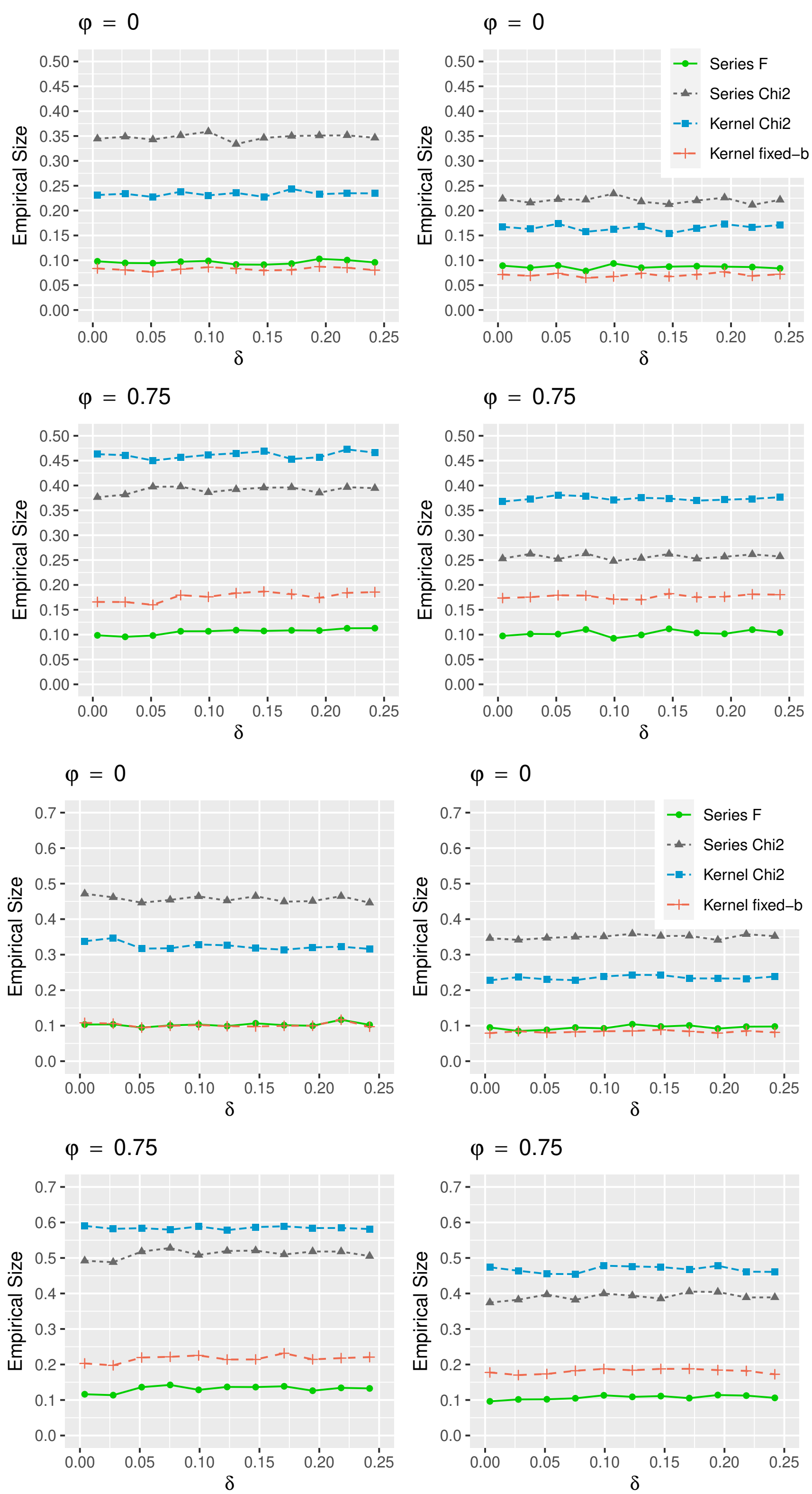

Figure 4: Empirical sizes, nonstationary setting. The left is $T = 30$ and the right is $T = 60$. In the top two rows, κ_u is multiplied by $1/2$, and in the bottom two rows, κ_u is multiplied by $1/4$.

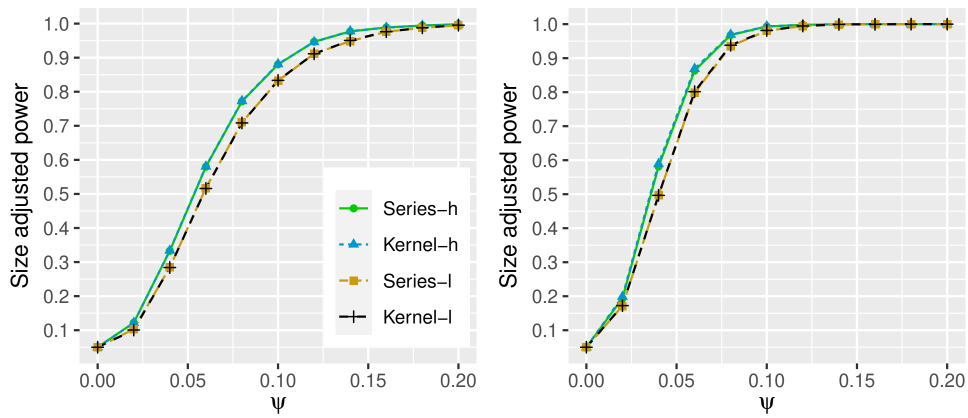

Figure 5: Size-adjusted powers in the stationary setting. The left panel corresponds to $T = 30$ and right to $T = 60$.

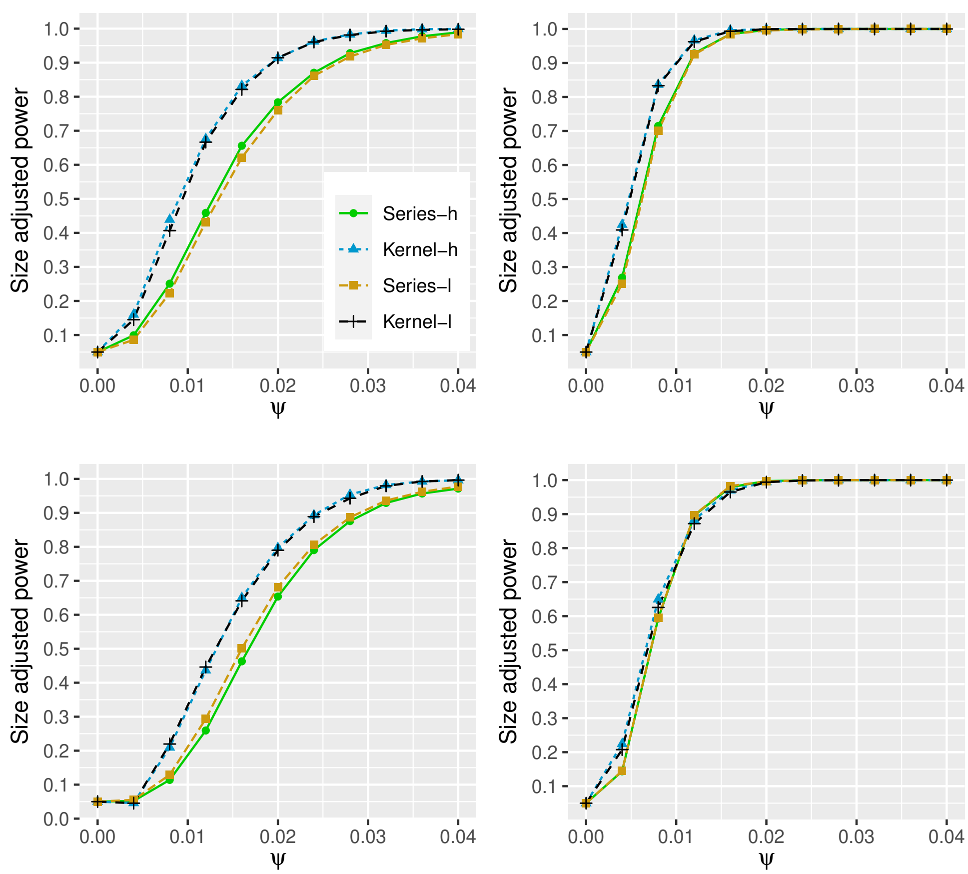

Figure 6: Size-adjusted powers in the nonstationary setting. The left panel corresponds to $T = 30$ and right to $T = 60$. In the upper row, the explanatory variables are exogenous ($\varphi = 0$). In the lower row the explanatory variables are endogenous ($\varphi = 0.75$).

8 Appendix of Proofs

Proof of Lemma 3.1. We have shown that $\sqrt{T}(\hat{\beta}_D - \beta) = \sqrt{T}(\hat{\beta}_C - \beta) + o_p(1)$. But

$$\sqrt{T}(\hat{\beta}_C - \beta) = \left[\frac{1}{T}\int_0^T X_t X_t' dt\right]^{-1}\left[\frac{1}{\sqrt{T}}\int_0^T X_t U_t dt\right] \overset{d}{\to} Q^{-1}\Omega^{1/2}W_d(1),$$

using Assumptions 3.1 and 3.2. Hence $\sqrt{T}(\hat{\beta}_D - \beta) \overset{d}{\to} Q^{-1}\Omega^{1/2}W_d(1)$.

For the second part of the lemma, we use the first part of the lemma to obtain

$$\begin{aligned}
&\frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_i\hat{u}_i \\
&= \frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_i\left[u_i - x_i'\left(\hat{\beta}_D - \beta\right)\right] \\
&= \frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_i u_i + \frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_i x_i' \cdot O_p\left(\frac{1}{\sqrt{T}}\right). \\
&= \frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_i u_i + \frac{1}{n}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_i x_i' \cdot O_p(1).
\end{aligned}$$

For the first term in the above expression, we let

$$S_{XU,i} = \frac{1}{\Lambda(n,\delta)}\sum_{j=1}^{i}x_j u_j \text{ for } i \geq 1 \text{ and } S_{XU,0} = 0.$$

Then, using the twice continuous differentiability of $\phi_j(\cdot)$, we obtain

$$\begin{aligned}
&\frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_i u_i \\
&= \sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)S_{XU,i} - \sum_{i=0}^{n-1}\phi_j\left(\frac{i+1}{n}\right)S_{XU,i} \\
&= \frac{1}{n}\sum_{i=1}^{n-1}\frac{\left[\phi_j(1/n) - \phi_j((i+1)/n)\right]}{1/n}S_{XU,i} + \phi_j(1)S_{XU,n} \\
&= \frac{1}{n}\sum_{i=1}^{n-1}\dot{\phi}_j\left(\frac{1}{n}\right)S_{XU,i} + \phi_j(1)S_{XU,n} + O_p\left(\frac{1}{n}\right)\frac{1}{n}\sum_{i=1}^{n-1}\|S_{XU,i}\| \\
&\overset{d}{\to} -\Omega^{1/2}\int_0^1\dot{\phi}_j(r)W_d(r)dr + \phi_j(1)\Omega^{1/2}W_d(1) \\
&= \Omega^{1/2}\int_0^1\phi_j(r)dW_d(r),
\end{aligned}$$

where the second last equality follows from the continuous mapping theorem and the last equality follows from integration by parts.

For the second term, we let

$$S_{XX,i} = \frac{1}{n}\sum_{j=1}^{i} x_i x_i' = \frac{1}{n}\sum_{j=1}^{i} X_{i\delta}X_{i\delta}' \text{ for } i \geq 1 \text{ and } S_{XX,0} = 0.$$

Also, let

$$\Delta_{XX,i} = S_{XX,i} - Q \cdot \left(\frac{i}{n}\right) \text{ for } i \geq 1 \text{ and } \Delta_{XX,0} = 0.$$

Then, using $\int_0^1 \phi_j(r)\, dr = 0$, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) x_i x_i' \\
&= \sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) \left[S_{XX,i} - S_{XX,(i-1)}\right] \\
&= Q\frac{1}{n}\sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) + \sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) \left(\Delta_{XX,i} - \Delta_{XX,i-1}\right) \\
&= \sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) \left(\Delta_{XX,i} - \Delta_{XX,i-1}\right) + o_p(1).
\end{aligned}$$

For some finite integer $L_j > 0$, let $\cup_{\ell=1}^{L_j} A_{j,\ell}$ be the partition of $[0,1]$ such that $\phi_j(\cdot)$ is monotonic on each interval $A_{j,\ell}$. Then

$$\begin{aligned}
&\left|\sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) \left(\Delta_{XX,i} - \Delta_{XX,i-1}\right)\right| \\
&\leq \left|\sum_{i=1}^{n-1} \left[\phi_j\left(\frac{i}{n}\right) - \phi_j\left(\frac{i+1}{n}\right)\right] \Delta_{XX,i}\right| + \left\|\phi_j(1)\, \Delta_{XX,n}\right\| \\
&= \sum_{i=1}^{n-1} \left|\phi_j\left(\frac{i}{n}\right) - \phi_j\left(\frac{i+1}{n}\right)\right| \left\|\Delta_{XX,i}\right\| + o_p(1) \\
&= \sum_{\ell=1}^{L_j} \sum_{i\in A_{j,\ell}} \left|\phi_j\left(\frac{i}{n}\right) - \phi_j\left(\frac{i+1}{n}\right)\right| \left\|\Delta_{XX,i}\right\| + o_p(1) \\
&\leq \sum_{\ell=1}^{L_j} 2 \max_{i\in A_{j,\ell}} \left|\phi_j\left(\frac{i}{n}\right)\right| \left\|\Delta_{XX,i}\right\| + o_p(1) = o_p(1),
\end{aligned}$$

where the last line holds because, due to the monotonicity, $\sum_{i\in A_{j,\ell}} \left|\phi_j\left(\frac{i}{n}\right) - \phi_j\left(\frac{i+1}{n}\right)\right|$ is equal to the absolute difference of $\phi_j(\cdot)$ evaluated at the two endpoints of $A_{j,\ell}$.

Therefore, $\frac{1}{n}\sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) x_i x_i' = o_p(1)$ and

$$\frac{1}{\Lambda(n,\delta)} \sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) x_i \hat{u}_i \xrightarrow{d} \Omega^{1/2} \int_0^1 \phi_j(r)\, dW_d(r).$$

■

Proof of Theorem 3.2. Note that

$$\begin{aligned}
&\operatorname{var}\left[\operatorname{vec}\left(\Omega^{1/2}\frac{1}{K}\sum_{j=1}^{K}\left[\int_0^1 \phi_j\left(r\right)dW_d\left(r\right)\right]^{\otimes 2}\Omega^{1/2}\right)\right]\\
&=\frac{1}{K^2}\operatorname{var}\left[\left(\Omega^{1/2}\otimes\Omega^{1/2}\right)\operatorname{vec}\left(\sum_{j=1}^{K}\left[\int_0^1 \phi_j\left(r\right)dW_d\left(r\right)\right]^{\otimes 2}\right)\right]\\
&=\frac{1}{K^2}\left(\Omega^{1/2}\otimes\Omega^{1/2}\right)\operatorname{var}\left[\operatorname{vec}\left(\sum_{j=1}^{K}\left[\int_0^1 \phi_j\left(r\right)dW_d\left(r\right)\right]^{\otimes 2}\right)\right]\left(\Omega^{1/2}\otimes\Omega^{1/2}\right)\\
&=\frac{1}{K}\left(\Omega^{1/2}\otimes\Omega^{1/2}\right)\left(\mathbb{I}_{d^2}+\mathbb{K}_{dd}\right)\left(\Omega^{1/2}\otimes\Omega^{1/2}\right)\\
&=\frac{1}{K}\left(\Omega^{1/2}\otimes\Omega^{1/2}\right)\left(\Omega^{1/2}\otimes\Omega^{1/2}\right)\left(\mathbb{I}_{d^2}+\mathbb{K}_{dd}\right)\\
&=\frac{1}{K}\left(\Omega\otimes\Omega\right)\left(\mathbb{I}_{d^2}+\mathbb{K}_{dd}\right).
\end{aligned}$$

Hence, under Assumption 3.4(i), we have

$$\operatorname{var}\left[\operatorname{vec}(\hat{\Omega}^*)\right]=\frac{1}{K}\left(\Omega\otimes\Omega\right)\left(\mathbb{I}_{d^2}+\mathbb{K}_{dd}\right)\left(1+o\left(1\right)\right).$$

To compute the bias of $\hat{\Omega}^*$, we denote $E\left[\left(x_i u_i\right)\left(x_\ell u_\ell\right)'\right]=\Gamma_{xu}\left(i-\ell\right)$. Note that

$$\begin{aligned}
&E\left(\hat{\Omega}^*\right)\\
&=\frac{1}{K}\sum_{j=1}^{K}\left[\frac{1}{\Lambda\left(n,\delta\right)^2}\sum_{i=1}^{n}\sum_{\ell=1}^{n}\phi_j\left(\frac{i}{n}\right)\phi_j\left(\frac{\ell}{n}\right)E\left(x_i u_i\right)\left(x_\ell u_\ell\right)'\right]\\
&=\frac{1}{K}\sum_{j=1}^{K}\frac{1}{\Lambda\left(n,\delta\right)^2}\sum_{i=1}^{n}\sum_{\ell=1}^{n}\phi_j\left(\frac{i}{n}\right)\phi_j\left(\frac{\ell}{n}\right)\Gamma_{xu}\left(i-\ell\right)\\
&=\frac{1}{K}\sum_{j=1}^{K}\frac{1}{\Lambda\left(n,\delta\right)^2}\sum_{i=1}^{n}\sum_{k=i-n}^{i-1}\phi_j\left(\frac{i}{n}\right)\phi_j\left(\frac{i-k}{n}\right)\Gamma_{xu}\left(k\right)\\
&=\frac{1}{K}\sum_{j=1}^{K}\frac{n}{\Lambda\left(n,\delta\right)^2}\sum_{k=-n+1}^{n-1}\left\{\frac{1}{n}\sum_{i=1}^{n}1\left\{\frac{1}{n}\leq\frac{i-k}{n}\leq 1\right\}\phi_j\left(\frac{i}{n}\right)\phi_j\left(\frac{i-k}{n}\right)\right\}\Gamma_{xu}\left(k\right)\\
&=\frac{1}{K}\sum_{j=1}^{K}\delta\sum_{k=-n+1}^{n-1}\omega_{j,n}\left(\frac{k}{n}\right)\Gamma_{xu}\left(k\right)
\end{aligned}$$

where

$$\omega_{j,n}\left(\frac{k}{n}\right)=\frac{1}{n}\sum_{i=1}^{n}1\left\{\frac{1}{n}\leq\frac{i-k}{n}\leq 1\right\}\phi_j\left(\frac{i}{n}\right)\phi_j\left(\frac{i-k}{n}\right).$$

The bias is then equal to

$$
\begin{aligned}
B_n &= E\left(\hat{\Omega}^*\right) - \Omega \\
&= \frac{1}{K}\sum_{j=1}^{K}\delta\sum_{k=-n+1}^{n-1}\left[\omega_{j,n}\left(\frac{k}{n}\right)-1\right]\Gamma_{xu}(k) + \delta\sum_{k=-n+1}^{n-1}\Gamma_{xu}(k) - \Omega \\
&:= B_{1n} + B_{2n}.
\end{aligned}
$$

For B_{2n}, we use Assumptions 3.4(ii) and 3.4(iii) to obtain:

$$
\begin{aligned}
B_{2n} &= \delta\sum_{k=-n+1}^{n-1}\Gamma_{xu}(k) - \Omega = \delta\sum_{k=-n+1}^{n-1}\Gamma_{XU}(k\delta) - \Omega \\
&= \delta\sum_{k=-n+1}^{n-1}\Gamma_{XU}(k\delta) - \int_{-T}^{T}\Gamma_{XU}(\tau)\,d\tau + O\left(\frac{1}{T^2}\right) \\
&= O(\delta) + O\left(\frac{1}{T^2}\right),
\end{aligned}
$$

where the $O\left(T^{-2}\right)$ term holds because

$$
\begin{aligned}
&\left\|\int_{-\infty}^{\infty}\Gamma_{XU}(\tau)\,d\tau - \int_{-T}^{T}\Gamma_{XU}(\tau)\,d\tau\right\| \\
&= \left\|\int_{-\infty}^{\infty}1\{|\tau|\geq T\}\,\Gamma_{XU}(\tau)\,d\tau\right\| \\
&\leq \frac{1}{T^2}\int_{-\infty}^{\infty}\tau^2 1\{|\tau|\geq T\}\,\|\Gamma_{XU}(\tau)\|\,d\tau \\
&\leq \frac{1}{T^2}\int_{-\infty}^{\infty}\tau^2\,\|\Gamma_{XU}(\tau)\|\,d\tau = O\left(\frac{1}{T^2}\right).
\end{aligned}
$$

For B_{1n}, we have, using Assumption 3.4(iv):

$$
\begin{aligned}
\omega_{j,n}(\zeta) &= \frac{1}{n}\sum_{i=1}^{n}1\left\{\frac{1}{n}\leq\frac{i}{n}-\zeta\leq 1\right\}\phi_j\left(\frac{i}{n}\right)\phi_j\left(\frac{i}{n}-\zeta\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}1\left\{\frac{1}{n}+\zeta\leq\frac{i}{n}\leq 1+\zeta\right\}\phi_j\left(\frac{i}{n}\right)\phi_j\left(\frac{i}{n}-\zeta\right) \\
&= \int_{\max(0,\zeta)}^{\min(1+\zeta,1)}\phi_j(r)\,\phi_j(r-\zeta)\,dr + O\left(\frac{j}{n}\right) \\
&:= \omega_j(\zeta) + O\left(\frac{j}{n}\right),
\end{aligned}
$$

uniformly over $j = 1, 2, ..., K$ and $\zeta\in[-1,1]$ where

$$
\omega_j(\zeta) = \int_{\max(0,\zeta)}^{\min(1+\zeta,1)}\phi_j(r)\,\phi_j(r-\zeta)\,dr.
$$

Note that $\omega_j(0) = 1$. Then we have, as $n \to \infty$,

$$
\begin{aligned}
B_{1n} &= \frac{1}{K}\sum_{j=1}^{K}\delta\sum_{k=-n+1}^{n-1}\left[\omega_{j,n}\left(\frac{k}{n}\right)-1\right]\Gamma_{xu}(k)\\
&= \frac{1}{K}\sum_{j=1}^{K}\delta\sum_{k=-n+1}^{n-1}\left[\omega_{j}\left(\frac{k}{n}\right)-1\right]\Gamma_{xu}(k) + \delta\sum_{k=-n+1}^{n-1}\left[\frac{1}{K}\sum_{j=1}^{K}O\left(\frac{j}{n}\right)\right]\Gamma_{xu}(k)\\
&= \frac{1}{K}\sum_{j=1}^{K}\delta\sum_{k=-n+1}^{n-1}\left[\omega_{j}\left(\frac{k}{n}\right)-1\right]\Gamma_{xu}(k) + O\left(\frac{K}{n}\right)\delta\sum_{k=-n+1}^{n-1}\left\|\Gamma_{XU}(k\delta)\right\|\\
&= \frac{1}{K}\sum_{j=1}^{K}\delta\sum_{k=-n+1}^{n-1}\left[\omega_{j}\left(\frac{k}{n}\right)-1\right]\Gamma_{xu}(k) + O\left(\frac{K}{n}\right)\\
&:= \tilde{B}_{1n} + O\left(\frac{K}{n}\right),
\end{aligned}
$$

where

$$
\tilde{B}_{1n} = \frac{1}{K}\sum_{j=1}^{K}\delta\sum_{k=-n+1}^{n-1}\left[\omega_{j}\left(\frac{k}{n}\right)-1\right]\Gamma_{xu}(k).
$$

Now,

$$
\begin{aligned}
\tilde{B}_{1n} &= \frac{1}{K}\sum_{j=1}^{K}\delta\sum_{k=-n+1}^{n-1}\left[\omega_{j}\left(\frac{k}{n}\right)-1\right]\Gamma_{xu}(k)\\
&= \delta\sum_{n/\log n<|k|\le n-1}\left[\frac{1}{K}\sum_{j=1}^{K}\omega_{j}\left(\frac{k}{n}\right)-1\right]\Gamma_{xu}(k) + \delta\sum_{|k|\le n/\log n}\left[\frac{1}{K}\sum_{j=1}^{K}\omega_{j}\left(\frac{k}{n}\right)-1\right]\Gamma_{xu}(k)\\
&= \tilde{B}_{11,n} + \tilde{B}_{12,n}
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{B}_{11,n} &= \delta\sum_{n/\log n<|k|\le n-1}\left[\frac{1}{K}\sum_{j=1}^{K}\omega_{j}\left(\frac{k}{n}\right)-1\right]\Gamma_{xu}(k)\\
&\le \delta\sum_{n/\log n<|k|\le n-1}\left|\frac{1}{K}\sum_{j=1}^{K}\omega_{j}\left(\frac{k}{n}\right)-1\right|\left(\frac{k}{n/\log n}\right)^{2}\left\|\Gamma_{XU}(k\delta)\right\|\\
&= C\left(\frac{\log n}{n}\right)^{2}\frac{1}{\delta^{2}}\left[\delta\sum_{k=-\infty}^{\infty}(k\delta)^{2}\left\|\Gamma_{XU}(k\delta)\right\|\right] = O\left(\frac{(\log n)^{2}}{T^{2}}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{B}_{12,n} &= \frac{1}{K}\sum_{j=1}^{K}\delta\sum_{|k|\le n/\log n}\left[\omega_j\left(\frac{k}{n}\right)-1\right]\Gamma_{xu}(k) \\
&= \frac{1}{K}\sum_{j=1}^{K}\delta\sum_{|k|\le n/\log n}\left[\omega_j\left(\frac{k}{n}\right)-1\right]\Gamma_{XU}(k\delta) \\
&= \frac{1}{K}\sum_{j=1}^{K}\delta\sum_{|k|\le n/\log n}\left[\dot{\omega}_j(0)\frac{k}{n}+\frac{1}{2}\ddot{\omega}_j\left(\frac{\tilde{k}}{n}\right)\left(\frac{k}{n}\right)^2\right]\Gamma_{XU}(k\delta) \\
&= \frac{1}{n\delta}\left[\frac{1}{K}\sum_{j=1}^{K}\dot{\omega}_j(0)\right]\delta\sum_{|k|\le n/\log n}k\delta\Gamma_{XU}(k\delta) \\
&+\left(\frac{1}{n\delta}\right)^2\frac{1}{K}\sum_{j=1}^{K}\delta\sum_{|k|\le n/\log n}\left[\frac{1}{2}\ddot{\omega}_j\left(\frac{\tilde{k}}{n}\right)\right](k\delta)^2\Gamma_{XU}(k\delta) \\
&= \frac{K^2}{T^2}\frac{1}{K^3}\sum_{j=1}^{K}\frac{1}{2}\ddot{\omega}_j(0)\,\delta\sum_{|k|\le n/\log n}(k\delta)^2\Gamma_{XU}(k\delta)(1+o(1))+O\left(\frac{1}{n\delta}\frac{1}{K}\sum_{j=1}^{K}\dot{\omega}_j(0)\right) \\
&= \frac{K^2}{T^2}\left(\frac{1}{K^3}\sum_{j=1}^{K}\frac{1}{2}\ddot{\omega}_j(0)\right)\left(\int_{-\infty}^{\infty}\tau^2\Gamma_{XU}(\tau)\,d\tau\right)(1+o(1))+O\left(\frac{1}{n\delta}\frac{1}{K}\sum_{j=1}^{K}\dot{\omega}_j(0)\right)
\end{aligned}
$$

Given that $\omega_j(\zeta)=\int_\zeta^1\phi_j(r)\,\phi_j(r-\zeta)\,dr$, we have

$$
\begin{aligned}
\dot{\omega}_j(\zeta) &= -\phi_j(\zeta)\,\phi_j(0)-\int_\zeta^1\phi_j(r)\,\dot{\phi}_j(r-\zeta)\,dr, \\
\ddot{\omega}_j(\zeta) &= -\dot{\phi}_j(\zeta)\,\phi_j(0)+\phi_j(\zeta)\,\dot{\phi}_j(0)+\int_\zeta^1\phi_j(r)\,\ddot{\phi}_j(r-\zeta)\,dr=\int_\zeta^1\phi_j(r)\,\ddot{\phi}_j(r-\zeta)\,dr.
\end{aligned}
$$

So

$$
\begin{aligned}
\dot{\omega}_j(0) &= -\phi_j^2(0)-\frac{1}{2}\left[\phi_j^2(1)-\phi_j^2(0)\right]=-\frac{1}{2}\left[\phi_j^2(1)+\phi_j^2(0)\right] \\
\ddot{\omega}_j(0) &= \int_0^1\phi_j(r)\,\ddot{\phi}_j(r)\,dr.
\end{aligned}
$$

Therefore, under Assumption 3.4(iii), we have

$$
\begin{aligned}
\tilde{B}_{11,n} &= \frac{K^2}{T^2}\frac{1}{K^3}\sum_{j=1}^{K}\frac{1}{2}\ddot{\omega}_j(0)\,\delta\sum_{|k|\le n/\log n}(k\delta)^2\Gamma_{XU}(k\delta) \\
&= \frac{K^2}{T^2}\left(\frac{1}{K^3}\sum_{j=1}^{K}\frac{1}{2}\int_0^1\phi_j(r)\,\ddot{\phi}_j(r)\,dr\right)\int_{-\infty}^{\infty}\tau^2\Gamma_{XU}(\tau)\,d\tau\,(1+o(1))+O\left(\frac{1}{T}\right) \\
&= -\frac{K^2}{T^2}c_\phi\int_{-\infty}^{\infty}\tau^2\Gamma_{XU}(\tau)\,d\tau\,(1+o(1))+O\left(\frac{1}{T}\right).
\end{aligned}
$$

Combining the above results yields the asymptotic bias formula. ■

Proof of Lemma 4.1. Consider $n^{-1/2}\mathbb{W}^x\Lambda_T^{-1}$. Let $g_n : \mathbb{D}^d[0,1] \to \mathbb{D}^d[0,1]$ map $f \in \mathbb{D}^d[0,1]$ to the cadlag simple function with $g_n(f)(t-) = f(t)$ at points $t \in \{1/n, 2/n, \ldots, 1\}$ and with jumps only at the points $\{1/n, 2/n \ldots, (n-1)/n\}$. If $f_n \in \mathbb{D}^d[0,1]$ are such that $f_n \to f$ where f is continuous, basic properties of the Skorokhod topology combined with the continuous differentiability in Assumption 3.3 give that $g_n(\phi_j f_n) \to \phi_j f$. Then Assumption 4.1 and the extended continuous mapping theorem (c.f. Theorem 1.11.1 of van der Vaart and Wellner (1996)) yield that $g_n(\phi_j(t)\left(\Lambda_T^{-1}X_{Tt}\right)) \Rightarrow \phi_j(t)X^\circ(t),\ t \in [0,1]$. Combining this with the continuous mapping theorem,

$$\begin{aligned}\frac{1}{\sqrt{n}}\Lambda_T^{-1}\mathbb{W}_j^x &= \frac{1}{n}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)\Lambda_T^{-1}x_i = \frac{1}{n}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)\Lambda_T^{-1}X_{i\delta} \\ &= \frac{1}{n}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)\Lambda_T^{-1}X_{\frac{i}{n}T} = \int_0^1 g_n\left(\phi_j(t)\Lambda_T^{-1}X_{Tt}\right)dt \\ &\xrightarrow{d} \int_0^1 \phi_j\left(r\right)X^\circ(r)dr = \eta_j.\end{aligned}$$

This holds jointly for $j = 1, \ldots, K$ and therefore,

$$\frac{1}{\sqrt{n}}\mathbb{W}^x\Lambda_T^{-1} \xrightarrow{d} \eta. \tag{13}$$

By Assumption 4.2, $\Lambda_T^{-1}X_0 = o_p(1)$ and by Assumption 3.3, $[\phi_j(t+1/n) - \phi_j(t)]/(1/n) = \dot{\phi}_j(t) + o(1)$ uniformly in $t \in [0, 1-1/n]$. Then, similarly to the above, for $\sqrt{\delta}\mathbb{W}_j^{\tilde{\Delta}x}$ we obtain

$$\begin{aligned}\sqrt{\delta}\mathbb{W}_j^{\tilde{\Delta}x} &= \sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)\Lambda_T^{-1}[x_i - x_{i-1}] \\ &= \frac{1}{n}\sum_{i=1}^{n-1}\frac{\left[\phi_j\left(\frac{i}{n}\right) - \phi_j\left(\frac{i+1}{n}\right)\right]}{\frac{1}{n}}\Lambda_T^{-1}x_i + \phi_j(1)\Lambda_T^{-1}x_n - \phi_j\left(\frac{1}{n}\right)\Lambda_T^{-1}x_0 \\ &= \frac{1}{n}\sum_{i=1}^{n-1}\frac{\left[\phi_j\left(\frac{i}{n}\right) - \phi_j\left(\frac{i+1}{n}\right)\right]}{\frac{1}{n}}\Lambda_T^{-1}X_{\frac{i}{n}T} + \phi_j(1)\Lambda_T^{-1}X_{\frac{n}{n}T} + o_p\left(1\right) \\ &\xrightarrow{d} -\int_0^1 \dot{\phi}_j(r)X^\circ(r)dr + \phi_j(1)X^\circ(1) = \int_0^1 \phi_j(r)dX^\circ(r) = \xi_j.\end{aligned}$$

This holds jointly for $j = 1, \ldots, K$ so that

$$\sqrt{\delta}\mathbb{W}^{\tilde{\Delta}x} \xrightarrow{d} \xi. \tag{14}$$

Finally, using standard arguments, we obtain, jointly for $j = 1, \ldots, K$,

$$\begin{aligned}
\sqrt{\delta}\mathbb{W}_j^{0\cdot x} &= \frac{\sqrt{\delta}}{\sqrt{n}} \sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right)[u_{0i} - \theta_0' \tilde{\Delta} x_i] \\
&= \frac{\sqrt{T}}{\sqrt{n\delta}} \sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) \frac{\delta}{\sqrt{T}} U_{0,\frac{i}{n}T} - \theta_0' \sqrt{T} \frac{\sqrt{\delta}}{\sqrt{n}} \frac{1}{\delta} \sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) \Lambda_T^{-1}[x_i - x_{i-1}] \\
&= \sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) \frac{\delta}{\sqrt{T}} U_{0,\frac{i}{n}T} - \theta_0' \sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) \Lambda_T^{-1}[x_i - x_{i-1}] \\
&\xrightarrow{d} \int_0^1 \phi_j(r) dB(r) - \int_0^1 \phi_j(r) \theta_0' dX^\circ(r) \\
&= \sigma_0 \int_0^1 \phi_j(r) dW_0(r) = \nu_j.
\end{aligned}$$

Therefore,

$$\sqrt{\delta}\mathbb{W}^{0\cdot x} \xrightarrow{d} \nu. \tag{15}$$

Assumption 4.2 and the joint convergence of $T^{-1/2} \int_0^{Tt} U_{0s} ds$ and $\Lambda_T^{-1} X_{Tt}$ in Assumption 4.1 yield that (13), (14), and (15) hold jointly, i.e., $\left(n^{-1/2}\mathbb{W}^x \Lambda_T^{-1}, \sqrt{\delta}\mathbb{W}^{\tilde{\Delta} x}, \sqrt{\delta}\mathbb{W}_j^{0\cdot x}\right) \xrightarrow{d} (\eta, \xi, \nu)$. ■

Proof of Theorem 4.1. We write

$$\mathbb{W}^y = \widetilde{\mathbb{W}}\gamma_0 + \mathbb{W}^{0\cdot x} + \alpha_0 \mathbb{W}^\alpha \tag{16}$$

where

$$\mathbb{W}_j^\alpha = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right), \quad \mathbb{W}^\alpha = (\mathbb{W}_1^\alpha, \ldots, \mathbb{W}_K^\alpha)'.$$

Note that for each $j = 1, \ldots, K$ we have

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) &= \sqrt{n} \frac{1}{n} \sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) \\
&= \sqrt{n}\left(\int_0^1 \phi_j(r)\, dr + O\left(\frac{1}{n}\right)\right) = O\left(\frac{1}{\sqrt{n}}\right) = o(1).
\end{aligned}$$

Therefore, combined with (7), (16) yields

$$\Upsilon_T(\hat{\gamma} - \gamma_0) = \left(\sqrt{\delta}\Upsilon_T^{-1}\widetilde{\mathbb{W}}'\widetilde{\mathbb{W}}\Upsilon_T^{-1}\sqrt{\delta}\right)^{-1} \sqrt{\delta}\Upsilon_T^{-1}\widetilde{\mathbb{W}}'\sqrt{\delta}\left[\mathbb{W}^{0\cdot x} + o(1)\right]. \tag{17}$$

By Lemma 4.1, recalling $\zeta = (\eta, \xi)$,

$$\left(\widetilde{\mathbb{W}}\Upsilon_T^{-1}\sqrt{\delta}, \sqrt{\delta}\mathbb{W}^{0\cdot x}\right) \xrightarrow{d} (\zeta, \nu), \tag{18}$$

where ζ and ν are independent as W_0 and X° are independent. From (17), (18), and the continuous mapping theorem,

$$\Upsilon_T(\hat{\gamma} - \gamma_0) \xrightarrow{d} (\zeta'\zeta)^{-1}\zeta'\nu \overset{d}{=} MN\left[0, \sigma_0^2 (\zeta'\zeta)^{-1}\right],$$

a mixed normal distribution.

The limiting distribution is the same as that in Theorem 1 of Hwang and Sun (2018). From (7), taking a block inverse and simplifying, it is straightforward to show that

$$\hat{\beta}_{TAOLS} - \beta_0 = \left[\mathbb{W}^{x\prime} M_{\tilde{\Delta}x} \mathbb{W}^{x}\right]^{-1} \mathbb{W}^{x\prime} M_{\tilde{\Delta}x} \mathbb{W}^{0\cdot x}.$$

Then, utilizing Lemma 4.1 and similar steps as above, we obtain that

$$\sqrt{T}\Lambda_T \left(\hat{\beta}_{TAOLS} - \beta_0\right) \xrightarrow{d} \left(\eta' M_\xi \eta\right)^{-1} \eta' M_\xi \nu,$$

which is identical to the decomposition Hwang and Sun (2018) obtains directly from $\left(\zeta'\zeta\right)^{-1}\zeta'\nu$. Hence, under H_0,

$$\begin{aligned} \sqrt{T}\tilde{\Lambda}_T \left(R\hat{\beta}_{TAOLS} - r\right) &= \tilde{\Lambda}_T R \Lambda_T^{-1} \sqrt{T}\Lambda_T (\hat{\beta}_{TAOLS} - \beta_0) \\ &\xrightarrow{d} A \left(\eta' M_\xi \eta\right)^{-1} \eta' M_\xi \nu. \end{aligned} \tag{19}$$

Now, by (7) and (16),

$$\hat{\mathbb{W}}^{0\cdot x} = \mathbb{W}^{y} - \widetilde{\mathbb{W}}\hat{\gamma} = \left[I_K - \widetilde{\mathbb{W}}\left(\widetilde{\mathbb{W}}'\widetilde{\mathbb{W}}\right)^{-1}\widetilde{\mathbb{W}}'\right]\mathbb{W}^{0\cdot x} + o(1).$$

Combining this with Lemma 4.1, the definition in (9) and denoting $M_\zeta = I_K - \zeta(\zeta'\zeta)^{-1}\zeta'$,

$$\begin{aligned} \delta \cdot \hat{\sigma}^2_{0\cdot x} &= \frac{1}{K}\sqrt{\delta}\mathbb{W}^{0\cdot x\prime}\left[I_K - \widetilde{\mathbb{W}}\left(\widetilde{\mathbb{W}}'\widetilde{\mathbb{W}}\right)^{-1}\widetilde{\mathbb{W}}'\right]\sqrt{\delta}\mathbb{W}^{0\cdot x} + o_p(1) \\ &= \frac{1}{K}\sqrt{\delta}\mathbb{W}^{0\cdot x\prime}\left\{I_K - \widetilde{\mathbb{W}}\Upsilon_T^{-1}\left[\left(\widetilde{\mathbb{W}}\Upsilon_T^{-1}\right)'\widetilde{\mathbb{W}}\Upsilon_T^{-1}\right]^{-1}\left(\widetilde{\mathbb{W}}\Upsilon_T^{-1}\right)'\right\}\sqrt{\delta}\mathbb{W}^{0\cdot x} + o_p(1) \\ &\xrightarrow{d} \frac{1}{K}\nu' M_\zeta \nu. \end{aligned} \tag{20}$$

Additionally,

$$\begin{aligned} &\frac{1}{\delta}\sqrt{T}\tilde{\Lambda}_T R\left\{\mathbb{W}^{x\prime} M_{\Delta x}\mathbb{W}^{x}\right\}^{-1} R'\tilde{\Lambda}_T\sqrt{T} \\ &= \tilde{\Lambda}_T R \Lambda_T^{-1}\left\{\frac{\sqrt{\delta}}{\sqrt{T}}\Lambda_T^{-1}\mathbb{W}^{x\prime}\left[I_K - \mathbb{W}^{\tilde{\Delta}x}(\mathbb{W}^{\tilde{\Delta}x\prime}\mathbb{W}^{\tilde{\Delta}x})^{-1}\mathbb{W}^{\tilde{\Delta}x\prime}\right]\frac{\sqrt{\delta}}{\sqrt{T}}\mathbb{W}^{x}\Lambda_T^{-1}\right\}^{-1}\Lambda_T^{-1}R'\tilde{\Lambda}_T \\ &\xrightarrow{d} A\left(\eta' M_\xi \eta\right)^{-1} A'. \end{aligned} \tag{21}$$

Via the joint convergence in (18), the analysis producing (19), (20), and (21), and the continuous mapping theorem, we obtain that under H_0,

$$\begin{aligned}
F_{TAOLS} &= \frac{1}{\hat{\sigma}^2_{0\cdot x}}\left(R\hat{\beta}_{TAOLS}-r\right)'\left[R\left(\mathbb{W}^{x\prime}M_{\tilde{\Delta}x}\mathbb{W}^x\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{TAOLS}-r\right)/p \\
&= \frac{1}{p}\frac{1}{\delta\hat{\sigma}^2_{0\cdot x}}\left(R\hat{\beta}_{TAOLS}-r\right)'\sqrt{T}\tilde{\Lambda}_T \\
&\qquad \times\left[\frac{1}{\delta}\sqrt{T}\tilde{\Lambda}_T R\left(\mathbb{W}^{x\prime}M_{\tilde{\Delta}x}\mathbb{W}^x\right)^{-1}R'\tilde{\Lambda}_T\sqrt{T}\right]^{-1}\sqrt{T}\tilde{\Lambda}_T\left(R\hat{\beta}_{TAOLS}-r\right) \\
&\overset{d}{\to} \frac{K}{p}\frac{\left[A(\eta'M_\xi\eta)^{-1}\eta'M_\xi\nu\right]'\left(A\left(\eta'M_\xi\eta\right)^{-1}A'\right)^{-1}\left[A(\eta'M_\xi\eta)^{-1}\eta'M_\xi\nu\right]}{\nu'M_\zeta\nu} \\
&= \frac{K}{p}\frac{Q'\left(A\left(\eta'M_\xi\eta\right)^{-1}A'\right)^{-1}Q}{\nu'M_\zeta\nu/\sigma_0^2}, \qquad (22)
\end{aligned}$$

where $Q = A(\eta'M_\xi\eta)^{-1}\eta'M_\xi\nu/\sigma_0$. Now, conditional on $\zeta = (\eta,\xi)$,

$$Q'\left(A\left(\eta'M_\xi\eta\right)^{-1}A'\right)^{-1}Q \overset{d}{=} \chi^2_p, \text{ and } \nu'M_\zeta\nu/\sigma^2_{0\cdot x} \overset{d}{=} \chi^2_{K-2d}.$$

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

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

Thus, conditional on ζ, the numerator and the denominator in (22) are independent chi-squared variates. This implies that

$$\frac{K}{p}\frac{Q'\left(R\left(\eta'M_\xi\eta\right)^{-1}R'\right)^{-1}Q}{\nu'M_\zeta\nu/\sigma_0^2} = \frac{K}{K-2d}\frac{Q'\left(R\left(\eta'M_\xi\eta\right)^{-1}R'\right)^{-1}Q/p}{\nu'M_\zeta\nu/\left[\sigma_0^2(K-2d)\right]} \overset{d}{=} \frac{K}{K-2d}F_{p,K-2d}$$

conditional on ζ. But the conditional distribution does not depend on the conditioning variable ζ, so it is also the unconditional distribution. This proves the second and final statement of the theorem. ■

Proof of Lemma 6.1. First, Assumptions 6.1 (ii) and (iii), imply (see, for example, Lemma 1, p. 166 of Billingsley (1968))

$$|\mathrm{cov}(\epsilon_t,\epsilon_{t+\ell})| \le 2\varphi_\ell^{1/2}\mathrm{var}(\epsilon_1).$$

Let C denote a constant greater than the absolute value of each component of $\Gamma_X(\tau)$ for all $\tau \ge 0$. It is sufficient to show that each coordinate of $\sum_{i=1}^n x_i\epsilon_i$ is $O_p(\sqrt{n})$. So without loss of generality, we can assume that $x_i \in \mathbb{R}$. We have

$$E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i\epsilon_i\right]^2 = E\left[\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2 x_i^2 + \frac{2}{n}\sum_{j=1}^{n-1}\sum_{i=1}^{n-j}\epsilon_i\epsilon_{i+j}x_i x_{i+j}\right]$$

$$\leq 2\text{var}(\epsilon_1)\left(\frac{1}{n}\sum_{i=1}^{n} E\left[X_{i\delta}^2\right] + \frac{2}{n}\sum_{j=1}^{n-1}\sum_{i=1}^{n-j}\varphi_j^{1/2}\left|E\left[X_{i\delta}X_{(i+j)\delta}\right]\right|\right)$$

$$\leq 2C\text{var}(\epsilon_1)\left(1 + 2\sum_{j=1}^{\infty}\varphi_j^{1/2}\right) < \infty.$$

Then $\sum_{i=1}^{n} x_i\epsilon_i = O_p(\sqrt{n})$ follows by Markov's inequality. ■

References

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

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2008). Ultra high frequency volatility estimation with dependent microstructure noise. Working paper, Department of Economics, Princeton University.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons.

Bilodeau, M. and Brenner, D. (2010). *Theory of Multivariate Statistics*. Springer-Verlag New York.

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.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

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. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21(6):1130—1164.

Lazarus, E., Lewis, D. J., Stock, J. H., and Watson, M. W. (2018). HAR inference: Recommendations for practice. *Journal of Business & Economic Statistics*, 36(4):541–559.

Lu, Y. and Park, J. Y. (2019). Estimation of longrun variance of continuous time stochastic process using discrete sample. *Journal of Econometrics*, 210(2):236–267.

Newey, W. K. and West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61(4):631–653.

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

Sun, Y. (2011). Robust trend inference with series variance estimator and testing-optimal smoothing parameter. *Journal of Econometrics*, 164(2):345–366.

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

Sun, Y. (2014a). Fixed-smoothing asymptotics in a two-step generalized method of moments framework. *Econometrica*, 82(6):2327–2370.

Sun, Y. (2014b). 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., Phillips, P. C. B., and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing. *Econometrica*, 76(1):175–194.

van der Vaart, A. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer Series in Statistics. Springer.

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.