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 2022

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 exogenous regressors that are general semimartingales. Endogeneous regressors are allowed in a nonstationary environment similar to cointegration models in the usual discrete-time setting. The F tests can be implemented in exactly the same way as in the 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 how we can automate inference procedures so that a researcher can make fewer technical and theoretical modeling decisions. At what sampling frequency should a researcher consider moving to an explicitly continuous-time framework? Should a researcher convert a

*Email: dpellatt@ucsd.edu; yisun@ucsd.edu. Address correspondence to: Department of Economics, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA. For very helpful comments, we thank Xiaohong Chen, an associate editor, and two anonymous referees.

high-frequency sample into a lower-frequency sample before conducting regression analysis in a discrete-time framework? If continuous-time modeling requires accounting for the sampling frequency, what measurement constitutes a single unit of time? An hour, a day, or a month? Designing trustworthy inference procedures in realistic sample sizes is also a concern.

In this paper we propose statistical tests 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 a 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 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.

[1]Typical examples of low-frequency discrete-time data include 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 frequency in the frequency domain that measures the speed that a process completes a cycle.

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 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 with exogenous regressors, 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 measurement noise that is of discrete-time nature. In both settings, the measurement noise is dominated by the continuous-time error component, and hence our asymptotic theory remains valid. Simulations show that moderate measurement noise has virtually no effect on the finite sample performance of our F test.

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 further justification for 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 consider the case where the regressors are stationary and consider a data-driven approach to selecting K. In Section 3 we consider the nonstationary case with cointegration. Section 4 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 The Case with Stationary Regressors

2.1 The basic setting

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 assume that U_t is stationary and $E(U_t|X_s, s \in [0, T]) = 0$ for any $t \in [0, T]$. In this section, we also assume that X_t is a stationary process and defer the case with a nonstationary X_t to Section 3. An intercept can be included in X_t in this section.

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

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

for $i = 1, \ldots, n$ and $n = T/\delta$. Here, for notational simplicity, we have assumed that T/δ is an integer. The discrete-time sample $\{(x_i, y_i)\}_{i=1}^n$ satisfies

$$y_i = x_i'\beta_0 + u_i, i = 1, 2, \ldots, n,$$

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 p.

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

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

Our test of H_0 against H_1 is based on the above estimator.

2.2 The test statistic

To test whether $R\beta_0$ is equal to r, 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

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

Based on this *approximate* variance formula, we construct the test statistic

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

where $\hat{u}_i = y_i - x_i'\hat{\beta}_D$ and $\widehat{\text{var}}(\sum_{i=1}^n x_i\hat{u}_i)$ is an estimator of the *approximate* variance of $\sum_{i=1}^n x_i u_i$. In the above, dividing by p does not affect the properties of the test.

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. When the basis functions can be paired naturally, we shall assume that K is even. Note that the basis functions are evaluated at i/n instead of i/T. This is an important point, and our asymptotic theory relies crucially on this construction. We have, therefore, effectively ignored the high-frequency nature of our time series observations that are sampled from continuous time processes. The test statistic is then

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

The form of the test statistic is exactly the same as what we would use for a standard regression with discrete time series. Importantly, there is no rescaling by n or T. 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.

2.3 Assumptions for the fixed-smoothing asymptotics

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 a 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 $T \to \infty$.

The fixed-smoothing asymptotics are developed under several assumptions. First and foremost, for any process $Z = \{Z_t : t \in [0, T]\}$ in this section, we assume that it can be decomposed into a continuous part and a pure-jump part:

$$Z_t = Z_t^c + Z_t^d$$

where $Z_t^d = \sum_{0 \le \tau \le t} \Delta Z_\tau$, $\Delta Z_\tau = Z_\tau - Z_{\tau-}$ and $Z_{\tau-} = \lim_{t \to \tau-} Z_t$. That is, we assume that $\{Z_t\}$ is the sum of a continuous local martingale (i.e., Z_t^c) and a sum of jump terms (i.e., Z_t^d).

Next, we present other technical assumptions and provide some discussion on each.

Assumption 2.1 *For $Z_t = X_t U_t$, $X_t' X_t$,*

$$\sum_{0 \le \tau \le T} E \|\Delta Z_\tau\| = O(T) \text{ as } T \to \infty,$$

where for a matrix M, $\|M\|$ is the Frobenius norm of M.

Assumption 2.1 is the same as the first part of Assumption A of Chang et al. (2018). It imposes a restriction on the number and sizes of the jumps in $\{Z_t\}$. The assumption is not stringent and is satisfied, for example, for processes with compound Poisson type jumps if the jump sizes are bounded in L_1 and the jump intensity is proportional to T.

Assumption 2.2 *For $j = 1, \ldots, K$, each function $\phi_j(\cdot)$ is twice continuously differentiable, and $\int_0^1 \phi_j(t)dt = 0$. Also, $\{\phi_j(\cdot)\}_{j=1}^K$ form an orthonormal set in $L^2[0, 1]$.*

Assumption 2.2 is very mild and is often maintained in the literature on orthonormal series variance estimation; see, for example, Assumption 1(b) in Sun (2014a). The Fourier basis functions given later in (2) satisfy this assumption. For ease of presentation, we set $\phi_0(\cdot) \equiv 1$, the constant function.

Lemma 2.1 *Let Assumptions 2.1 and 2.2 hold. For $Z_t = X_t U_t$, $X_t' X_t$ and $z_i = Z_{i\delta}$,*

$$\frac{1}{n}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)z_i = \frac{1}{T}\int_0^T \phi_j\left(\frac{t}{T}\right)Z_t dt + O_p\left(e_{\delta,T}(Z)\right),\ j = 0, 1, \ldots, K$$

as $\delta \to 0$ and $T \to \infty$,[2] *where*

$$e_{\delta,T}(Z) = \Delta_{\delta,T}(Z) + \frac{\delta}{T}\sup_{t\in[0,T]}\|Z_t\| + \delta$$

and

$$\Delta_{\delta,T}(Z) = \sup_{\tau,t\in[0,T]}\ \sup_{|\tau-t|\le\delta}\|Z_\tau^c - Z_t^c\|$$

is the modulus of continuity of the continuous part of Z.

Lemma 2.1 shows that the discrete-time average is an approximation to the continuous-time integral with the approximation error controlled by the modulus of continuity of Z, a technical term $\delta\sup_{t\in[0,T]}\|Z_t\|/T$ that captures the edge effects, and the sampling interval δ. In the proof of Lemma 2.1, we show that under Assumption 2.1, the effect of jumps on the approximation error is of order $O_p(\delta)$.

Assumption 2.3 *For $\{\phi_j\}$ satisfying Assumption 2.2,*

$$\frac{1}{T}\int_0^T \phi_j\left(\frac{t}{T}\right)X_t X_t' dt = o_p(1) \text{ for } j = 1, \ldots, K,$$

and

$$\frac{1}{T}\int_0^T X_t X_t' dt = S + o_p(1)$$

for a positive definite matrix S as $T \to \infty$.

To understand the assumption, let X_{tk} be the k-th element of X_t. Suppose X_t is stationary and $E|X_{tk}X_{tl}X_{sk}X_{sl}| < \infty$ for any $k, l = 1, 2, \ldots, d$ and any $t, s \in [0, T]$. Assume further that $cov(X_{tk}X_{tl}, X_{sk}X_{sl}) = f_{kl}(t-s)$ for some bounded function $f_{kl}(\cdot)$ satisfying $f_{kl}(\tau) \to 0$ as $|\tau| \to \infty$. Then, by the Fubini–Tonelli theorem,

$$E\frac{1}{T}\int_0^T \phi_j\left(\frac{t}{T}\right)X_t X_t' dt = E\left(X_t X_t'\right)\cdot\frac{1}{T}\int_0^T \phi_j\left(\frac{t}{T}\right)dt = E\left(X_t X_t'\right)\int_0^1 \phi_j(r)\,dr,$$

[2] This should be understood in the following way: $\left\|n^{-1}\sum_{i=1}^n \phi_j(i/n)\,z_i - T^{-1}\int_0^T \phi_j(t/T)\,Z_t dt\right\| = O_p\left(e_{\delta,T}(Z)\right)$. We use the same convention when O_p or o_p is used in matrix equalities.

for all $j = 0, 1, \ldots, K$. By the Fubini–Tonelli theorem and the dominated convergence theorem,

$$
\begin{aligned}
& var\left(\frac{1}{T}\int_0^T \phi_j\left(\frac{t}{T}\right) X_{ti}X_{tk}dt\right) \\
&= \frac{1}{T^2}\int_0^T\int_0^T \phi_j\left(\frac{t}{T}\right)\phi_j\left(\frac{s}{T}\right) cov(X_{ti}X_{tk}, X_{si}X_{sk})dtds \\
&= \frac{1}{T^2}\int_0^T\int_0^T \phi_j\left(\frac{t}{T}\right)\phi_j\left(\frac{s}{T}\right) f_{kl}\left(t-s\right)dtds \\
&= \int_0^1\int_0^1 \phi_j\left(s\right)\phi_j\left(t\right) f_{kl}\left(T\left(t-s\right)\right)dtds \to 0
\end{aligned}
$$

for $j = 0, 1, \ldots, K$. Hence Assumption 2.3 holds for $S = E\left(X_t X_t'\right)$.

Assumption 2.4 *For $\{\phi_j\}$ satisfying Assumption 2.2,*

$$\frac{1}{\sqrt{T}}\int_0^T \phi_j\left(\frac{t}{T}\right) X_t U_t dt \Rightarrow \Omega^{1/2}\int_0^1 \phi_j\left(r\right)dW_d\left(r\right) \text{ jointly for } j = 0, 1, 2, \ldots, K$$

as $\delta \to 0$ and $T \to \infty$, where $W_d\left(r\right)$ is the $d \times 1$ standard Brownian motion process,

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

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

Assumption 2.4 is a multivariate CLT in the continuous-time setting. As in the discrete time setting, there is a large body of literature on CLT's for additive functionals in a continuous time setting. For example, Rozanov (1960) establishes a CLT for additive functionals such as $T^{-1/2}\int_0^T \phi_j\left(t/T\right) X_t U_t dt$. The sufficient conditions, which include a mixing condition and a moment condition, are similar to those in the discrete time setting.

If a functional CLT (FCLT) holds such that $T^{-1/2}\int_0^{[Tr]} X_t U_t dt \Rightarrow \Omega^{1/2} W_d\left(r\right)$, then using integration by parts and the continuous mapping theorem, we can show that Assumption 2.4 holds. Sufficient conditions for the FCLT for the class of functions of continuous-time stationary ergodic Markov processes can be founded in Bhattacharya (1982). For more discussions, see Equations 1–3 and remarks in Section 2 of Lu and Park (2019). Note that an FCLT is stronger than necessary, but the gap between an FCLT and the above multivariate CLT may be of theoretical interest only. Here we only need a multivariate CLT. This is an advantage of using a series LRV estimator. If we use a kernel LRV estimator, then an FCLT is needed for developing fixed-smoothing asymptotics.

Assumption 2.5 *(i)* $\sqrt{T}e_{\delta,T}\left(XU\right) = o_p\left(1\right)$ *and (ii)* $e_{\delta,T}\left(XX'\right) = o_p\left(1\right)$.

Assumption 2.5 is the same as Assumption D1 of Chang et al. (2018). Assumption 2.5(i) holds if $\sqrt{T}\delta = o\left(1\right)$, $\sqrt{T}\Delta_{\delta,T}\left(XU\right) = o_p\left(1\right)$ and $\sup_{t\in[0,T]}\|XU\| = o_p(\sqrt{T}/\delta)$. The first condition, namely $\sqrt{T}\delta = o\left(1\right)$, requires that $\delta \to 0$ fast enough as $T \to \infty$, that is, the continuous-time process has to be sampled frequently enough. The second condition, namely $\sqrt{T}\Delta_{\delta,T}\left(XU\right) = o_p\left(1\right)$, requires that the continuous part of $\{X_t U_t\}$ does not fluctuate too much over the sampling intervals of length δ. Using the moment bounds in Fischer and Nappo (2009) and the Markov inequality, we can obtain that

$$\Delta_{\delta,T}(XU) = O_p\left[\left(\delta \log \frac{2T}{\delta}\right)^{1/2}\right]$$

if $(X_tU_t)^c$ is an Ito process whose drift and diffusion coefficients satisfy some mild conditions. So $\sqrt{T}\Delta_{\delta,T}(XU) = o_p(1)$ if $T\delta\log(T/\delta) = o(1)$. The third condition, namely $\sup_{t\in[0,T]}\|XU\| = o_p(\sqrt{T}/\delta)$, requires that the maximum value of the process $\{X_tU_t\}$ over $[0,T]$ does not explode too quickly as T grows. For example, if $\sup_{t\in[0,T]}\|XU\| = O_p(T)$ and $\sqrt{T}\delta = o(1)$, then $\sup_{t\in[0,T]}\|XU\| = O_p(T) = O_p(\sqrt{T}\delta\cdot\sqrt{T}/\delta) = o_p(\sqrt{T}/\delta)$ and the third condition holds. Assumption 2.5(ii) is of the same form as Assumption 2.5(i). With some obvious modifications, our discussions on Assumption 2.5(i) can be applied to Assumption 2.5(ii).

2.4 Fixed-smoothing asymptotics

Define

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

which is the least-square analogue of $\hat{\beta}_D$ in the space $L^2[0,T]$ using the continuous-time 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.

We first show that $\sqrt{T}[\hat{\beta}_D-\beta]$ and $\sqrt{T}[\hat{\beta}_C-\beta]$ are asymptotically equivalent. Letting $Z_t = X_tU_t$ and $j=0$ in Lemma 2.1, we have

$$\frac{1}{n}\sum_{i=1}^n x_iu_i = \frac{1}{T}\int_0^T X_tU_tdt + O_p\left(e_{\delta,T}(XU)\right).$$

Multiplying the above equation by $\sqrt{T}$, we obtain

$$\frac{1}{\Lambda(n,\delta)}\sum_{i=1}^n x_iu_i = \frac{1}{\sqrt{T}}\int_0^T X_tU_tdt + o_p(1),$$

where $\Lambda(n,\delta) = \sqrt{n/\delta}$ and we have used Assumption 2.5(i).

Using Lemma 2.1 with $Z_t = X_tX_t'$ and $j=0$ and Assumption 2.5(ii), we have

$$\frac{1}{n}\sum_{i=1}^n x_ix_i' = \frac{1}{T}\int_0^T X_tX_t'dt + o_p(1).$$

Hence,

$$\begin{aligned}
\sqrt{T}\left[\hat{\beta}_D-\beta_0\right] &= [n/\Lambda(n,\delta)]\left[\hat{\beta}_D-\beta_0\right] \\
&= \left(\frac{1}{n}\sum_{i=1}^n x_ix_i'\right)^{-1}\left(\frac{1}{\Lambda(n,\delta)}\sum_{i=1}^n x_iu_i\right) \\
&= \left(\frac{1}{T}\int_0^T X_tX_t'dt\right)^{-1}\frac{1}{\sqrt{T}}\int_0^T X_tU_tdt + o_p(1) \\
&= \sqrt{T}(\hat{\beta}_C-\beta_0) + o_p(1).
\end{aligned}$$

The above derivations show that Assumptions 2.1 and 2.5 ensure that $\sqrt{T}(\hat{\beta}_D - \beta_0)$ and $\sqrt{T}(\hat{\beta}_C - \beta_0)$ are asymptotically equivalent. Invoking Assumptions 2.3 and 2.4, we obtain the asymptotic distribution of $\sqrt{T}(\hat{\beta}_D - \beta)$. We present this and another key result, which requires Assumption 2.2, in the lemma below.

Lemma 2.2 *Let Assumptions 2.1–2.5 hold. Then*

$$\sqrt{T}(\hat{\beta}_D - \beta_0) = \sqrt{T}(\hat{\beta}_C - \beta_0) + o_p(1) \Rightarrow S^{-1}\Omega^{1/2}W_d(1)$$

and

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

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

Lemma 2.2 shows that $\hat{\beta}_D$ converges to β_0 at the rate of $\sqrt{T}$. 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. We do not obtain the rate of $\sqrt{n}$, which is the typical rate for the discrete-time data with a fixed sampling interval (e.g., δ is fixed to be 1) and n weakly dependent observations. The difference can be traced back to the unusual rate in the weak convergence result:

$$\frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}x_iu_i \Rightarrow \Omega^{1/2}W_d(1).$$

Because $\{x_iu_i\}$ becomes highly correlated as $\delta \to 0$, in order to obtain a well-defined weak limit, we need to normalize the sum $\sum_{i=1}^{n}x_iu_i$ by $\Lambda(n,\delta) := \sqrt{n/\delta}$, which is larger than the usual normalization factor $\sqrt{n}$ by an order of magnitude.

Using Lemma 2.2, we have, under the null hypothesis:

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

In the above, rescalings by $\delta\Lambda(n,\delta)$, $1/\Lambda(n,\delta)$ or $1/(\delta\Lambda(n,\delta)^2)$ in the first equality are for theoretical arguments only. In practice, the test statistic F_T is computed according to the definition in (1) without using any rescaling.

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

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

Under Assumption 2.2, $\left[\int_0^1 \phi_j(r)\, dW_p(r)\right]^{\otimes 2}$ 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 2.1 *Let Assumptions 2.1 – 2.5 hold. Then, for a fixed $K \geq p$,*

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

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

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

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. Such a result is also related to the following result in Sun (2004): the t-statistic can be made convergent in a spurious regression when high-order autocorrelations are properly accounted for.

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

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

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-order autocorrelations 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 $T \to \infty$.

2.5 Choice of the smoothing parameter K

In this subsection, we propose a rule for choosing K. Part of our theoretical analysis is the high-frequency continuous-time counterparts of Phillips (2005), which develops a rule for choosing 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).

To abstract away the technical issues that will not affect the practical implementation of the proposed rule, 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 2.6 *The following hold:*

(i) $var\left[\text{vec}(\hat{\Omega}^*)\right] = var\left[\text{vec}\left(\Omega^{1/2}\frac{1}{K}\sum_{j=1}^{K}\left[\int_0^1\phi_j(r)\,dW_d(r)\right]^{\otimes 2}\Omega^{1/2}\right)\right](1+o(1))$ *as* $T\to\infty$ *for both a fixed* K *and a growing* K *(i.e.,* $K\to\infty$*).*

(ii) Let $\Gamma_{XU}(\tau) = E\left(X_t U_t U_{t-\tau} X'_{t-\tau}\right)$*. For some* $\iota > 0$*, there exists positive constants* C_1 *and* C_2 *such that*

$$\|\Gamma_{XU}(\tau)\| \le C_1 \textit{ for all } \tau \textit{ and } \|\Gamma_{XU}(\tau)\| \le C_1\tau^{-(3+\iota)} \textit{ for all } |\tau| \ge C_2.$$

(iii) $\delta\sum_{k=-n+1}^{n-1}(k\delta)^m\,\Gamma_{XU}(k\delta) - \int_{-T}^{T}\tau^m\Gamma_{XU}(\tau)\,d\tau = O(\delta)$ *for* $m = 0, 2$.

(iv) For some constant $C > 0$, $\sup_{j\in[K]}\sup_{r\in[0,1]}\max\left\{|\phi_j(r)|, |\dot{\phi}_j(r)|/j\right\} \le C$ *where* $\dot{\phi}_j$ *is the first order derivative of* ϕ_j *and* $[K] := \{1,\ldots,K\}$.

(v) If $K\to\infty$ *as* $T\to\infty$*, then, for some constant* $c_{\phi,2}\neq 0$,

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

where $\ddot{\phi}_j$ *is the second order derivative of* ϕ_j.

Assumption 2.6(i) is a high-level assumption. When K is fixed and Assumptions 2.1–2.5 hold,

$$\hat{\Omega}^* \Rightarrow \Omega^{1/2}\frac{1}{K}\sum_{j=1}^{K}\left[\int_0^1\phi_j(r)\,dW_d(r)\right]^{\otimes 2}\Omega^{1/2}.$$

So Assumption 2.6(i) says 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. For both a fixed K and a growing K, we can show that an assumption similar to Assumption 2.3(b) in Lu and Park (2019), Assumption 2.2, and Assumptions 2.6(ii)-(iv) are sufficient for Assumption 2.6(i). The details and proof are given in the online supplementary appendix.

Assumption 2.6(ii) imposes that the covariance $\|\Gamma_{XU}(\tau)\|$ is bounded above and decays to zero at a certain rate. The assumption ensures that $\delta\sum_{k=-\infty}^{\infty}(k\delta)^2\|\Gamma_{XU}(k\delta)\| < \infty$ and $\int_{-\infty}^{\infty}\tau^2\|\Gamma_{XU}(\tau)\| < \infty$ (see the proof Theorem 2.2). The summability condition can be regarded as the continuous counterpart of the integrability condition. These conditions are often imposed directly in the literature. For the latter condition, see, for example, Assumption 2.2 in Lu and Park (2019) (pp. 239).

Assumption 2.6(iii) assumes that the discrete sum is a good approximation to the integral. Note that

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

Therefore, Assumption 2.6(iii) holds if $\delta\sum_{k=-n+1}^{n-1}\max_{t\in[k\delta,(k+1)\delta]}\left\|\frac{\partial[t^m\Gamma_{XU}(t)]}{\partial t}\right\| < \infty$.

Assumptions 2.6(iv) and (v) contain some additional mild conditions on the basis functions. The assumptions are satisfied for the sine and cosine basis functions (i.e., Fourier bases) given by

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

For the above set of Fourier bases, we have

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

and hence

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

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" $c_{k,q}$ defined by

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

The parameter $c_{\phi,2}$ in Assumption 2.6(v) plays the same role in series LRV estimation as $c_{k,q}$ 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) discusses why the set of Legendre polynomials may not be a good choice. We focus on second-order series LRV estimators in this paper.

Theorem 2.2 *Let Assumption 2.6 hold.*

(*a*) *Under Assumption 2.6(i), as $T \to \infty$, 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.

(*b*) *Under Assumptions 2.6(ii)–(v), as $T \to \infty$ and $K \to \infty$, the bias of $\hat{\Omega}^*$ satisfies*

$$E(\hat{\Omega}^* - \Omega) = -c_{\phi,2}\frac{K^2}{T^2}B_2 + o\left(\frac{K^2}{T^2}\right) + O\left(\delta + \frac{(\log n)^2}{T^2} + \frac{1}{T}\right),$$

where

$$B_2 = \int_{-\infty}^{\infty} \tau^2 \Gamma_{XU}\left(\tau\right) d\tau.$$

(*c*) *Under Assumptions 2.6(ii)–(iv), as $T \to \infty$ for a fixed K, the bias of $\hat{\Omega}^*$ satisfies*

$$E(\hat{\Omega}^* - \Omega) = -\frac{1}{T}c_{\phi,1}B_1 + o\left(\frac{1}{T}\right) + O\left(\delta + \frac{(\log n)^2}{T^2} + \frac{1}{n}\right),$$

where

$$c_{\phi,1} = c_{\phi,1}\left(K\right) := \frac{1}{2}\frac{1}{K}\sum_{j=1}^{K}\left[\phi_j^2\left(1\right) + \phi_j^2\left(0\right)\right] \text{ and } B_1 = \int_{-\infty}^{\infty} \tau \Gamma_{XU}\left(\tau\right) d\tau.$$

When $K \to \infty$ and $T \to \infty$, the variance and bias expressions are similar to those in the case with discrete-time data. Their interpretations are also similar. For example, when X_tU_t is positively autocorrelated such that $\Gamma_{XU}(\tau) > 0$ for all τ, then $B_2 > 0$ and $\hat{\Omega}^*$ is biased downward. This is analogous to the discrete-time case. Note that the dominating bias is equal to $-c_\phi K^2 T^{-2} B_2$ instead of $-c_\phi K^2 n^{-2} B_2$. The latter can be shown to be the dominating bias in the usual discrete-time case for a fixed time interval (e.g., $\delta = 1$) with n observations. A takeaway from this comparison is that the effective sample size of a high-frequency sample (i.e., $\delta \to 0$) from a continuous-time process 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_2, 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 $-c_\phi\left(K^2 n^{-2}\right)\left(B_2\delta^{-2}\right)$. 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]

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

which is the mean square error of $\mathrm{vec}(\hat{\Omega}^*)$. It follows from Theorems 2.2 (i) and (ii) that

$$\begin{aligned}&\mathrm{MSE}(\hat{\Omega}^*)\\&= \mathrm{tr}\left[\{\Omega \otimes \Omega\}\left(\mathbb{I}_{d^2} + \mathbb{K}_{dd}\right)\right]\frac{1}{K} + c_{\phi,2}^2 \mathrm{vec}\left(B_2\right)' \mathrm{vec}\left(B_2\right)\frac{K^4}{T^4}\\&+ o\left(\frac{1}{K} + \frac{K^4}{T^4}\right) + O\left(\delta^2 + \frac{(\log n)^4}{T^4} + \frac{1}{T^2}\right).\end{aligned}$$

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

$$K = \kappa\left(\Omega, B_2\right)^{1/5} T^{4/5}, \tag{3}$$

where

$$\kappa\left(\Omega, B_2\right) := \left(\frac{tr\left[\{\Omega \otimes \Omega\}\left(\mathbb{I}_{d^2} + \mathbb{K}_{dd}\right)\right]}{4c_{\phi,2}^2 \mathrm{vec}\left[B_2\right]' \mathrm{vec}\left[B_2\right]}\right).$$

When $K = \kappa\left(\Omega, B_2\right)^{1/5} T^{4/5}$, the first two terms in $\mathrm{MSE}(\hat{\Omega}^*)$ are of order $T^{-4/5}$. To ensure the terms that we ignore are indeed of a 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 τ to be large enough. Such a requirement is compatible with the sufficient conditions for Assumption 2.5(i).

In the case of usual discrete time series data with a fixed sampling time interval and n observations, the optimal choice of K is given by

$$K_D = \kappa\left(\Omega_D, B_{2D}\right)^{1/5} n^{4/5}, \tag{4}$$

where

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

The formula is the same as that in (3) but with T replaced by n. See, for example, Phillips (2005). In the above, Ω_D and B_{2D} are the discrete analogues of Ω and B_2. If we use the formula for K in (4) 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 with a shrinking sample interval (i.e., $\delta \to 0$). The choice of $K = cn^{4/5}$ is

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

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

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

[4] Given that K is an integer, we should round $\kappa\left(\Omega, B_2\right)^{1/5} T^{4/5}$ up to the next integer and use it as K. We ignore this for the theoretical analysis but implement it in the simulation study.

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 discrete-time process with a fixed time interval (e.g., $\delta = 1$) and n observations, and we use a parametric AR(1) plug-in approach to implement (4). We fit an AR(1) model to each component $z_{i,j}$ of z_i :

$$z_{i,j} = \rho_j z_{i-1,j} + e_{zj} \text{ for } j = 1, 2, \ldots, d$$

with the AR parameter and error variance estimated by

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

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

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

and let

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

The above data-driven choice does not require the value of δ, and hence we do not need to pin down the unit of time in measuring the sampling intervals. Whether the length of the sampling intervals is measured in seconds, hours, days, or months does not affect how we compute $\hat{K}_D$. The value of $\hat{K}_D$ is invariant to the unit of time, and an applied researcher does not have to choose a unit of time.

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,

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

$$= \left[\frac{1}{8c_{\phi,2}^2} \left(\sum_{j=1}^{d} \frac{c_{2j}^2}{c_{1j}^8} \right)^{-1} \left(\sum_{j=1}^{d} \frac{c_{2j}^2}{c_{1j}^4} \right) \right]^{1/5} T^{4/5} \left(1 + o_p(1)\right).$$

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 implementation 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. In practice, $\{u_i\}_{i=1}^n$ is of course not observed, so $\hat{K}_D$ in (5) is computed utilizing $\{\hat{z}_i = x_i \hat{u}_i\}_{i=1}^n$ where $\hat{u}_i = y_i - x_i' \hat{\beta}_D$.

Note that the high-frequency compatible rate of K is of order $T^{4/5}$, 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.

The above MSE-optimal choice of K is obtained under the rate assumption that $K \to \infty$ but at a slower rate than T. The so-obtained choice rule in (3) satisfies the rate assumption. One may wonder whether we can obtain an MSE-optimal choice of K under the "fixed-K" assumption that K is held fixed. The answer is no. Under the fixed-K asymptotics, Theorem 2.2 shows that the variance of $\hat{\Omega}^*$ is proportional to $1/K$ and the squared-bias is proportional to $1/T^2$. To minimize the dominating terms in the MSE, we would make K as large as possible. Such an approach would then drive K to infinity and make it incompatible with the "fixed-K" assumption to begin with. As an example, consider the case when $d = 1$ and the Fourier basis functions in (2) are used. By Theorem 2.2 (i) and (iii), the dominating terms in the MSE are

$$\frac{1}{T^2} B_1 + \frac{2}{K} \Omega^2,$$

as $c_{\phi,1}(K) = \frac{1}{2} \frac{1}{K} \sum_{j=1}^{K} \left[\phi_j^2(1) + \phi_j^2(0) \right] = 1$. It is now clear that there is no fixed-value of K that minimizes the above: any fixed value of K is dominated by a larger value.

The above analysis shows that only the large-K asymptotic framework is theoretically coherent with an asymptotically optimal choice of K. Such an optimal choice of K is seemingly incompatible with the distributional approximation obtained under the fixed-K asymptotic theory. This is not the case, and we provide a justification here. Let $\mathcal{C}_\alpha(p, K)$ be the $(1-\alpha)$-quantile of the fixed-K asymptotic distribution of F_T, that is

$$\Pr\left(\frac{K}{K-p+1} F_{p,K-p+1} > \mathcal{C}_\alpha(p, K) \right) = \alpha.$$

Note that $K/(K+p-1)F_{p,K-p+1} \Rightarrow \chi_p^2/p$ as $K \to \infty$. Letting $K \to \infty$ in the above equation and using the dominated convergence theorem, we obtain

$$\Pr\left(\chi_p^2/p > \lim_{K\to\infty} \mathcal{C}_\alpha(p,K)\right) = \alpha.$$

This shows that $\lim_{K\to\infty} \mathcal{C}_\alpha(p,K) = \chi_{p,\alpha}^2/p$, where $\chi_{p,\alpha}^2$ is the $(1-\alpha)$-quantile of the chi-squared distribution χ_p^2. Therefore, under the large K asymptotics, $\mathcal{C}_\alpha(p,K)$ is an asymptotically valid critical value, even though it is based on the fixed-K asymptotic distribution. In the literature on the fixed-smoothing asymptotics for discrete-time data with a fixed δ, it has been proved that for the location models and linear regression models, critical values based on the fixed-smoothing asymptotic distribution (i.e., K is fixed for series LRV estimation) are second-order correct under the increasing-smoothing asymptotics (i.e., $K \to \infty$). See, for example, Sun (2013) for the case with series LRV estimation and Sun (2014a) and Sun et al. (2008) for the case with kernel LRV estimation.

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 fact that our observations are collected at a high frequency with the sampling interval δ going to zero. We can use exactly the same approach as we would do in the case with discrete-time observations where the time distance between neighboring observations is fixed: the test statistic is constructed in the same way, and the smoothing parameter is chosen in the same way. We do not need to choose a unit of time to measure the sampling duration. 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.

3 The Nonstationary Case

3.1 Exogenous Regressors

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

$$Y_t = \alpha_0 + X_t'\beta_0 + U_{0t} \tag{6}$$

where $X_t \in \mathbb{R}^{d\times 1}$ is a nonstationary process, $U_{0t} \in \mathbb{R}$ is a stationary process, $\{X_t\}$ and $\{U_{0t}\}$ are independent.[5] As in the case with stationary regressors, only a discrete set of points $\{x_i = X_{i\delta}, y_i = Y_{i\delta}\}_{i=1}^n$ are observed. The discrete-time model is

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

where $u_{0i} = U_{0,i\delta}$. The object of interest is the slope parameter β_0, and we aim at testing $H_0 : R\beta_0 = r$ against $H_1 : R\beta_0 \neq r$ 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.

We consider the same limiting experiment where $\delta \to 0$ and $T \to \infty$ for a fixed K.

[5]We use U_{0t} instead of U_t to denote the error process because in the next subsection we will use U_t to denote $(U_{0t}', U_{xt}')'$. We shall use $U_{0\cdot}$ to denote $\{U_{0t} : t \in [0,T]\}$.

Assumption 3.1 *For $e_{\delta,T}(U_{0\cdot})$ defined in the same way as in Lemma 2.1,*

$$\sum_{0\leq\tau\leq T} E\left|\Delta U_{0\tau}\right| = O\left(T\right) \text{ and } e_{\delta,T}(U_{0\cdot}) = o_p\left(1\right).$$

The above assumption is similar to Assumptions 2.1 and 2.5(i). It ensures that

$$\Lambda\left(n,\delta\right)^{-1}\sum_{i=1}^{n} u_{0i} = T^{-1/2}\int_0^T U_{0t}dt + o_p\left(1\right).$$

Assumption 3.2 *For a sequence of $d\times d$ diagonal matrices (Λ_T) with diverging diagonal elements*

$$\begin{pmatrix} \Lambda_T^{-1}X_{Tr} \\ T^{-1/2}\int_0^{Tr} U_{0s}ds \end{pmatrix} \Rightarrow \begin{pmatrix} X^{\circ}\left(r\right) \\ \sigma_0 W_0(r) \end{pmatrix} \text{ for } \sigma_0 > 0 \text{ and } r\in(0,1]$$

as $T\to\infty$, where $X^{\circ}(\cdot)$ is a continuous (a.s.) semimartingale, $W_0(\cdot)$ is standard Brownian motion, and $X^{\circ}(\cdot)$ and $W_0(\cdot)$ are independent.

The weak convergence in Assumption 3.2 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. The assumption is the continuous-time analogue of the traditional invariance principles. It is similar to Assumption C2 in Chang et al. (2018) which points out that the assumption is satisfied for a wide class of continuous-time processes. For general null recurrent diffusions and jump diffusions, Kim and Park (2017) provides sufficient conditions under which $\Lambda_T^{-1}X_{Tr}\Rightarrow X^{\circ}\left(r\right)$. As discussed after Assumption 2.4, Lu and Park (2019) provides sufficient conditions under which $T^{-1/2}\int_0^{Tr} U_{0s}ds \Rightarrow \sigma_0 W_0(r)$.

For $j=1,\ldots,K$, let

$$\eta_j = \int_0^1 \phi_j(r)X^{\circ}(r)dr,$$

and

$$\eta = (\eta_1,\ldots,\eta_K)' \in \mathbb{R}^{K\times d}.$$

Assumption 3.3 *With probability one, $\eta'\eta$ is of full rank d.*

Assumption 3.3 requires that, with probability one, the $L^2[0,1]$ projection coefficients of components of X° in the directions $\phi_j, j=1,\ldots,K$, form d 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 when K is large enough.

Now we detail the testing procedure. Assume that $K\geq d+1$. The testing steps are as follows:

1. Create the transformed data $\{\mathbb{W}_j^y,\mathbb{W}_j^x\}_{j=1}^K$ where

$$\mathbb{W}_j^y = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)y_i,\ \mathbb{W}_j^x = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_i. \tag{7}$$

   Denote the matrix forms of transformed data by

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

2. Regress $\mathbb{W}^y$ on $\mathbb{W}^x$ without an intercept by OLS. This yields the transformed OLS estimator $\hat{\beta}_{TOLS}$ and the residual vector $\hat{\mathbb{W}}^{u_0}$:

$$\hat{\beta}_{TOLS} = \left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1} \mathbb{W}^{x\prime}\mathbb{W}^y, \ \hat{\mathbb{W}}^{u_0} = \mathbb{W}^y - \mathbb{W}^x \hat{\beta}_{TOLS}. \tag{8}$$

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

$$F_{TOLS} = \frac{1}{\hat{\sigma}_0^2}(R\hat{\beta}_{TOLS} - r)' \left[R\left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1} R'\right]^{-1} (R\hat{\beta}_{TOLS} - r)/p, \tag{9}$$

where

$$\hat{\sigma}_0^2 = \frac{1}{K}\sum_{j=1}^{K}(\hat{\mathbb{W}}_j^{u_0})^2 = \frac{1}{K}\hat{\mathbb{W}}^{u_0\prime}\hat{\mathbb{W}}^{u_0}. \tag{10}$$

Define

$$\mathbb{W}_j^{u_0} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \phi_j\left(\frac{i}{n}\right) u_{0i}, \ \underset{K\times 1}{\mathbb{W}^{u_0}} = (\mathbb{W}_1^{u_0}, \ldots, \mathbb{W}_K^{u_0})'.$$

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 $\mathbb{W}^x$, $\mathbb{W}^{u_0}$, and $\hat{\beta}_{TOLS}$.

Lemma 3.1 *Let Assumptions 2.2, 3.1–3.3 hold. Then, as* $\delta \to 0$ *and* $T \to \infty$*,*
(a) $(n^{-1/2}\mathbb{W}^x\Lambda_T^{-1}, \sqrt{\delta}\mathbb{W}^{u_0}) \Rightarrow (\eta, \nu)$;
(b) $\sqrt{T}\Lambda_T(\hat{\beta}_{TOLS} - \beta_0) \Rightarrow (\eta'\eta)^{-1}(\eta'\nu)$.

Let $R(\ell, \cdot)$ and r_ℓ be the ℓ-th rows of R and r, respectively. Since we do not require that all elements of (X_{T_r}) converge at the same rate, the rate of convergence of $R(\ell, \cdot)\hat{\beta}_{TOLS}$ depends on the element of $\hat{\beta}_{TOLS}$ that has the slowest rate of convergence among those elements appearing in the ℓ-th restriction. To capture this, for $\ell = 1, \ldots, p$, we define the sets

$$\mathcal{I}_\ell := \{j : \text{ for } j \in \{1, 2, \ldots, d\} \text{ such that } R(\ell, j) \neq 0\},$$

which consists of the indices of the coefficients that appear in the ℓ-th restriction. When T is large enough, the rate of convergence of $R(\ell, \cdot)\hat{\beta}_{TOLS}$ is given by $\sqrt{T}\min_{j\in\mathcal{I}_\ell}\Lambda_T(j, j)$. Let

$$\tilde{\Lambda}_T = \text{diag}\left(\min_{j\in\mathcal{I}_1}\Lambda_T(j, j), \ldots, \min_{j\in\mathcal{I}_p}\Lambda_T(j, j)\right),$$

which is a $p \times p$ diagonal matrix.[6] Then $\lim_{T\to\infty}\tilde{\Lambda}_T R\Lambda_T^{-1} = R_\circ$ for a matrix $R_\circ \in \mathbb{R}^{p\times d}$ whose (ℓ, j)-th element $R_\circ(\ell, j)$ is equal to

$$R_\circ(\ell, j) = \lim_{T\to\infty}\tilde{\Lambda}_T(\ell, \ell) R(\ell, j)/\Lambda_T(j, j) = R(\ell, j)\lim_{T\to\infty}\left[\min_{m\in\mathcal{I}_\ell}\Lambda_T(m, m)/\Lambda_T(j, j)\right]. \tag{11}$$

[6] $\min_{j\in\mathcal{I}_\ell}\Lambda_T(j, j)$ should be interpreted as the minimum of $\Lambda_T(j, j)$ over $j \in \mathcal{I}_\ell$ when T is large enough.

That is, $R_\circ$ is the same as R after we zero out the elements in each row of R for which the corresponding coefficients can be estimated at a faster rate than the slowest rate for the coefficients involved in this row. We require that $R_\circ$ is of row rank p, a condition that is clearly satisfied when there is no heterogeneity in the rates of convergence, for example, $R_\circ = R$ when Λ_T is a scalar matrix.

Theorem 3.1 *Let Assumptions 2.2, 3.1–3.3 hold. If $K \geq d+1$ and $\lim_{T\to\infty} \tilde{\Lambda}_T R \Lambda_T^{-1}$ is of rank p, then*

$$F_{TOLS} \Rightarrow \frac{K}{K-d} \cdot F_{p,K-d},$$

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

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. Asymptotic F theory in a regression with nonstationary and exogenous regressors has been recently developed in Sun (2021) for discrete time series. Since the limiting process $X^\circ(\cdot)$ can be highly nonstandard and goes beyond what has been considered in Sun (2021), Theorem 3.1 has widened the applicability of the asymptotic F theory. See Kim and Park (2017) for the nonstandard forms that $X^\circ(\cdot)$ can take when $\{X_t\}$ is a null recurrent diffusion process.

To implement the F test, we 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 (10) takes a form similar to that in the stationary case. The infeasible variance estimator can be written as

$$\hat{\sigma}_0^2 = \frac{1}{K} \sum_{j=1}^{K} \left[\frac{1}{\Lambda(n,\delta)} \sum_{i=1}^{n} \phi_j \left(\frac{i}{n}\right) u_{0i} \right]^2.$$

As a practical rule of thumb, we can adapt the data-driven procedure in the stationary case and 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 $\{u_{0i}\}$, computing the AR(1) data-driven $\hat{K}_D$ using the formula in (5).

3. Let $\hat{K}^* = \max(\hat{K}_D, d+3)$ and use $\hat{K}^*$ to construct the transformed regression. Taking the maximum between $\hat{K}_D$ and $d+3$ ensures that the limiting F distribution has a finite mean.

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

We note in passing that an asymptotic F theory may also be developed based on the usual OLS estimator in step 1 above rather than the transformed OLS estimator, but then a series variance estimator with judiciously crafted basis functions has to be used. See Sun (2021) for more details in the discrete-time setting. We will not pursue this extension and choose to use a transformed regression, which can be regarded as a special case of the transformed and augmented regression in the next subsection. Hwang and Sun (2018) provides some discussion on the advantages of the transformed approach, including its robustness to contaminations whose energy is concentrated at high frequencies in the frequency domain.

3.2 Endogenous Regressors

We consider the same model $Y_t = \alpha_0 + X_t'\beta_0 + U_{0t}$ as in the previous subsection, but we now allow $\{X_t\}$ to be endogenous. The cost of admitting endogeneity comes in the form of less flexibility for the data generating process of the weak limit of $\Lambda_T^{-1} X_{Tr}$, $r \in [0,1]$. Namely, we require that $\Lambda_T^{-1} = T^{-1/2} I_d$ and that the limiting process be Brownian motion. As we discuss shortly, this requirement is a natural adaptation of the discrete time literature on inference in cointegrating regressions. For example, it is similar to the discrete time framework adopted in Vogelsang and Wagner (2014) and Hwang and Sun (2018). It is an open question whether an asymptotic F theory can still be developed for other forms of nonstationarity. As before, we only observe a discrete set of points $\{(x_i, y_i)\}_{i=1}^n$ satisfying $y_i = \alpha_0 + x_i'\beta_0 + u_{0i}$. Again we want to test $H_0 : R\beta_0 = r$ against $H_1 : R\beta_0 \neq r$.

We maintain Assumption 3.1 regarding the stationary process $\{U_{0t}\}$ but now allow for some forms of dependence between $\{X_t\}$ and $\{U_{0t}\}$. Towards this end, the assumption below is similar to and replaces Assumption 3.2.

Assumption 3.4 *As $T \to \infty$, the following functional central limit theorem holds:*

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \int_0^{Tr} U_{0s} ds \\ \frac{1}{\sqrt{T}} X_{Tr} \end{pmatrix} \Rightarrow \begin{pmatrix} B_0(r) \\ B_x(r) \end{pmatrix} := \Omega^{1/2} \begin{pmatrix} W_0(r) \\ W_x(r) \end{pmatrix} \text{ for } r \in [0,1]$$

where $\Omega^{1/2} \left(\Omega^{1/2}\right)' = \Omega$,

$$\Omega = \begin{pmatrix} \underset{1\times 1}{\sigma_0^2} & \underset{1\times d}{\sigma_{0x}} \\ \underset{d\times 1}{\sigma_{x0}} & \underset{d\times d}{\Omega_{xx}} \end{pmatrix},$$

and $W_0(\cdot)$ and $W_x(\cdot)$ are independent standard Brownian motions.

The weak convergence requirement in Assumption 3.4 is a natural counterpart to conditions in the discrete-time literature on co-integrating regressions. For example, replacing a sum with an integral in the discrete time setting of Vogelsang and Wagner (2014) might suggest modeling

$$X_t = X_0 + \int_0^t U_{x\tau} d\tau. \tag{12}$$

for some stationary process $\left\{U_{xt} \in \mathbb{R}^{d\times 1}, t \in [0,T]\right\}$. Then Assumption 3.4 is equivalent to an FCLT for the stationary process $\left\{U_t = (U_{0t}', U_{xt}')' \in \mathbb{R}^{d+1}, t \in [0,T]\right\}$ provided that $X_0 = o_p(T^{1/2})$. However, the form in (12) is not particularly desirable, and Assumption 3.4 is more flexible. For example, the data generating process in the non-stationary simulation environment of Section 4 satisfies Assumption 3.4. There, $\{X_t\}$ follows a two-dimensional Brownian motion and $\{U_{0,t}\}$ is a stationary Ornstein Uhlenbeck process that may not be independent of $\{X_t\}$. Alternatively to (12), we may view the continuous-time generalization of the setting in Vogelsang and Wagner (2014) and Hwang and Sun (2018) as requiring that, up to terms that are $o_p(T^{1/2})$, $\{X_t\}$ possesses some form of stationary increments that may be correlated with U_{0t} such that Assumption 3.4 holds. Viewing continuous time I(1) processes as nonstationary processes with stationary increments is adopted, for example, in Comte (1999).

In our asymptotic development, it is convenient to use the Cholesky form of $\Omega^{1/2}$ so that

$$B(\cdot) = \begin{pmatrix} B_0(\cdot) \\ B_x(\cdot) \end{pmatrix} = \begin{pmatrix} \sigma_{0\cdot x} W_0(\cdot) + \sigma_{0x}\Omega_{xx}^{-1/2} W_x(\cdot) \\ \Omega_{xx}^{1/2} W_x(\cdot) \end{pmatrix}, \tag{13}$$

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

For $j = 1, \ldots, K$, define

$$\eta_j = \int_0^1 \phi_j(r) B_x(r) dr, \ \xi_j = \int_0^1 \phi_j(r) dB_x(r),$$
$$\tilde{\nu}_j = \int_0^1 \phi_j(r) dW_0(r), \ \nu_j = \int_0^1 \phi_j(r) dB_0(r) = \sigma_{0\cdot x}\tilde{\nu}_j + \xi_j'\theta_0,$$

for $\theta_0 = \Omega_{xx}^{-1}\sigma_{x0}$ and

$$\eta = (\eta_1, \ldots, \eta_K)' \in \mathbb{R}^{K\times d}, \ \xi = (\xi_1, \ldots, \xi_K)' \in \mathbb{R}^{K\times d}, \ \zeta = (\eta, \xi) \in \mathbb{R}^{K\times 2d},$$
$$\tilde{\nu} = (\tilde{\nu}_1, \ldots, \tilde{\nu}_K)' \in \mathbb{R}^{K\times 1}, \ \nu = (\nu_1, \ldots, \nu_K)' \in \mathbb{R}^{K\times 1}.$$

Then $\nu = \xi\theta_0 + \sigma_{0\cdot x}\tilde{\nu}$.

Next, we make an assumption similar to Assumption 3.3.

Assumption 3.5 *With probability one, $\zeta'\zeta$ is of full rank $2d$.*

Let $\tilde{\Delta} x_i = (x_i - x_{i-1})/\delta$. 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 $u_{0\cdot x,i} = u_{0i} - \tilde{\Delta} x_i'\theta_0$. Using the transformed variables $\{\mathbb{W}_j^y, \mathbb{W}_j^\alpha, \mathbb{W}_j^x, \mathbb{W}_j^{\tilde{\Delta} x}, \mathbb{W}_j^{u_{0\cdot x}}\}_{j=1}^K$ defined similarly as in (7), we have

$$\mathbb{W}_j^y = \mathbb{W}_j^\alpha \alpha_0 + \mathbb{W}_j^x \beta_0 + \mathbb{W}_j^{\tilde{\Delta} x}\theta_0 + \mathbb{W}_j^{u_{0\cdot x}}$$

where, for example, $\mathbb{W}_j^\alpha = n^{-1/2}\sum_{i=1}^n \phi_j(i/n)$ and $\mathbb{W}_j^{u_{0\cdot x}} = n^{-1/2}\sum_{i=1}^n \phi_j(i/n)\, u_{0\cdot x,i} = \mathbb{W}_j^{u_0} - \mathbb{W}_j^{\tilde{\Delta} x}\theta_0$. Our test of $H_0 : R\beta_0 = r$ is based on estimating the above transformed and augmented regression by OLS. We call the estimator the TAOLS estimator. We outline the steps below:

1. Create the transformed variables $\{\mathbb{W}_j^y, \mathbb{W}_j^x, \mathbb{W}_j^{\tilde{\Delta} x}\}_{j=1}^K$ and stack them to form the data matrices $\mathbb{W}^y$, $\mathbb{W}^x$, and $\mathbb{W}^{\tilde{\Delta} x}$. For example, $\mathbb{W}^{\tilde{\Delta} x} = (\mathbb{W}_1^{\tilde{\Delta} x}, \ldots, \mathbb{W}_K^{\tilde{\Delta} x})' \in \mathbb{R}^{K\times d}$.

2. Regress $\mathbb{W}^y$ on $\mathbb{W}^x$ and $\mathbb{W}^{\tilde{\Delta} x}$ by OLS. 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 $\hat{\mathbb{W}}^{u_{0\cdot x}}$ 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

$$\underset{2d\times 1}{\hat{\gamma}} \equiv \begin{pmatrix} \hat{\beta}_{TAOLS} \\ \hat{\theta}_{TAOLS} \end{pmatrix} = (\widetilde{\mathbb{W}}'\widetilde{\mathbb{W}})^{-1}\widetilde{\mathbb{W}}'\mathbb{W}^y, \quad \hat{\mathbb{W}}^{u_{0\cdot x}} := \mathbb{W}^y - \widetilde{\mathbb{W}}\hat{\gamma}. \tag{14}$$

3. Calculate the test statistic

$$F_{TAOLS} = \frac{1}{\hat{\sigma}^2_{0\cdot x}} (R\hat{\beta}_{TAOLS} - r)' \left[R \left(\mathbb{W}^{x\prime} M_{\tilde{\Delta}x} \mathbb{W}^x \right)^{-1} R' \right]^{-1} (R\hat{\beta}_{TAOLS} - r)/p, \tag{15}$$

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

$$\hat{\sigma}^2_{0\cdot x} = \frac{1}{K}\sum_{j=1}^{K}(\hat{\mathbb{W}}^{u_{0\cdot x}}_j)^2 = \frac{1}{K}(\hat{\mathbb{W}}^{u_{0\cdot x}})'\hat{\mathbb{W}}^{u_{0\cdot x}}. \tag{16}$$

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 $\tilde{\Delta}x_i$. There is no need to know the value of δ or its unit. We note that the test statistic in (15) is constructed in the same way as in the discrete-time setting.

Theorem 3.2 *Let Assumptions 2.2, 3.1, 3.4, and 3.5 hold. Denote $\gamma_0 = (\beta_0', \theta_0')'$ and*

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

(a) As $T \to \infty$ for a fixed K,

$$\left[(nT)^{-1/2}\, \mathbb{W}^x, \delta^{1/2}\mathbb{W}^{\tilde{\Delta}x}, \delta^{1/2}\mathbb{W}^{u_0} \right] \Rightarrow (\eta, \xi, \nu).$$

(b) As $T \to \infty$ for a fixed K,

$$\Upsilon_T\,(\hat{\gamma} - \gamma_0) \Rightarrow \sigma_{0\cdot x}\left(\zeta'\zeta\right)^{-1}\zeta'\tilde{\nu}.$$

In particular,

$$T(\hat{\beta}_{TAOLS} - \beta_0) \Rightarrow \sigma_{0\cdot x}\left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \tilde{\nu} \overset{d}{=} MN\left[0, \sigma^2_{0\cdot x}\left(\eta' M_\xi \eta\right)^{-1}\right],$$

where $M_\xi = \mathbb{I}_K - \xi(\xi'\xi)^{-1}\xi'$ and "MN" stands for "mixed normal".

(c) If $K \geq 2d + 1$, then, as $T \to \infty$ for a fixed K,

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

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

Theorem 3.2 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 with a shrinking sampling interval (i.e., $\delta \to 0$) or discrete-time data with a fixed sampling interval (e.g., $\delta = 1$). This gives us much practical convenience.

To implement the above F test, we follow the procedure below, which is similar to that in the exogenous case.

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 $\{u_{0i}\}$, computing the AR(1) data-driven $\hat{K}_D$ using the formula in (5).

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

4 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. As an alternative to an OU explanatory variable process, we also consider the process $X_t = C_t - \mu_c$ where

$$dC_t = \kappa_x (\mu_c - C_t)\, dt + \sigma_x \sqrt{C_t} dV_t,$$

and V_t is again standard Brownian motion. This corresponds to Feller's Square Root (SR) process. In this setting, we keep the OU process $\{U_t\}$ as described above (again with $\{W_t\}$ independent of V_t) and $(\mu_c, \kappa_x, \sigma_x) = (4.8196, 0.1794, 0.9367)$ where these parameters come from fitting the SR process to 3-month T-bill rates from 1971 to 2016.

In the nonstationary setting, we consider the model

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

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_{0t}) as a stationary OU process. In particular, for $j \in \{1,2\}$, we have

$$dX_{j,t} = \sigma_j dZ_{j,t} \ \text{ and } \ dU_{0t} = -\kappa_u U_{0t} dt + \sigma_u dZ_{3,t},$$

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 $\varphi \geq 0$. 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 from 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 and 1/4, allowing for 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 (or SR) process becomes more persistent and in the OU case 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, OU, and SR processes except in the nonstationary 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 (2) and choose K via the data driven procedures described in Sections 2 and 3. 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 χ_p^2/p. The scaled chi-squared approximation can also be obtained by letting $K \to \infty, \delta \to 0$ and $T \to \infty$ jointly[7]. Utilizing the critical values from this distribution with our test statistics, we denote the resulting results by "Series Chi2." In the figures for the nonstationary setting, "Series F" and "Series Chi2" are reserved for the procedure outlined in Subsection 3.2 that can accommodate endogeneity. These labels are replaced by "S-EXO F" and "S-EXO Chi2" , respectively, for the procedures designed where $\{U_t\}$ is assumed exogenous described in Subsection 3.1.

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 (QS) 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 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

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

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.

4.1 Size study

Figures 2 – 5 display the empirical sizes (i.e., the null rejection probabilities) in the different simulation scenarios.

Figures 2 and 3 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 or SR 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 is the reference distribution used in a testing procedure.

In the nonstationary setting with exogenous regressors, the performance of the F tests 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 tests 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. Similarly, the F-test of Subsection 3.2 designed for potential endogeneity also relies on a Brownian motion limiting process for its validity. While this does not cause problems in our simulation setting where the premise holds, the fixed-b asymptotic distribution and that associated with the F-test in Subsection 3.2 are, in general, functionals of (X°), which may contain additional nuisance parameters beyond its scale. A benefit of our approach in Subsection 3.1 is that the conditioning argument in the proof of Theorem 3.1 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 of our knowledge, the F test in Subsection 3.2 appears to be the only asymptotically valid test in the literature. Unsurprisingly, it exhibits better size properties than the alternative tests from the pre-existing literature, including the fixed-b version of the test in Chang et al. (2018). While the F-test of Subsection 3.1 which assumes the error process $\{U_t\}$ is exogenous appears to maintain competitiveness against the F test of Subsection 3.2, this unfortunately is an artifact of the particular DGPs in our simulation setting. In this simulation environment, it can be shown that the limiting distribution of the exogeneity-based test is a noncentral F distribution that depends on nuisance parameters. The F distribution used happens to be relatively close to the finite sample distribution but will result in a poor approximation in general. 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 2 – 5 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.

4.2 Power study

Figures 6 – 8 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 there are only two different test statistics, ours and that in Chang et al. (2018) and the power is size-adjusted, there are only two different sized-adjusted power curves. The reported figures only display the comparison for the series-based approach in Sections 2 and 3 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."

Figures 6 and 7 show 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 most noticeably in the procedure that allows for endogeneity. 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.

Figures 6 and 8 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.

5 Empirical Application

Here we examine the series-based F test in an application to interest rate data that are available at multiple sampling frequencies. In particular, we revisit an application appearing in Chang et al. (2018), which focuses on characterizing the co-movements of interest rates among securities with different times to maturity. As discussed in Chang et al. (2018), the ability of the U.S. Federal Reserve System (FED) to influence long-term interest rates via the short-term Federal Funds Rate (FFR) was challenged during the Global Financial Crisis (GFC) of 2008 when the zero lower bound for the FFR was reached. This partially motivated the FED's adoption of non-conventional policies such as quantitative easing. To investigate the dynamics between short and long rates within their linear hypothesis testing methodology, Chang et al. (2018) test for "parallel shifts" among securities with varying maturities. Here, "parallel shifts" refers to changes in the yields of securities with different maturities tending to be of the same size and direction. Chang et al. (2018) regress 10-year U.S. Treasury bond (T-bond) yields on 3-month Treasury bill (T-bill) yields and consider data before and after the GFC separately. The existence of "parallel shifts" would imply a slope coefficient near one, and Chang et al. (2018) find that, prior to the GFC, there is no strong evidence against the null hypothesis of a unit slope coefficient. This is consistent with the view that the FED was able to successfully influence long rates via short rate policies prior to the GFC.

This regression setting, detailed below, is useful for evaluating our testing procedure because it is simple and allows for the consideration of several hypothesis tests of varying theoretical credibility. For example, the additional null hypothesis that the intercept coefficient is zero states that, on average, the yield spread is zero. If the yields of U.S. government securities of different duration differ based on compensation for interest rate risks and the expectations of future interest rates, we may expect to reject this hypothesis. Additionally, as the setting has been analyzed in Chang et al. (2018), we may contrast our methodology and results with theirs. We find that the conclusions stemming from the F tests are largely in line with those from the testing procedures of Chang et al. (2018). We observe, however, that the F test for one hypothesis test of interest produces a less ambiguous result at the daily sampling frequency and also bypasses a subjective modeling decision that can inflate one of the test statistics analyzed in Chang et al. (2018).

The continuous-time regression of interest is given by

$$Y_t = \alpha + X_t\beta + U_t,$$

where Y_t is the yield (in percent) of 10-year T-bonds at time t and X_t is the yield of 3-month T-bills. We observe $\{X_{i\delta}\}_{i=1}^n$ and $\{Y_{i\delta}\}_{i=1}^n$ at three fixed sampling interval lengths, δ, corresponding to daily, monthly, and quarterly frequencies. The number of observations n varies with δ as each sample is derived from a fixed time span, but we do not complicate the notation here. Recall, additionally, that the F test would be valid if applied to discrete time series under the standard discrete-time assumptions. The two yield series of different maturities are available from the Federal Reserve Economic Data (FRED) of the St. Louis FED. As in Chang et al. (2018), we consider three null hypotheses independently of one another and we consider two different sample windows. All hypothesis tests are performed twice, once utilizing data from each sample window separately. The null hypotheses are $H_0^\alpha : \alpha = 0$, $H_0^\beta : \beta = 1$, and $H_0^{\alpha,\beta} : \alpha = 0$ and $\beta = 1$ jointly. The first sample window includes data from 1962 to 2007 while the second contains observations from 2008 to 2019.

The two interest rate series plotted at the various sampling frequencies are presented in Figure 1. In Table 1, we present the test statistics associated with the various null hypotheses for each sample window. Test statistics titled "Series-F" refer to the F test described in Section 2 designed around the stationary regression setting. Those under the header "Kernel-χ^2" are performed utilizing the kernel-based χ^2 test of Chang et al. (2018) which they refer to as the H-test. The χ^2 tests (i.e., H-tests) in Figure 1 are calculated using the Andrews (1991) bandwidth procedure which is "high-frequency-compatible" as discussed in Chang et al. (2018) and utilizing the QS kernel. Rejection of a null hypothesis at the 5% level is indicated by "*" and rejection at the 1% level is indicated by "**". P-values are included in brackets for testing the null of "parallel shifts" H_0^β.

We can see from Table 1 that the results of the F tests are stable across all sampling frequency choices. This is consistent with the theory developed earlier in the paper, namely that the tests are valid for high-frequency observations over a long span and have direct counterparts that are valid and familiar in the discrete-time setting when the sampling frequency is lower. For the F tests, the statistical conclusions reached for each null hypothesis and sampling window remain the same for each sampling frequency: all null hypotheses are rejected at the 1% level except that we are unable to reject the "parallel shifts" hypothesis H_0^β at even the 5% significance level in any frequency using data prior to the GFC. This evidence is consistent with the view that the FED was able to control long rates via short-term policy rates prior to the GFC. Additionally, there is evidence against the hypothesis of a zero average yield spread (H_0^α, which is included in $H_0^{\alpha,\beta}$) before and after the GFC of 2008. This is consistent with the stylized fact that the yield curve tends to be upward sloping. The results and conclusions of the F tests are thus in agreement with the findings of Chang et al. (2018) where χ^2-based tests with "high-frequency compatible" bandwidths are utilized. Note that their findings are mirrored by those for the kernel-based χ^2 tests reported in Figure 1 which are computed according to their methodology. In contrast, Chang et al. (2018) show that in this regression setting, tests that are not robust to the sampling frequency or utilize a bandwidth choice that is not "high-frequency compatible" will reject H_0^β at the daily frequency.

Lastly, we discuss some differences between the F test and the kernel-based χ^2 test of Chang et al. (2018) in this application that may be indicative of the benefits of the F test. First, note that for the kernel-based χ^2 test using pre-GFC observations at the daily sampling frequency, the test statistic surpasses the critical value for a 5% test but not that of a 1% test. Chang et al. (2018) choose to view this as failing to reject the null hypothesis, requiring that the test statistic surpass the 1% critical value to take a more conservative stance. To this end, they note that the nominal size may understate the empirical rejection probability as observed in their (and our) simulations. On the other hand, the F test statistic here fails to surpass the critical value for a 5% test, corresponding to a p-value of 0.0787. As seen in our simulations and discussed in relation to the fixed-smoothing literature in Subsection 4.1, the F test can result in tests with more accurate size. This example may be a case where some ambiguity regarding test significance is avoided.

Another point of interest for the F test in this example is as follows. Some of the test statistics considered in Chang et al. (2018) may require/allow the researcher to determine a continuous-time modeling parameter that could influence the test statistic's magnitude. Such a test statistic utilizes a (kernel-based) LRV estimator that, when utilizing the discrete-time counterpart LRV estimator, requires a "high-frequency compatible" bandwidth parameter b_n in order to produce a valid test. One choice they consider is their continuous-time rule of thumb (CRT). This is given by

$$b_n = cn^a/\delta^{1-a},$$

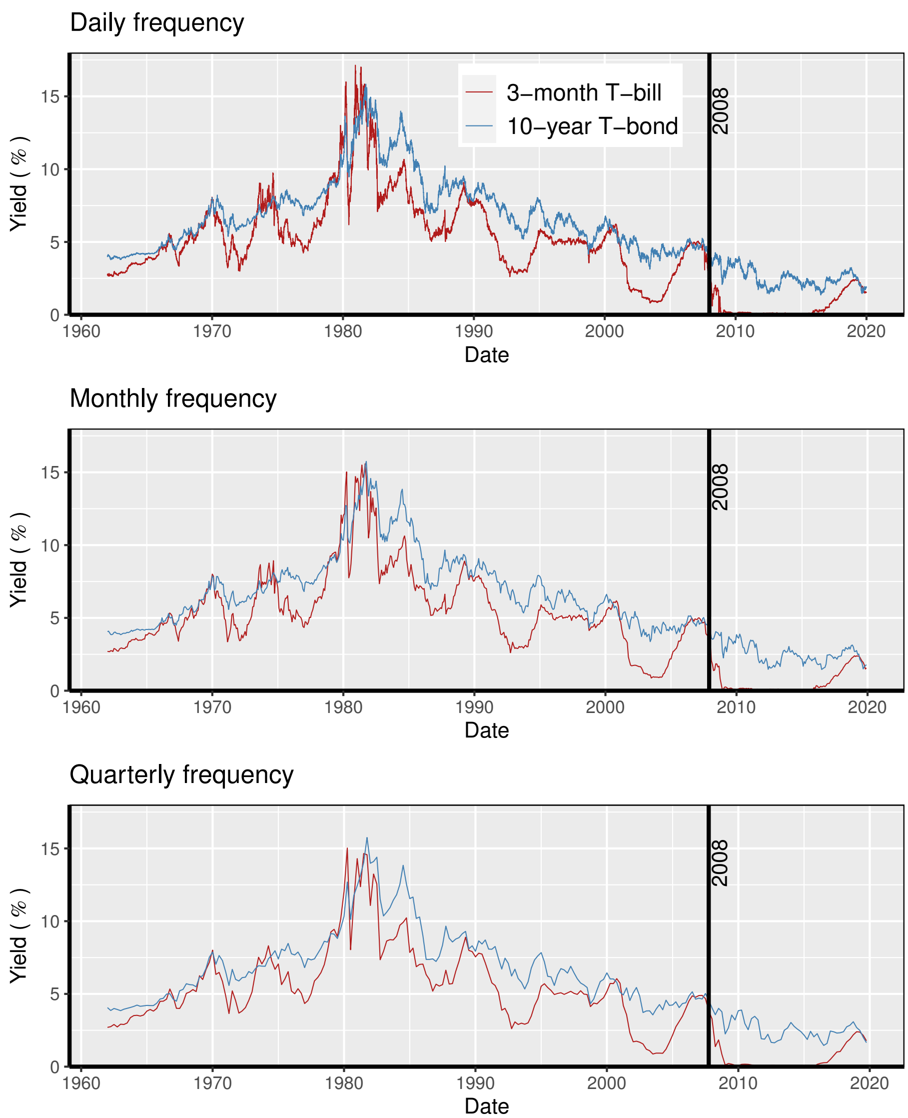

Figure 1: 10-year Treasury bond and 3-month Treasury bill yields at the sampling frequencies analyzed. A line at the beginning of 2008 demarcates the two sample windows.

Sample: 1962-2007						
Sampling Freq.	Daily		Monthly		Quarterly	
Test Stat.	Series-F	Kernel-χ^2	Series-F	Kernel-χ^2	Series-F	Kernel-χ^2
H_0^α	18.73**	22.10**	18.77**	20.74**	19.96**	19.93**
H_0^β	3.69	4.30*	3.42	3.67	3.07	3.15
	[0.0787]	[0.03801*]	[0.0874]	[0.0553]	[0.1000]	[0.0760]
$H_0^{\alpha,\beta}$	17.75**	38.26**	18.97**	38.95**	21.26**	41.43**
Sample: 2008-2019						
Sampling Freq.	Daily		Monthly		Quarterly	
Test Stat.	Series-F	Kernel-χ^2	Series-F	Kernel-χ^2	Series-F	Kernel-χ^2
H_0^α	87.19**	106.62**	87.68**	107.52**	129.07**	124.77**
H_0^β	81.44**	32.29**	81.59**	27.48**	37.94**	20.38**
$H_0^{\alpha,\beta}$	46.25**	113.73**	48.11**	116.19**	65.96**	127.73**

Table 1: Test statistics computed with observations collected at different sampling frequencies. Brackets contain p-values. Rejection of a null hypothesis at the 5% level is indicated by "*" and rejection at the 1% level is indicated by "**". p-values are included in brackets for testing the null of "parallel shifts" H_0^β based on the pre-GFC observations.

where $c > 0$ and $0 < a < 1$. In contrast to discrete-time rules of thumb for kernel-based LRV bandwidth parameters, there is now a division by δ^{1-a}. However, δ depends on the unit of time that T is measured in, which may be subjective. Suppose we set $c = 2.3019$ and $a = 1/5$ and wish to test H_0^β with daily observations between 1962 and 2007. These choices for a and c correspond to a guideline in Andrews (1991) for the QS kernel in a discrete-time setting when considering an AR(1) process with coefficient 0.5. (The observation below also holds with similar test statistics and p-values if we choose the alternative discrete-time rule of thumb choices $c = 3/4$ and $a = 1/3$, suggested in the undergraduate textbook Stock and Watson (2019); see equation (16.17) there). If we assume T is measured in years, i.e., $T = 46$ years between 1962 and 2007, then $\delta = 1/252$ for about 252 trading days in a year. Alternatively, suppose we think that T should be measured in months so that $T = 552$ months. Then we may set $\delta = 1/21$ for roughly 21 trading days in a month. As we see below, this distinction changes the test conclusion.

Table 2 contains the test statistics computed from daily observations between 1962 and 2007 for the null hypothesis H_0^β. In addition to the test statistics considered earlier, it includes two alternative calculations for the kernel-based χ^2 test statistic, denoted by "Kernel-χ^2-CRT, $\delta = 1/252$" and "Kernel-χ^2-CRT, $\delta = 1/21$." These correspond to the choice of δ described above. The corresponding test statistics from Table 1 are also included. The kernel-based χ^2 test of Chang et al. (2018) reported earlier in Table 1 that is calculated utilizing the procedure of Andrews (1991) is now denoted "Kernel-χ^2-AD." Note that, as the F test statistic, this version of the test statistic does not feature a direct reliance on a user inputted δ. From Table 2, we see that changing δ from 1/252 to 1/21 increases the CRT-based test statistic to surpass the critical value for a 1% test. If δ is chosen too large, we get a bandwidth that is too small for a continuous-time process that varies slowly at higher frequency observations. The effect is similar to using a

Sample: 1962-2007, Daily Frequency				
Test Stat.	Series-F	Kernel-χ^2-AD	Kernel-χ^2-CRT, $\delta = 1/252$	Kernel-χ^2-CRT, $\delta = 1/21$
H_0^β	3.69	4.30*	4.46*	10.03**
	[0.0787]	[0.0380]	[0.0347]	[0.0015**]

Table 2: Test statistics computed with observations collected at a daily sampling frequency during 1962-2007 for the "parallel shifts" hypothesis H_0^β. In addition to the test statistics from earlier, additional kernel-based χ^2 test statistics of Chang et al. (2018) are presented when computed with the CRT using $\delta = 1/252$ and $\delta = 1/21$. Rejection of a null hypothesis at the 5% level is indicated by "*" and rejection at the 1% level is indicated by "**". p-values are included in brackets.

"high-frequency incompatible" bandwidth, a setting explored in Chang et al. (2018) that leads to spurious tests with divergent test statistics. This example suggests that tests which do not rely on a user choosing δ, such as the F test or the test of Chang et al. (2018) that utilizes the Andrews (1991) bandwidth procedure, may be more robust against debatable modeling decisions that could impact statistical significance. In addition to potential size-accuracy benefits, the F test adds to the available tests with this feature, and only one such test is discussed in Chang et al. (2018).

6 Robustness to Additive Measurement Noise

Here we consider the implications of including additional noises in the regression error. We show that, under reasonable assumptions, the additive noises do not affect much of the theory underlying 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 realized volatility measurement in asset returns.

One way to model the noise is to assume that at each observation time $i\delta$, $i = 1, \ldots, n$, 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 $y_i = Y_{i\delta} + \epsilon_i$. The additive noises $\{\epsilon_i\}_{i=1}^n$ are different from $\{u_i\}_{i=1}^n$ in that there is no continuous-time process that governs the sequence $\{\epsilon_i\}_{i=1}^n$ no matter how small δ is. At each time we measure the value of Y_t, there is an additive noise attached to its true value. For easy reference, we will call this additive noise the measurement noise, although the noise may come from other sources.

For $i = 1, \ldots, n$ and $n = T/\delta$, the *observed* discretized model in the stationary case is now

$$y_i = x_i'\beta_0 + u_i + \epsilon_i. \tag{17}$$

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

$$y_i = \alpha_0 + x_i'\beta_0 + u_{0i} + \epsilon_i. \tag{18}$$

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 in nature, $\{\epsilon_i\}_{i=1}^n$ is stationary and strongly mixing with mixing coefficients $\{\varphi_\ell\}_{\ell=1}^\infty$ that satisfy $\sum_{\ell=1}^\infty \varphi_\ell^{1/2} < \infty$. (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 Assumptions 2.2 and 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 $\delta \to 0$ and $T \to \infty$,*

$$\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right) x_i\epsilon_i = O_p(\sqrt{n}) \text{ for } j = 0, 1, \ldots, K.$$

Whereas high serial correlation in $\{x_i u_i\}_{i=1}^n$ for small δ leads to $\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right) x_i u_i = O_p(\sqrt{n/\delta})$, the tie between the noise dependence structure and the sampling frequency yields that $\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right) x_i\epsilon_i$ is an order of magnitude smaller (in probability) than $\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right) x_i u_i$ despite the persistence in $\{x_i\}_{i=1}^n$. Consequently, Lemma 2.2 and Theorem 2.1 from Section 2 continue to hold. More specifically, when the observed $\{y_i\}_{i=1}^n$ take the form in (17), we have

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

In the series LRV estimator of Section 2, a key object now becomes

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

It follows from Lemma 6.1 that

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

Therefore, under the conditions of Lemma 6.1, the objects in Lemma 2.2 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 2.1 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 2.1 remains valid when the observed discretized model contains additive measurement noises as in* (17)*.*

In the nonstationary setting, it is easy to see that arguments similar to those in the proof of Lemma 6.1 yield the following lemma.

Lemma 6.2 *Let Assumptions 2.2 and 6.1 hold. Then for* $j = 0, 1, \ldots, K$*,*

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

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 effect of the additive measurement noise is negligible. Lemma 3.1 remains valid provided that Assumption 6.1 holds. Consequently, Theorem 3.1 still holds. Similarly, Theorem 3.2 remains valid.

Corollary 6.2 *Consider the nonstationary setting and let Assumption 6.1 hold. Then Theorems 3.1 and 3.2 remain valid when the observed discretized model contains additive measurement noises as in* (18)*.*

Corollaries 6.1 and 6.2 show that key components of our asymptotic theory are robust to the presence of additive measurement noises in $\{y_i\}_{i=1}^n$. The estimation and inference procedures of Sections 2 and 3 can be carried out without any modification. There are some caveats, however. First, in finite samples, the magnitude of $\sum_{i=1}^{n} \phi_j \left(i/n \right) x_i \epsilon_i$ in the stationary setting, or $\sum_{i=1}^{n} \phi_j \left(i/n \right) \epsilon_i$ in the nonstationary setting, depends on the size of the measurement noise. Only if $var(\epsilon_i)$ is not too large relative to the long run variance $\int_{-\infty}^{\infty} \Gamma_{XU} \left(\tau \right) d\tau$ of $\{X_t U_t\}$ or $\int_{-\infty}^{\infty} \Gamma_{U} \left(\tau \right) d\tau$ of $\{U_t\}$ can we safely ignore the terms due to the additive measurement noise. Second, when $var(\epsilon_i)$ is much larger than the variance of $x_i u_i$ or u_i at high frequencies, this can impact the procedure for selecting a high-frequency-compatible smoothing parameter K in our F tests or the high-frequency-compatible bandwidth parameters in alternative testing methodologies such as those in Chang et al. (2018).

In the supplementary online appendix, we revisit the simulation study in Section 4, now including additive measurement noise. We design the simulations so that a measure of the proportion of variation in the total regression error term $u_i + \epsilon_i$ stemming from ϵ_i is 35%. This measure is defined in the online appendix similarly to a noise-to-signal ratio analyzed in Aït-Sahalia and Yu (2009). There, it is reported that a typical noise-to-signal ratio for high-frequency stock returns with microstructure noise is 36.6%. In our simulations reported in the online appendix, the size and power properties of our tests in the presence of additive noise remain very similar to those reported in Section 4, with only a very slight increase in the null rejection probabilities when δ is very small. Additionally, the F test maintains its improved size properties in the various simulation environments. We note here, however, that it is possible to increase the variance of the additive noises enough that the size properties of all tests considered are negatively impacted. What threshold of the additive noise variance is too large? How to choose a high-frequency-compatible smoothing parameter when the additive noise may be large? We leave these questions to future research.

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 with exogeneous regressors, 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 measurement noises in the regressor error.

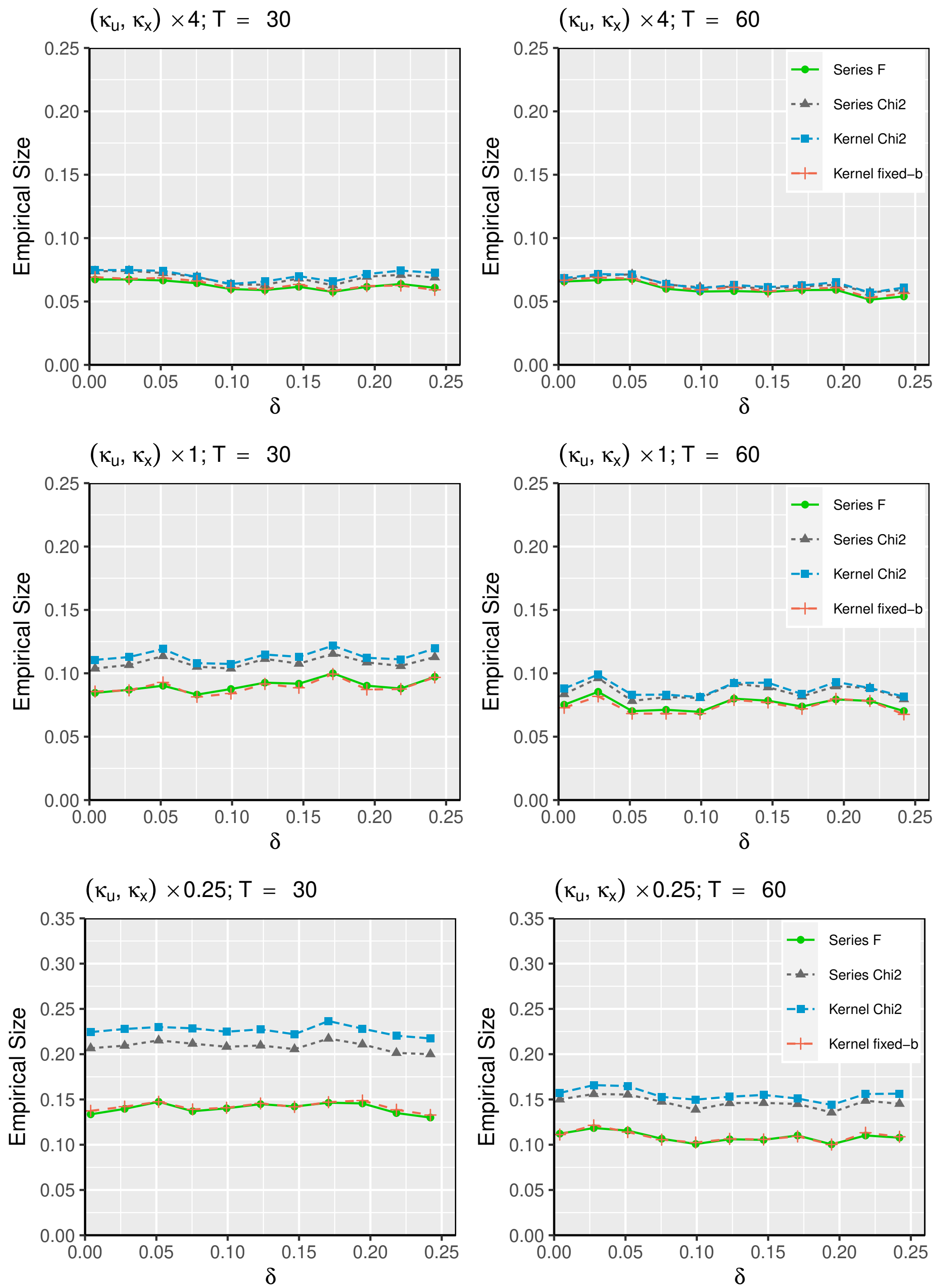

Figure 2: Empirical sizes in the stationary simulation setting when X_t follows an OU process and (κ_u, κ_x) are multiplied by factors of 4, 1 and 1/4.

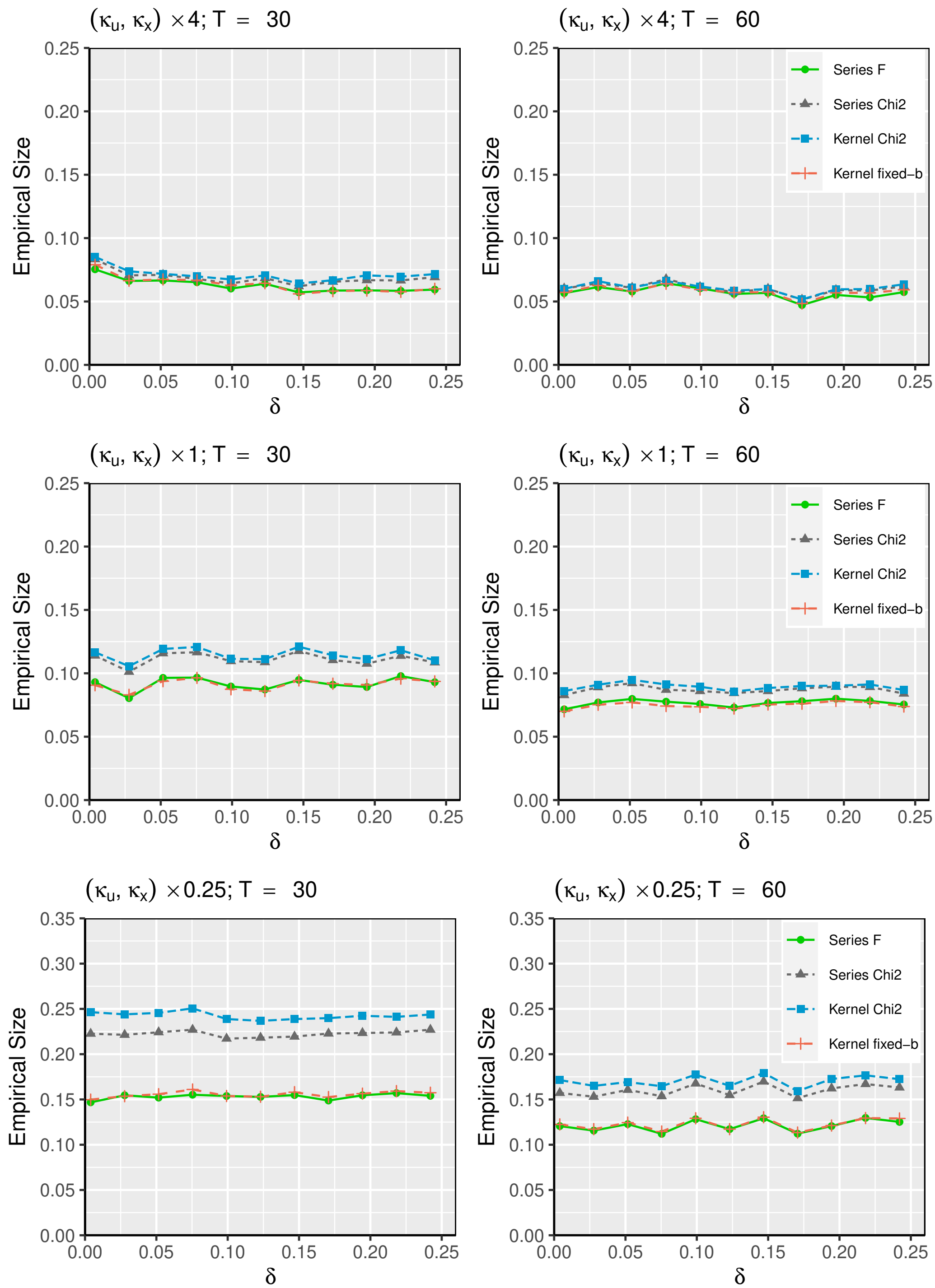

Figure 3: Empirical sizes in the stationary simulation setting when X_t follows an SR process and (κ_u, κ_x) are multiplied by factors of 4, 1, and 1/4.

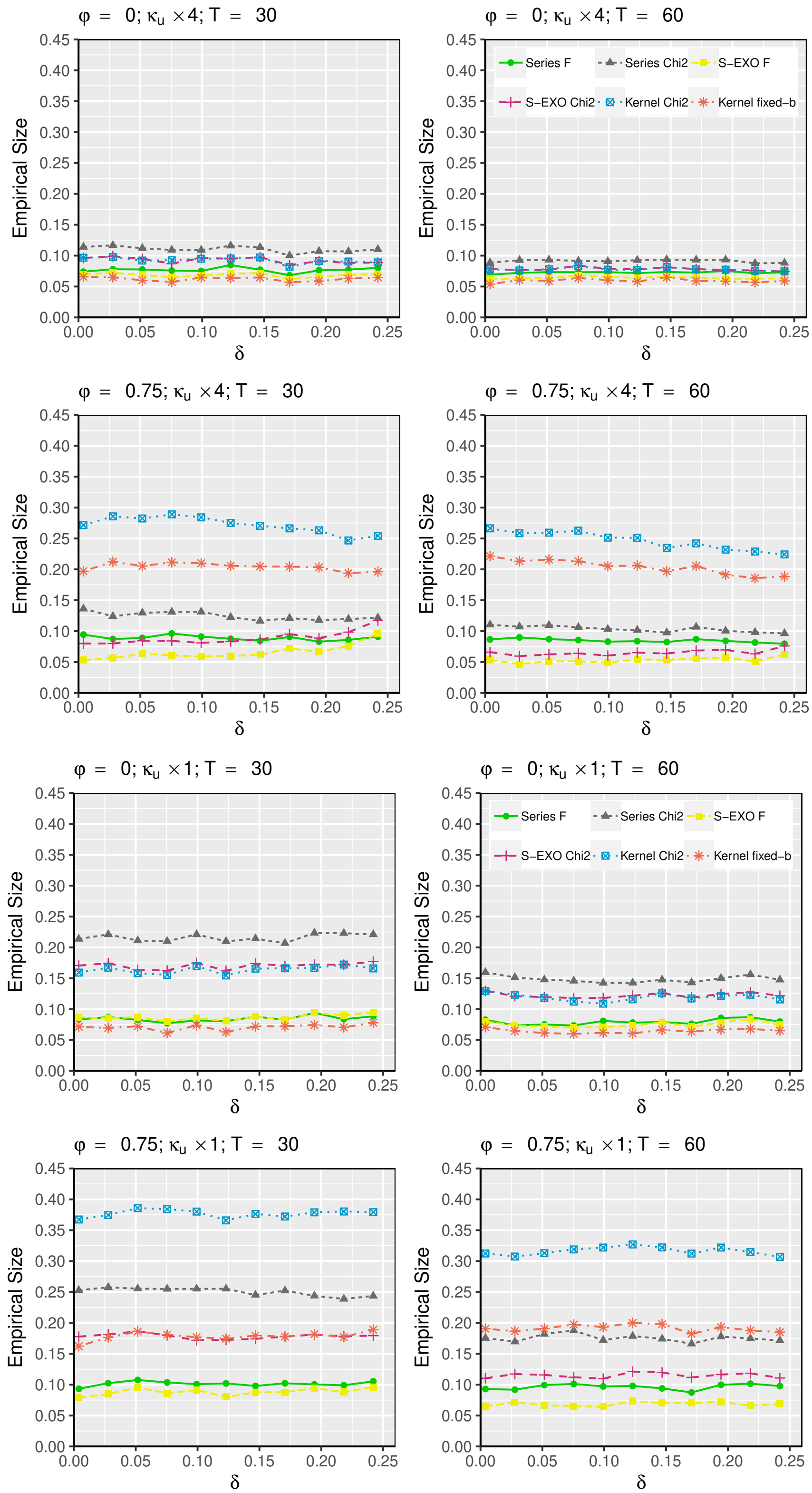

Figure 4: Empirical sizes in the nonstationary simulation setting when κ_u is multiplied by factors of 4 and 1.

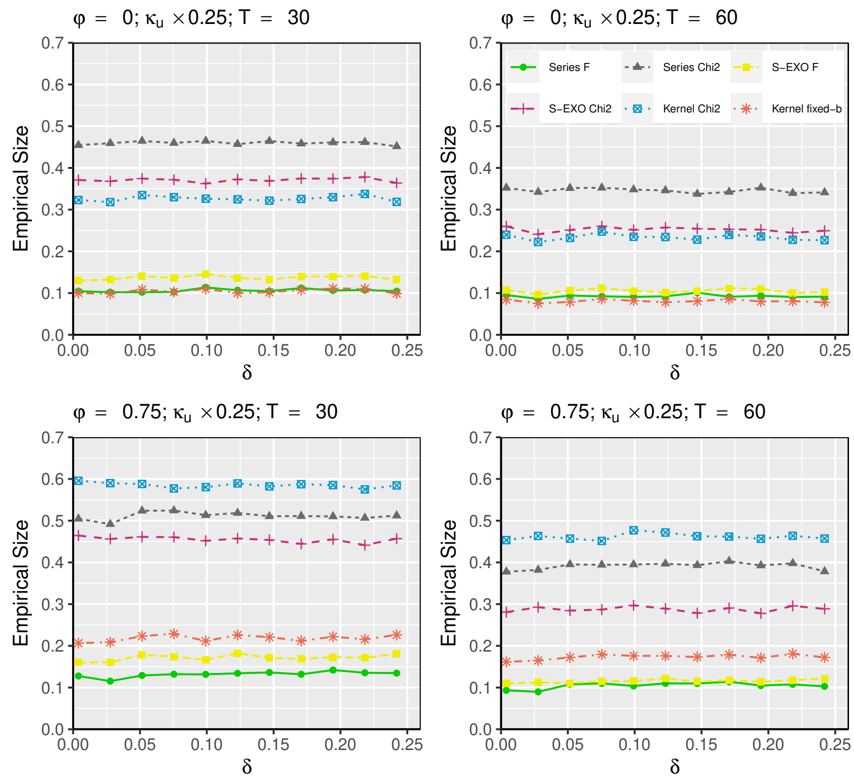

Figure 5: Empirical sizes in the nonstationary simulation setting when κ_u is multiplied by 1/4.

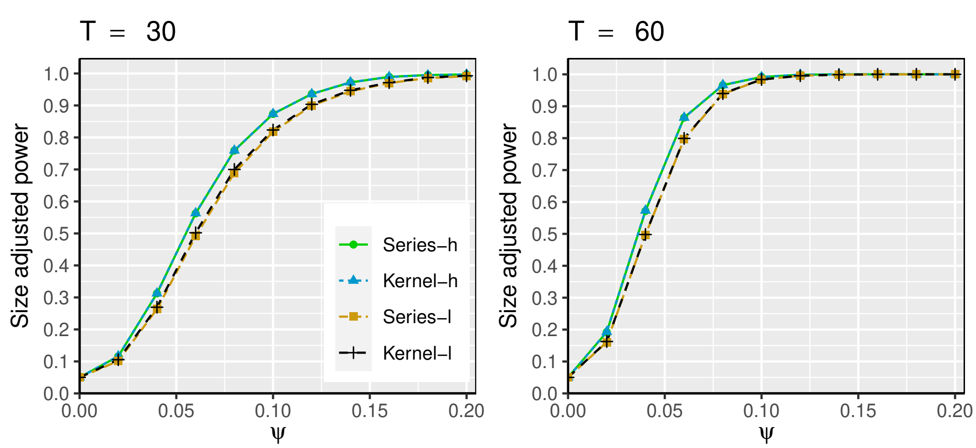

Figure 6: Size-adjusted powers in the stationary setting when X_t is distributed according to the OU process described in Section 4

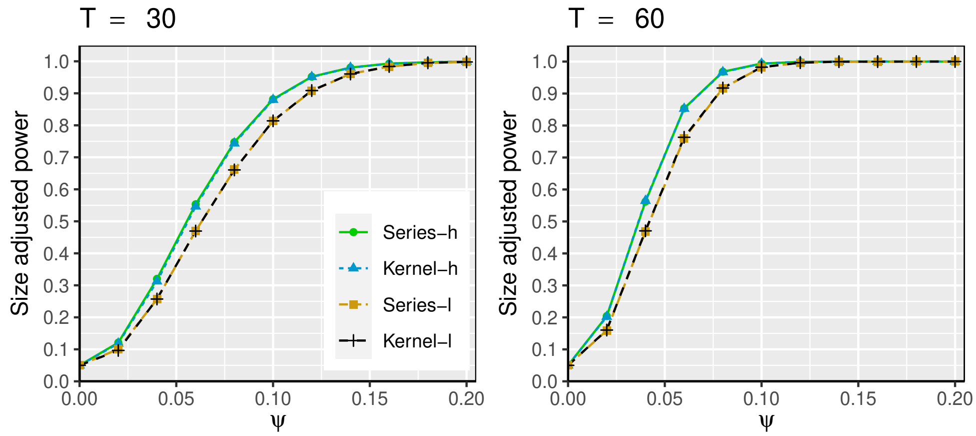

Figure 7: Size-adjusted powers in the stationary setting when X_t is distributed according to the SR process described in Section 4

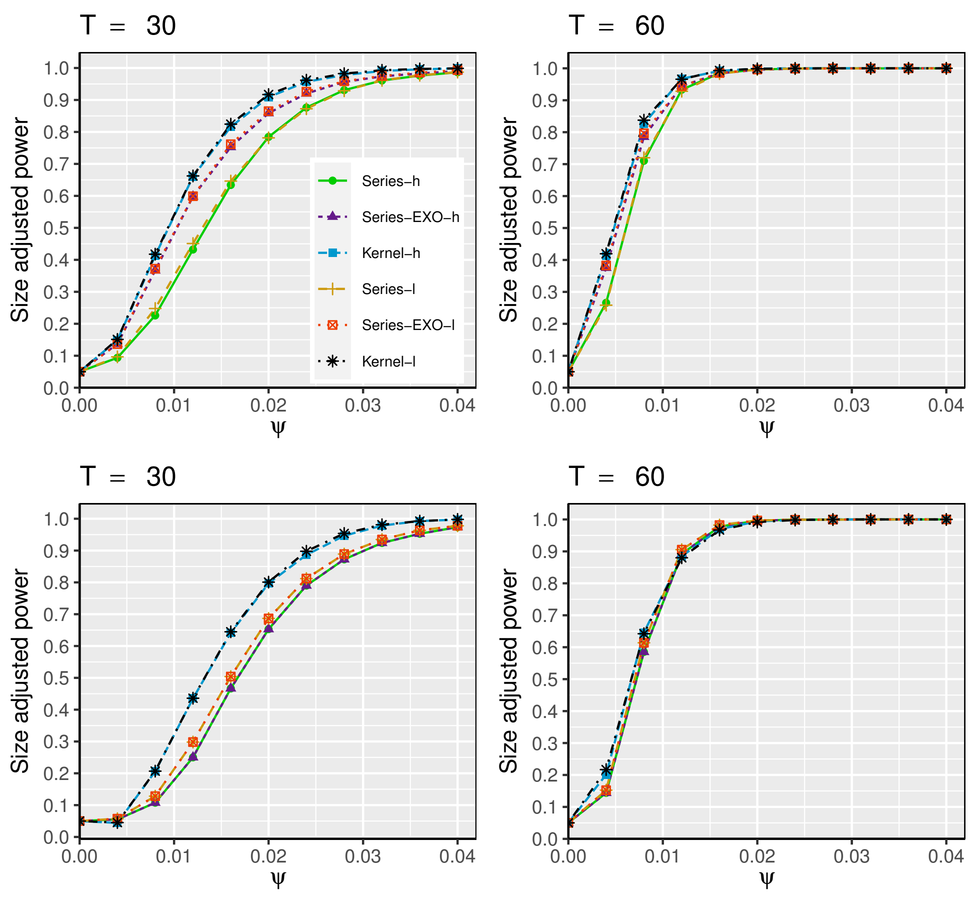

Figure 8: Size-adjusted powers in the nonstationary setting. In the upper row, the explanatory variables are exogenous ($\varphi = 0$). In the lower row the explanatory variables are endogenous ($\varphi = 0.75$).

8 Appendix of Proofs

Proof of Lemma 2.1. We start by writing

$$\frac{1}{T}\int_0^T \phi_j\left(\frac{t}{T}\right)Z_t dt = \frac{1}{T}\sum_{i=1}^n \int_{(i-1)\delta}^{i\delta} \phi_j\left(\frac{t}{T}\right)Z_t dt, \tag{19}$$

and

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right)z_i &= \frac{1}{T}\sum_{i=1}^n \delta\phi_j\left(\frac{i-1}{n}\right)Z_{(i-1)\delta} + \frac{\delta}{T}\left[\phi_j\left(1\right)Z_T - \phi_j\left(0\right)Z_0\right] \\ &= \frac{1}{T}\sum_{i=1}^n \delta\phi_j\left(\frac{i-1}{n}\right)Z_{(i-1)\delta} + O_p\left(\frac{\delta}{T}\sup_{t\in[0,T]}\|Z_t\|\right).\end{aligned}$$

So,

$$\begin{aligned}&\frac{1}{T}\int_0^T \phi_j\left(\frac{t}{T}\right)Z_t dt - \frac{1}{n}\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right)z_i \\ &= \frac{1}{T}\sum_{i=1}^n \int_{(i-1)\delta}^{i\delta}\left[\phi_j\left(\frac{t}{T}\right)Z_t - \phi\left(\frac{i-1}{n}\right)Z_{(i-1)\delta}\right]dt + O_p\left(\frac{\delta}{T}\sup_{t\in[0,T]}\|Z_t\|\right) \\ &= \frac{1}{T}\sum_{i=1}^n \int_{(i-1)\delta}^{i\delta}\phi_j\left(\frac{t}{T}\right)\left[Z_t - Z_{(i-1)\delta}\right]dt \\ &+ \frac{1}{T}\sum_{i=1}^n \int_{(i-1)\delta}^{i\delta}\left[\phi_j\left(\frac{t}{T}\right) - \phi_j\left(\frac{i-1}{n}\right)\right]Z_{(i-1)\delta}dt + O_p\left(\frac{\delta}{T}\sup_{t\in[0,T]}\|Z_t\|\right).\end{aligned}$$

Using

$$\left\|Z_t - Z_{(i-1)\delta}\right\| \le \left\|Z_t^c - Z_{(i-1)\delta}^c\right\| + \sum_{(i-1)\delta<\tau\le t}\|\Delta Z_\tau\|$$

and Assumptions 2.1 and 2.2, we have

$$\begin{aligned}&\frac{1}{T}\sum_{i=1}^n \int_{(i-1)\delta}^{i\delta}\left\|\phi_j\left(\frac{t}{T}\right)\left[Z_t - Z_{(i-1)\delta}\right]\right\|dt \\ &\le \frac{1}{T}\sum_{i=1}^n \int_{(i-1)\delta}^{i\delta}\left|\phi_j\left(\frac{t}{T}\right)\right|\sup_{\|\tilde{\tau}-\tau\|\le\delta}\|Z_{\tilde{\tau}}^c - Z_\tau^c\|dt + \frac{1}{T}\sum_{i=1}^n \int_{(i-1)\delta}^{i\delta}\left|\phi_j\left(\frac{t}{T}\right)\right|\sum_{(i-1)\delta<\tau\le i\delta}\|\Delta Z_\tau\|dt \\ &\le \frac{1}{T}\sum_{i=1}^n \int_{(i-1)\delta}^{i\delta}\sup_{\|\tilde{\tau}-\tau\|\le\delta}\|Z_{\tilde{\tau}}^c - Z_\tau^c\|dt + \frac{\delta}{T}\sum_{\tau=0}^T\|\Delta Z_\tau\|\max_{r\in[0,1]}|\phi_j(r)| \\ &= O_p\left(\Delta_{\delta,T}(Z)\right) + O_p(\delta).\end{aligned}$$

In addition, for some $i^* \in (i-1, i]$,

$$\begin{aligned}
&\frac{1}{T}\sum_{i=1}^{n}\int_{(i-1)\delta}^{i\delta}\left\|\left[\phi_j\left(\frac{t}{T}\right)-\phi_j\left(\frac{i-1}{n}\right)\right]Z_{(i-1)\delta}\right\|dt \\
&\leq \frac{1}{T}\sum_{i=1}^{n}\int_{(i-1)\delta}^{i\delta}\frac{1}{n}\left|\dot{\phi}_j\left(\frac{t^*}{n}\right)\right|\left\|Z_{(i-1)\delta}\right\|dt \\
&\leq \max_{r\in[0,1]}\left|\dot{\phi}_j\left(r\right)\right|\cdot\frac{\delta}{T}\sup_{t\in[0,T]}\|Z_t\| = O_p\left(\frac{\delta}{T}\sup_{t\in[0,T]}\|Z_t\|\right),
\end{aligned}$$

where $\dot{\phi}_j(\cdot)$ is the first order derivative of $\phi_j(\cdot)$. Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)z_i-\frac{1}{T}\int_0^T\phi_j\left(\frac{t}{T}\right)Z_t dt = O_p\left(\Delta_{\delta,T}(Z)+\frac{\delta}{T}\sup_{t\in[0,T]}\|Z_t\|+\delta\right)=O_p\left(e_{\delta,T}(Z)\right).$$

■

Proof of Lemma 2.2. We have shown that $\sqrt{T}(\hat{\beta}_D-\beta)=\sqrt{T}(\hat{\beta}_C-\beta)+o_p(1)$. But

$$\sqrt{T}(\hat{\beta}_C-\beta)=\left[\frac{1}{T}\int_0^T X_tX_t'dt\right]^{-1}\left[\frac{1}{\sqrt{T}}\int_0^T X_tU_tdt\right]\Rightarrow S^{-1}\Omega^{1/2}W_d(1),$$

using Assumptions 2.3 and 2.4. Hence $\sqrt{T}(\hat{\beta}_D-\beta)\Rightarrow S^{-1}\Omega^{1/2}W_d(1)$.

For the second part of the lemma, we use the first part of the lemma and Lemma 2.1 to obtain

$$\begin{aligned}
&\frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_i\hat{u}_i \\
&=\frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_i\left[u_i-x_i'\left(\hat{\beta}_D-\beta\right)\right] \\
&=\frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_iu_i+\frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_ix_i'\cdot O_p\left(\frac{1}{\sqrt{T}}\right). \\
&=\frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_iu_i+\frac{1}{n}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)x_ix_i'\cdot O_p(1) \\
&=\frac{1}{\sqrt{T}}\int_0^T\phi_j\left(\frac{t}{T}\right)X_tU_tdt+o_p(1)
\end{aligned}$$

where we have used $\Lambda(n,\delta)\sqrt{T}=n$, Assumption 2.3, and Assumption 2.5(i). Under Assumption 2.4, we then have

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

for each $j=1,2,\ldots,K$. The joint convergence over $j=1,2,\ldots,K$ holds by the Cramér–Wold theorem. ■

Proof of Theorem 2.2. Part (a): For notational simplicity, we assume that $\Omega^{1/2}$ is symmetric. 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 2.6(i), we have

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

Part (b): Before computing the bias when $T\to\infty$ and $K\to\infty$, we first show that $\delta\sum_{k=-\infty}^{\infty}|k\delta|^m\left\|\Gamma_{XU}\left(k\delta\right)\right\|<\infty$ for $m=0,1,2$. Using Assumption 2.6(ii), we have

$$\begin{aligned}
\delta\sum_{|k|<n}|k\delta|^m\left\|\Gamma_{XU}\left(k\delta\right)\right\|&=\delta\sum_{|k\delta|\le C_2}|k\delta|^m\left\|\Gamma_{XU}\left(k\delta\right)\right\|+\delta\sum_{C_2<|k\delta|<n}|k\delta|^m\left\|\Gamma_{XU}\left(k\delta\right)\right\|\\
&\le\delta\sum_{|k\delta|\le C_2}C_2^mC_1+C_1\delta\sum_{C_2<|k\delta|<n}|k\delta|^m\left(k\delta\right)^{-(3+\iota)}\\
&=2C_2^mC_1\delta\cdot\frac{C_2}{\delta}+C_1\delta^{m-2-\epsilon}\sum_{C_2<|k\delta|<n}|k|^{-(3-m+\iota)}\\
&\le 2C_2^{m+1}C_1+C_1\delta^{m-2-\epsilon}\cdot O\left(\frac{C_2}{\delta}\right)^{1-(3-m+\iota)}\\
&=2C_2^{m+1}C_1+C_1\cdot O\left(C_2^{1-(3-m+\iota)}\right)=O\left(1\right).
\end{aligned}$$

So we have $\delta\sum_{k=-\infty}^{\infty}|k\delta|^m\left\|\Gamma_{XU}\left(k\delta\right)\right\|<\infty$. By the same argument, Assumption 2.6(ii) implies that $\int_{-\infty}^{\infty}|\tau|^m\left\|\Gamma_{XU}\left(\tau\right)\right\|<\infty$ for $m=0,1,2$.

Next, we compute the bias of $\hat{\Omega}^*$ when $T\to\infty$ and $K\to\infty$. Denote $E\left[\left(x_iu_i\right)\left(x_\ell u_\ell\right)'\right]=$

$\Gamma_{xu}(i-\ell)$. Note that

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

where

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

The bias is then equal to

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

For $\mathcal{B}_{2n}$, we use Assumption 2.6(iii) with $m=0$ to obtain:

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

where the $O\left(T^{-2}\right)$ term holds because under Assumption 2.6(ii),

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

For $\mathcal{B}_{1n}$, we have, using $\sup_{r\in[0,1]}\left|\dot{\phi}_j(r)\right| \leq jC$ in Assumption 2.6(iv):

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

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

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

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

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

where

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

Now,

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

where

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

and

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

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

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

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 Assumptions 2.6(iv) and (v), we have

$$\begin{aligned}\tilde{\mathcal{B}}_{12,n} &= \frac{K^2}{T^2}\frac{1}{K^3}\sum_{j=1}^{K}\frac{1}{2}\ddot{\omega}_j(0)\,\delta\sum_{|k|\le n/\log n}(k\delta)^2\,\Gamma_{XU}(k\delta)\\
&= \frac{K^2}{T^2}\left(\frac{1}{K^3}\sum_{j=1}^{K}\frac{1}{2}\int_0^1\phi_j(r)\,\ddot{\phi}_j(r)\,dr\right)\int_{-\infty}^{\infty}\tau^2\Gamma_{XU}(\tau)\,d\tau\,(1+o(1))+O\left(\frac{1}{T}\right)\\
&= -\frac{K^2}{T^2}c_\phi\int_{-\infty}^{\infty}\tau^2\Gamma_{XU}(\tau)\,d\tau\,(1+o(1))+O\left(\frac{1}{T}\right)\end{aligned}$$

as $K\to\infty$ and $T\to\infty$.

Combining the above results yields the asymptotic bias formula for the case where $K\to\infty$ and $T\to\infty$.

Part (c): As in the proof of Part (b), we have

$$\mathcal{B}_n = E(\hat{\Omega}^*)-\Omega := \tilde{B}_{12,n}+O\left(\frac{K}{n}+\delta+\frac{1}{T^2}+\frac{(\log n)^2}{T^2}\right),$$

where

$$\tilde{\mathcal{B}}_{12,n}=\frac{1}{K}\sum_{j=1}^{K}\delta\sum_{|k|\le n/\log n}\left[\omega_j\left(\frac{k}{n}\right)-1\right]\Gamma_{XU}(k\delta).$$

For the rest of the proof, we use arguments different from that for Part (b). Using $\omega_j(0)=1$ and $\dot{\omega}_j(0)=-\frac{1}{2}\left[\phi_j^2(1)+\phi_j^2(0)\right]$, we have

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

Therefore,

$$\begin{aligned}\mathcal{B}_n = &-\frac{1}{2}\frac{1}{T}\left(\frac{1}{K}\sum_{j=1}^{K}\left[\phi_j^2(1)+\phi_j^2(0)\right]\right)\int_{-\infty}^{\infty}\tau\Gamma_{XU}(\tau)\,d\tau\\
&+o\left(\frac{1}{T}\right)+O\left(\frac{1}{n}+\delta+\frac{(\log n)^2}{T^2}\right).\end{aligned}$$

■

Proof of Lemma 3.1. Part (a). We first consider $n^{-1/2}\mathbb{W}^x\Lambda_T^{-1}$. Let $g_n : \mathbb{D}^d[0,1] \to \mathbb{D}^d[0,1]$ be defined by

$$g_n(f)(t) = \sum_{i=1}^{n} f\left(\frac{i}{n}\right) 1\left\{t \in \left[\frac{i-1}{n}, \frac{i}{n}\right)\right\} + f(1)1\{t=1\}.$$

If the functions $f_n \in \mathbb{D}^d[0,1]$ are such that $f_n \to f$ for a continuous function f, then the continuity of ϕ_j in Assumption 2.2 implies that $\phi_j(\cdot) f_n(\cdot) \to \phi_j(\cdot) f(\cdot)$ in $\mathbb{D}^d[0,1]$ and $\phi_j(\cdot) f(\cdot)$ is a continuous function. It follows from basic properties of the Skorokhod topology that $g_n(\phi_j f_n) \to \phi_j f$. Using the weak convergence $\Lambda_T^{-1} X_{Tr} \Rightarrow X^\circ(r)$ in Assumption 3.2 and the extended continuous mapping theorem (c.f. Theorem 1.11.1 of van der Vaart and Wellner (1996)), we have $g_n(\phi_j(t)\left(\Lambda_T^{-1}X_{Tt}\right)) \Rightarrow \phi_j(t)X^\circ(t),\ t \in [0,1]$. Combining this with the continuous mapping theorem, we have

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

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

$$\frac{1}{\sqrt{n}}\mathbb{W}^x\Lambda_T^{-1} \Rightarrow \eta. \tag{20}$$

Next, under Assumption 3.1, Lemma 2.1 holds with $Z_t = U_{0t}$. Hence,

$$\begin{aligned}
\sqrt{\delta}\mathbb{W}_j^{u_0} &= \frac{\sqrt{\delta}}{\sqrt{n}}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)u_{0i} = \frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)u_{0i} \\
&= \frac{1}{\sqrt{T}}\int_0^T \phi_j\left(\frac{t}{T}\right)U_{0t}dt + o_p(1).
\end{aligned}$$

Let $S_t = T^{-1/2}\int_0^t U_{0r}dr$ for $t \in (0,T]$ and $S_0 = 0$. Using the continuous mapping theorem and integration by parts, we obtain, jointly for $j = 1, \ldots, K$,

$$\begin{aligned}
&\sqrt{\delta}\mathbb{W}_j^{u_0} = \int_0^T \phi_j\left(\frac{t}{T}\right)dS_t + o_p(1) \\
&= \int_0^1 \phi_j(r)\, dS_{Tr} + o_p(1) = \phi_j(1)\, S_T - \phi_j(0)\, S_0 - \int_0^1 S_{Tr}\dot{\phi}_j(r)\, dr + o_p(1) \\
&\Rightarrow \sigma_0\phi_j(1)\, W_0(1) - \sigma_0\phi_j(0)\, W_0(0) - \sigma_0\int_0^1 \dot{\phi}_j(r)\, W_0(r)dr \\
&= \sigma_0\int_0^1 \phi_j(r)\, dW_0(r),
\end{aligned}$$

where the weak convergence follows from Assumption 3.2. Therefore,

$$\sqrt{\delta}\mathbb{W}^{u_0} \Rightarrow \nu. \tag{21}$$

The joint convergence of $\Lambda_T^{-1} X_{Tt}$ and $T^{-1/2}\int_0^{Tt} U_{0r}dr$ in Assumption 3.2 yields that (20) and (21) hold jointly, i.e., $(n^{-1/2}\mathbb{W}^x\Lambda_T^{-1}, \sqrt{\delta}\mathbb{W}^{u_0}) \Rightarrow (\eta, \nu)$.

Part (b). We write

$$\mathbb{W}^y = \mathbb{W}^x\beta_0 + \mathbb{W}^{u_0} + \alpha_0\mathbb{W}^\alpha \tag{22}$$

where

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

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,

$$\mathbb{W}^y = \mathbb{W}^x\beta_0 + \mathbb{W}^{u_0} + o_p(1). \tag{23}$$

It then follows that

$$\hat\beta_{TOLS} = \left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1}\left(\mathbb{W}^{x\prime}\left[\mathbb{W}^x\beta_0 + \mathbb{W}^{u_0} + o_p(1)\right]\right), \tag{24}$$

and so

$$\hat\beta_{TOLS} - \beta_0 = \left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1}\left(\mathbb{W}^{x\prime}\left[\mathbb{W}^{u_0} + o_p(1)\right]\right).$$

By Part (a) and Assumption 3.3, we then have

$$\begin{aligned}&\sqrt{T}\Lambda_T\left[\hat\beta_{TOLS} - \beta_0\right] \\ &= n^{1/2}\Lambda_T\sqrt{\delta}\left[\hat\beta_{TOLS} - \beta_0\right] \\ &= \left[\left(n^{1/2}\Lambda_T\right)^{-1}\left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)\left(n^{1/2}\Lambda_T\right)^{-1}\right]^{-1}\left(n^{1/2}\Lambda_T\right)^{-1}\mathbb{W}^{x\prime}\mathbb{W}^{u_0}\sqrt{\delta}\left(1 + o_p(1)\right) \\ &\Rightarrow \left(\eta'\eta\right)^{-1}\left(\eta'\nu\right).\end{aligned}$$

■

Proof of Theorem 3.1. By definition, $\hat{\mathbb{W}}^u = \mathbb{W}^y - \mathbb{W}^x\hat\beta_{TOLS}$. Using (23) and (24), we then have

$$\begin{aligned}\hat{\mathbb{W}}^u &= \mathbb{W}^x\beta_0 + \mathbb{W}^{u_0} + o_p(1) - \mathbb{W}^x\left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1}\mathbb{W}^{x\prime}\left[\mathbb{W}^x\beta_0 + \mathbb{W}^{u_0} + o_p(1)\right] \\ &= \left[\mathbb{I}_K - \mathbb{W}^x\left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1}\mathbb{W}^{x\prime}\right]\left(\mathbb{W}^{u_0} + o_p(1)\right).\end{aligned} \tag{25}$$

Hence, by Lemma 3.1(i),

$$\begin{aligned}\delta\cdot\hat\sigma_0^2 &= \frac{1}{K}\sqrt{\delta}\left(\mathbb{W}^{u_0} + o_p(1)\right)'\left[\mathbb{I}_K - \mathbb{W}^x\left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1}\mathbb{W}^{x\prime}\right]\sqrt{\delta}\left(\mathbb{W}^{u_0} + o_p(1)\right) \\ &\Rightarrow \frac{1}{K}\nu' M_\eta\nu.\end{aligned}$$

where $M_\eta = \mathbb{I}_K - \eta(\eta'\eta)^{-1}\eta'$. Using Lemma 3.1(ii), we have, under H_0,

$$\sqrt{T}\tilde{\Lambda}_T(R\hat{\beta}_{TOLS} - r) = (\tilde{\Lambda}_T R \Lambda_T^{-1})\sqrt{T}\Lambda_T(\hat{\beta}_{TOLS} - \beta_0) \Rightarrow R_\circ \left(\eta'\eta\right)^{-1}\left(\eta'\nu\right)$$

and

$$\begin{aligned}
& n\tilde{\Lambda}_T^{-1}\left[R\left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1}R'\right]^{-1}\tilde{\Lambda}_T^{-1} \\
&= n\tilde{\Lambda}_T^{-1}\left\{Rn^{1/2}\Lambda_T^{-1}\left[\left(\mathbb{W}^x\Lambda_T^{-1}n^{-1/2}\right)'\mathbb{W}^x\Lambda_T^{-1}n^{-1/2}\right]^{-1}\left(Rn^{1/2}\Lambda_T^{-1}\right)'\right\}^{-1}\tilde{\Lambda}_T^{-1} \\
&= \left\{\tilde{\Lambda}_T R\Lambda_T^{-1}\left[\left(\mathbb{W}^x\Lambda_T^{-1}n^{-1/2}\right)'\mathbb{W}^x\Lambda_T^{-1}n^{-1/2}\right]^{-1}\left(\tilde{\Lambda}_T R\Lambda_T^{-1}\right)'\right\}^{-1} \\
&\Rightarrow [R_\circ\left(\eta'\eta\right)^{-1}R_\circ']^{-1}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
F_{TOLS} &= \frac{1}{\hat{\sigma}_0^2}(R\hat{\beta}_{TOLS} - r)'\left[R\left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1}R'\right]^{-1}(R\hat{\beta}_{TOLS} - r)/p \\
&= \frac{1}{p}\frac{1}{\delta\hat{\sigma}_0^2}(R\hat{\beta}_{TOLS} - r)'\sqrt{T}\tilde{\Lambda}_T \\
&\quad \times n\tilde{\Lambda}_T^{-1}\left[R\left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1}R'\right]^{-1}\tilde{\Lambda}_T^{-1} \times \sqrt{T}\tilde{\Lambda}_T(R\hat{\beta}_{TOLS} - r) \\
&\Rightarrow \frac{K}{p}\frac{\left[R_\circ(\eta'\eta)^{-1}\eta'\nu\right]'\left(R_\circ\left(\eta'\eta\right)^{-1}R_\circ'\right)^{-1}\left[R_\circ(\eta'\eta)^{-1}\eta'\nu\right]}{\nu' M_\eta \nu} \\
&= \frac{K}{p}\frac{Q'\left(R_\circ\left(\eta'\eta\right)^{-1}R_\circ'\right)^{-1}Q}{\nu' M_\eta\nu/\sigma_0^2}, \qquad (26)
\end{aligned}$$

where $Q = R_\circ(\eta'\eta)^{-1}\eta'\nu/\sigma_0$. Now, conditional on η,

$$Q'\left(R_\circ\left(\eta'\eta\right)^{-1}R_\circ'\right)^{-1}Q \overset{d}{=} \chi_p^2, \text{ and } \nu' M_\eta \nu/\sigma_0^2 \overset{d}{=} \chi_{K-d}^2.$$

Additionally, conditional on η, $M_\eta\nu$ and $\eta'\nu$ are independent, as both $M_\eta\nu$ and $\eta'\nu$ are normal and the conditional covariance is

$$cov\left(M_\eta\nu, \eta'\nu\right) = M_\eta\eta = 0.$$

Thus, conditional on η, the numerator and the denominator in (26) are independent chi-squared variates. This implies that

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

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

■

Proof of Theorem 3.2. Part (a): Setting $\Lambda_T = \sqrt{T}\mathbb{I}_d$ and $X^{\circ}(r) = B_x(r)$ we can proceed nearly identically to the proof of Lemma 3.1(a) to obtain that

$$\left[(nT)^{-1/2}\,\mathbb{W}^x, \delta^{1/2}\mathbb{W}^{u_0}\right] \Rightarrow (\eta, \nu).$$

It remains to show that $\delta^{1/2}\mathbb{W}^{\tilde{\Delta}x} \Rightarrow \xi$ jointly with the above convergence. The joint convergence holds by the Cramér–Wold theorem. It remains to prove the marginal convergence $\delta^{1/2}\mathbb{W}^{\tilde{\Delta}x} \Rightarrow \xi$. We have

$$\begin{aligned}
\delta^{1/2}\mathbb{W}_j^{\tilde{\Delta}x} &= \frac{1}{\sqrt{n\delta}}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)[x_i - x_{i-1}] = \sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)T^{-1/2}[x_i - x_{i-1}] \\
&= \frac{1}{n}\sum_{i=1}^{n-1}\frac{\left[\phi_j\left(\frac{i}{n}\right) - \phi_j\left(\frac{i+1}{n}\right)\right]}{1/n}T^{-1/2}x_i + \phi_j(1)T^{-1/2}x_n - \phi_j\left(\frac{1}{n}\right)T^{-1/2}x_0 \\
&= -\frac{1}{n}\sum_{i=1}^{n-1}\dot{\phi}_j\left(\frac{i}{n}\right)T^{-1/2}X_{\frac{i}{n}T} + \phi_j(1)T^{-1/2}X_T - \phi_j\left(\frac{1}{n}\right)T^{-1/2}X_0 \\
&+ O_p\left(\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n-1}T^{-1/2}\left\|X_{\frac{i}{n}T}\right\|\right). \qquad (27)
\end{aligned}$$

Using the continuous mapping theorem and Assumption 3.4, we have $n^{-1}\sum_{i=1}^{n-1}T^{-1/2}\left\|X_{\frac{i}{n}T}\right\| \Rightarrow \int_0^1\|B_x(r)\|\,dr$ and hence the last term in (27) is of order $O_p(1/n) = o_p(1)$. Therefore, using integration by parts,

$$\begin{aligned}
&\delta^{1/2}\mathbb{W}_j^{\tilde{\Delta}x} \\
&= \frac{1}{n}\sum_{i=1}^{n-1}\dot{\phi}_j\left(\frac{i}{n}\right)T^{-1/2}X_{\frac{i}{n}T} + \phi_j(1)T^{-1/2}X_T - \phi_j\left(\frac{1}{n}\right)T^{-1/2}X_0 + o_p(1) \\
&\Rightarrow -\int_0^1\dot{\phi}_j(r)B_x(r)dr + \phi_j(1)B_x(1) - \phi_j(0)B_x(0) \\
&= \int_0^1\phi_j(r)dB_x(r) = \xi_j.
\end{aligned}$$

This holds jointly for $j = 1, \ldots, K$ so that $\delta^{1/2}\mathbb{W}^{\tilde{\Delta}x} \Rightarrow \xi$.

Part (b). Following the same argument as in the proof Theorem 3.1, we can ignore the intercept. To simplify the notation, we assume from the outset that there is no intercept in the model so that

$$\mathbb{W}^y = \mathbb{W}^x\beta_0 + \mathbb{W}^{u_0}.$$

Given this, we have

$$\hat{\gamma} - \begin{pmatrix}\beta_0 \\ 0\end{pmatrix} = \begin{pmatrix}\mathbb{W}^{x\prime}\mathbb{W}^x & \mathbb{W}^{x\prime}\mathbb{W}^{\tilde{\Delta}x} \\ \mathbb{W}^{\tilde{\Delta}x\prime}\mathbb{W}^x & \mathbb{W}^{\tilde{\Delta}x\prime}\mathbb{W}^{\tilde{\Delta}x}\end{pmatrix}^{-1}\begin{pmatrix}\mathbb{W}^{x\prime}\mathbb{W}^{u_0} \\ \mathbb{W}^{\tilde{\Delta}x\prime}\mathbb{W}^{u_0}\end{pmatrix}.$$

Recall that $\Upsilon_T = diag\left(T\mathbb{I}_d, \mathbb{I}_d\right)$. Using Part (a) and noting that $\delta^{1/2}/T = (nT)^{-1/2}$, we have

$$
\begin{aligned}
&\Upsilon_T\left[\hat{\gamma} - \begin{pmatrix} \beta_0 \\ 0 \end{pmatrix}\right] \\
&= \left[\delta^{1/2}\Upsilon_T^{-1}\begin{pmatrix} (\mathbb{W}^x)'\,\mathbb{W}^x & \mathbb{W}^{x\prime}\mathbb{W}^{\tilde{\Delta}x} \\ (\mathbb{W}^{\tilde{\Delta}x})'\mathbb{W}^x & \mathbb{W}^{\tilde{\Delta}x\prime}\mathbb{W}^{\tilde{\Delta}x} \end{pmatrix}\Upsilon_T^{-1}\delta^{1/2}\right]^{-1}\Upsilon_T^{-1}\delta^{1/2}\begin{pmatrix} \mathbb{W}^{x\prime}\mathbb{W}^{u_0}\delta^{1/2} \\ \mathbb{W}^{\tilde{\Delta}x\prime}\mathbb{W}^{u_0}\delta^{1/2} \end{pmatrix} \\
&= \begin{pmatrix} (nT)^{-1/2}\,\mathbb{W}^{x\prime}\mathbb{W}^x\,(nT)^{-1/2} & (nT)^{-1/2}\,\mathbb{W}^{x\prime}\mathbb{W}^{\tilde{\Delta}x}\delta^{1/2} \\ \delta^{1/2}\mathbb{W}^{\tilde{\Delta}x}\mathbb{W}^x\,(nT)^{-1/2} & \delta^{1/2}\mathbb{W}^{\tilde{\Delta}x\prime}\mathbb{W}^{\tilde{\Delta}x}\delta^{1/2} \end{pmatrix}^{-1}\begin{pmatrix} (nT)^{-1/2}\,\mathbb{W}^{x\prime}\mathbb{W}^{u_0}\delta^{1/2} \\ \delta^{1/2}\mathbb{W}^{\tilde{\Delta}x\prime}\mathbb{W}^{u_0}\delta^{1/2} \end{pmatrix} \\
&\Rightarrow \begin{pmatrix} \eta'\eta & \eta'\xi \\ \xi'\eta & \xi'\xi \end{pmatrix}^{-1}\begin{pmatrix} \eta'\nu \\ \xi'\nu \end{pmatrix}.
\end{aligned}
$$

Plugging $\nu = \sigma_{0\cdot x}\tilde{\nu} + \xi\theta_0$ into the above limit, we have

$$
\begin{aligned}
\Upsilon_T\left[\hat{\gamma} - \begin{pmatrix} \beta_0 \\ 0 \end{pmatrix}\right] &\Rightarrow \begin{pmatrix} \eta'\eta & \eta'\xi \\ \xi'\eta & \xi'\xi \end{pmatrix}^{-1}\begin{pmatrix} \eta'\xi \\ \xi'\xi \end{pmatrix}\theta_0 + \sigma_{0\cdot x}\begin{pmatrix} \eta'\eta & \eta'\xi \\ \xi'\eta & \xi'\xi \end{pmatrix}^{-1}\begin{pmatrix} \eta'\tilde{\nu} \\ \xi'\tilde{\nu} \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ \theta_0 \end{pmatrix} + \sigma_{0\cdot x}\left(\zeta'\zeta\right)^{-1}\zeta'\tilde{\nu}.
\end{aligned}
$$

That is, $\Upsilon_T\left(\hat{\gamma} - \gamma_0\right) \Rightarrow \sigma_{0\cdot x}\left(\zeta'\zeta\right)^{-1}\zeta'\tilde{\nu}$. The first block of this result is $T(\hat{\beta}_{TAOLS} - \beta_0) \Rightarrow \sigma_{0\cdot x}\left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \tilde{\nu}$.

Part (c). First, it follows from Part (b) that under H_0,

$$
T(R\hat{\beta}_{TAOLS} - r) = RT(\hat{\beta}_{TAOLS} - \beta_0) \Rightarrow \sigma_{0\cdot x}R\left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \tilde{\nu}. \tag{28}
$$

Next,

$$
\begin{aligned}
\delta(\hat{\mathbb{W}}^{u_{0\cdot x}})'\hat{\mathbb{W}}^{u_{0\cdot x}} &= \delta\mathbb{W}^{u_0\prime}\left[I - \widetilde{\mathbb{W}}(\widetilde{\mathbb{W}}'\widetilde{\mathbb{W}})^{-1}\widetilde{\mathbb{W}}'\right]\mathbb{W}^{u_0} \\
&= (\sqrt{\delta}\mathbb{W}^{u_0})'\left[I - (\widetilde{\mathbb{W}}\delta^{1/2}\Upsilon_T^{-1})\left[(\widetilde{\mathbb{W}}\delta^{1/2}\Upsilon_T^{-1})(\widetilde{\mathbb{W}}\delta^{1/2}\Upsilon_T^{-1})'\right]^{-1}(\widetilde{\mathbb{W}}\delta^{1/2}\Upsilon_T^{-1})'\right]\sqrt{\delta}\mathbb{W}^{u_0} \\
&\Rightarrow \nu'\left(I - \zeta\left(\zeta'\zeta\right)^{-1}\zeta'\right)\nu = \sigma_{0\cdot x}^2\tilde{\nu}'\left(I - \zeta\left(\zeta'\zeta\right)^{-1}\zeta'\right)\tilde{\nu} = \sigma_{0\cdot x}^2\tilde{\nu}' M_\zeta \tilde{\nu}.
\end{aligned}
$$

Hence,

$$
\delta\hat{\sigma}_{0\cdot x}^2 = \frac{1}{K}\delta(\hat{\mathbb{W}}^{u_{0\cdot x}})'\hat{\mathbb{W}}^{u_{0\cdot x}} \Rightarrow \frac{1}{K}\sigma_{0\cdot x}^2\tilde{\nu}' M_\zeta \tilde{\nu}. \tag{29}
$$

Combining (28) and (29), we have

$$
\begin{aligned}
&F_{TAOLS} \\
&= \frac{1}{\hat{\sigma}_{0\cdot x}^2}(R\hat{\beta}_{TAOLS} - r)'\left[R\left(\mathbb{W}^{x\prime}M_{\tilde{\Delta}x}\mathbb{W}^x\right)^{-1}R'\right]^{-1}(R\hat{\beta}_{TAOLS} - r)/p \\
&= \frac{1}{p\delta\hat{\sigma}_{0\cdot x}^2}[RT(\hat{\beta}_{TAOLS} - \beta_0)]'\{R[(nT)^{-1/2}\,\mathbb{W}^{x\prime}M_{\tilde{\Delta}x}\mathbb{W}^x\,(nT)^{-1/2}]^{-1}R'\}^{-1}RT(\hat{\beta}_{TAOLS} - \beta_0) \\
&\Rightarrow \frac{\left[R\left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \tilde{\nu}\right]'\left[R\left(\eta' M_\xi \eta\right)^{-1}R'\right]^{-1}\left[R\left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \tilde{\nu}\right]/p}{\tilde{\nu}' M_\zeta \tilde{\nu}/K} \\
&= \frac{K}{p}\frac{Q'\left(R\left(\eta' M_\xi \eta\right)^{-1}R'\right)^{-1}Q}{\tilde{\nu}' M_\zeta \tilde{\nu}},
\end{aligned}
$$

where $Q = R(\eta' M_\xi \eta)^{-1}\eta' M_\xi \tilde{\nu}$. Following the argument similar to that in the proof of Theorem 3.1, we can then show that $F_{TAOLS} \Rightarrow \frac{K}{K-2d} \cdot F_{p,K-2d}$. ■

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

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

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

$$\begin{aligned}
& E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_j\left(\frac{i}{n}\right) x_i\epsilon_i\right]^2 \\
&= E\left\{\frac{1}{n}\sum_{i=1}^n \left[\phi_j\left(\frac{i}{n}\right)\right]^2 \epsilon_i^2 x_i^2 + \frac{2}{n}\sum_{\ell=1}^{n-1}\sum_{i=1}^{n-\ell} \epsilon_i\epsilon_{i+\ell}x_i x_{i+\ell}\phi_j\left(\frac{i}{n}\right)\phi_j\left(\frac{i+\ell}{n}\right)\right\} \\
&\leq 2\max_{r\in[0,1]}\phi_j^2(r)\,\text{var}(\epsilon_1)\left(\frac{1}{n}\sum_{i=1}^n E\left[X_{i\delta}^2\right] + \frac{2}{n}\sum_{\ell=1}^{n-1}\sum_{i=1}^{n-\ell}\varphi_\ell^{1/2}\left|E\left[X_{i\delta}X_{(i+\ell)\delta}\right]\right|\right) \\
&= 2\max_{r\in[0,1]}\phi_j^2(r)\,\text{var}(\epsilon_1)\left(\Gamma_X(0) + \frac{2}{n}\sum_{\ell=1}^{n-1}\sum_{i=1}^{n-\ell}\varphi_\ell^{1/2}\left|\Gamma_X(\ell\delta)\right|\right) \\
&\leq 2C\max_{r\in[0,1]}\phi_j^2(r)\,\text{var}(\epsilon_1)\left[1 + 2\sum_{\ell=1}^{n-1}\left(1-\frac{\ell}{n}\right)\varphi_\ell^{1/2}\right] \\
&\leq 2C\max_{r\in[0,1]}\phi_j^2(r)\,\text{var}(\epsilon_1)\left(1 + 2\sum_{\ell=1}^{\infty}\varphi_\ell^{1/2}\right) < \infty.
\end{aligned}$$

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

Aït-Sahalia, Y. and Yu, J. (2009). High frequency market microstructure noise estimates and liquidity measures. *The Annals of Applied Statistics*, 3(1):422 – 457.

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.

Bhattacharya, R. N. (1982). On the functional central limit theorem and the law of the iterated logarithm for markov processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 60(2):185–201.

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.

Comte, F. (1999). Discrete and continuous time cointegration. *Journal of Econometrics*, 88(2):207–226.

Fischer, M. and Nappo, G. (2009). On the moments of the modulus of continuity of itô processes. *Stochastic Analysis and Applications*, 28(1):103–122.

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.

Kim, J. and Park, J. Y. (2017). Asymptotics for recurrent diffusions with application to high frequency regression. *Journal of Econometrics*, 196(1):37–54.

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.

Rozanov, Y. A. (1960). A central limit theorem for additive random functions. *Theory of Probability & Its Applications*, 5(2):221–223.

Stock, J. H. and Watson, M. W. (2019). *Introduction to Econometrics (Fourth Edition)*. Pearson.

Sun, Y. (2004). A convergent t-statistic in spurious regressions. *Econometric Theory*, 20(5):943–962.

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. (2021). Some extensions of asymptotic F and t theory in nonstationary regressions. Working paper, Department of Economics, UC San Diego.

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.

Online Supplementary Appendix

In this appendix, we provide sufficient conditions for Assumption 2.6(i) and report the simulation results for the setting with additive measurement noise.

S.1 Sufficient conditions for Assumption 2.6(i)

For notational simplicity, we consider the case that $v_i = x_i u_i$ is a scalar. The vector case requires only additional matrix algebra. The underlying continuous time process is $V_t = X_t U_t$. Let $v^* = (v_1^*, ..., v_n^*)$ be a zero-mean Gaussian sequence with the same covariance as $v = (v_1, ..., v_n)$. Then the fourth-order cumulant $\kappa_{v,4}(\ell_1, \ell_2, \ell_3, \ell_4)$ of $\{v_i\}_{i=1}^n$ is defined to be

$$\kappa_{v,4}(\ell_1, \ell_2, \ell_3, \ell_4) = E\left(v_{\ell_1} v_{\ell_1+\ell_2} v_{\ell_1+\ell_3} v_{\ell_1+\ell_4}\right) - E\left(v_{\ell_1}^* v_{\ell_1+\ell_2}^* v_{\ell_1+\ell_3}^* v_{\ell_1+\ell_4}^*\right).$$

We need the following assumption.

Assumption S.1 *(i) v_i is fourth-order stationary with covariance $\Gamma_v(k) = E(v_i v_{i-k})$ and fourth-order cumulant $\kappa_{v,4}(\ell_1, \ell_2, \ell_3, \ell_4)$; (ii) there is a constant C that does not depend on δ or n such that*

$$\delta^3 \sum_{\ell_1=-n+1}^{n-1} \sum_{\ell_2=-n+1}^{n-1} \sum_{\ell_3=-n+1}^{n-1} \left|\kappa_{v,4}(0, \ell_1, \ell_2, \ell_3)\right| < C.$$

Assumption S.1(ii) is the discrete analogue of its continuous counterpart

$$\int_{-T}^{T} \int_{-T}^{T} \int_{-T}^{T} \kappa_{V,4}(0, r_1, r_2, r_3)\, dr_1 dr_2 dr_3 < \infty,$$

where $\kappa_{V,4}$ is the fourth order cumulant of $\{V_t\}$. The above condition is the same as Assumption 2.3(b) in Lu and Park (2019).

Proposition S.1 *Let Assumptions 2.2, 2.6(ii)-(iv), and S.1 hold. If $K^2 = o(n)$ and $K = o(T)$, then as $\delta \to 0$ and $T \to \infty$,*

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

for both a fixed K and a growing K (i.e., $K \to \infty$).

Proof of Proposition S.1. In the following, we write $\sum_{j_1=1}^K \sum_{j_2=1}^K$ as $\sum_{j_1,j_2}$ when there is no possibility of confusion. All results in this proof hold for both a fixed K and large K unless stated otherwise. We have

$$\begin{aligned}
var(\hat{\Omega}^*) &= var\left\{\frac{1}{K}\sum_{j=1}^{K}\left[\frac{1}{\Lambda(n,\delta)}\sum_{i=1}^{n}\phi_j\left(\frac{i}{n}\right)v_i\right]^2\right\} \\
&= \frac{1}{K^2\Lambda(n,\delta)^2}\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4}^{n}\phi_{j_1}(\frac{i_1}{n})\phi_{j_1}(\frac{i_2}{n})\phi_{j_2}(\frac{i_3}{n})\phi_{j_2}(\frac{i_4}{n})E\left[(v_{i_1}v_{i_2} - Ev_{i_1}v_{i_2})(v_{i_3}v_{i_4} - Ev_{i_3}v_{i_4})\right] \\
&= \frac{1}{K^2\Lambda(n,\delta)^2}\sum_{j_1,j_2}E\sum_{i_1=1}^{n}\sum_{k_1=i_1-n}^{i_1-1}\sum_{i_2=1}^{n}\sum_{k_2=i_2-n}^{i_2-1}\phi_{j_1}(\frac{i_1}{n})\phi_{j_1}(\frac{i_1-k_1}{n})\phi_{j_2}(\frac{i_2}{n})\phi_{j_2}(\frac{i_2-k_2}{n}) \\
&\times (v_{i_1}v_{i_1-k_1} - Ev_{i_1}v_{i_1-k_1})(v_{i_2}v_{i_2-k_2} - Ev_{i_2}v_{i_2-k_2}).
\end{aligned}$$

Let

$$\phi_{j_1,j_2,j_3,j_4}(i_1,i_2,k_1,k_2)=\phi_{j_1}(\frac{i_1}{n})\phi_{j_2}(\frac{i_2}{n})\phi_{j_3}(\frac{k_1}{n})\phi_{j_4}(\frac{k_2}{n}),\ \phi_{j_1,j_2,}(i_1,i_2)=\phi_{j_1}(\frac{i_1}{n})\phi_{j_2}(\frac{i_2}{n}),$$
$$\mu_4(i_1,i_2,k_1,k_2)=E\left(v_{i_1}v_{i_1-k_1}v_{i_2}v_{i_2-k_2}\right),\ \mu_4^*(i_1,i_2,k_1,k_2)=E\left(v_{i_1}^*v_{i_1-k_1}^*v_{i_2}^*v_{i_2-k_2}^*\right).$$

Recall that $v^*=(v_1^*,...,v_n^*)$ is a zero-mean Gaussian sequence with the same covariance as $v=(v_1,...,v_n)$. We have

$$\begin{aligned}
&\mu_4^*(i_1,i_2,k_1,k_2)\\
&:=E\left(v_{i_1}^*v_{i_1-k_1}^*v_{i_2}^*v_{i_2-k_2}^*\right)\\
&=E\left(v_{i_1}^*v_{i_1-k_1}^*\right)E\left(v_{i_2}^*v_{i_2-k_2}^*\right)+E\left(v_{i_1}^*v_{i_2}^*\right)E\left(v_{i_1-k_1}^*v_{i_2-k_2}^*\right)+E\left(v_{i_1}^*v_{i_2-k_2}^*\right)E\left(v_{i_1-k_1}^*v_{i_2}^*\right)\\
&=E\left(v_{i_1}v_{i_1-k_1}\right)E\left(v_{i_2}v_{i_2-k_2}\right)+E\left(v_{i_1}v_{i_2}\right)E\left(v_{i_1-k_1}v_{i_2-k_2}\right)+E\left(v_{i_1}v_{i_2-k_2}\right)E\left(v_{i_1-k_1}v_{i_2}\right).
\end{aligned}$$

By definition, $\mu_4(i_1,i_2,k_1,k_2)-\mu_4^*(i_1,i_2,k_1,k_2)=\kappa_{v,4}(i_1,-k_1,i_2-i_1,i_2-k_2-i_1)$. So

$$\begin{aligned}
&var(\hat{\Omega}^*)\\
&=\frac{1}{K^2\Lambda(n,\delta)^4}\sum_{j_1,j_2}\sum_{i_1=1}^{n}\sum_{k=i_1-n}^{i_1-1}\sum_{i_2=1}^{n}\sum_{k=i_2-n}^{i_2-1}\phi_{j_1,j_1,j_2,j_2}(i_1,i_1-k_1,i_2,i_2-k_2)\kappa_{v,4}(i_1,-k_1,i_2-i_1,i_2-k_2-i_1)\\
&+\frac{1}{K^2\Lambda(n,\delta)^4}\sum_{j_1,j_2}\sum_{i_1=1}^{n}\sum_{k=i_1-n}^{i_1-1}\sum_{i_2=1}^{n}\sum_{k=i_2-n}^{i_2-1}\phi_{j_1,j_1,j_2,j_2}(i_1,i_1-k_1,i_2,i_2-k_2)E\left(v_{i_1}v_{i_2}\right)E\left(v_{i_1-k_1}v_{i_2-k_2}\right)\\
&+\frac{1}{K^2\Lambda(n,\delta)^4}\sum_{j_1,j_2}\sum_{i_1=1}^{n}\sum_{k=i_1-n}^{i_1-1}\sum_{i_2=1}^{n}\sum_{k=i_2-n}^{i_2-1}\phi_{j_1,j_1,j_2,j_2}(i_1,i_1-k_1,i_2,i_2-k_2)E\left(v_{i_1}v_{i_2-k_2}\right)E\left(v_{i_1-k_1}v_{i_2}\right)\\
&:=I_1+I_2+I_3.
\end{aligned}$$

Using $|\phi_{j_1,j_1,j_2,j_2}(i_1,i_2,k_1,k_2)|\leq C$ for some constant C, which holds under Assumption 2.2, we obtain

$$\begin{aligned}
&|I_1|\\
&\leq\frac{1}{K^2\Lambda(n,\delta)^4}\sum_{j_1,j_2}\sum_{i_1=1}^{n}\sum_{k_1=i_1-n}^{i_1-1}\sum_{i_2=1}^{n}\sum_{k=i_2-n}^{i_2-1}|\phi_{j_1,j_1,j_2,j_2}(i_1,i_2,k_1,k_2)|\,|\kappa_{v,4}(i_1,-k_1,i_2-i_1,i_2-k_2-i_1)|\\
&\leq\frac{C}{K^2\Lambda(n,\delta)^4}\sum_{j_1,j_2}\sum_{i_1=1}^{n}\sum_{k_1=i_1-n}^{i_1-1}\sum_{i_2=1}^{n}\sum_{k=i_2-n}^{i_2-1}|\kappa_{v,4}(i_1,-k_1,i_2-i_1,i_2-k_2-i_1)|\\
&=\frac{C}{K^2\Lambda(n,\delta)^4}\sum_{j_1,j_2}\sum_{i_1=1}^{n}\sum_{k_1=i_1-n}^{i_1-1}\sum_{\ell_1=i_1-n}^{i_1-1}\sum_{k_2=i_1+\ell_1-n}^{i_1+\ell_1-1}|\kappa_{v,4}(0,-k_1-i_1,k_1-2i_1,i_2-k_2-2i_1)|\\
&\leq\frac{n/\delta^3}{\Lambda(n,\delta)^4}\frac{C}{K^2}\sum_{j_1,j_2}\left(\delta^3\sum_{\ell_1=-n+1}^{n-1}\sum_{\ell_2=-n+1}^{n-1}\sum_{\ell_3=-n+1}^{n-1}|\kappa_{v,4}(0,\ell_1,\ell_2,\ell_3)|\right)\\
&=O\left(\frac{1}{T}\right),
\end{aligned}$$

where we have used Assumption S.1.

It remains to consider I_2 and I_3. Using change of variables repeatedly, we have

$$I_2 = \frac{1}{K^2 \Lambda (n,\delta)^4} \sum_{j_1,j_2} \sum_{i_1=1}^{n} \sum_{k_1=i_1-n}^{i_1-1} \sum_{i_2=1}^{n} \sum_{k_2=i_2-n}^{i_2-1} \phi_{j_1,j_1,j_2,j_2}(i_1, i_1-k_1, i_2, i_2-k_2)$$
$$\times \Gamma_v(i_2-i_1)\Gamma_v(i_2-i_1-(k_2-k_1))$$
$$= \frac{1}{K^2 \Lambda (n,\delta)^4} \sum_{j_1,j_2} \sum_{i_1=1}^{n} \sum_{k_1=i_1-n}^{i_1-1} \sum_{i_2=1}^{n} \sum_{k_2=i_2-n}^{i_2-1} \phi_{j_1,j_2}(i_1, i_2)\phi_{j_1,j_2}(i_1-k_1, i_2-k_2)$$
$$\times \Gamma_v(i_2-i_1)\Gamma_v(i_2-i_1-(k_2-k_1))$$
$$= \frac{1}{K^2 \Lambda (n,\delta)^4} \sum_{j_1,j_2} \sum_{i_1=1}^{n} \sum_{i=i_1-1}^{i_1-n} \sum_{k_1=i_1-n}^{i_1-1} \sum_{k_2=i_1-i-n}^{i_1-i-1} \phi_{j_1,j_2}(i_1, i_1-i)\phi_{j_1,j_2}(i_1-k_1, i_1-i-k_2)$$
$$\times \Gamma_v(-i)\Gamma_v(-i-(k_2-k_1))$$
$$= \frac{1}{K^2 \Lambda (n,\delta)^4} \sum_{j_1,j_2} \left\{ \sum_{i_1=1}^{n} \sum_{i=i_1-n}^{i_1-1} \phi_{j_1,j_2}(i_1, i_1-i)\Gamma_v(i) \right\}^2$$
$$= \frac{\delta^2}{K^2} \sum_{j_1,j_2} \left\{ \sum_{i=-n+1}^{n-1} \left[\frac{1}{n} \sum_{i_1=1}^{n} 1\left\{ \frac{1}{n} \le \frac{i_1-i}{n} \le 1 \right\} \phi_{j_1}\left(\frac{i_1}{n}\right)\phi_{j_2}\left(\frac{i_1-i}{n}\right) \right] \Gamma_v(i) \right\}^2.$$

For any $\varsigma \in [0,1]$, define

$$\omega_{j_1,j_2,n}(\varsigma) = \frac{1}{n} \sum_{i_1=1}^{n} 1\left\{ \frac{1}{n} \le \frac{i_1}{n} - \varsigma \le 1 \right\} \phi_{j_1}\left(\frac{i_1}{n}\right)\phi_{j_2}\left(\frac{i_1}{n} - \varsigma\right).$$

Then

$$I_2 = \frac{\delta^2}{K^2} \sum_{j_1,j_2} \left[\sum_{i=-n+1}^{n-1} \omega_{j_1,j_2,n}\left(\frac{i}{n}\right)\Gamma_v(i) \right]^2.$$

Under Assumptions 2.2 and 2.6(iv), we have

$$\omega_{j_1,j_2,n}(\varsigma) = \frac{1}{n} \sum_{i_1=1}^{n} 1\left\{ \frac{1}{n} + \varsigma \le \frac{i_1}{n} \le 1 + \varsigma \right\} \phi_{j_1}\left(\frac{i_1}{n}\right)\phi_{j_2}\left(\frac{i_1}{n} - \varsigma\right)$$
$$= \int_{\max(0,\varsigma)}^{\min(1+\varsigma,1)} \phi_{j_1}(r)\phi_{j_2}(r-\varsigma)\,dr + O\left(\frac{\max(j_1,j_2)}{n}\right)$$
$$:= \omega_{j_1,j_2}(\varsigma) + O\left(\frac{\max(j_1,j_2)}{n}\right),$$

uniformly over $j_1, j_2 \in [K]$. That is, there exists a constant C not dependent on j_1, j_2, ς, or K such that $|\omega_{j_1,j_2,n}(\varsigma) - \omega_{j_1,j_2,n}(\varsigma)| \le C(j_1+j_2)/n$. We can choose C large enough so that $\sup_{j_1,j_2,\varsigma} |\omega_{j_1,j_2}(\varsigma)| \le C$. Hence, for

$$I_{21} = \frac{\delta^2}{K^2} \sum_{j_1,j_2} \left[\sum_{i=-n+1}^{n-1} \omega_{j_1,j_2}\left(\frac{i}{n}\right)\Gamma_v(i) \right]^2,$$

we have

$$
\begin{aligned}
I_2 &= I_{21} + O\left[\frac{\delta^2}{K^2}\sum_{j_1,j_2}\left(\frac{\max\left(j_1,j_2\right)}{n}\right)\sum_{i=-n+1}^{n-1}\left|\Gamma_v\left(i\right)\right|\right] \\
&= I_{21} + O\left[\frac{\delta}{K^2}\sum_{j_1,j_2}\left(\frac{\max\left(j_1,j_2\right)}{n}\right)\delta\sum_{i=-n+1}^{n-1}\left|\Gamma_v\left(i\right)\right|\right] \\
&= I_{21} + O\left(\frac{\delta}{K^2}\frac{K^3}{n}\right) = I_{21} + O\left(\frac{1}{K}\frac{TK^2}{n^2}\right) \\
&= I_{21} + o\left(\frac{1}{K}\right),
\end{aligned}
$$

as $TK^2/n^2 = o\left(1\right)$. In the above, we have used $\delta\sum_{i=-n+1}^{n-1}\left|\Gamma_v\left(i\right)\right| < \infty$, which holds under Assumption 2.6(ii).

Note that under Assumptions 2.2 and 2.6(iv), we have

$$
\begin{aligned}
&\omega_{j_1,j_2}\left(0\right) = 1\left\{j_1 = j_2\right\}, \\
&\dot{\omega}_{j_1,j_2}\left(\varsigma\right) = -\phi_{j_1}\left(\varsigma\right)\phi_{j_2}\left(0\right) - \int_{\varsigma}^{1}\phi_{j_1}\left(r\right)\dot{\phi}_{j_2}\left(r-\varsigma\right)dr,
\end{aligned}
$$

where $\omega_{j_1,j_2}\left(0\right) = 1\left\{j_1 = j_2\right\}$, which follows from the orthonormality of $\left\{\phi_j\right\}$. So $\sup_{j_1,j_2,\varsigma}\left|\dot{\omega}_{j_1,j_2}\left(\varsigma\right)\right| < Cj_2$ for some constant $C > 0$.

Using the above expressions of the derivatives and taking a Taylor expansion, we have, for $i^*\left(i\right) \in [0,i]$,

$$
\begin{aligned}
I_{21} &= \frac{\delta^2}{K^2}\sum_{j_1,j_2}\left[\sum_{i=-n+1}^{n-1}\omega_{j_1,j_2}\left(\frac{i}{n}\right)\Gamma_v\left(i\right)\right]^2 \\
&= \frac{\delta^2}{K^2}\sum_{j_1,j_2}\left[\sum_{i=-n+1}^{n-1}\Gamma_v\left(i\right)1\left\{j_1 = j_2\right\} + \sum_{i=-n+1}^{n-1}\dot{\omega}_{j_1,j_2}\left(\frac{i^*\left(i\right)}{n}\right)\frac{i}{n}\Gamma_v\left(i\right)\right]^2 \\
&= \frac{\delta^2}{K^2}\sum_{j_1,j_2}\left(\sum_{i=-n+1}^{n-1}\Gamma_v\left(i\right)1\left\{j_1 = j_2\right\}\right)^2 \\
&+ \frac{2}{K^2}\sum_{j_1=j_2}\left(\sum_{i=-n+1}^{n-1}\Gamma_v\left(i\right)\right)\left(\delta\sum_{i=-n+1}^{n-1}\dot{\omega}_{j_1,j_2}\left(\frac{i^*\left(i\right)}{n}\right)\frac{i}{n}\delta\Gamma_v\left(i\right)\right) \\
&+ \frac{\delta^2}{K^2}\sum_{j_1,j_2}\left[\sum_{i=-n+1}^{n-1}\dot{\omega}_{j_1,j_2}\left(\frac{i^*\left(i\right)}{n}\right)\frac{i}{n}\Gamma_v\left(i\right)\right]^2 \\
&= \frac{1}{K^2}\sum_{j=1}^{K}\left(\delta\sum_{i=-n+1}^{n-1}\Gamma_v\left(i\right)\right)^2 + O\left(\frac{1}{K^2}\frac{K^2}{n}\right) + O\left(\frac{1}{K}\frac{K^2}{T^2}\right) \\
&= \left(\delta\sum_{i=-n+1}^{n-1}\Gamma_v\left(i\right)\right)^2\frac{1}{K} + o\left(\frac{1}{K}\right),
\end{aligned}
$$

where we have used $\delta \sum_{i=-n+1}^{n-1} |i| \, \delta \, |\Gamma_v(i)| < \infty$, which holds under Assumption 2.6(ii).

Now under Assumption 2.6(iii) with $m = 0$ and Assumption 2.6(ii), we have

$$\delta \sum_{i=-n+1}^{n-1} \Gamma_v(i) = \delta \sum_{i=-n+1}^{n-1} \Gamma_V(i\delta) \to \Omega.$$

Therefore, we have proved that $I_2 = \frac{1}{K}\Omega(1 + o(1))$. Similar arguments can be invoked to show that $I_3 = \frac{1}{K}\Omega(1 + o(1))$. Details are omitted here. Combining the results for I_1, I_2, and I_3 yields the desired result: $var(\hat{\Omega}^*) = \frac{1}{K}2\Omega^2(1 + o(1))$. ■

S.2 Simulations with Additive Measurement Noise

Here we report the impact of including an additive noise component in the simulation environments of Section 4. All simulations and the corresponding figures are reproduced. Now, however, each observation y_i also includes an additive noise component ϵ_i as described in Section 6. The choice of data generating process for $\{\epsilon_i\}_{i=1}^n$ in the various simulations environments is described below. This is then followed by Figures S.1–S.7, which correspond to Figures 2–8 in the main text, respectively.

Each simulation setting (endogenous, exogenous, etc.) considered in Section 4, now with additive noise as in Section 6, involves a regression model of the form

$$y_i = x_i'\beta_0 + u_i + \epsilon_i, \; i = 1, \ldots, n, \tag{S.1}$$

where the continuous-time processes $\{Y_t\}$, $\{X_t\}$ and $\{U_t\}$ are sampled at time $t = i\delta$. Note that in the nonstationary setting $u_i = U_{i\delta}$ should be replaced by $u_{0i} = U_{0i\delta}$, a distinction we ignore here to simplify the notation.

In each simulation setting considered in Section 4, $\{U_t\}$ follows a stationary Ornstein Uhlenbeck (OU) process:

$$dU_t = -\kappa_u U_t dt + \sigma_u dW_t,$$

where $\{W_t\}$ is standard Brownian motion. There are different values of (κ_u, σ_u) for different simulation settings.

In each setting, we let $\{\epsilon_i\}_{i=1}^n$ be an i.i.d. sequence of random variables with $\epsilon_i \sim N(0, a^2)$ for all i. We choose a differently for the various settings as follows. Let

$$z_i = (u_{i+1} + \epsilon_{i+1}) - (u_i + \epsilon_i) = U_{(i+1)\delta} - U_{i\delta} + \epsilon_{i+1} - \epsilon_i.$$

z_i is then the change in all "error" terms of the regression model (S.1) between successive measurements. It encapsulates evolution in the "error" terms coming from both the continuous-time process $\{U_t\}$ and the additive measurement noise $\{\epsilon_i\}$. For a given sampling frequency δ, we define the error noise to signal ratio (ENSR_δ or ENSR) by

$$\text{ENSR}_\delta = \frac{\text{Var}[\epsilon_{i+1} - \epsilon_i]}{\text{Var}[z_i]} = \frac{2a^2}{\frac{\sigma_u^2}{\kappa_u}(1 - \exp(-\kappa_u\delta)) + 2a^2},$$

which is independent of i as both $\{U_t\}$ and $\{\epsilon_t\}$ are stationary. We have used basic properties of the OU process to derive the variance of z_i. ENSR_δ tells us how much of the variation in the regression error terms between successive observations comes from the additive noise component when observations are recorded at time increments of length δ. In each simulation environment, a is chosen so that the ENSR at the highest sampling frequency (i.e., at the smallest δ) is equal to 0.35, corresponding to 35% of this variation.

The simulation results mirroring those reported in Section 6 are reported in the figures below.

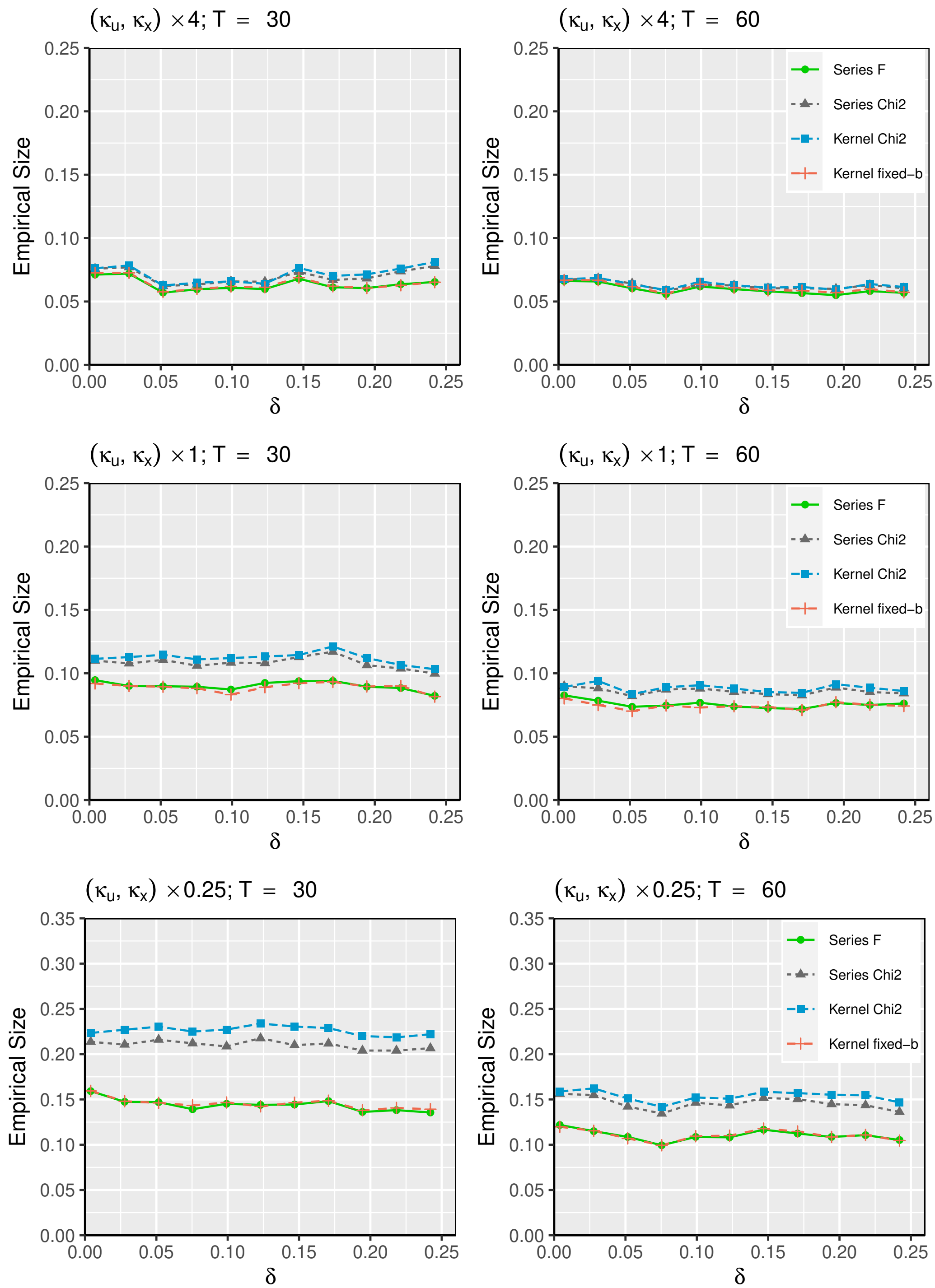

Figure S.1: Empirical sizes in the stationary simulation setting with additive noise when X_t follows an OU process and (κ_u, κ_x) are multiplied by factors of 4, 1 and 1/4.

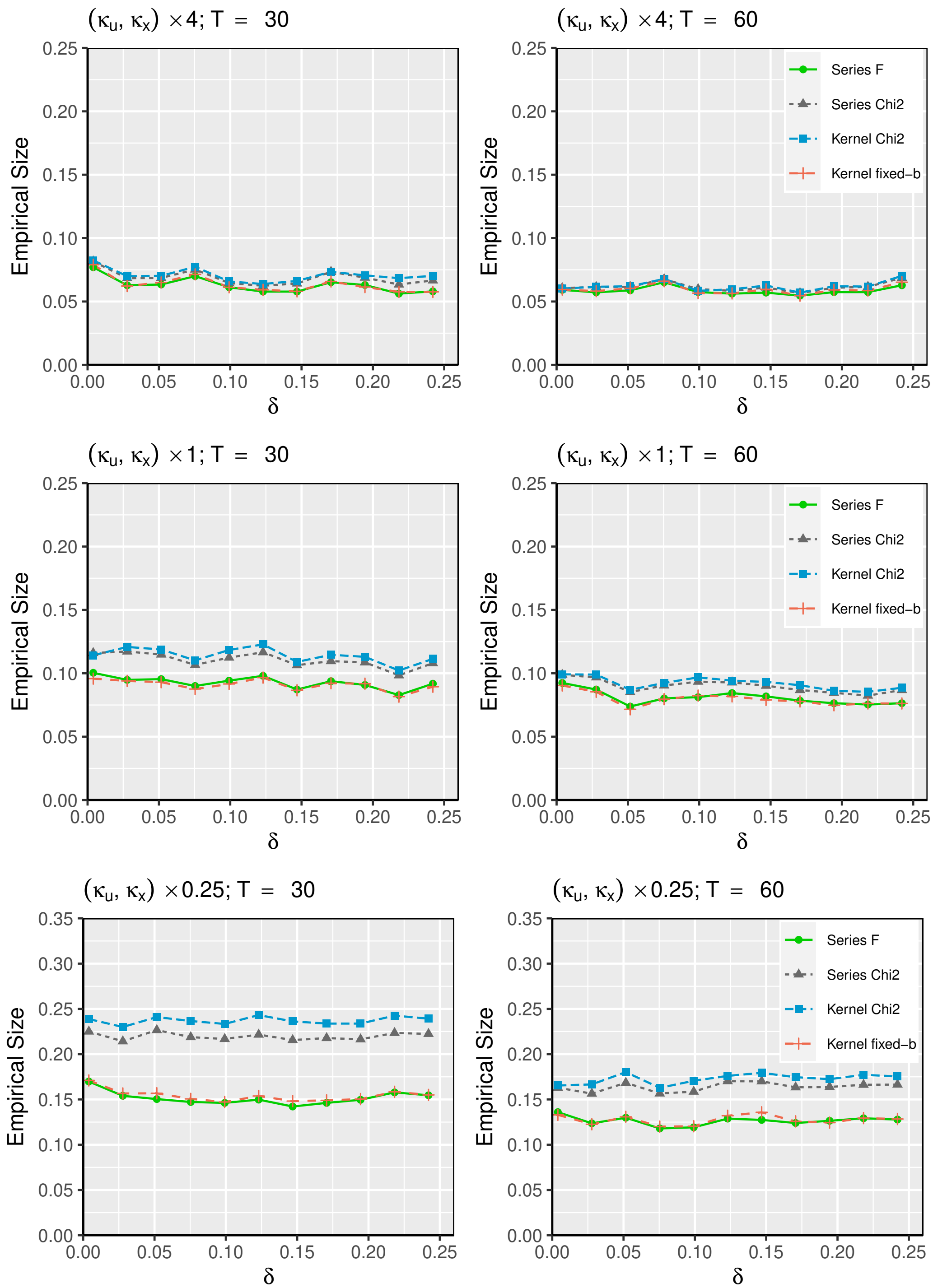

Figure S.2: Empirical sizes in the stationary simulation setting with additive noise when X_t follows an SR process and (κ_u, κ_x) are multiplied by factors of 4, 1, and 1/4.

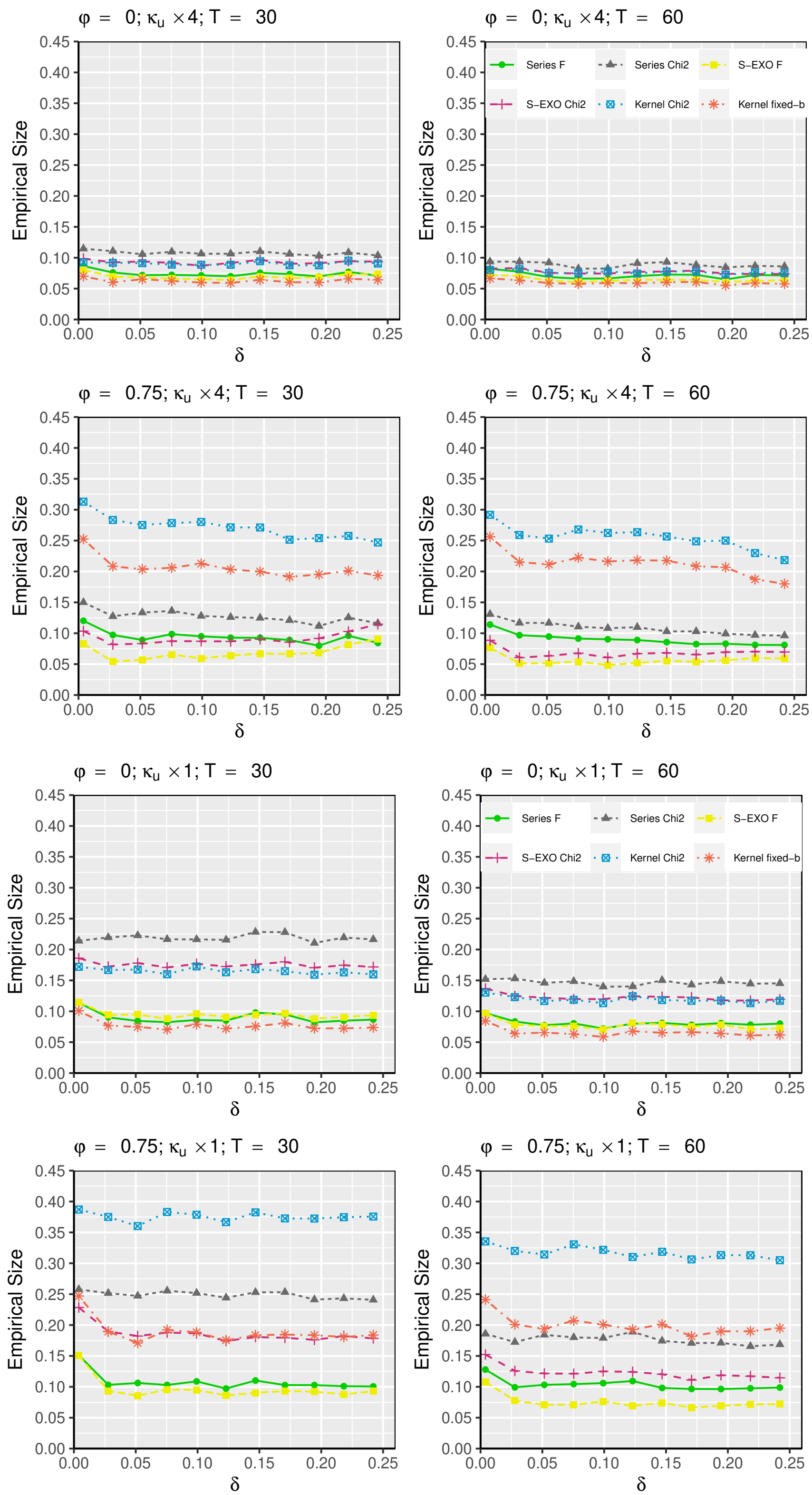

Figure S.3: Empirical sizes in the nonstationary simulation setting with additive noise when κ_u is multiplied by factors of 4 and 1.

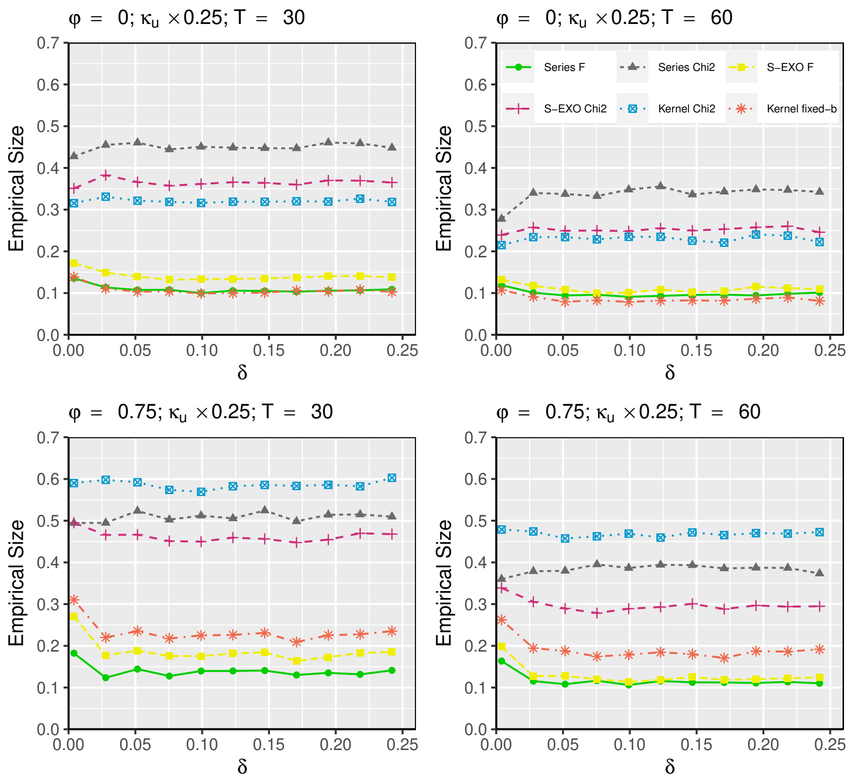

Figure S.4: Empirical sizes in the nonstationary simulation setting when κ_u is multiplied by $1/4$.

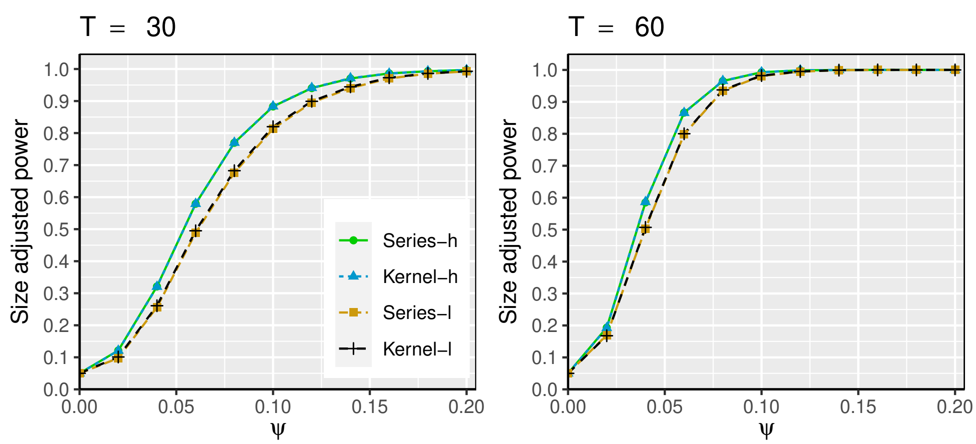

Figure S.5: Size-adjusted powers in the stationary setting with additive noise when X_t is distributed according to the OU process described in Section 4

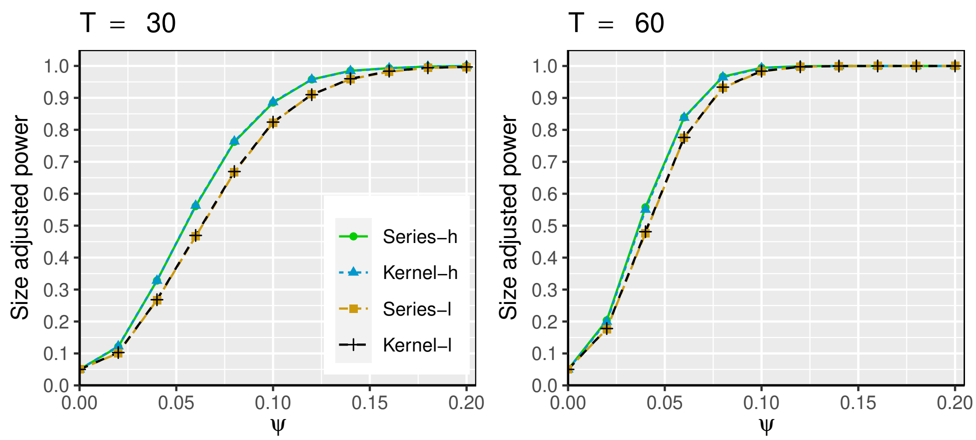

Figure S.6: Size-adjusted powers in the stationary setting with additive noise when X_t is distributed according to the SR process described in Section 4

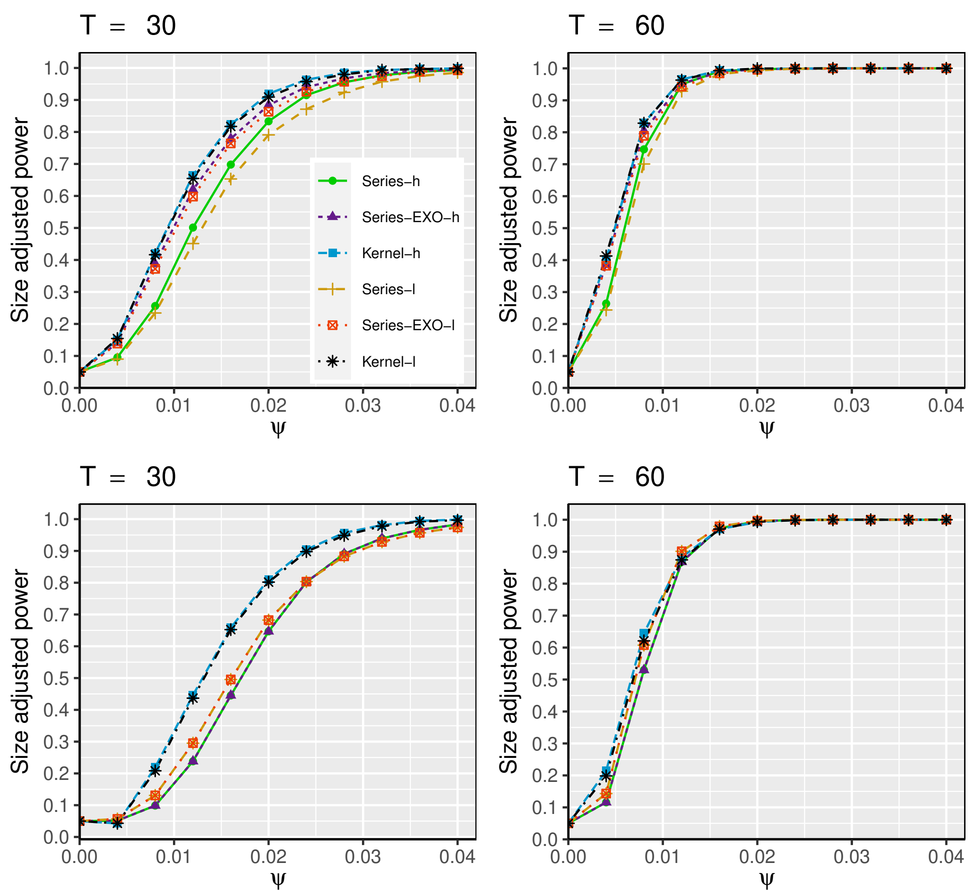

Figure S.7: Size-adjusted powers in the nonstationary setting with additive noise. In the upper row, the explanatory variables are exogenous ($\varphi = 0$). In the lower row the explanatory variables are endogenous ($\varphi = 0.75$).