Simple and Trustworthy Asymptotic t Tests in Difference-in-Differences Regressions*

Cheng Liu
Economics and Management School
Wuhan University, China

Yixiao Sun
Department of Economics
UC San Diego, USA

Abstract

The paper proposes two asymptotically valid t tests in a difference-in-differences (DD) regression when the number of time periods is large while the number of individuals can be small or large. Each of two t tests is based on a special heteroscedasticity and autocorrelation robust (HAR) variance estimator that is tailored to inference problems in the DD setting. The difference between the two t tests is that one is based on a general form of the sandwich variance estimator while the other is based on a special form of that estimator. The asymptotic distributions of both t tests depend on the smoothing parameter K in the HAR variance estimator. A testing-optimal procedure for choosing K for the t test with a special sandwich variance estimator is developed through minimizing the type II error subject to a constraint on the type I error of the t test. By capturing the estimation uncertainty of the HAR variance estimators, both t tests have more accurate size than the corresponding normal tests and are just as powerful as the latter. Compared to the nonstandard tests that are designed to reduce the size distortion of the normal tests, the proposed t tests are just as accurate but much more convenient to use, as the critical values are from the standard t table.

Keywords: Basis Functions, Difference-in-Differences, Fixed-smoothing Asymptotics, Heteroscedasticity and Autocorrelation Robust, Panel Data, Student's t distribution, t test

JEL Classification Number: C12, C33

1 Introduction

The paper considers estimation and inference in a difference-in-differences (DD) regression. To make trustworthy inferences, we have to obtain a reliable estimator of the standard error. In the presence of both temporal and cross-sectional dependence, the basic clustered standard error estimator is inconsistent. If one clusters by individual, observations may be correlated for the same individual, but they are often required to be independent for different individuals. See, for example, Bertrand, Duflo, and Mullainathan (2004). If one clusters by time, then observations in the same time period can have arbitrary correlation, but they are often required to be independent across time. In this paper, we consider clustering by time but allow the clusters to be temporally dependent. Our approach is in the spirit of Driscoll and Kraay (1998), but we employ a different

*Email: chengliu_eco@whu.edu.cn and yisun@ucsd.edu. Sun gratefully acknowledges partial research support from NSF under Grant No. SES-1530592. Address correspondence to Yixiao Sun, Department of Economics, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA.

heteroscedasticity and autocorrelation robust variance estimator. In principle, we could consider clustering by individual and allow for spatial dependence across individuals, but this requires an extra variable to indicate the direction and strength of the spatial dependence. In fact, if such a variable is available, we can use the approach of Kim and Sun (2013), which treats the temporal and cross-sectional dependence symmetrically. An advantage of the clustering-by-time approach is that no additional information is needed, as the time index provides a natural yardstick for the measurement of temporal dependence.

For the DD regression, the clustering-by-time approach amounts to collapsing the panel data into time series data. Cross-sectional dependence affects the variance of the collapsed time series but has no effect on its temporal dependence. To estimate the asymptotic variance of the DD estimator, we need to estimate only the long-run variance (LRV) of some collapsed time series. There are many nonparametric LRV estimators, among which kernel LRV estimators are popular in applied research (see, for example, Andrews (1991)). A recent study by Yu Sun (2016) adopts the kernel approach. In this paper, we consider the series approach to LRV estimation. The most primitive version of this estimator is the simple averaged periodogram estimator, which involves taking a simple average of the first few periodograms. The number of periodograms is the smoothing parameter underlying this series LRV estimator. Equivalently, this approach involves first projecting the time series onto a sequence of Fourier basis functions (i.e., sine and cosine functions) and then taking the simple average of the squared projection coefficients as the LRV estimator. More general basis functions can be used. In fact, one of the advantages of the series LRV approach is that we have the freedom to choose any sequence of basis functions. Each basis function delivers a direct estimator of the LRV, and the series LRV estimator is a simple average of these direct estimators. The number of terms in the average, K, which can be regarded as the effective sample size, characterizes the amount of smoothing.

A main contribution of the paper is to establish the fixed-smoothing asymptotics of the Studentized t statistic. The fixed-smoothing asymptotics is obtained under the assumption that K is fixed as T goes to infinity. The cross-sectional sample size n can be fixed or grow with T. We also assume that the policy change takes place in the middle of the time series so that the number of pre-treatment periods is comparable to the number of post-treatment periods. The asymptotic approximation so obtained captures the randomness of the nonparametric variance estimator. It reflects the effect of the basis functions, the level of smoothing, and the effect of the trend function if a trend is present in the DD regression. Moreover, it is more accurate than the widely used standard normal approximation, which fails to capture these effects. The fixed-smoothing asymptotic distribution is nonstandard. Nevertheless, it is free from any nuisance parameter and can be simulated without too much difficulty.

Another contribution of the paper is the design of a new set of basis functions such that the t statistic follows the standard t distribution under the fixed-smoothing asymptotics. This is achieved by transforming any given set of basis functions in $L^2[0,1]$. The transformation, a type of Gram-Schmidt orthonormalization, ensures that the asymptotic variance estimator is equal in distribution to an average of *iid* chi-square variates in large samples, which is necessary for the asymptotic t approximation theory. The asymptotic t test is very convenient to use, as critical values are readily available from standard statistical tables and programming environments.

The regressor of interest in the DD regression is a special regressor. For the treatment group, this regressor takes the value 0 in the pre-treatment periods and switches to the value 1 in the post-treatment periods. From a time series perspective, it has energy concentrated at the origin. The special form of the regressor allows for two different approaches to be used in estimating the asymptotic variance of the DD estimator. In the first approach, we ignore the fact that the regressor is special and use a general sandwich variance estimator. In a time series regression with stationary data, this approach entails estimating the long-run variance of a product process: the product of the regressor and the regression error. In the second approach, we take advantage of the special form of the regressor, and we estimate the long-run variance of the regression error process only. The resulting sandwich variance estimator collapses to a special form, which we call the collapsed sandwich variance estimator.

Our fixed-smoothing asymptotics and t limit theory apply to both forms of variance estimators, but the fixed-smoothing limiting distribution is different for different variance estimators. This should be regarded as an attractive property, as the finite sample distribution may be sensitive to the form of the variance estimator. The transformation used to obtain the standard t limiting distribution also depends on the form of the variance estimators. In both cases, the transformation is easy to implement and requires only the computation the Cholesky decomposition of a positive-definite matrix.

The smoothing parameter K plays the important role of determining the size and power tradeoff of the asymptotic t tests. In the literature on LRV estimation and heteroscedasticity and autocorrelation robust (HAR) inference, Phillips (2005) proposes to choose K by minimizing the asymptotic MSE of the LRV estimator. However, the MSE-based choice of K may not be optimal for testing problems. In hypothesis testing, the main objects of interest are the type I and type II errors. The choice of K should then be targeted at these fundamental quantities. Following Sun (2011), we develop a selection procedure that is optimal for the testing problem at hand. In particular, we consider one of the two asymptotic t tests and choose K to minimize its type II error while controlling its type I error.

In our simulations, we compare the performances of the fixed-smoothing tests with those of the asymptotic normal tests. Each type of tests actually consists of four tests, reflecting the different combinations of whether a transformation is applied to the Fourier bases or not and which of the two forms of the variance estimator is used. In all cases, a fixed-smoothing test is found to be more accurate than the corresponding asymptotic normal test. Among the fixed-smoothing tests, the t test based on the transformed bases is just as accurate as the corresponding nonstandard test based on the original bases. These observations remain valid regardless of whether K is fixed *a priori* or data driven. Power study under data-driven K-values shows that all tests have similar power properties. In view of the size accuracy of the asymptotic t tests and their convenience to use, we recommend using the asymptotic t tests in empirical applications.

This paper contributes to the literature on the fixed-smoothing asymptotics in general and the asymptotic F and t test theory in particular. The asymptotic F and t tests have been developed in Sun (2011) for trend regression, in Sun (2013) for stationary moment processes, in Sun (2014c) for highly persistent moment processes, in Hansen (2007) for stationary panel time series, and in Hwang and Sun (2017) for stationary data in an overidentified GMM framework. Lazarus, Lewis, Stock, and Watson (2016) provide some practical guidance on the F and t tests for time series regressions. See also Sun and Kim (2012, 2015) for the F limit theory for the J statistic, and the F and t limit theory for the Wald statistic and t statistic in a spatial setting. None of these papers considers the DD regression where the regressor of interest is a special deterministic function and is hence nonstationary by definition. More broadly, the paper is related to the fixed-b asymptotic theory where kernel LRV estimators are used. See Kiefer and Vogelsang (2002a, 2002b, 2005) and Sun (2014a) and the references therein. A paper that is closest to this paper is the paper by Yu Sun (2016), who considers the fixed-b asymptotic theory for the DD regression. However, the asymptotic distribution there is complicated and not as easy to use as Student's t distribution.

The rest of the paper is organized as follows. Section 2 presents the basic setting and introduces the DD estimator and the general sandwich variance estimator. Section 3 establishes the fixed-smoothing asymptotics of the t statistic based on the general sandwich variance estimator, and Section 4 develops an asymptotically valid t test. Section 5 considers the collapsed sandwich variance estimator and the associated t limit theory. Section 6 proposes a data-driven and testing-optimal approach to choosing the smoothing parameter K. Section 7 reports the simulation evidence. The last section concludes. Proofs are given in the appendix.

2 The Basic Setting and DD Estimator

We consider the difference-in-differences regression

$$Y_{it} = \lambda_t + \tau(t)' \alpha_i + Treat_i \cdot \beta_{10} + Post_t \cdot \beta_{20} + Treat_i \cdot Post_t \cdot \theta_{10} + Z_{it}'\theta_{20} + \epsilon_{it},$$

for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, where λ_t is the time fixed effect and $\tau(t)' \alpha_i$ is the individual-specific time trend. If $\tau(t) = (1, t)'$ and $\alpha_i = (\alpha_{i0}, \alpha_{i1})'$, for example, we have $\tau(t)' \alpha_i = \alpha_{i0} + \alpha_{i1} \cdot t$, where α_{i0} is the individual fixed effect and α_{i1} is the individual-specific linear trend coefficient. We assume that the first element of $\tau(t)$ is 1 so that individual fixed effects are always included. $Treat_i$ is a dummy variable indicating the treatment or control group. Individual i belongs to the treatment group if $Treat_i$ is equal to 1; otherwise, individual i belongs to the control group. Without loss of generality, we assume that observations are sorted along the cross-sectional dimension so that $Treat_i = 1\{i \leq \mu n\}$ for some $\mu \in (0, 1)$. $Post_t$ is a dummy variable indicating the post-treatment periods. That is, $Post_t = 1\{t \geq \nu T\}$ for some $\nu \in (0, 1)$. For notational convenience, we assume that μn and νT are positive integers. Z_{it} is a $d_Z \times 1$ vector of other covariates. The parameter of interest is θ_{10}, which captures the effect of the training program.

To estimate θ_{10}, we first remove the trend component $\tau(t)' \alpha_i$. In view of individual heterogeneity in the intercept and slope coefficient, we detrend each time series individually. Let

$$Y_{it}^{\tau} = Y_{it} - \left(\sum_{s=1}^{T} Y_{is} \tau(s)' \right) \left(\sum_{s=1}^{T} \tau(s) \tau(s)' \right)^{-1} \tau(t),$$
$$Z_{it}^{\tau} = Z_{it} - \left(\sum_{s=1}^{T} Z_{is} \tau(s)' \right) \left(\sum_{s=1}^{T} \tau(s) \tau(s)' \right)^{-1} \tau(t)$$

be the detrended variables, and define λ_t^{τ}, $Post_t^{\tau}$, and ϵ_{it}^{τ} similarly. Then

$$Y_{it}^{\tau} = \lambda_t^{\tau} + Post_t^{\tau} \cdot \beta_{20} + Treat_i \cdot Post_t^{\tau} \cdot \theta_{10} + (Z_{it}^{\tau})'\theta_{20} + \epsilon_{it}^{\tau}.$$

Note that the group-specific effect $Treat_i \cdot \beta_{10}$ has been eliminated by detrending.

Next, we remove the time fixed effect λ_t^{τ} using the cross-sectional fixed-effect transformation. Let

$$\tilde{Y}_{it}^{\tau} = Y_{it}^{\tau} - \frac{1}{n} \sum_{j=1}^{n} Y_{jt}^{\tau}, \tag{1}$$

and define other variables such as $\tilde{Z}_{it}^{\tau}$, $\widetilde{Treat}_i$, and $\tilde{\epsilon}_{it}^{\tau}$ similarly. Then

$$\tilde{Y}_{it}^{\tau} = \widetilde{Treat}_i \cdot Post_t^{\tau} \cdot \theta_{10} + (\tilde{Z}_{it}^{\tau})'\theta_{20} + \tilde{\epsilon}_{it}^{\tau}. \tag{2}$$

Note that the cross-sectional fixed-effect transformation eliminates both λ_t^{τ} and $Post_t^{\tau} \cdot \beta_{20}$.

Let

$$X_{it} = \begin{pmatrix} Treat_i \cdot Post_t \\ Z_{it} \end{pmatrix}, \tilde{X}_{it}^{\tau} = \begin{pmatrix} \widetilde{Treat_i} \cdot Post_t^{\tau} \\ \tilde{Z}_{it}^{\tau} \end{pmatrix}, \tag{3}$$

and $\theta_0 = (\theta_{10}, \theta_{20}')'$. Then the OLS estimator $\hat{\theta}$ of $\theta_0 = (\theta_{10}, \theta_{20}')'$ is given by

$$\hat{\theta} = \left[\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{X}_{it}^{\tau} (\tilde{X}_{it}^{\tau})' \right]^{-1} \left[\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{X}_{it}^{\tau} \tilde{Y}_{it}^{\tau} \right]. \tag{4}$$

The estimator $\hat{\theta}$ is numerically identical to the fixed-effects OLS estimator based on the original equation, that is, the OLS estimator with time dummies, individual dummies, and their interactions with the trend function.

Since the coefficients associated with Z_{it} may not have any causal interpretation, and are often not the parameters of interest in empirical applications, we focus only on the parameter θ_{10} in this paper. As an estimator of θ_{10}, the first element $\hat{\theta}_1$ of $\hat{\theta}$ is often referred to as the difference-in-differences estimator, as it is equal to a difference in differences in the simple case with only two periods.

In this paper, we consider the asymptotics along the direction in which $T \to \infty$. The cross-sectional sample size n can be fixed or grow with T. In the latter case, the sufficient conditions for Assumption 3.4 below require that $n/T \to 0$. For simplicity, we will denote the asymptotic direction as "$T \to \infty$" in both cases. Given the asymptotic direction we consider, we write

$$\sqrt{nT}(\hat{\theta} - \theta_0) = \left[\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{X}_{it}^{\tau} (\tilde{X}_{it}^{\tau})' \right]^{-1} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{nt} \right],$$

where

$$u_{nt} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{X}_{it}^{\tau} \tilde{\epsilon}_{it}^{\tau}$$

is in the "clustering-by-time" format.

Let

$$\hat{\epsilon}_{it}^{\tau} = \tilde{Y}_{it}^{\tau} - (\tilde{X}_{it}^{\tau})' \hat{\theta} \text{ and } \hat{u}_{nt} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{X}_{it}^{\tau} \hat{\epsilon}_{it}^{\tau}.$$

Then the variance of $\sum_{t=1}^{T} u_{nt}/\sqrt{T}$ can be estimated by

$$\Omega = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_K\left(\frac{t}{T}, \frac{s}{T}\right) \hat{u}_{nt} \hat{u}_{ns}',$$

where $Q_K(\cdot, \cdot)$ is a symmetric weighting function and K is the smoothing parameter. The above estimator belongs to the general class of quadratic long-run variance estimators, which includes most if not all commonly used nonparametric LRV estimators as special cases.

In this paper, we focus on the series LRV estimator with $Q_K(r, s)$ given by

$$Q_K(r, s) = \frac{1}{K} \sum_{k=1}^{K} \Phi_k(r) \Phi_k(s),$$

where $\{\Phi_k(r)\}$ are basis functions in $L^2[0,1]$. In the econometrics literature, the series LRV estimator has been recently used, for example, in Phillips (2005), Müller (2007), and Sun (2011, 2013, 2014a, 2014b). Plugging the above weighting function into $\hat{\Omega}$, we obtain

$$\hat{\Omega} = \frac{1}{K}\sum_{k=1}^{K}\hat{\Omega}_k$$

for

$$\hat{\Omega}_k = \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\hat{u}_{n,t}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\hat{u}_{n,t}\right]'.$$

Thus $\hat{\Omega}$ is a simple average of some "direct" estimators $\hat{\Omega}_k$, and K is the effective sample size. If K is even and $\{\Phi_k(r)\} = \{\sqrt{2}\sin(2\pi kr), \sqrt{2}\cos(2\pi kr), k = 1, 2, \ldots, K/2\}$, then the series LRV estimator is proportional to the spectral density estimator at the origin that takes a simple average of the first $K/2$ periodograms. The averaged periodogram estimator is a common spectral density estimator. In the traditional asymptotic framework, Phillips (2005) has shown that the averaged periodogram estimator is asymptotically equivalent to the kernel LRV estimator based on the Daniell kernel. For further discussion of series LRV estimation, see Sun (2013). A necessary condition for $\hat{\Omega}$ to be positive definite is that $K \geq d_Z + 1$, which will be assumed throughout the rest of the paper.

Using $\hat{\Omega}$ as the middle term in the sandwich variance estimator, we obtain the HAR variance estimator:

$$\hat{V} = \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^{\tau}(\tilde{X}_{it}^{\tau})'\right]^{-1}\hat{\Omega}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^{\tau}(\tilde{X}_{it}^{\tau})'\right]^{-1}.$$

The above estimator is in the spirit of the estimator of Driscoll and Kraay (1998), who use a kernel LRV estimator instead of our series LRV estimator. One advantage of using a series LRV estimator is that an asymptotic t approximation theory can be developed; see Sections 4 and 5 for details.

To test the null hypothesis $\theta_{10} = r_0$, we construct the t statistic

$$\mathbb{T} = \frac{\sqrt{nT}(R\hat{\theta} - r_0)}{\sqrt{\hat{V}_R}},$$

where $R = (1, 0, \ldots, 0) \in \mathbb{R}^{d_Z+1}$ and

$$\hat{V}_R = R\hat{V}R'.$$

3 Fixed-Smoothing Asymptotics

To investigate the asymptotic properties of $\hat{\theta}$ and the associated t statistic, we make the following assumptions on the basis functions and the trend function.

Assumption 3.1 *The basis functions $\Phi_k(\cdot)$, $k = 1, 2, \ldots, K$, are piecewise monotonic and continuously differentiable.*

Assumption 3.2 *There exists a $d_\tau \times d_\tau$ diagonal matrix D_τ such that*

$$\tau_D([Tr]) := D_\tau \times \tau([Tr]) \to \tau(r)$$

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

$$\frac{1}{T}\sum_{t=1}^{T}\tau_D(t)\,\tau_D(t)' \to \int_0^1 \tau(r)\tau(r)'dr \text{ as } T\to\infty,$$

where $\int_0^1 \tau(r)\tau(r)'dr$ *is positive definite.*

For commonly used polynomial trend functions, Assumption 3.2 holds trivially. For example, when $\tau(t) = 1$, we can choose $D_\tau = 1$, in which case

$$\frac{1}{T}\sum_{t=1}^{T}\tau_D(t)\,\tau_D(t)' = 1.$$

When $\tau(t) = (1,t)'$, we can choose $D_\tau = diag(1, 1/T)$, in which case

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\tau_D(t)\,\tau_D(t)' &= \frac{1}{T}\begin{pmatrix}1 & 0\\ 0 & \frac{1}{T}\end{pmatrix}\begin{pmatrix}\sum_{t=1}^{T}1 & \sum_{t=1}^{T}t\\ \sum_{t=1}^{T}t & \sum_{t=1}^{T}t^2\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & \frac{1}{T}\end{pmatrix}\\ &= \begin{pmatrix}T^{-1}\sum_{t=1}^{T}1 & T^{-2}\sum_{t=1}^{T}t\\ T^{-2}\sum_{t=1}^{T}t & T^{-3}\sum_{t=1}^{T}t^2\end{pmatrix} \to \int_0^1 \tau(r)\tau(r)'dr.\end{aligned}$$

Given that the first element of $\tau(t)$ is a constant, the (1,1)-th element of D_τ is always 1.

Next, we decompose Z_{it} into a sum of three terms:

$$Z_{it} = \lambda_{zt} + \alpha_{zi}\cdot\tau(t) + \mathcal{Z}_{it},$$

where λ_{zt} and $\alpha_{zi}\cdot\tau(t)$ represent time fixed effects and linear trend effects, respectively. Let

$$\bar{\mathcal{Z}}_{\cdot,t}^{treat} = \frac{1}{n\mu}\sum_{i=1}^{\mu n}\mathcal{Z}_{it} \text{ and } \bar{\mathcal{Z}}_{\cdot,t}^{control} = \frac{1}{n(1-\mu)}\sum_{j=\mu n+1}^{n}\mathcal{Z}_{jt}$$

be the averaged time series of $\mathcal{Z}$ for the treatment group and the control group, respectively. Define

$$\tilde{\mathcal{Z}}_{it} = \mathcal{Z}_{it} - \bar{\mathcal{Z}}_{\cdot,t} \text{ and } \tilde{\mathcal{Z}}_{it}^{\tau} = \tilde{\mathcal{Z}}_{it} - \left(\sum_{s=1}^{T}\tilde{\mathcal{Z}}_{is}\tau(s)'\right)\left(\sum_{s=1}^{T}\tau(s)\,\tau(s)'\right)^{-1}\tau(t).$$

We make the following assumptions on $\mathcal{Z}_{it}$.

Assumption 3.3

$$\frac{1}{T}\sum_{t=1}^{[Tr]}\bar{\mathcal{Z}}_{\cdot,t}^{treat}\cdot\tau_D(t)' = \frac{1}{T}\sum_{t=1}^{[Tr]}\bar{\mathcal{Z}}_{\cdot,t}^{control}\cdot\tau_D(t)' + o_p(1)$$

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

Assumption 3.4 $T^{-1}\sum_{t=1}^{[Tr]}\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathcal{Z}}_{it}^{\tau}(\tilde{\mathcal{Z}}_{it}^{\tau})' \to rG$ *uniformly over* $r\in[0,1]$ *for some positive-definite matrix* G.

Assumption 3.3 requires that, in terms of their projections onto the trend function, the averaged time series $\{\bar{\mathcal{Z}}_{\cdot,t}^{treat}\}$ and $\{\bar{\mathcal{Z}}_{\cdot,t}^{control}\}$ do not differ systematically across the treatment and control groups. More precisely, if for any block of the time series spanning $t = [Tr_1], [Tr_1] + 1, \ldots, [Tr_2]$, the projections of $\{\bar{\mathcal{Z}}_{\cdot,t}^{treat}\}$ and $\{\bar{\mathcal{Z}}_{\cdot,t}^{control}\}$ onto the trend function are approximately the same, then Assumption 3.3 holds. This is similar to the "parallel paths" assumption that is often imposed in a difference-in-differences regression.

Assumption 3.4 is similar to a standard assumption in the fixed-smoothing asymptotics. If

(i) $T^{-1}\sum_{t=1}^{[Tr]} \tilde{\mathcal{Z}}_{it}\tilde{\mathcal{Z}}_{it}' \to rG$ uniformly over $r \in [0,1]$ and $i = 1, 2, \ldots, n$

(ii) $T^{-1}\sum_{s=1}^{[Tr]} \tilde{\mathcal{Z}}_{is}\tau_D(s)' = o_p(1)$ uniformly over $r \in [0,1]$ and $i = 1, 2, \ldots, n$,

then Assumption 3.4 is satisfied. Uniformity over r in the above conditions can be obtained by invoking a uniform law of large numbers for time series data, which typically requires some mixing and moment conditions. Uniformity over $i = 1, 2, \ldots, n$ can be obtained by using a classical argument. Consider condition (i) as an example. We have

$$\Pr\left(\max_{i\in\{1,2,\ldots,n\}} \sup_r ||T^{-1}\sum_{s=1}^{[Tr]}(\tilde{\mathcal{Z}}_{is}\tilde{\mathcal{Z}}_{it}' - G)|| > \varepsilon\right) \leq \sum_{i=1}^{n}\Pr\left(\sup_r ||T^{-1}\sum_{s=1}^{[Tr]}(\tilde{\mathcal{Z}}_{is}\tilde{\mathcal{Z}}_{it}' - G)|| > \varepsilon\right).$$

Therefore, if the summand in the above upper bound is of order $O(1/T)$ uniformly over i and $n/T \to 0$, then condition (i) holds. The uniform $O(1/T)$ bound on the summand will hold if the mixing and moment conditions for the time series ULLN hold uniformly over i. This entails imposing some restriction on the degree of cross-sectional heterogeneity.

To investigate the rate of information accumulation in $\tilde{X}_{it}^{\tau}$, the serially detrended and cross-sectionally demeaned regressor, we write

$$\frac{1}{nT}\sum_{t=1}^{[Tr]}\sum_{i=1}^{n}\tilde{X}_{it}^{\tau}(\tilde{X}_{it}^{\tau})' = S(r) := \begin{pmatrix} S_{11}(r) & S_{12}(r) \\ S_{21}(r) & S_{22}(r) \end{pmatrix}, \tag{5}$$

where

$$S_{11}(r) = \frac{1}{T}\sum_{t=1}^{[Tr]}[Post_t^{\tau}]^2 \cdot \frac{1}{n}\sum_{i=1}^{n}[\widetilde{Treat_i}]^2,$$

$$S_{21}(r) = \frac{1}{T}\sum_{t=1}^{[Tr]} Post_t^{\tau} \cdot \frac{1}{n}\sum_{i=1}^{n}\tilde{Z}_{it}^{\tau} \cdot \widetilde{Treat_i},$$

$$S_{22}(r) = \frac{1}{T}\sum_{t=1}^{[Tr]}\frac{1}{n}\sum_{i=1}^{n}\tilde{Z}_{it}^{\tau}(\tilde{Z}_{it}^{\tau})'.$$

Let

$$H_{\nu}(r) = 1(r \geq \nu) - \left[\int_0^1 1(s \geq \nu)\tau(s)'ds\right]\left[\int_0^1 \tau(s)\tau(s)'ds\right]^{-1}\tau(r)$$

be the projection of $1(r \geq \nu)$ onto the orthogonal complement of the space spanned by the trend function $\tau(r)$. $H_{\nu}(r)$ is the limit of $Post_{[Tr]}^{\tau}$ as $T \to \infty$.

The following lemma establishes the limits of $S_{11}(r)$, $S_{21}(r)$, and $S_{22}(r)$.

Lemma 3.1 *Let Assumptions 3.2–3.4 hold. Then*
(*a*) $S_{11}(r) = \mathcal{S}_{11}(r) + O\left(T^{-1}\right)$,
(*b*) $S_{21}(r) = \mathcal{S}_{21}(r) + o_p(1)$,
(*c*) $S_{22}(r) = \mathcal{S}_{22}(r) + o_p(1)$,
uniformly over $r \in [0,1]$, *where*

$$\mathcal{S}_{11}(r) = \mu(1-\mu)\int_0^r H_\nu^2(s)\,ds$$

and

$$\mathcal{S}_{21}(r) = 0,\ \mathcal{S}_{22}(r) = rG.$$

Given that $\mathcal{S}_{21}(r) = 0$, Lemma 3.1 shows that the regressor of interest in the serially detrended and cross-sectionally demeaned regression is orthogonal to other regressors. The reason to include Z_{it} in the regression is to reduce the regression error so that we can have a more efficient estimator. The crucial assumption that drives this result is Assumption 3.3. Without this assumption, $\mathcal{S}_{21}(r)$ will not be zero. When $\mathcal{S}_{21}(r) \neq 0$, the additional control variable Z_{it} may help achieve the key identification assumption: there is no systematic difference in ϵ_{it} across the treatment and control groups after controlling for Z_{it}. However, if Z_{it} is causally affected by the policy change, then controlling for Z_{it} may block a channel through which the policy change exerts its effect, leading to a biased causal effect estimator.

To establish the limiting distribution of $\hat{\theta}$ and the asymptotic variance estimator $\hat{V}_R$, we maintain the following functional central limit theorem (FCLT).

Assumption 3.5 $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat}_i \cdot \epsilon_{it}\right) \to^d \Lambda B(r)$ *for some* $\Lambda > 0$.

When n is fixed, Assumption 3.5 is an FCLT for the time series $\{\sum_{i=1}^{n}\widetilde{Treat}_i \cdot \epsilon_{it}/\sqrt{n}\}$. When n grows with T, then $\{\sum_{i=1}^{n}\widetilde{Treat}_i \cdot \epsilon_{it}/\sqrt{n}\}$ should be regarded as a triangular array, and Assumption 3.5 is an FCLT for a triangular array. There is a vast literature on time series FCLT, both for cases where the underlying time series is a triangular array and for cases where it is not. Assumption 3.5 is a high-level assumption. Sufficient conditions often involve some moment and mixing conditions. For example, when n grows with T, we can invoke Theorem 7.18 of White (2001) to show that the following conditions are sufficient.

Condition 3.1 (*i*) $E\frac{1}{nT}\sum_{t=1}^{[Tr]}\sum_{i=1}^{n}\widetilde{Treat}_i \cdot \epsilon_{it} = 0$.
(*ii*) $E(|n^{-1/2}\sum_{i=1}^{n}\widetilde{Treat}_i \cdot \epsilon_{it}|^{\delta}) \leq \Delta < \infty$ *for some* $\delta > 2$.
(*iii*) *The sequence* $\{n^{-1/2}\sum_{i=1}^{n}\widetilde{Treat}_i \cdot \epsilon_{it}\}_{t=1}^{T}$ *is* α*-mixing with* α*-mixing coefficient satisfying* $\alpha(m) = O\left(m^{-\delta/(\delta-2)+e}\right)$ *for some* $\delta > 2$ *and* $e > 0$.
(*iv*) $var[(nT)^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{Treat}_i \cdot \epsilon_{it}] > C > 0$ *for sufficiently large* T.

To verify Condition 3.1(i), we note that

$$\begin{aligned}
&\frac{1}{nT}\sum_{t=1}^{[Tr]}\left(\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}\right)\\
&=(1-\mu)\,\mu\cdot\frac{1}{T}\sum_{t=1}^{[Tr]}\left[\frac{1}{n\mu}\sum_{i=1}^{n\mu}\epsilon_{it}-\frac{1}{n\,(1-\mu)}\sum_{i=n\mu+1}^{n}\epsilon_{it}\right]\\
&=(1-\mu)\,\mu\cdot\frac{1}{T}\sum_{t=1}^{[Tr]}\left(\bar{\epsilon}_{\cdot,t}^{treat}-\bar{\epsilon}_{\cdot,t}^{control}\right).
\end{aligned}$$

So Condition 3.1(i) holds if

$$E\frac{1}{T}\sum_{t=1}^{[Tr]}\left(\bar{\epsilon}_{\cdot,t}^{treat}-\bar{\epsilon}_{\cdot,t}^{control}\right)=0, \tag{6}$$

that is, if there is no systematic difference in the averages of $\bar{\epsilon}_{\cdot,t}^{treat}$ and $\bar{\epsilon}_{\cdot,t}^{control}$ over $t=[Tr_1],\ldots,[Tr_2]$ for any $r_2>r_1$. This is a version of the "parallel paths" assumption in the DD regression.

Condition 3.1(ii) is a type of Rosenthal inequality. It holds if the cross-sectional dependence is weak enough and ϵ_{it} has enough moments. See, for example, Doukhan (1994, Sec 1.4.1). Condition 3.1(iii) is a standard mixing condition. If each time series ϵ_{it} satisfies the given mixing condition, then Condition 3.1(iii) holds. Condition 3.1(iv) rules out the degenerate case in which the variance goes to zero.

Lemma 3.2 *Let Assumptions 3.3–3.5 hold. Then*

$$\sqrt{nT}(\hat{\theta}_1-\theta_{10})\to^d\frac{\Lambda}{\mu\,(1-\mu)}\frac{\int_0^1 H_\nu(r)\,dB(r)}{\int_0^1 H_\nu^2(r)\,dr}\overset{d}{=}\frac{\Lambda}{\mu\,(1-\mu)\sqrt{\int_0^1 H_\nu^2(r)\,dr}}N(0,1). \tag{7}$$

For Lemma 3.2 to hold, we need Assumption 3.5 for only $r=\nu$ and 1 and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tau_D(t)\,\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}\to^d\Lambda\int_0^1\tau_D(r)\,dB(r).$$

In this case, (6) needs to hold for only $r=\nu$ and 1. That is, the averages of $\bar{\epsilon}_{\cdot,t}^{treat}$ and $\bar{\epsilon}_{\cdot,t}^{control}$ over the pre-treatment periods (and post-treatment periods) are the same in the mean sense. This is the usual "parallel paths" assumption for identification in the absence of a linear trend. We maintain the stronger Assumption 3.5 for technical convenience and for establishing the asymptotic distribution of the asymptotic variance estimator to be defined later.

Note that we obtain the $\sqrt{nT}$ rate of convergence of $\hat{\theta}_1$ when both T and n approach infinity, because we have implicitly assumed weak cross-sectional dependence. The Rosenthal-type inequality in Condition 3.1(ii) holds only if the cross-sectional dependence is weak enough. If there is a group effect in ϵ_{it} such that $\epsilon_{it}=Treat_i\times e_t^{(1)}+(1-Treat_i)\times e_t^{(2)}+\tilde{\epsilon}_{it}$ for some sequences $e_t^{(1)}$ and $e_t^{(2)}$ where $\{\tilde{\epsilon}_{it}\}$ are independent for different i or t, then Condition 3.1(ii) cannot hold when $n\to\infty$. In this case, we have to use a different argument. Instead of requiring that $n^{-1/2}T^{-1/2}\sum_{t=1}^{[Tr]}(\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it})$ satisfy an FCLT, we require that

$n^{-1}T^{-1/2}\sum_{t=1}^{[Tr]}(\sum_{i=1}^{n}\widetilde{Treat}_i \cdot \epsilon_{it})$ satisfy an FCLT. As a consequence, $\sqrt{nT}$ asymptotic normality in Lemma 3.2 will be reduced to $\sqrt{T}$ asymptotic normality. To reflect this, we need to make some minor changes to our theoretical results and their proofs, but our proposed testing procedures remain asymptotically valid without any modification.

Lemma 3.3 *Let Assumptions 3.1–3.5 hold. Then*

$$\hat{V}_R \to^d \Lambda^2 \frac{1}{K}\sum_{k=1}^{K}\left[\int_0^1 \Phi_k^H(r)\, dB(r)\right]^2 \left[\mu(1-\mu)\int_0^1 H_\nu^2(s)\, ds\right]^{-2}$$

jointly with (7), where

$$\Phi_k^H(r) = \left[\Phi_k(r) - \bar{\Phi}_{H^2,k}\right] H_\nu(r) - \left[\int_0^1 \Phi_k(s) H_\nu(s)\tau(s)'\, ds\right]\left[\int_0^1 \tau(s)\tau(s)'\, ds\right]^{-1}\tau(r)$$

and

$$\bar{\Phi}_{H^2,k} = \left[\int_0^1 \Phi_k(s) H_\nu^2(s)\, ds\right]\left[\int_0^1 H_\nu^2(s)\, ds\right]^{-1}.$$

The term $\bar{\Phi}_{H^2,k} H_\nu(r)$ in $\Phi_k^H(r)$ reflects the effect of the estimation uncertainty in $\hat{\theta}_1$ on the asymptotic distribution of $\hat{V}_R$. If $\bar{\Phi}_{H^2,k} = 0$ for all k, that is, the basis functions are orthogonal to the information accumulation process as reflected in $H_\nu^2(\cdot)$, then this term vanishes, and the estimation error in $\hat{\theta}_1$ has no effect on the asymptotic distribution of $\hat{V}_R$. The remaining terms in $\Phi_k^H(r)$ are the L^2 projection of $\Phi_k(r) H_\nu(r)$ onto the orthogonal complement of the space spanned by the trend function $\tau(r)$. The projection is present because we do not observe ϵ_{it}. Even if we know the true θ_0, we can only hope to recover ϵ_{it}^τ, the projected version of ϵ_{it}. For this reason, the stochastic approximation of the series LRV estimator involves the term $\sum_{t=1}^{T}\Phi_k(t/T)\, Post_t^\tau \cdot \epsilon_{it}^\tau$. Rearranging the projecting operation, this term is numerically identical to $\sum_{t=1}^{T}(\Phi_k(t/T)\, Post_t^\tau)^\tau \cdot \epsilon_{it}$, where $(\Phi_k(t/T)\, Post_t^\tau)^\tau$ is the projected version of $\Phi_k(t/T)\, Post_t^\tau$. As $T \to \infty$, the effect of this projection is manifested in the L^2 projection of $\Phi_k(r) H_\nu(r)$. It is useful to point out that, by construction, $\Phi_k^H(r)$ is orthogonal to $1\{r \geq \nu\}$, $\tau(r)$, and $H_\nu(r)$.

Note that the estimation uncertainty in $\hat{\theta}_2$ has no effect on the asymptotic distribution of $\hat{V}_R$. This is due to the information orthogonality given in Lemma 3.1.

Theorem 3.1 *Let Assumptions 3.1–3.5 hold. Then*

$$\mathbb{T} \to^d \mathcal{T}_\infty := \frac{\int_0^1 H_\nu(r)\, dB(r)}{\left\{\frac{1}{K}\sum_{k=1}^{K}\left[\int_0^1 \Phi_k^H(r)\, dB(r)\right]^2\right\}^{1/2}}.$$

Like the finite sample distributions, the limiting distribution of $\mathbb{T}$ depends on the trend function included in the regression, the basis functions used in the asymptotic variance estimation, and the number of basis functions used. This is an attractive feature of the fixed-smoothing approximation, as it captures the effects of the trend function and the variance estimator, which clearly affect the finite sample distribution of $\mathbb{T}$.

The limiting distribution $\mathcal{T}_\infty$ is the same regardless of whether time fixed effects or individual fixed effects are included in the regression. Moreover, it does not depend on the relative sizes of the two groups. These features make the limiting distribution easy to use. However, it does depend on the length of the post-treatment periods relative to that of the pre-treatment periods.

Figure 1 plots the nonstandard critical values against the value of K. The critical values are for a two-sided 5% test. We consider two choices of $\tau(t) : \tau(t) = 1$ and $\tau(t) = (1, t)'$, leading to a model with no any trend and a model with linear trends, respectively. It is clear that the critical values depend on ν, which characterizes the time at which the policy change takes place. They also depend on the form of the trend function $\tau(t)$ and the number of basis functions used. In all cases, the critical value decreases with K and approaches the standard normal critical value, i.e., 1.96, as K increases. While the standard normal critical value stays the same regardless of the time at which the policy change takes place, the form of the trend function, and the number of basis functions, the nonstandard critical value is tailored to each specific case. That is why the asymptotic nonstandard test has more accurate size than the asymptotic normal test.

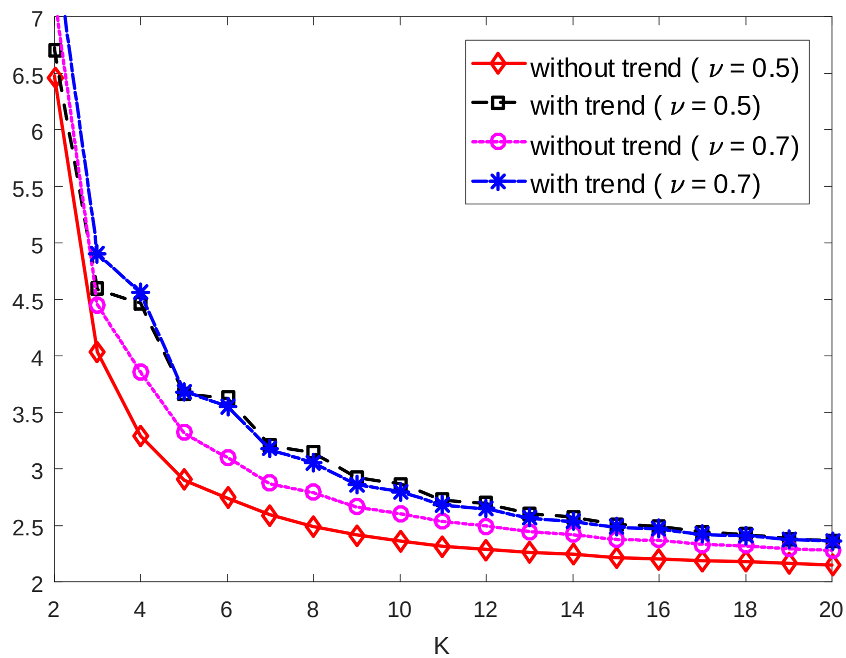

Figure 1: Nonstandard fixed-smoothing critical values for models with and without linear trends and for different values of ν

4 Asymptotic t Test

The limiting distribution is pivotal but nonstandard. One advantage of using the series LRV estimator is that we have the freedom to choose the basis functions. We hope to choose a set of basis functions such that $\mathcal{T}_\infty$ becomes the standard t distribution.

Define

$$\xi_0 = \frac{1}{\|H_\nu\|_2} \int_0^1 H_\nu(r)\, dB(r) \sim N(0,1),$$
$$\xi_k = \frac{1}{\|H_\nu\|_2} \int_0^1 \Phi_k^H(r)\, dB(r) \text{ for } k = 1, \ldots, K$$

where $\|H_\nu\|_2 = (\int_0^1 H_\nu^2(s)\, ds)^{1/2}$. Then

$$\mathcal{T}_\infty = \frac{\xi_0}{\left(\frac{1}{K}\sum_{k=1}^K \xi_k^2\right)^{1/2}}.$$

Note that for $k = 1, \ldots, K$,

$$cov(\xi_0, \xi_k) = \|H_\nu\|_2^{-2} \int_0^1 \Phi_k^H(r)\, H_\nu(r)\, dr = \|H_\nu\|_2^{-2} \int_0^1 \left[\Phi_k(r) - \bar{\Phi}_{H^2,k}\right] H_\nu^2(r)\, dr = 0,$$

using the definition of $\bar{\Phi}_{H^2,k}$ and $\int_0^1 H_\nu(s)\, \tau(s)'\, ds = 0$. So ξ_0 is independent of ξ_k for $k = 1, 2, \ldots, K$. If $\xi_k \sim iid\ N(0,1)$ for $k = 1, \ldots, K$, then $\mathcal{T}_\infty$ follows the standard t distribution with K degrees of freedom.

Some simple calculations show that for $k_1, k_2 = 1, 2, \ldots, K$,

$$cov(\xi_{k_1}, \xi_{k_2}) = \int_0^1 \int_0^1 \Phi_{k_1}(r)\, C_\nu^H(r,s)\, \Phi_{k_2}(s)\, drds,$$

where

$$C_\nu^H(r,s) = \frac{H_\nu(r)}{\|H_\nu\|_2} \left\{ \delta(r-s) - H_\nu(r)\, H_\nu(s) - \tau(r)' \left[\int_0^1 \tau(t)\, \tau(t)'\, dt\right]^{-1} \tau(s) \right\} \frac{H_\nu(s)}{\|H_\nu\|_2}$$

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

$$\int_0^1 \int_0^1 \Phi_{k_1}(r)\, H_\nu(r)\, \delta(r-s)\, \Phi_{k_2}(s)\, H_\nu(s)\, drds = \int_0^1 \Phi_{k_1}(r)\, \Phi_{k_2}(r)\, H_\nu^2(r)\, dr.$$

To ensure that $\xi_k \sim iid\ N(0,1)$ for $k = 1, 2, \ldots, K$, we require that

$$\int_0^1 \int_0^1 \Phi_{k_1}(r)\, C_\nu^H(r,s)\, \Phi_{k_2}(s)\, drds = 1\{k_1 = k_2\} \text{ for } k_1, k_2 = 1, \ldots, K. \tag{8}$$

Instead of searching for the basis functions that satisfy (8), we search for their discrete versions: the basis vectors. For each basis function $\Phi_k(r)$, the corresponding basis vector is defined as

$$\mathbf{\Phi}_k = \left(\Phi_k\left(\frac{1}{T}\right), \Phi_k\left(\frac{2}{T}\right), \ldots, \Phi_k\left(\frac{T}{T}\right)\right)'.$$

We focus on the basis vectors for two reasons. First, it is computationally more convenient to obtain the basis vectors. Second, it is the basis vectors that are actually used in the variance estimation.

Let $\mathbf{C}_H$ be the $T \times T$ matrix whose (i, j)-th element is equal to

$$C_\nu^H\left(\frac{i}{T}, \frac{j}{T}\right) = \left[\frac{1}{T}\sum_{t=1}^{T} H_\nu^2\left(\frac{t}{T}\right)\right]^{-1} \left\{T H_\nu^2\left(\frac{i}{T}\right) 1\{i=j\} - H_\nu^2\left(\frac{i}{T}\right) H_\nu^2\left(\frac{j}{T}\right)\right.$$
$$\left. - H_\nu\left(\frac{i}{T}\right)\tau\left(\frac{i}{T}\right)\left[\frac{1}{T}\sum_{\ell=1}^{T}\tau\left(\frac{\ell}{T}\right)\tau\left(\frac{\ell}{T}\right)'\right]^{-1}\tau\left(\frac{j}{T}\right) H_\nu\left(\frac{j}{T}\right)\right\}.$$

By definition, $\mathbf{C}_H$ is a positive-definite symmetric matrix. For any two vectors $\ell_1, \ell_2 \in \mathbb{R}^T$, we define the inner product

$$\langle \ell_1, \ell_2 \rangle = \ell_1' \mathbf{C}_H \ell_2 / T^2, \tag{9}$$

which makes $\mathbb{R}^T$ a Hilbert space. The discrete analogue of (8) is

$$\langle \mathbf{\Phi}_{k_1}, \mathbf{\Phi}_{k_2} \rangle = 1\{k_1 = k_2\} \text{ for } k_1, k_2 = 1, \ldots, K. \tag{10}$$

Note that (10) is different from the usual orthonormality in the Euclidean sense. In general, the basis vectors $\{\mathbf{\Phi}_k\}$ do not satisfy (10) even if they are orthonormal according to the usual inner product in $\mathbb{R}^T$. However, given any set of candidate basis functions or vectors $\{\mathbf{\Phi}_k, k = 1, 2, \ldots, K\}$, we can make them satisfy the above conditions via the Gram-Schmidt orthogonalization.

More specifically, we let

$$\begin{aligned}
&\tilde{\mathbf{\Phi}}_1 = \mathbf{\Phi}_1, \\
&\tilde{\mathbf{\Phi}}_2 = \mathbf{\Phi}_2 - \tfrac{\langle \mathbf{\Phi}_2, \tilde{\mathbf{\Phi}}_1 \rangle}{\langle \tilde{\mathbf{\Phi}}_1, \tilde{\mathbf{\Phi}}_1 \rangle} \tilde{\mathbf{\Phi}}_1, \\
&\cdots \\
&\tilde{\mathbf{\Phi}}_K = \mathbf{\Phi}_K - \tfrac{\langle \mathbf{\Phi}_K, \tilde{\mathbf{\Phi}}_{K-1,} \rangle}{\langle \tilde{\mathbf{\Phi}}_{K-1}, \tilde{\mathbf{\Phi}}_{K-1} \rangle} \tilde{\mathbf{\Phi}}_{K-1} - \ldots - \tfrac{\langle \mathbf{\Phi}_K, \tilde{\mathbf{\Phi}}_1 \rangle}{\langle \tilde{\mathbf{\Phi}}_1, \tilde{\mathbf{\Phi}}_1 \rangle} \tilde{\mathbf{\Phi}}_1.
\end{aligned}$$

By construction, $\left\langle \tilde{\mathbf{\Phi}}_{k_1}, \tilde{\mathbf{\Phi}}_{k_2} \right\rangle = 0$ for $k_1 \neq k_2$. Let

$$\tilde{\mathbf{\Phi}}_{k,H} = \frac{\tilde{\mathbf{\Phi}}_k}{\sqrt{\left\langle \tilde{\mathbf{\Phi}}_k, \tilde{\mathbf{\Phi}}_k \right\rangle}},$$

then $\{\tilde{\mathbf{\Phi}}_{1,H}, \ldots, \tilde{\mathbf{\Phi}}_{K,H}\}$ is a set of bases in $\mathbb{R}^T$ that satisfies the conditions in (10).

Let $\mathbf{\Phi} = (\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K)$. To obtain $\tilde{\mathbf{\Phi}}_H = (\tilde{\mathbf{\Phi}}_{1,H}, \ldots, \tilde{\mathbf{\Phi}}_{K,H})$ in a matrix programming environment, we first compute the upper triangular factor R_H of the Cholesky decomposition of $\mathbf{\Phi}' \mathbf{C}_H \mathbf{\Phi} / T^2$ such that $\mathbf{\Phi}' \mathbf{C}_H \mathbf{\Phi} / T^2 = R_H' R_H$. We can then let

$$\tilde{\mathbf{\Phi}}_H = \mathbf{\Phi} (R_H)^{-1}.$$

For such a choice of $\tilde{\mathbf{\Phi}}_H$, we have

$$(\tilde{\mathbf{\Phi}}_H)' \mathbf{C}_H \tilde{\mathbf{\Phi}}_H / T^2 = \left(R_H'\right)^{-1} \mathbf{\Phi} \mathbf{C}_H \mathbf{\Phi} (R_H)^{-1} / T^2 = \left(R_H'\right)^{-1} R_H' R_H (R_H)^{-1} = I_K,$$

so the conditions in (10) are satisfied.

As $T \to \infty$, $\mathbf{\Phi}' \mathbf{C}_H \mathbf{\Phi}/T^2$ converges to the variance Σ_ξ of $\xi = (\xi_1, \ldots, \xi_K)'$. This implies that R_H converges to the upper triangular factor of the Cholesky decomposition of Σ_ξ. As a result, every transformed basis vector is approximately equal to a linear combination of the original basis vectors. The implied basis function is thus equal to a linear combination of the original basis functions. Therefore, if Assumption 3.1 holds for the original basis functions, it also holds for the transformed basis functions.

Using $\{\tilde{\mathbf{\Phi}}_{k,H}\}$ as the basis vectors for construction of the asymptotic variance estimator, we have

$$\mathcal{T}_\infty =^d \frac{\xi_0}{\left(\frac{1}{K}\sum_{k=1}^K \xi_k^2\right)^{1/2}} =^d t_K.$$

That is, the t statistic $\mathbb{T}$ is asymptotically distributed as the standard t distribution with K degrees of freedom.

To sum up, the asymptotic t test consists of the following steps:

1. Estimate the parameter of interest.

   (a) Detrend each time series separately, and then remove the cross-sectional average from each detrended variable.

   (b) Estimate θ_{10} and θ_{20} by running the OLS regression

   $$\tilde{Y}_{it}^\tau = \tilde{X}_{it}^\tau \theta_0 + \tilde{\epsilon}_{it}^\tau,$$

   where $\tilde{Y}_{it}^\tau$ and $\tilde{X}_{it}^\tau$ are the transformed variables given in (1) and (3), respectively. Denote the estimates by $\hat{\theta}_1$ and $\hat{\theta}_2$ and the residual by $\hat{\epsilon}_{it}^\tau$.

2. Construct the transformed basis vectors.

   (a) Letting $Post_t^\tau$ be the detrended "*Post*" dummy

   $$Post_t^\tau = Post_t - \left(\sum_{s=1}^T Post_s \cdot \tau(s)'\right)\left(\sum_{s=1}^T \tau(s)\tau(s)'\right)^{-1} \tau(t)$$

   and

   $$\|Post^\tau\| = \left\{\frac{1}{T}\sum_{t=1}^T \left[Post_t^\tau\right]^2\right\}^{1/2} \tag{11}$$

   be its empirical norm, construct the following vectors and matrices:

   $$\mathbf{H}_T = \left(\left[Post_1^\tau\right]^2, \left[Post_2^\tau\right]^2, \ldots, \left[Post_T^\tau\right]^2\right)' \in \mathbb{R}^{T\times 1},$$

   $$\mathbf{A} = \frac{1}{T}\sum_{t=1}^T \tau(t)\tau(t)' \in \mathbb{R}^{d_\tau \times d_\tau},$$

   $$\mathbf{B} = \begin{pmatrix} Post_1^\tau \cdot \tau(1)' \\ \cdots \\ Post_T^\tau \cdot \tau(T)' \end{pmatrix} \in \mathbb{R}^{T\times d_\tau},$$

   $$\mathbf{C}_H = \frac{1}{\|Post^\tau\|^2}\left[T \cdot \mathrm{diag}(\mathbf{H}_T) - \mathbf{H}_T\mathbf{H}_T' - \mathbf{B}\mathbf{A}^{-1}\mathbf{B}'\right] \in \mathbb{R}^{T\times T}.$$

(b) Let $\mathbf{\Phi} = [\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K] \in \mathbb{R}^{T\times K}$ be the matrix of the original basis vectors. For example, when K is even, we can take the columns of $\mathbf{\Phi}$ to be

$$\mathbf{\Phi}_{2j-1} = \left[\sqrt{2}\cos(2j\pi \cdot 1/T), \sqrt{2}\cos(2j\pi \cdot 2/T), \ldots, \sqrt{2}\cos(2j\pi \cdot T/T)\right]', \tag{12}$$

$$\mathbf{\Phi}_{2j} = \left[\sqrt{2}\sin(2j\pi \cdot 1/T), \sqrt{2}\sin(2j\pi \cdot 2/T), \ldots, \sqrt{2}\sin(2j\pi \cdot T/T)\right]', \tag{13}$$

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

Compute the upper triangular factor R_H of the Cholesky decomposition of $\mathbf{\Phi}'\mathbf{C}_H\mathbf{\Phi}/T^2$ such that $\mathbf{\Phi}'\mathbf{C}_H\mathbf{\Phi}/T^2 = R_H' R_H$.

(c) Compute the matrix $\tilde{\mathbf{\Phi}}_H = [\tilde{\mathbf{\Phi}}_{1,H}, \ldots, \tilde{\mathbf{\Phi}}_{K,H}] = \mathbf{\Phi}\,(R_H)^{-1}$. Each column of $\tilde{\mathbf{\Phi}}_H$ consists of a transformed basis vector.

3. Compute the variance estimator and the t-statistic.

(a) Letting

$$\hat{\Omega} = \frac{1}{K}\sum_{k=1}^{K}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{\mathbf{\Phi}}_{k,H,t}\hat{u}_{n,t}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{\mathbf{\Phi}}_{k,H,t}\hat{u}_{n,t}\right]',$$

where $\hat{u}_{nt} = n^{-1/2}\sum_{i=1}^{n}\tilde{X}_{it}^{\tau}\hat{\epsilon}_{it}^{\tau}$ and $\tilde{\mathbf{\Phi}}_{k,H,t}$ is the t-th element of $\tilde{\mathbf{\Phi}}_{k,H}$, compute

$$\hat{V} = \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^{\tau}(\tilde{X}_{it}^{\tau})'\right]^{-1}\hat{\Omega}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^{\tau}(\tilde{X}_{it}^{\tau})'\right]^{-1}.$$

(b) Construct the t statistic for testing the null $H_0 : \theta_{10} = r_0$:

$$\mathbb{T} = \frac{\sqrt{nT}(\hat{\theta}_1 - r_0)}{\sqrt{\hat{V}_{11}}},$$

where $\hat{V}_{11}$ is the (1,1)-th element of $\hat{V}$.

(c) On the basis of $\mathbb{T}$, perform the asymptotic t test using the critical values from Student's t distribution with K degrees of freedom.

5 Alternative Variance Estimator and Asymptotic t Test

Lemma 3.2 shows that

$$\sqrt{nT}(\hat{\theta}_1 - \theta_{10}) \to^d N\left(0, \frac{\Lambda^2}{\mu(1-\mu)}\frac{1}{\mu(1-\mu)\int_0^1 H_\nu^2(r)\,dr}\right). \tag{14}$$

All the components in the asymptotic variance other than Λ^2 can be estimated easily. More specifically, $\mu(1-\mu)\int_0^1 H_\nu^2(r)\,dr$ can be estimated by

$$S_{11}(1) = \frac{1}{T}\sum_{t=1}^{T}[Post_t^{\tau}]^2 \cdot \frac{1}{n}\sum_{i=1}^{n}[\widetilde{Treat_i}]^2,$$

and $\mu(1-\mu)$ can be estimated by $n^{-1}\sum_{i=1}^{n}[\widetilde{Treat_i}]^2$. It suffices to estimate Λ^2, the long-run variance of $\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}/\sqrt{n}$, in order to make inference about θ_{10}.

The series estimator of Λ^2 can be constructed as

$$\hat{\Lambda}^2=\frac{1}{K}\sum_{k=1}^{K}(\hat{\Lambda}_k)^2,$$

where

$$\hat{\Lambda}_k=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\hat{\epsilon}_{it}^{\tau}.$$

The asymptotic variance of $\hat{\theta}_1$ can then be estimated by

$$\hat{\sigma}^2=\hat{\Lambda}^2\cdot\left[\frac{1}{n}\sum_{i=1}^{n}(\widetilde{Treat_i})^2\right]^{-2}\left\{\frac{1}{T}\sum_{t=1}^{T}[Post_t^{\tau}]^2\right\}^{-1}.$$

The corresponding t statistic is

$$\tilde{\mathbb{T}}=\frac{\sqrt{nT}(\hat{\theta}_1-\theta_{10})}{\hat{\sigma}}.$$

While $\hat{V}_R$ is the sandwich variance estimator in the most general form, $\hat{\sigma}^2$ can be regarded as a special form of the sandwich estimator: a "collapsed" sandwich estimator. They are asymptotically equivalent when K approaches infinity. However, when K is fixed, they have different limiting distributions.

The above construction, which provides an alternative to the one in Section 4, is specific to the DD setting where the regressor of interest $Post_t^{\tau}$ is a deterministic function whose energy is concentrated at the origin. This setting is similar to trend regressions or cointegrating regressions where the regressors have energy concentrated at the origin.

Theorem 5.1 *Let Assumptions 3.1–3.5 hold. Then*

(*a*)

$$\hat{\sigma}^2\to^d\Lambda^2\frac{1}{K}\sum_{k=1}^{K}\left[\int_0^1\Phi_k^{\mathcal{H}}(r)\,dB(r)\right]^2\cdot\left[\mu(1-\mu)\int_0^1 H_\nu^2(s)\,ds\right]^{-2}$$

jointly with (7), where

$$\Phi_k^{\mathcal{H}}(r)=\Phi_k(r)-(P_H\Phi_k)\cdot H_\nu(r)-\left[\int_0^1\Phi_k(s)\tau(s)'\right]\left[\int_0^1\tau(s)\tau(s)'\,ds\right]^{-1}\tau(r)$$

and

$$P_H\Phi_k=\left[\int_0^1\Phi_k(r)H_\nu(r)\,dr\right]\left[\int_0^1 H_\nu^2(s)\,ds\right]^{-1}.$$

(*b*)

$$\tilde{\mathbb{T}}\to^d\tilde{\mathcal{T}}_\infty:=\frac{\int_0^1 H_\nu(r)\,dB(r)}{\left\{\frac{1}{K}\sum_{k=1}^{K}\left[\int_0^1\Phi_k^{\mathcal{H}}(r)\,dB(r)\right]^2\right\}^{1/2}\left(\int_0^1 H_\nu^2(s)\,ds\right)^{1/2}}.$$

Theorem 5.1(a) is analogous to Lemma 3.3. The term $(P_H\Phi_k) H_\nu(r)$ in $\Phi_k^{\mathcal{H}}(r)$ reflects the effect of the estimation uncertainty in $\hat{\theta}_1$. If the projection of $\Phi_k(r)$ onto $H_\nu(r)$ is zero, then this term disappears. The remaining terms in $\Phi_k^{\mathcal{H}}(r)$ are the L^2 projection of $\Phi_k(r)$ onto the orthogonal complement of the space spanned by the trend functions in $\tau(r)$. We can also write

$$\Phi_k^{\mathcal{H}}(r) = \Phi_k(r) - \tilde{c}_k \cdot 1(r \geq \nu) - \tilde{d}_k' \cdot \tau(r)$$

for

$$\tilde{c}_k = P_H\Phi_k \text{ and } \tilde{d}_k = \left[\int_0^1 \tau(s)\tau(s)'ds\right]^{-1}\left\{\int_0^1 \left[\Phi_k(s) - (P_H\Phi_k)\cdot 1(s \geq \nu)\right]\tau(s)ds\right\}.$$

So, $\Phi_k^{\mathcal{H}}(r)$ is the L^2 projection of $\Phi_k(r)$ onto the orthogonal complement of the space spanned by $1(r \geq \nu)$ and the trend function $\tau(r)$. Theorem 5.1(b) is analogous to Theorem 3.1. Like $\mathcal{T}_\infty$, the asymptotic distribution $\tilde{\mathcal{T}}_\infty$ is pivotal, and it captures the effects of the trend function, the basis functions, and the time at which the policy change takes place.

While it is not hard to simulate the nonstandard critical values, it is more convenient to use critical values that are readily available from standard statistical tables and programming environments. Let

$$\eta_0 = \frac{\int_0^1 H_\nu(r)\, dB(r)}{\left(\int_0^1 H_\nu^2(s)\, ds\right)^{1/2}}$$

and

$$\eta_k = \int_0^1 \Phi_k^{\mathcal{H}}(r)\, dB(r), k = 1, ..., K,$$

which are all normal. Then

$$\tilde{\mathcal{T}}_\infty = \frac{\eta_0}{\left(\frac{1}{K}\sum_{k=1}^K \eta_k^2\right)^{1/2}}.$$

Since

$$cov(\eta_0, \eta_k) = \left(\int_0^1 H_\nu^2(s)\, ds\right)^{-1/2} \int_0^1 H_\nu(r)\, \Phi_k^{\mathcal{H}}(r)\, dr = 0, \text{ for } k = 1, 2, \ldots, K,$$

η_0 and η_k are independent for $k = 1, 2, \ldots, K$. Also some calculations show that

$$cov(\eta_{k_1}, \eta_{k_2}) = \int_0^1 \Phi_{k_1}^{\mathcal{H}}(r) \cdot \Phi_{k_2}^{\mathcal{H}}(r)\, dr = \int_0^1\int_0^1 \Phi_{k_1}(r)\, C_\nu^{\mathcal{H}}(r, s)\, \Phi_{k_2}(s)\, drds,$$

where

$$C_\nu^{\mathcal{H}}(r, s) = \delta(r - s) - \frac{H_\nu(r)\, H_\nu(s)}{\int_0^1 H_\nu^2(t)\, dt} - \tau(r)'\left[\int_0^1 \tau(t)\,\tau(t)'\, dt\right]^{-1}\tau(s). \tag{15}$$

We can orthonormalize the basis functions $\{\Phi_k(r)\}$ so that η_k becomes $iid\ N(0, 1)$. In this case, the asymptotic distribution of $\tilde{\mathbb{T}}$ becomes a t distribution, leading to an alternative asymptotic t test.

The steps for the alternative asymptotic t test are similar to those given at the end of Section 4. For completeness, we outline the steps below.

1. Follow the same step as before to estimate the parameter of interest.

2. Transform the original basis vectors.

   (a) Let $\boldsymbol{\tau} = (\tau(1)', \ldots, \tau(T)')' \in \mathbb{R}^{T\times d_\tau}$ and $Post^\tau = (Post_1^\tau, \ldots, Post_T^\tau)' \in \mathbb{R}^{T\times 1}$. Construct the projection matrix

$$\mathbf{C}_\mathcal{H} = T\left[I_{T\times T} - Post^\tau \cdot \left(Post^{\tau\prime} Post^\tau\right)^{-1}(Post^\tau)' - \boldsymbol{\tau}\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}'\right] := T \cdot M_{Post,\tau}. \tag{16}$$

   (b) Let $R_\mathcal{H}$ be the upper triangular factor of the Cholesky decomposition of $\mathbf{\Phi}'\mathbf{C}_\mathcal{H}\mathbf{\Phi}/T^2$, where, as before, $\mathbf{\Phi} = [\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K]$ is the matrix of the original basis vectors. Compute the matrix

$$\mathbf{\Phi}_\mathcal{H} = [\mathbf{\Phi}_{1,\mathcal{H}}, \ldots, \mathbf{\Phi}_{K,\mathcal{H}}] = \mathbf{\Phi}\left(R_\mathcal{H}\right)^{-1}.$$

3. Compute the variance estimator and perform the t test.

   (a) Estimate the asymptotic variance of $\hat{\theta}$ by

$$\hat{\sigma}^2 = \hat{\Lambda}^2 \cdot \left[\frac{1}{n}\sum_{i=1}^n (\widetilde{Treat_i})^2\right]^{-2}\left\{\frac{1}{T}\sum_{t=1}^T [Post_t^\tau]^2\right\}^{-1},$$

   where

$$\hat{\Lambda}^2 = \frac{1}{K}\sum_{k=1}^K\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbf{\Phi}_{k,\mathcal{H},t}\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{Treat_i}\cdot\hat{\epsilon}_{it}^\tau\right]^2$$

   and $\mathbf{\Phi}_{k,\mathcal{H},t}$ is the t-th element of the vector $\mathbf{\Phi}_{k,\mathcal{H}}$.

   (b) Perform the test using $\tilde{\mathbb{T}} = \sqrt{nT}(\hat{\theta}_1 - \theta_{10})/\hat{\sigma}$ as the test statistic and Student's t distribution with K degrees of freedom as the reference distribution.

6 Testing-Optimal Choice of K

In this section, we propose a testing-optimal choice of the smoothing parameter K. The proposed method is based on high-order approximations of the type I and type II errors of the asymptotic t test in the previous section.

We consider the DD regression without additional covariates Z_{it} and assume that the error term ϵ_{it} is Gaussian. More general models with non-Gaussian errors or with covariates that can vary in arbitrary ways across both the time dimension and the cross-sectional dimension require highly technical arguments. For example, when the errors are not Gaussian, we have to follow the most general approach to develop Edgeworth expansions for time series data. This often requires highly technical assumptions that are difficult to verify. See, for example, Sun and Phillips (2009) for the technical assumptions and a full-fledged Edgeworth expansion. While the asymptotic testing-optimal rule for smoothing-parameter choice that we develop for the special case may not be theoretically optimal for more general cases in large samples, it may still be quite informative in finite samples. The results of our simulations lend some support to this possibility.

In the absence of Z_{it}, the DD estimator $\hat{\theta}_1$ is numerically identical to the OLS estimator based on the regression model

$$M_\tau \mathcal{Y}_t = M_\tau \cdot Post_t \cdot \sqrt{n}\mu(1-\mu)\theta_{10} + M_\tau e_t, \tag{17}$$

where

$$\mathcal{Y}_t = \sqrt{n}\mu\,(1-\mu)\left(\frac{1}{n\mu}\sum_{i=1}^{n\mu} Y_{it} - \frac{1}{n(1-\mu)}\sum_{i=n\mu+1}^{n} Y_{it}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat}_i \cdot Y_{it},$$

$$e_t = \sqrt{n}\mu\,(1-\mu)\left(\frac{1}{n\mu}\sum_{i=1}^{n\mu} \epsilon_{it} - \frac{1}{n\,(1-\mu)}\sum_{i=n\mu+1}^{n} \epsilon_{it}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat}_i \cdot \epsilon_{it}, \tag{18}$$

and $M_\tau = I_T - \boldsymbol{\tau}\left(\boldsymbol{\tau}\boldsymbol{\tau}'\right)^{-1}\boldsymbol{\tau}'$. In fact, it is easy to rigorously establish the numerical equivalence. To highlight the estimation method behind $\hat{\theta}_1$, in this section we write

$$\hat{\theta}_1 = \hat{\theta}_{1,\mathrm{OLS}} = \frac{1}{\sqrt{n}\mu\,(1-\mu)}\left(Post'\cdot M_\tau \cdot Post\right)^{-1}\left(Post'\cdot M_\tau\cdot \mathcal{Y}\right),$$

where $Post = (Post_1, Post_2, ..., Post_T)'$ and $\mathcal{Y} = (\mathcal{Y}_1, \mathcal{Y}_2, ..., \mathcal{Y}_T)'$.

Denote the variance matrix of $e = (e_1, e_2, ..., e_T)'$ by Ω. On the basis of (17), we can also estimate θ_1 by the generalized least-squares estimator:

$$\hat{\theta}_{1,\mathrm{GLS}} = \frac{1}{\sqrt{n}\mu\,(1-\mu)}\left[(M_\tau Post)'\left(M_\tau \Omega M_\tau'\right)^{-} M_\tau Post\right]^{-1}\left[(M_\tau Post)'\left(M_\tau\Omega M_\tau'\right)^{-} M_\tau \mathcal{Y}\right],$$

where $(M_\tau \Omega M_\tau')^{-}$ is the Moore-Penrose pseudoinverse of $M_\tau \Omega M_\tau'$.

By direct calculation, it's easy to show that $E(\hat{\theta}_{1,\mathrm{GLS}} - \theta_{10})(\hat{\theta}_{1,\mathrm{GLS}} - \hat{\theta}_{1,\mathrm{OLS}}) = 0$. In addition, letting

$$\hat{e}^\tau = [I_T - M_\tau \cdot Post \cdot \left(Post' \cdot M_\tau \cdot Post\right)^{-1} Post' \cdot M_\tau] M_\tau e$$

be the OLS residual, we can show that

$$E(\hat{\theta}_{1,\mathrm{GLS}} - \theta_{10})\,(\hat{e}^\tau)' = 0.$$

Hence $\hat{\theta}_{1,\mathrm{GLS}} - \theta_{10}$ is independent of both $\hat{\theta}_{1,\mathrm{GLS}} - \hat{\theta}_{1,\mathrm{OLS}}$ and $\hat{e}^\tau$. Using the definition of e_t given in (18), we can show that $\hat{e}^\tau = \sum_{i=1}^{n}\widetilde{Treat}_i \cdot \hat{\epsilon}_{it}^\tau/\sqrt{n}$. It then follows that

$$\hat{\sigma}^2 = \frac{1}{K}\sum_{k=1}^{K}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{\Phi}_{k,\mathcal{H},t}\hat{e}_t^\tau\right]^2 [\mu\,(1-\mu)]^{-2}\left\{\frac{1}{T}\sum_{t=1}^{T}[Post_t^\tau]^2\right\}^{-1}, \tag{19}$$

which is a quadratic form in $\hat{e}_t^\tau$. Therefore, $\hat{\theta}_{1,\mathrm{GLS}} - \theta_{10}$ is also independent of $\hat{\sigma}^2$.

Let Ψ and ψ be the cdf and pdf of the standard norm distribution, respectively. Denote $\sigma^2_{\mathrm{GLS}} = var\left[\sqrt{nT}(\hat{\theta}_{1,\mathrm{GLS}} - \theta_{10})\right]$. Using the independence of $\hat{\theta}_{1,\mathrm{GLS}} - \theta_{10}$ from $\hat{\theta}_{1,\mathrm{GLS}} - \hat{\theta}_{1,\mathrm{OLS}}$

and $\hat{\sigma}^2$, we obtain, for any $z \in \mathbb{R}$,

$$
\begin{aligned}
&P\left(\frac{\sqrt{nT}(\hat{\theta}_{1,\mathrm{OLS}}-\theta_{10})}{\hat{\sigma}} \le z\right) = P\left(\frac{\sqrt{nT}(\hat{\theta}_{1,\mathrm{OLS}}-\theta_{10})}{\sigma_{\mathrm{GLS}}}\frac{\sigma_{\mathrm{GLS}}}{\hat{\sigma}} \le z\right) \\
&= P\left(\frac{\sqrt{nT}(\hat{\theta}_{1,\mathrm{GLS}}-\theta_{10})}{\sigma_{\mathrm{GLS}}} \le \frac{z\hat{\sigma}}{\sigma_{\mathrm{GLS}}} + \frac{\sqrt{nT}\left(\hat{\theta}_{1,\mathrm{GLS}}-\hat{\theta}_{1,\mathrm{OLS}}\right)}{\sigma_{\mathrm{GLS}}}\right) \\
&= E\Psi\left(\frac{z\hat{\sigma}}{\sigma_{\mathrm{GLS}}} + \frac{\sqrt{nT}\left(\hat{\theta}_{1,\mathrm{GLS}}-\hat{\theta}_{1,\mathrm{OLS}}\right)}{\sigma_{\mathrm{GLS}}}\right) \\
&= E\Psi\left(\frac{z\hat{\sigma}}{\sigma_{\mathrm{GLS}}}\right) + E\left[\psi\left(\frac{z\hat{\sigma}}{\sigma_{\mathrm{GLS}}}\right)\frac{\sqrt{nT}\left(\hat{\theta}_{1,\mathrm{GLS}}-\hat{\theta}_{1,\mathrm{OLS}}\right)}{\sigma_{\mathrm{GLS}}}\right] + O\left(E[\sqrt{nT}(\hat{\theta}_{1,\mathrm{GLS}}-\hat{\theta}_{1,\mathrm{OLS}})]^2\right) \\
&= E\Psi\left(\frac{z\hat{\sigma}}{\sigma_{\mathrm{GLS}}}\right) + O\left(E[\sqrt{nT}(\hat{\theta}_{1,\mathrm{GLS}}-\hat{\theta}_{1,\mathrm{OLS}})]^2\right),
\end{aligned}
$$

where the last equation holds because $\hat{\sigma}$ does not change and $\hat{\theta}_{1,\mathrm{GLS}} - \hat{\theta}_{1,\mathrm{OLS}}$ changes sign when e is replaced by $-e$. Similarly, we have

$$
P\left(\frac{\sqrt{nT}(\hat{\theta}_{1,\mathrm{OLS}}-\theta_{10})}{\hat{\sigma}} \ge z\right) = E\Psi\left(-\frac{z\hat{\sigma}}{\sigma_{\mathrm{GLS}}}\right) + O\left(E[\sqrt{nT}(\hat{\theta}_{1,\mathrm{GLS}}-\hat{\theta}_{1,\mathrm{OLS}})]^2\right).
$$

Let $G(\cdot)$ be the cdf of the χ_1^2 distribution. Then

$$
P\left(\left|\frac{\sqrt{nT}(\hat{\theta}_{1,\mathrm{OLS}}-\theta_{10})}{\hat{\sigma}}\right| \le z\right) = EG\left(\frac{z^2\hat{\sigma}^2}{\sigma_{\mathrm{GLS}}^2}\right) + O\left(E[\sqrt{nT}(\hat{\theta}_{1,\mathrm{GLS}}-\hat{\theta}_{1,\mathrm{OLS}})]^2\right).
$$

Our asymptotic expansion is based on the above approximation. Further expansions require us to approximate the asymptotic bias and variance of $\hat{\sigma}^2$ and establish the rate $O\left(E[\sqrt{nT}(\hat{\theta}_{1,\mathrm{GLS}}-\hat{\theta}_{1,\mathrm{OLS}})]^2\right)$. To this end, we maintain the following assumption.

Assumption 6.1 *(a) $\{e_t\}$ is a stationary Gaussian process with a spectral density that is twice continuously differentiable and bounded above and away from zero uniformly over n in a neighborhood around the origin.*

(b) For $\Phi_F^{\mathcal{H}}(r) = \left[\Phi_1^{\mathcal{H}}(r), ..., \Phi_K^{\mathcal{H}}(r)\right]'$, the smallest eigen value of $\int_0^1 \Phi_F^{\mathcal{H}}(r)\Phi_F^{\mathcal{H}}(r)'dr$ is bounded away from zero uniformly over K.

(c) The basis functions $\{\Phi_k(r)\}$ and $\tau(r)$ are twice continuously differentiable.

(d) For $\Phi_F(r) = [\Phi_1(r), ..., \Phi_K(r)]'$, $\dot{\Phi}_F(i) = \left[\dot{\Phi}_1(r), ..., \dot{\Phi}_K(r)\right]$, and $\dot{\Phi}_k(r) = d\Phi_k(r)/dr$, the following holds:

$$
\begin{aligned}
\int_0^1 \|\Phi_F(r)\|^2 dr &= O(K) \\
\|\Phi_F(i)\|^2 &= O(K),\ i = 0,\ 1 \\
\left\|\dot{\Phi}_F(i)\right\|^2 &= O(K^3),\ i = 0,\ \nu,\ \text{and } 1,
\end{aligned}
$$

where $\|\cdot\|$ is the Euclidean norm.

The conditions on the spectral density in Assumption 6.1 ensure that $E[\sqrt{nT}(\hat{\theta}_{1,\text{GLS}} - \hat{\theta}_{1,\text{OLS}})]^2 = O(1/T)$. They are also needed for evaluating the asymptotic bias and variance of $\hat{\sigma}^2$. The other conditions in Assumption 6.1 are further restrictions on the basis functions and trend functions. It is not hard to show that they are satisfied for Fourier basis functions and polynomial trend functions.

Let $t_K^{\alpha/2}$ be the $1-\alpha/2$ quantile of Student's t-distribution with K degrees of freedom, and let χ_1^{α} be the $1-\alpha$ quantile of the χ_1^2 distribution. Let $G_{\delta^2}(\cdot)$ and $G_{3,\delta^2}(\cdot)$ be the cdf's of the noncentral χ_1^2 and χ_3^2 distributions with noncentrality parameter δ^2. The following theorem establishes high-order approximations to the type I and type II errors of the asymptotic t test based on $\tilde{\mathbb{T}}$.

Theorem 6.1 *Let Assumptions 3.2 and 6.1 hold. Consider the asymptotics under which $K \to \infty$ such that $K/T + T/K^2 \to 0$.*

(*a*) *The type I error of the t test based on $\tilde{\mathbb{T}}$ satisfies*

$$P(|\tilde{\mathbb{T}}| > t_K^{\alpha/2}|H_0) = \alpha - \frac{K^2\bar{B}}{T^2}G'(\chi_1^{\alpha})\chi_1^{\alpha} + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right). \tag{20}$$

(*b*) *Under the local alternative $H_1(\delta^2): \theta_1 - \theta_{10} = (nT)^{-1/2}\sigma\varrho$, where $\varrho = \pm\delta$ with equal probability 1/2, the type II error of the t test based on $\tilde{\mathbb{T}}$ satisfies*

$$\begin{aligned} P(|\tilde{\mathbb{T}}| < t_K^{\alpha/2}|H_1(\delta^2)) &= G_{\delta^2}(\chi_1^{\alpha}) + \frac{K^2\bar{B}}{T^2}G'_{\delta^2}(\chi_1^{\alpha})\chi_1^{\alpha} \\ &+ \frac{\delta^2}{2K}G'_{3,\delta^2}(\chi_1^{\alpha})\chi_1^{\alpha} + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right), \end{aligned} \tag{21}$$

where $\bar{B} = B/\Lambda^2$,

$$B = -\omega^{(2)}(0)\sum_{p=-\infty}^{\infty} p^2\sigma_{e,p}^2,\ \Lambda^2 = \sum_{p=-\infty}^{\infty}\sigma_{e,p}^2, \sigma_{e,p}^2 = E(e_t e_{t-p}),$$

$$\begin{aligned} \omega^{(2)}(0) &= \frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}\int_0^1 \dot{\Phi}_F(s)'\left[\int_0^1 \Phi_F^{\mathcal{H}}(s)\left[\Phi_F^{\mathcal{H}}(s)\right]'ds\right]^{-1}\dot{\Phi}_F(s)\,ds \\ &= \frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}tr\left(\left[\int_0^1 \Phi_F^{\mathcal{H}}(s)\left[\Phi_F^{\mathcal{H}}(s)\right]'ds\right]^{-1}\int_0^1 \dot{\Phi}_F(s)\dot{\Phi}_F(s)'\,ds\right). \end{aligned}$$

The above results are similar to Theorem 5 in Sun (2011) with $p = 1$ but with a different $\bar{B}$. Suppose we use the Fourier basis functions $\Phi_{2j-1} = \sqrt{2}\cos(2\pi j)$ and $\Phi_{2j} = \sqrt{2}\sin(2\pi j)$ for $j = 1, ..., K/2$. If $\tau(t)$ is a vector of polynomial trend functions, then Proposition 9.1 in the appendix shows that $\omega^{(2)}(0) = \pi^2/6$. This gives rise to a $\hat{B}$ that is different from what is obtained in Sun (2011). The difference is due to the use of cosine basis functions in Sun (2011), while we use both cosine and sine basis functions.

Following Sun (2011), we can ignore the high-order terms and approximate the type I and type II errors by

$$\begin{aligned} e_I &= \alpha - \frac{K^2\bar{B}}{T^2}G'(\chi_1^{\alpha})\chi_1^{\alpha}, \\ e_{II} &= G_{\delta^2}(\chi_1^{\alpha}) + \frac{K^2\bar{B}}{T^2}G'_{\delta^2}(\chi_1^{\alpha})\chi_1^{\alpha} + \frac{\delta^2}{2K}G'_{3,\delta^2}(\chi_1^{\alpha})\chi_1^{\alpha}. \end{aligned}$$

To obtain an optimal smoothing parameter K for testing, we propose to choose K by minimizing the type II error while controlling the type I error. More specifically, we solve the following problem:

$$\min \; e_{II} \;\; s.t. \;\; e_I \leq \kappa\alpha,$$

where $\kappa > 1$ is a tolerance parameter. We allow the type I error to be different from the nominal type I error α, but it cannot be larger than $\kappa\alpha$. For example, when $\kappa = 1.2$ and $\alpha = 5\%$, the upper bound is 6% rather than 5%. Our approach to selecting K has a decision-theoretic basis, as it amounts to selecting K to minimize a loss function that is a weighted average of type I and type II errors with the weight given by the implied Lagrangian multiplier for the constraint $e_I \leq \kappa\alpha$. See Sun, Phillips, and Jin (2011) for related ideas.

Following an argument similar to that in Sun (2011), we find that the optimal K for the above problem is

$$K_{\text{opt}} = \left\{ \frac{\delta^2 G'_{3,\delta^2}(\chi_1^\alpha)}{4\bar{B}\left[G'_{\delta^2}(\chi_1^\alpha) - \lambda_{\text{opt}} G'(\chi_1^\alpha)\right]} \right\}^{1/3} T^{2/3}, \tag{22}$$

where

$$\lambda_{\text{opt}} = \begin{cases} 0, & \text{if } \bar{B} > 0 \\ \frac{G'_{\delta^2}(\chi_1^\alpha)}{G'(\chi_1^\alpha)} + \delta^2 \frac{|\bar{B}|^{1/2} G'_{3,\delta^2}(\chi_1^\alpha)[\chi_1^\alpha]^{3/2}[G'(\chi_1^\alpha)]^{1/2}}{4[(\kappa-1)\alpha]^{3/2} T}, & \text{if } \bar{B} \leq 0. \end{cases} \tag{23}$$

The optimal K_{opt} in (22) depends on the noncentrality parameters κ and δ. As in Sun (2011), we allow κ to depend on the sample size T. For a larger T, we may require κ to be closer to 1. We suggest choosing δ^2 so that the first-order power of the asymptotic two-sided t test is 75%, that is, choosing δ^2 so that $1 - G_{\delta^2}(\chi_1^\alpha) = 75\%$ for a given significance level α. We refer to Sun (2011) for more detailed discussions on how to choose κ and δ^2.

For practical implementation, we use the parametric plug-in approach to estimate the unknown B and Λ^2. Suppose we use the simple AR(1) plug-in by fitting an AR(1) model to $\hat{e}_t = \sum_{i=1}^n \widetilde{Treat}_i \cdot \hat{\epsilon}_{it}/\sqrt{n}$. Let $\hat{\rho}_e$ be the estimated AR coefficient and $\hat{\sigma}_e^2$ be the estimated error variance. Then the plug-in estimators of Λ^2 and $\bar{B}$ are

$$\hat{\Lambda}^2 = \frac{\hat{\sigma}_e^2}{(1-\hat{\rho}_e)^2}, \quad \text{and} \quad \bar{B}^{est} = -\frac{2\omega^{(2)}(0)\hat{\rho}_e}{(1-\hat{\rho}_e)^2},$$

and the plug-in estimator of K is

$$\hat{K}_{\text{opt}} = \begin{cases} \left(\frac{(1-\hat{\rho}_e)^2}{8\omega^{(2)}(0)|\hat{\rho}_e|}\right)^{1/3} \left(\frac{G'_{3,\delta^2}(\chi_1^\alpha)\delta^2}{G'_{\delta^2}(\chi_1^\alpha)}\right)^{1/3} T^{2/3}, & \text{if } \bar{B}^{est} > 0 \\ \left(\frac{(1-\hat{\rho}_e)^2}{2\omega^{(2)}(0)|\hat{\rho}_e|}\right)^{1/2} \left(\frac{(\kappa-1)\alpha}{G'(\chi_1^\alpha)\chi_1^\alpha}\right)^{1/2} T, & \text{if } \bar{B}^{est} < 0. \end{cases} \tag{24}$$

It is clear that for $|\hat{\rho}_e| \in (0,1)$, $\hat{K}$ decreases as $|\hat{\rho}_e|$ increases. A smaller K is desired in the presence of stronger autocorrelation. Intuitively, when the autocorrelation is high, we should use only very few periodogram coordinates that are close to the origin. We do so in order to avoid smoothing bias, which can be large if smoothing is taken over a wide window in the frequency domain. For a given window size K, the larger the value of $|\rho_e|$, the larger the absolute smoothing bias.

7 Simulation Evidence

In our simulations, we consider the following data generating process

$$Y_{it} = \lambda_t + \tau(t)' a_i + Treat_i \cdot \beta_{10} + Post_t \cdot \beta_{20} + Treat_i \cdot Post_t \cdot \theta_{10} + \epsilon_{it},$$

for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$ where $Treat_i = 1\{i \leq 0.5n\}$ and $Post_t = 1\{t \geq 0.5T\}$. The error term follows independent $AR(1)$ processes with AR parameter ρ:

$$\epsilon_{it} = \rho\epsilon_{it-1} + e^{\epsilon}_{it} \text{ with } \epsilon_{i0} = 0.$$

While $\{e^{\epsilon}_{it}\}$ is *iid* over time, there is cross-sectional dependence. We consider the case with $n = m^2$ for some positive integer m. Individuals are assumed to be located on a regular $m \times m$ integer lattice so that we can write

$$e^{\epsilon}_{it} = e^{\epsilon}_{i_1,i_2,t} \text{ for } 1 \leq i_1, i_2 \leq m,$$

where (i_1, i_2) is the location of the i-th individual. For each time period t, e^{ϵ}_{it} is a spatial average of *iid* innovations:

$$\begin{aligned} e^{\epsilon}_{(i_1,i_2),t} = {} & \phi\left(v_{i_1-1,i_2,t} + v_{i_1,i_2-1,t} + v_{i_1+1,i_2,t} + v_{i_1,i_2+1,t}\right) \\ & + \phi^2\left(v_{i_1-2,i_2,t} + v_{i_1,i_2-2,t} + v_{i_1+2,i_2,t} + v_{i_1,i_2+2,t}\right) \\ & + \phi^2\left(v_{i_1+1,i_2+1,t} + v_{i_1-1,i_2-1,t} + v_{i_1+1,i_2-1,t} + v_{i_1-1,i_2+1,t}\right) + v_{i_1,i_2,t}, \end{aligned}$$

where $v_{i_1,i_2,t}$ is *iid* $N(0,1)$ across i_1, i_2, and t. That is, $e^{\epsilon}_{it} \sim SMA(2)$, a spatial moving average of order 2 according to the taxicab distance.

For the trend component, we consider two common cases. In the first case, $\tau(t) = 1$, i.e., there is no trending function, and only individual fixed effects are included. In this case, time series detrending reduces to demeaning. In the second case, $\tau(t) = (1, t)'$, i.e., there are both individual fixed effects and linear time trends. For other model parameters, we take $\rho = -0.6, -0.3, 0, 0.3, 0.6,$ and 0.9 and set ϕ to be $\phi = 0$ and 0.5. We set all other parameters to zero, as all the tests we consider are invariant to them. The (n, T) combinations under consideration are $(4^2, 50)$, $(4^2, 100)$, $(4^2, 200)$, $(8^2, 50)$, $(8^2, 100)$, $(8^2, 200)$, $(16^2, 50)$, $(16^2, 100)$, and $(16^2, 200)$.

We are interested in testing $H_0 : \theta_{10} = 0$ with two-sided alternatives so that each test rejects the null when the absolute value of the t statistic is large enough. We consider two significance levels: $\alpha = 5\%$ and $\alpha = 10\%$. We first examine the finite sample performances of the tests with the general sandwich variance estimator and then examine those of the tests with the alternative variance estimator presented in Section 5.

7.1 Size accuracy with the general sandwich variance estimator

There are two groups of tests with the general sandwich variance estimator. The first group of tests uses the sine and cosine basis functions: $\Phi_{2j-1}(x) = \sqrt{2}\cos(2j\pi x)$, $\Phi_{2j}(x) = \sqrt{2}\sin(2j\pi x)$, $j = 1, \ldots, K/2$. There are two tests in this group: the nonstandard fixed-smoothing test and the standard normal test. The former uses nonstandard critical values which are obtained by simulations, as discussed at the end of Section 3. The latter uses standard normal critical values. The second group of tests is based on the sine and cosine basis functions, but we do not use them directly.

Instead, we employ the transformation given in Section 4 to obtain a new set of basis vectors, which is then used in the asymptotic variance estimation. There are also two tests in this group: the standard fixed-smoothing t test, which uses critical values from the t distribution with K degrees of freedom and the standard normal test, which uses standard normal critical values.

Figures 2 and 3 plot the empirical null rejection probabilities of 5% tests against the values of K when $n = 64$, $T = 100$, and when the model contains no linear trend, i.e., $\tau(t) = 1$. The difference between these two figures is that Figure 2 reports the case with no cross-sectional dependence (i.e., $\phi = 0$) while Figure 3 reports the case with cross-sectional dependence (i.e., $\phi = 0.5$). Several patterns emerge from both figures. First, the nonstandard test and the standard t test have more accurate size than the standard normal tests, especially when K is small. Second, as K increases, the difference between a fixed-smoothing test and the corresponding standard normal test diminishes. This is well expected: when K is large, the t critical value becomes close to the standard normal critical value. When the plain vanilla sine and cosine bases are used, the fixed-smoothing critical value also approaches the standard normal critical value as K increases. See Figure 1. Intuitively, when K is large, the estimation uncertainty in the asymptotic variance estimator becomes small, and there is not much room for the fixed-smoothing asymptotics — which is designed to capture this estimation uncertainty — to achieve a better result. Third, for the two fixed-smoothing tests, the size accuracy of the standard t test is very close to that of the nonstandard test. When the time series dependence is not strong, e.g., $\rho \leq 0.6$, the empirical null rejection probabilities are virtually the same across these two fixed-smoothing tests. When the time series dependence is strong, e.g., $\rho = 0.9$, the nonstandard test is slightly more accurate than the standard t test. Fourth, comparing these two figures, we can see that the cross-sectional dependence does not affect the size properties of any of the four tests: the empirical rejection probabilities are almost the same no matter whether cross-sectional dependence is present or not. Finally, for the asymptotic normal tests, it pays to employ the transformed basis vectors rather than the original basis vectors. Transformation is desirable even if one does not want to use the fixed-smoothing asymptotic approximations. However, in terms of size accuracy, the asymptotic normal tests are dominated by the nonstandard test and the standard t test, even if transformed basis vectors are used.

Figures 4 and 5 are similar to Figures 2 and 3, respectively, but the data generating process for each individual contains a linear trend. It is clear that all the patterns identified from Figures 2 and 3 are still applicable. Comparing Figures 4 and 5 with Figures 2 and 3, we see that the size properties depend on whether a linear trend is included in the data generating process or not. It appears that the effect of the linear trend interacts with the strength of the temporal dependence. When the AR parameter is large, e.g., $\rho = 0.9$, it is beneficial to have a linear trend. A possible explanation is that detrending can help reduce strong temporal dependence without introducing too much extra variation from the trend estimation.

We also investigate the effect of the sample size on test performance. Simulation results not reported here show that the null rejection probabilities remain more or less the same for different values of n when the time series sample size stays the same. This is compatible with the asymptotic results that the cross-sectional dependence does not affect the size properties of any of the tests. To estimate the asymptotic variance, we essentially collapse the panel data into time series data. The cross-sectional dependence and cross-sectional sample size do not affect the persistence of the collapsed time series. As a result, they do not affect the size properties of any of the tests. On the other hand, when there is substantial temporal dependence, all tests become more accurate as T increases. In addition, as T increases, the difference between the standard t

test and the nonstandard test diminishes. When $T = 200$ and for all the cross-sectional sample sizes considered, the two tests have almost identical null rejection probabilities.

7.2 Size accuracy with the alternative variance estimator

We now turn to the tests based on the alternative variance estimator. We consider the same set of four tests as before, but each test statistic is based on the "collapsed" sandwich variance estimator constructed in Section 5. To save space, we report only the case with a linear trend but no cross-sectional dependence in Figure 6. This figure is representative of all other cases. The qualitative observations for the cases with a general sandwich variance estimator remain valid. In particular, the fixed-smoothing nonstandard test and the t test are more accurate than the asymptotic normal tests, especially when K is small and the AR parameter is large.

Comparing Figure 6 with Figure 4, we see that it does not matter much whether we use the general sandwich variance estimator or the collapsed sandwich variance estimator: for each test, the difference in the null rejection probability is small. Nevertheless, for the nonstandard test based on the original sine and cosine bases, the general sandwich variance estimator leads to somewhat improved size performance. For the other three tests, the collapsed sandwich variance estimator performs slightly better. When the collapsed sandwich variance estimator is used in place of the general sandwich variance estimator, the size difference between the standard t test and the nonstandard test becomes smaller.

To sum up the simulation results discussed thus far, we see that the fixed-smoothing tests (both the t test and the nonstandard test) are more accurate than the asymptotic normal tests. The difference between the t test and the nonstandard test is small, especially when the temporal dependence is weak or the time series is long. The cross-sectional dependence and the sample size n do not have much effect on the size properties of any of the tests. For each type of test, the form of the sandwich variance estimator has a small effect on the size accuracy.

7.3 Size and power under data-driven choice of K

Table 1 reports the size of each of the eight tests we considered in Sections 7.1 and 7.2. We use the data-driven $\hat{K}_{\text{opt}}$ given in (24), but we make two adjustments. First, we use the truncated LS estimator

$$\tilde{\rho}_e = \frac{\hat{\rho}_e}{|\hat{\rho}_e|}0.97 + \left(\hat{\rho}_e - \frac{\hat{\rho}_e}{|\hat{\rho}_e|}0.97\right) 1\left\{|\hat{\rho}_e| \leq 0.97\right\}$$

instead of the original estimator $\hat{\rho}_e$ in computing $\hat{K}_{\text{opt}}$. Second, we truncate $\hat{K}_{\text{opt}}$ to be between 2 and $T/2$, and we round it to the greatest even number less than $\hat{K}_{\text{opt}}$. Rounding is used to speed up the computation. It has a minimal effect on test performances and is not necessary in practical implementation. κ is chosen to be 1.2 when $T = 100$. It is clear from Table 1 that the fixed-smoothing tests are more accurate than the standard normal tests in almost all cases, especially when the AR parameter is positive. The only exceptions are the cases with a negative AR parameter, in which case the fixed-smoothing tests tend to under-reject. Overall, the fixed-smoothing tests have quite accurate size. Among the fixed-smoothing tests, the standard t test appears to be more accurate than the corresponding nonstandard test.

Figure 7 presents the size-adjusted power of the eight tests considered. Note that the tests with the same test statistic have identical size-adjusted power. In our setting, given the same basis vectors and the form of the asymptotic variance estimator, the fixed-smoothing test and the normal test have the same size-adjusted power. Thus it suffices to consider four tests that reflect the two type of basis vectors and the two form of the variance estimator. Figure 7 reports the case with a linear trend but no cross-sectional dependence (i.e., $\tau(t) = (1,t)'$ and $\phi = 0$). The figures for the other cases are similar. The basic observation is that all four tests have more or less the same size-adjusted power function. In other words, the basis transformation and the form of the asymptotic variance estimator have almost no effect on the size-adjusted power. This, coupled with its size accuracy and convenience to use, suggests that we use the t tests in empirical applications.

8 Conclusion

This paper develops two asymptotically valid t tests in the DD regressions when T is relatively large. These t tests employ standard t critical values and are thus easy to use. They are more accurate than the normal tests but have the same power properties. The cross-sectional sample size n can be fixed or grow with T. Simulations show that the proposed t tests work well even when n is comparable to T. Given these attractive properties, we recommend using the t tests in place of the normal tests in empirical applications.

There are a few possible extensions. First, when the underlying process is persistent, we can use prewhitening to reduce the size distortion of the proposed t tests. This extension is straightforward. Second, while the paper considers only panel data, it is easy to see that the proposed procedures would work for repeated cross-section data as well. In that case, the only change needed would be to switch the order of detrending and averaging. Instead of first detrending each time series and then taking an average within each group, as we do in this paper, for repeated cross-section data we would first take an average within each group and then detrend the averaged data for each group. Finally, we consider the case where there is only one policy change. We do not imagine that there would be much difficulty in allowing for multiple policy changes, with possibly heterogeneous effects, but we leave the details to future research.

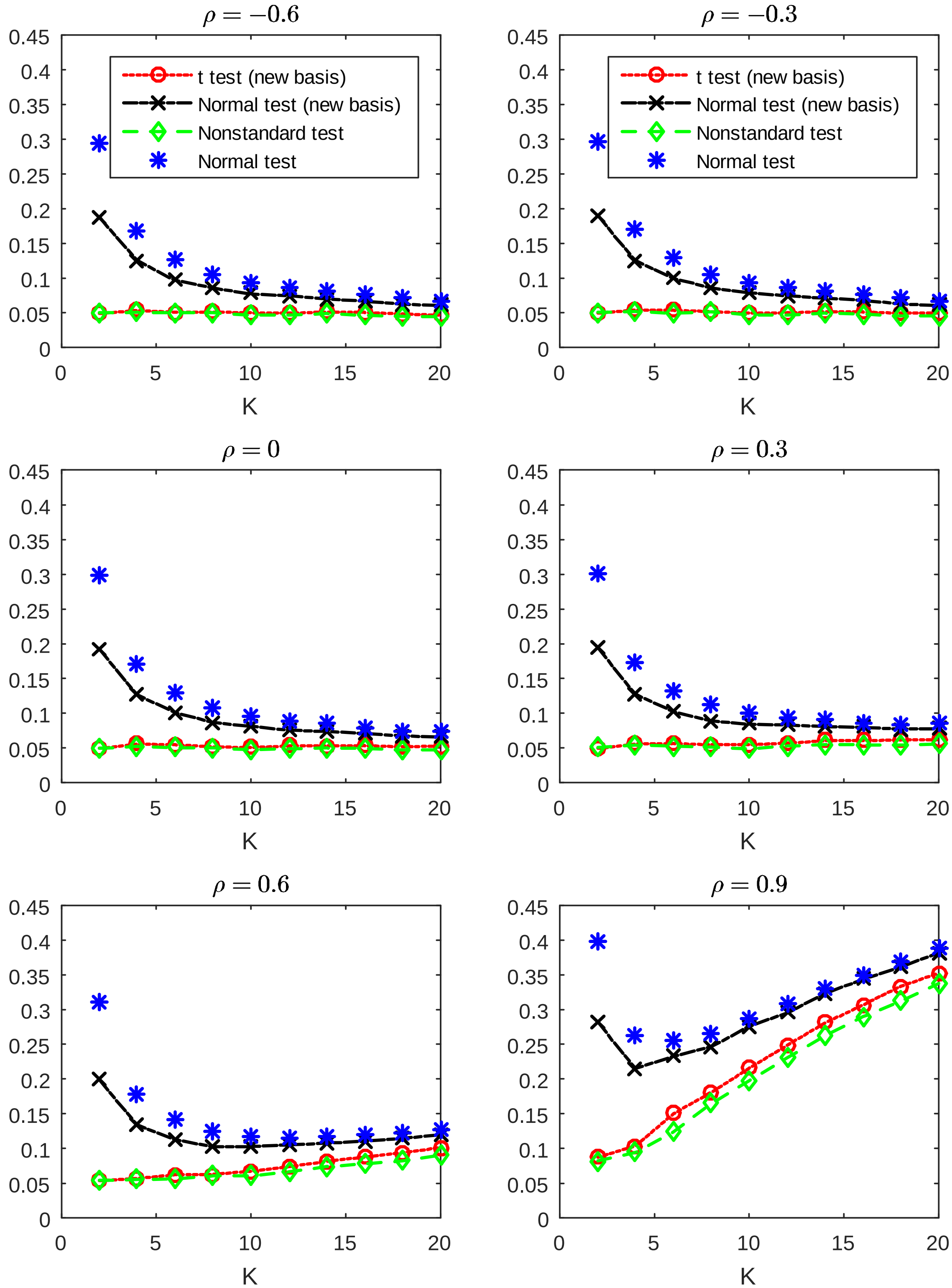

Figure 2: Empirical null rejection probabilities of 5% two-sided tests with the general sandwich variance estimator when $n = 64, T = 100, \phi = 0$, and there is no time trend

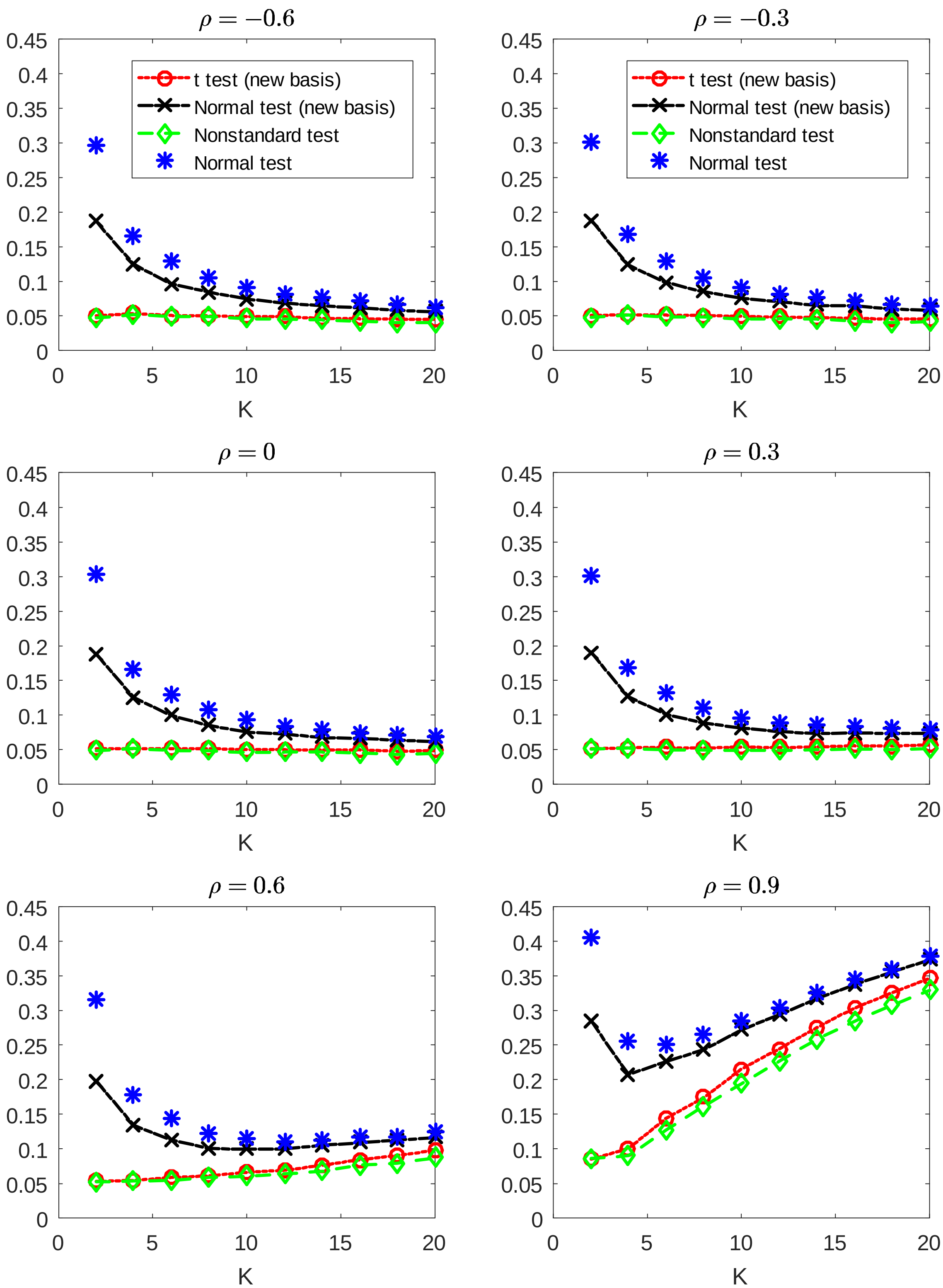

Figure 3: Empirical null rejection probabilities of 5% two-sided tests with the general sandwich variance estimator when $n = 64, T = 100, \phi = 0.5$, and there is no time trend

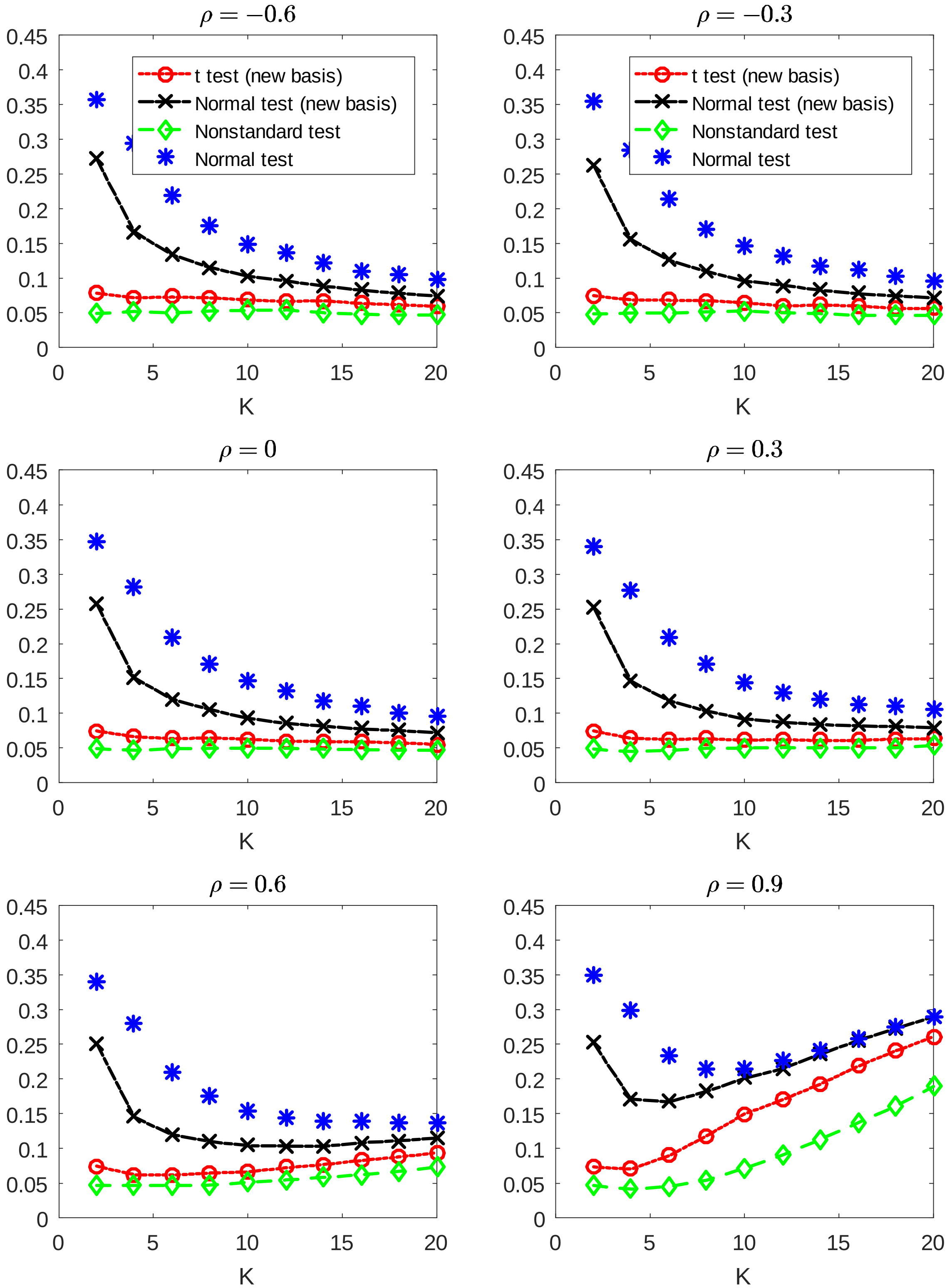

Figure 4: Empirical null rejection probabilities of 5% two-sided tests with the general sandwich variance estimator when $n = 64, T = 100, \phi = 0$, and there are linear time trends

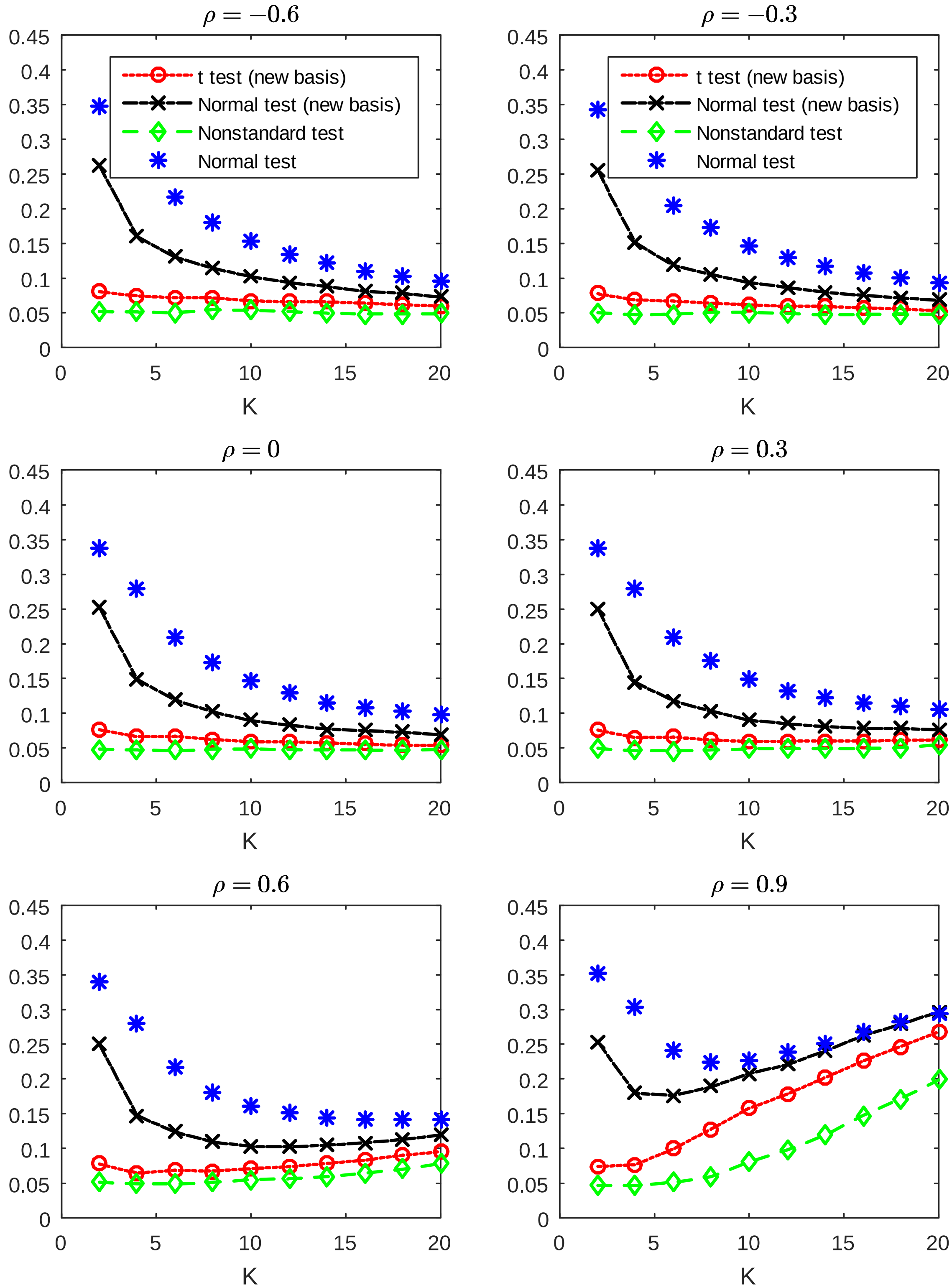

Figure 5: Empirical null rejection probabilities of 5% two-sided tests with the general sandwich variance estimator when $n = 64, T = 100, \phi = 0.5$, and there are linear time trends

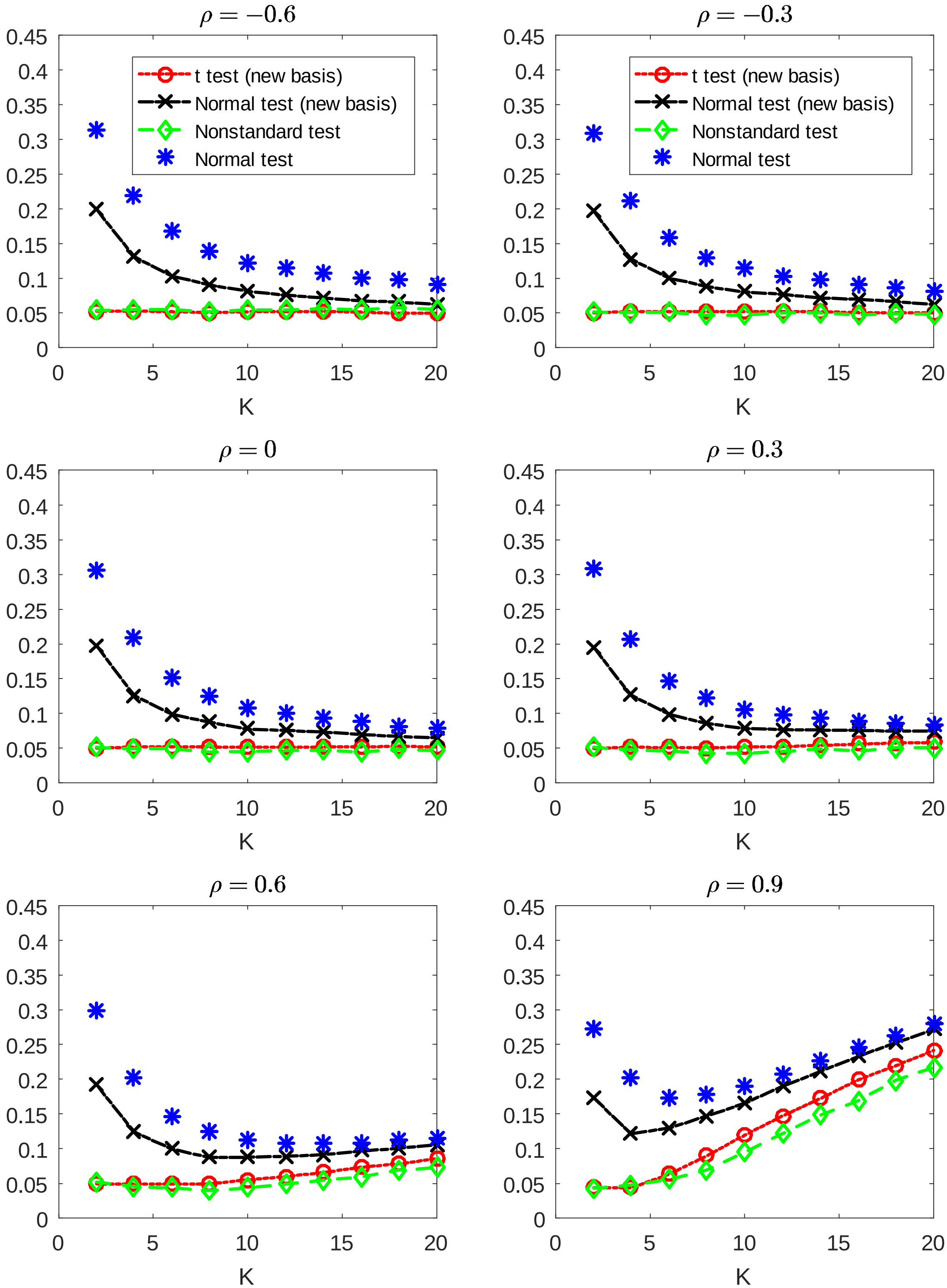

Figure 6: Empirical null rejection probabilities of 5% two-sided tests with the collapsed sandwich variance estimator when $n = 64, T = 100, \phi = 0$, and there are linear time trends

Table 1: Empirical size of different 5% tests in DD regression with sample size $n = 64, T = 100$, and data-driven choice of K

	General Sandwich Variance				Collapsed Sandwich Variance			
	Transformed Bases		Fourier Bases		Transformed Bases		Fourier Bases	
	t_1	$\mathcal{N}_1$	$\mathcal{T}_{1\infty}$	$\mathcal{N}_1$	t_2	$\mathcal{N}_2$	$\mathcal{T}_{2,\infty}$	$\mathcal{N}_2$
				$\tau = 1, \phi = 0$				
$\rho = -0.6$	0.034	0.053	0.044	0.059	0.034	0.051	0.047	0.062
$\rho = -0.3$	0.036	0.056	0.046	0.060	0.037	0.056	0.046	0.062
$\rho = 0$	0.047	0.061	0.053	0.064	0.047	0.060	0.054	0.064
$\rho = 0.3$	0.054	0.081	0.066	0.085	0.055	0.080	0.065	0.083
$\rho = 0.6$	0.045	0.102	0.067	0.115	0.046	0.102	0.064	0.113
$\rho = 0.9$	0.042	0.255	0.107	0.344	0.036	0.226	0.092	0.302
				$\tau(t) = 1, \phi = 0.5$				
$\rho = -0.6$	0.034	0.052	0.044	0.058	0.033	0.054	0.046	0.064
$\rho = -0.3$	0.037	0.054	0.045	0.058	0.036	0.056	0.045	0.061
$\rho = 0$	0.045	0.058	0.051	0.061	0.045	0.058	0.051	0.061
$\rho = 0.3$	0.053	0.079	0.064	0.084	0.053	0.079	0.062	0.083
$\rho = 0.6$	0.045	0.105	0.067	0.119	0.042	0.103	0.063	0.114
$\rho = 0.9$	0.040	0.262	0.110	0.347	0.035	0.229	0.091	0.305
				$\tau(t) = (1, t)', \phi = 0$				
$\rho = -0.6$	0.042	0.063	0.045	0.076	0.034	0.054	0.057	0.078
$\rho = -0.3$	0.041	0.058	0.047	0.073	0.036	0.054	0.048	0.067
$\rho = 0$	0.047	0.059	0.051	0.072	0.044	0.056	0.050	0.062
$\rho = 0.3$	0.053	0.076	0.060	0.098	0.049	0.072	0.057	0.079
$\rho = 0.6$	0.044	0.099	0.053	0.155	0.035	0.087	0.052	0.111
$\rho = 0.9$	0.029	0.183	0.047	0.297	0.020	0.131	0.056	0.215
				$\tau = (1, t)', \phi = 0.5$				
$\rho = -0.6$	0.049	0.072	0.051	0.083	0.040	0.063	0.064	0.083
$\rho = -0.3$	0.044	0.066	0.051	0.082	0.040	0.060	0.052	0.072
$\rho = 0$	0.049	0.065	0.056	0.080	0.046	0.061	0.052	0.068
$\rho = 0.3$	0.057	0.081	0.069	0.109	0.054	0.077	0.062	0.085
$\rho = 0.6$	0.048	0.106	0.059	0.162	0.041	0.092	0.059	0.115
$\rho = 0.9$	0.034	0.195	0.055	0.310	0.023	0.138	0.058	0.223

Note: The t tests, denoted by "t", are based on t critical values. The normal test, denoted by "$\mathcal{N}$", are based on standard normal critical values. The nonstandard tests, denoted by "$\mathcal{T}$ ", are based on simulated nonstandard critical values.

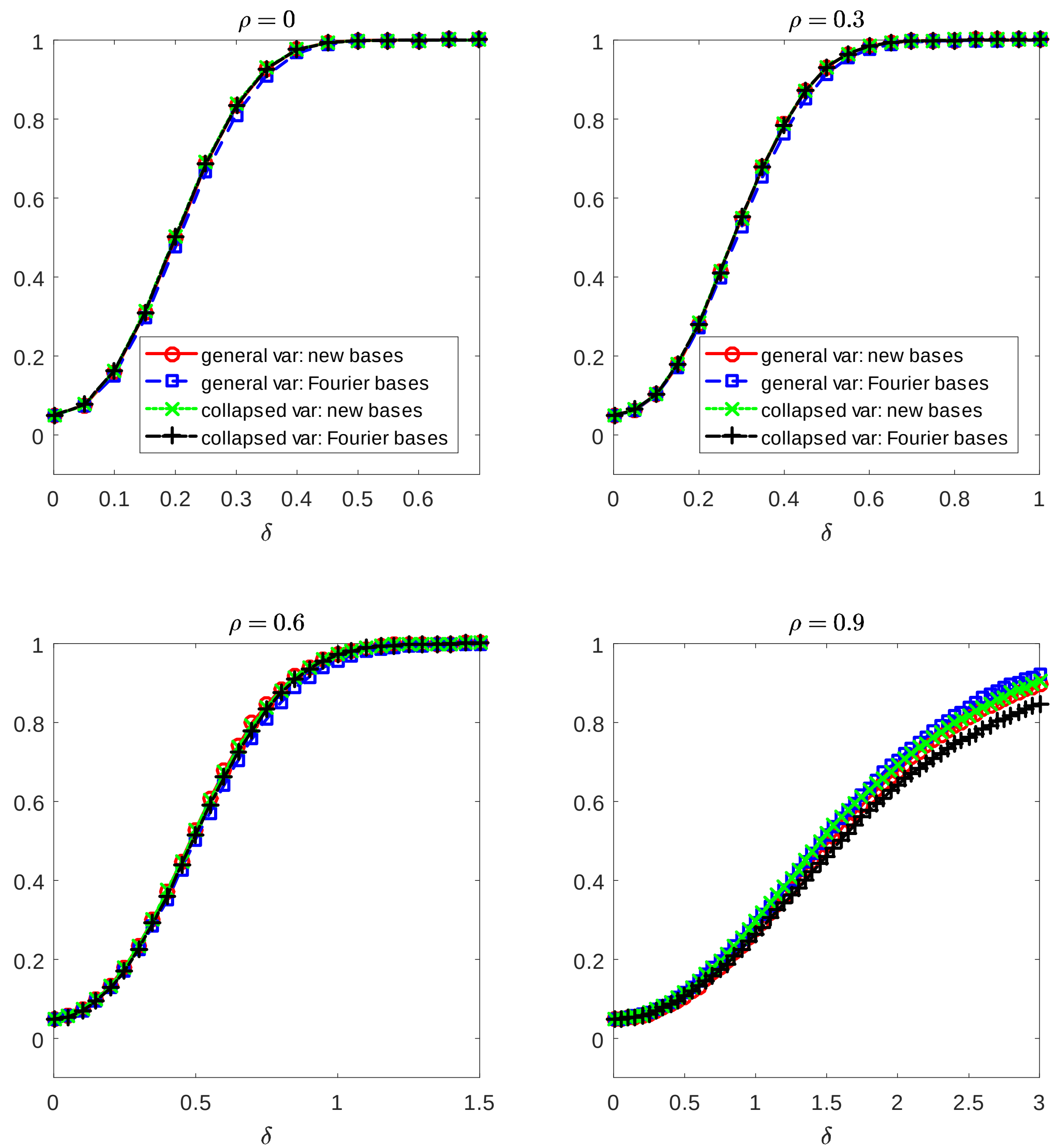

Figure 7: Size-adjusted power of the tests with different variance estimators and basis functions for $n = 64$, $T = 100$ in the presence of linear trends but no cross sectional dependence.

9 Appendix of Proofs

Proof of Lemma 3.1. Part (a). We have

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{[Tr]}[Post_t^{\tau}]^2 \\
&= \frac{1}{T}\sum_{t=1}^{[Tr]}\left[1\left\{\frac{t}{T}\geq\nu\right\}-\left(\sum_{s=1}^{T}Post_s\cdot\tau(s)'\right)\left(\sum_{s=1}^{T}\tau(s)\tau(s)'\right)^{-1}\tau(t)\right]^2 \\
&= \frac{1}{T}\sum_{t=1}^{[Tr]}\left[1\left\{\frac{t}{T}\geq\nu\right\}-\left(\frac{1}{T}\sum_{s=1}^{T}Post_s\cdot\tau_D(s)'\right)\left(\frac{1}{T}\sum_{s=1}^{T}\tau_D(s)\tau_D(s)'\right)^{-1}\tau_D(t)\right]^2 \\
&\to \int_0^r\left\{1\{t\geq\nu\}-\left[\int_0^1 1\{s\geq\nu\}\tau(s)'\,ds\right]\left[\int_0^1\tau(s)\tau(s)'\,ds\right]^{-1}\tau(t)\right\}^2 dt+O\left(T^{-1}\right)
\end{aligned}
$$

uniformly over $r\in[0,1]$. Some elementary calculation shows that

$$
\frac{1}{n}\sum_{i=1}^{n}[\widetilde{Treat_i}]^2=\frac{1}{n}\sum_{i=1}^{n}\left(1\{i\leq n\mu\}-\mu\right)^2=\mu(1-\mu).
$$

Combining the above results yields Lemma 3.1(a).

Part (b). We have

$$
\begin{aligned}
S_{21}(r) &= \frac{1}{T}\sum_{t=1}^{[Tr]}Post_t^{\tau}\cdot\frac{1}{n}\left(\sum_{i=1}^{n}(Treat_i-\mu)\cdot\tilde{Z}_{it}^{\tau}\right) \\
&= \frac{1}{T}\sum_{t=1}^{[Tr]}Post_t^{\tau}\cdot\frac{1}{n}\sum_{i=1}^{n}Treat_i\cdot\tilde{Z}_{it}^{\tau}=\frac{1}{T}\sum_{t=1}^{Tr}Post_t^{\tau}\cdot\frac{1}{n}\sum_{i=1}^{\mu n}\tilde{Z}_{it}^{\tau} \\
&= \frac{1}{T}\sum_{t=1}^{[Tr]}Post_t^{\tau}\cdot\frac{1}{n}\sum_{i=1}^{\mu n}\left(Z_{it}^{\tau}-\frac{1}{n}\sum_{j=1}^{n}Z_{jt}^{\tau}\right) \\
&= \frac{1}{T}\sum_{t=1}^{[Tr]}Post_t^{\tau}\cdot\frac{1}{n}\sum_{i=1}^{\mu n}Z_{it}^{\tau}-\frac{1}{T}\sum_{t=1}^{[Tr]}Post_t^{\tau}\cdot\frac{\mu}{n}\sum_{j=1}^{n}Z_{jt}^{\tau} \\
&= \mu(1-\mu)\left[\frac{1}{T}\sum_{t=1}^{[Tr]}Post_t^{\tau}\cdot\frac{1}{n\mu}\sum_{i=1}^{\mu n}Z_{it}^{\tau}-\frac{1}{T}\sum_{t=1}^{[Tr]}Post_t^{\tau}\cdot\frac{1}{n(1-\mu)}\sum_{j=\mu n+1}^{n}Z_{jt}^{\tau}\right],
\end{aligned}
$$

where the last line follows from some simple calculations.

Note that

$$
\begin{aligned}
Z_{it}^{\tau} &= Z_{it} - \left(\sum_{s=1}^{T} Z_{is}\tau(s)'\right)\left(\sum_{s=1}^{T}\tau(s)\tau(s)'\right)^{-1}\tau(t) \\
&= \lambda_{zt} + \alpha_{zi}\tau(t) + \mathcal{Z}_{it} \\
&- \left(\sum_{s=1}^{T} \lambda_{zs}\tau(s)'\right)\left(\sum_{s=1}^{T}\tau(s)\tau(s)'\right)^{-1}\tau(t) \\
&- \left(\sum_{s=1}^{T} \alpha_{zi}\tau(s)\tau(s)'\right)\left(\sum_{s=1}^{T}\tau(s)\tau(s)'\right)^{-1}\tau(t) \\
&- \left(\sum_{s=1}^{T} \mathcal{Z}_{is}\tau(s)'\right)\left(\sum_{s=1}^{T}\tau(s)\tau(s)'\right)^{-1}\tau(t) = \lambda_{zt}^{\tau} + \mathcal{Z}_{it}^{\tau}
\end{aligned}
\tag{25}
$$

for

$$
\begin{aligned}
\lambda_{zt}^{\tau} &= \lambda_{zt} - \left(\sum_{s=1}^{T} \lambda_{zs}\tau(s)'\right)\left(\sum_{s=1}^{T}\tau(s)\tau(s)'\right)^{-1}\tau(t), \\
\mathcal{Z}_{it}^{\tau} &= \mathcal{Z}_{it} - \left(\sum_{s=1}^{T} \mathcal{Z}_{is}\tau(s)'\right)\left(\sum_{s=1}^{T}\tau(s)\tau(s)'\right)^{-1}\tau(t).
\end{aligned}
$$

We have

$$
\begin{aligned}
S_{21}(r) &= \mu(1-\mu)\left[\frac{1}{T}\sum_{t=1}^{[Tr]} Post_t^{\tau}\cdot\frac{1}{n\mu}\sum_{i=1}^{\mu n}\mathcal{Z}_{jt}^{\tau} - \frac{1}{T}\sum_{t=1}^{[Tr]} Post_t^{\tau}\cdot\frac{1}{n(1-\mu)}\sum_{i=\mu n+1}^{n}\mathcal{Z}_{it}^{\tau}\right] \\
&= \mu(1-\mu)\left[\frac{1}{T}\sum_{t=1}^{[Tr]} Post_t^{\tau}\cdot\left(\bar{\mathcal{Z}}_{\cdot,t}^{treat}\right)^{\tau} - \frac{1}{T}\sum_{t=1}^{[Tr]} Post_t^{\tau}\cdot\left(\bar{\mathcal{Z}}_{\cdot,t}^{control}\right)^{\tau}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
\bar{\mathcal{Z}}_{\cdot,t}^{treat} &= \frac{1}{n\mu}\sum_{i=1}^{\mu n}\mathcal{Z}_{jt}, \quad \left(\bar{\mathcal{Z}}_{\cdot,t}^{treat}\right)^{\tau} = \frac{1}{n\mu}\sum_{i=1}^{\mu n}\mathcal{Z}_{jt}^{\tau}, \\
\bar{\mathcal{Z}}_{\cdot,t}^{control} &= \frac{1}{n(1-\mu)}\sum_{i=\mu n+1}^{n}\mathcal{Z}_{it}, \text{ and } \left(\bar{\mathcal{Z}}_{\cdot,t}^{control}\right)^{\tau} = \frac{1}{n(1-\mu)}\sum_{i=\mu n+1}^{n}\mathcal{Z}_{it}^{\tau}.
\end{aligned}
$$

In the above expression for $S_{21}(r)$, the time effect λ_{zt}^{τ} has been cancelled out.

Denoting

$$
A_{\tau\tau} = \frac{1}{T}\sum_{s=1}^{T}\tau_D(s)\tau_D(s)',
$$

we have

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{[Tr]} Post_t^{\tau}\cdot\left(\bar{\mathcal{Z}}_{\cdot,t}^{treat}\right)^{\tau}\\
&=\frac{1}{T}\sum_{t=1}^{[Tr]}\left[Post_t-\left(\frac{1}{T}\sum_{s=1}^{T} Post_s\cdot\tau_D\left(s\right)'\right)A_{\tau\tau}^{-1}\tau_D\left(t\right)\right]\\
&\times\left[\bar{\mathcal{Z}}_{\cdot,t}^{treat}-\left(\frac{1}{T}\sum_{s=1}^{T}\bar{\mathcal{Z}}_{\cdot,s}^{treat}\tau_D\left(s\right)'\right)A_{\tau\tau}^{-1}\tau_D\left(t\right)\right]\\
&=\frac{1}{T}\sum_{t=T\nu}^{[Tr]}\bar{\mathcal{Z}}_{\cdot,t}^{treat}-\left(\frac{1}{T}\sum_{t=1}^{[Tr]}\bar{\mathcal{Z}}_{\cdot,t}^{treat}\cdot\tau_D\left(t\right)'\right)A_{\tau\tau}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\tau_D\left(s\right)\cdot Post_s\right)\\
&-\left(\frac{1}{T}\sum_{s=1}^{T}\bar{\mathcal{Z}}_{\cdot,s}^{treat}\cdot\tau_D\left(s\right)'\right)A_{\tau\tau}^{-1}\left(\frac{1}{T}\sum_{t=1}^{[Tr]} Post_t\cdot\tau_D\left(t\right)\right)\\
&+\frac{1}{T}\sum_{t=1}^{[Tr]}\tau_D\left(t\right)'A_{\tau\tau}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\tau_D\left(s\right)Post_s\right)\left(\frac{1}{T}\sum_{s=1}^{T}\bar{\mathcal{Z}}_{\cdot,s}^{treat}\tau_D\left(s\right)'\right)A_{\tau\tau}^{-1}\tau_D\left(t\right),
\end{aligned}
$$

where we have used

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{[Tr]}\left[\left(\frac{1}{T}\sum_{s=1}^{T} Post_s\cdot\tau_D\left(s\right)'\right)A_{\tau\tau}^{-1}\tau_D\left(t\right)\right]\bar{\mathcal{Z}}_{\cdot,t}^{treat}\\
&=\frac{1}{T}\sum_{t=1}^{[Tr]}\bar{\mathcal{Z}}_{\cdot,t}^{treat}\left[\left(\frac{1}{T}\sum_{s=1}^{T} Post_s\cdot\tau_D\left(s\right)'\right)A_{\tau\tau}^{-1}\tau_D\left(t\right)\right]\\
&=\frac{1}{T}\sum_{t=1}^{[Tr]}\bar{\mathcal{Z}}_{\cdot,t}^{treat}\left[\tau_D\left(t\right)'A_{\tau\tau}^{-1}\frac{1}{T}\sum_{s=1}^{T}\tau_D\left(s\right)\cdot Post_s\right]\\
&=\left(\frac{1}{T}\sum_{t=1}^{[Tr]}\bar{\mathcal{Z}}_{\cdot,t}^{treat}\tau_D\left(t\right)'\right)A_{\tau\tau}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\tau_D\left(s\right)\cdot Post_s\right).
\end{aligned}
$$

The above holds because $(T^{-1}\sum_{s=1}^{T} Post_s\cdot\tau_D\left(s\right)')A_{\tau\tau}^{-1}\tau_D\left(t\right)$ is a scalar.

A similar expression can be obtained for $\frac{1}{T}\sum_{t=1}^{[Tr]} Post_t^{\tau}\cdot(\bar{\mathcal{Z}}_{\cdot,t}^{control})^{\tau}$. It then follows from Assumption 3.3 that

$$S_{21}\left(r\right)=o_p\left(1\right).$$

Part (c). Using (25) and Assumption 3.4, we have

$$
\begin{aligned}
S_{22}\left(r\right)&=\frac{1}{T}\sum_{t=1}^{[Tr]}\frac{1}{n}\sum_{i=1}^{n}\tilde{Z}_{it}^{\tau}(\tilde{Z}_{it}^{\tau})'=\frac{1}{T}\sum_{t=1}^{[Tr]}\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathcal{Z}}_{it}^{\tau}\left(\tilde{\mathcal{Z}}_{it}^{\tau}\right)'\\
&=rG+o_p\left(1\right)
\end{aligned}
$$

uniformly over $r\in[0,1]$. ■

Proof of Lemma 3.2. In view of Lemma 3.1, we have

$$\sqrt{nT}\left(\hat{\theta}_1-\theta_{10}\right) \tag{26}$$

$$=\left[\frac{1}{nT}\sum_{i,t}(\widetilde{Treat_i})^2\cdot(Post_t^{\tau})^2\right]^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Post_t^{\tau}\cdot\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}^{\tau}\right)+o_p(1) \tag{27}$$

$$=\frac{1}{\mathcal{S}_{11}(1)}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Post_t^{\tau}\cdot\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}^{\tau}\right]+o_p(1). \tag{28}$$

By Assumption 3.5, we have

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Post_t^{\tau}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}^{\tau}\right)\\
&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(Post_t^{\tau})^{\tau}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}\right)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Post_t^{\tau}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}\right)\\
&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[1\left\{\frac{t}{T}\geq\nu\right\}-\left(\frac{1}{T}\sum_{s=1}^{T}Post_s\cdot\tau_D(s)'\right)\left(\frac{1}{T}\sum_{s=1}^{T}\tau_D(s)\tau_D(s)'\right)^{-1}\tau_D(t)\right]\\
&\times\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}^{\tau}\right)\to^d\Lambda\int_0^1 H_{\nu}(r)\,dB(r).
\end{aligned}$$

Therefore,

$$\sqrt{nT}\left(\hat{\theta}_1-\theta_{10}\right)\to^d\frac{\Lambda}{\mu(1-\mu)}\frac{\int_0^1 H_{\nu}(r)\,dB(r)}{\int_0^1 H_{\nu}^2(r)\,dr}\overset{d}{=}\frac{\Lambda}{\mu(1-\mu)\sqrt{\int_0^1 H_{\nu}^2(r)\,dr}}\cdot N(0,1),$$

as desired. ■

We will need the following lemma to prove Lemma 3.3.

Lemma 9.1 *Let Assumptions 3.1–3.4 hold. Then*

$$\frac{1}{T}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\left[\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_{it}^{\tau}(\tilde{X}_{it}^{\tau})'\right]=\begin{pmatrix}\mu(1-\mu)\int_0^1\Phi_k(r)H_{\nu}^2(r)\,dr & 0\\ 0 & \int_0^1\Phi_k(r)\,dr\cdot G\end{pmatrix}+o_p(1).$$

Proof of Lemma 9.1. Define $S(0)=0$. Recalling the definition of $S(r)$ in 5, we have

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_{it}^{\tau}(\tilde{X}_{it}^{\tau})'\\
&=\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\left[S\left(\frac{t}{T}\right)-S\left(\frac{t-1}{T}\right)\right]\\
&=\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)S\left(\frac{t}{T}\right)-\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)S\left(\frac{t-1}{T}\right)\\
&=\sum_{t=1}^{T-1}\left[\Phi_k\left(\frac{t}{T}\right)-\Phi_k\left(\frac{t+1}{T}\right)\right]S\left(\frac{t}{T}\right)+\Phi_k(1)\,S(1).
\end{aligned}$$

Let

$$\mathcal{S}(r) := \begin{pmatrix} \mathcal{S}_{11}(r) & \mathcal{S}_{12}(r) \\ \mathcal{S}_{21}(r) & \mathcal{S}_{22}(r) \end{pmatrix} \text{ with } \mathcal{S}_{21}(r) = \mathcal{S}_{12}'(r),$$

and define

$$\Delta_t = S\left(\frac{t}{T}\right) - \mathcal{S}\left(\frac{t}{T}\right).$$

It follows from Lemma 3.1 that $\sup_t \|\Delta_t\| = o_p(1)$. Using this, we obtain

$$\begin{aligned}
&\sum_{t=1}^{T-1}\left[\Phi_k\left(\frac{t}{T}\right) - \Phi_k\left(\frac{t+1}{T}\right)\right] S\left(\frac{t}{T}\right) + \Phi_k(1) S(1) \\
&= \sum_{t=1}^{T-1}\left[\Phi_k\left(\frac{t}{T}\right) - \Phi_k\left(\frac{t+1}{T}\right)\right]\left[\Delta_t + \mathcal{S}\left(\frac{t}{T}\right)\right] + \Phi_k(1)\left(\Delta_T + \mathcal{S}(1)\right) \\
&= \sum_{t=1}^{T-1}\left[\Phi_k\left(\frac{t}{T}\right) - \Phi_k\left(\frac{t+1}{T}\right)\right]\mathcal{S}\left(\frac{t}{T}\right) + \Phi_k(1)\mathcal{S}(1) \\
&+ \sum_{t=1}^{T-1}\left[\Phi_k\left(\frac{t}{T}\right) - \Phi_k\left(\frac{t+1}{T}\right)\right]\Delta_t + \Phi_k(1)\Delta_T \\
&= \sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\left[\mathcal{S}\left(\frac{t}{T}\right) - \mathcal{S}\left(\frac{t-1}{T}\right)\right] \\
&+ \sum_{t=1}^{T-1}\left[\Phi_k\left(\frac{t}{T}\right) - \Phi_k\left(\frac{t+1}{T}\right)\right]\Delta_t + o_p(1).
\end{aligned}$$

Under the piecewise monotonicity condition in Assumption 3.1 , for some finite κ we can partition the set $\{1, 2, \ldots, T-1\}$ into κ maximal non-overlapping subsets $\cup_{j=1}^{\kappa}\mathcal{I}_j$ such that $\Phi_k(t/T)$ is monotonic on each $\mathcal{I}_j := \{\mathcal{I}_{jL}, \ldots, \mathcal{I}_{jU}\}$. Now

$$\begin{aligned}
&\left\|\sum_{t=1}^{T-1}\left[\Phi_k\left(\frac{t}{T}\right) - \Phi_k\left(\frac{t+1}{T}\right)\right]\Delta_t\right\| \\
&\leq \left\|\sum_{j=1}^{\kappa}\sum_{t\in\mathcal{I}_j}\left[\Phi_k\left(\frac{t}{T}\right) - \Phi_k\left(\frac{t+1}{T}\right)\right]\Delta_t\right\| + o_p(1) \\
&\leq \sum_{j=1}^{\kappa}\sum_{t\in\mathcal{I}_j}\left|\Phi_k\left(\frac{t}{T}\right) - \Phi_k\left(\frac{t+1}{T}\right)\right|\sup_s\|\Delta_s\| + o_p(1) \\
&= \sum_{j=1}^{\kappa}(\pm)_j\sum_{t\in\mathcal{I}_j}\left[\Phi_k\left(\frac{t}{T}\right) - \Phi_k\left(\frac{t+1}{T}\right)\right]\sup_s\|\Delta_s\| + o_p(1) \\
&= \sum_{j=1}^{\kappa}\left|\left[\Phi_k\left(\frac{\mathcal{I}_{jU}}{T}\right) - \Phi_k\left(\frac{\mathcal{I}_{jL}}{T}\right)\right]\right|\sup_s\|\Delta_s\| + o_p(1) \\
&= O(1)\sup_s\|\Delta_s\| + o_p(1) = o_p(1),
\end{aligned}$$

where the $o_p(1)$ term in the first inequality reflects the case when t and $t+1$ belong to different partitions and "$(\pm)_j$" takes "+" or "−" depending on whether $\Phi_k(t/T)$ is increasing or decreasing on the interval $[\mathcal{I}_{jL}, \mathcal{I}_{jU}]$. As a result,

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\left[\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_{it}^{\tau}(\tilde{X}_{it}^{\tau})'\right] \\
&= \sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\left[\mathcal{S}\left(\frac{t}{T}\right)-\mathcal{S}\left(\frac{t-1}{T}\right)\right]+o_p(1) \\
&= \int_0^1 \Phi_k(r)\, d\mathcal{S}(r)+o_p(1) \\
&= \begin{pmatrix} \mu(1-\mu)\int_0^1 \Phi_k(r) H_\nu^2(r)\,dr & 0 \\ 0 & \int_0^1 \Phi_k(r)\,dr\cdot G \end{pmatrix}+o_p(1).
\end{aligned}
$$

■

Proof of Lemma 3.3. (a) Using Lemma 9.1, we have

$$
\begin{aligned}
& \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\hat{u}_{n,t} \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{X}_{it}^{\tau}\left[\tilde{Y}_{it}^{\tau}-(\tilde{X}_{it}^{\tau})'\hat{\theta}\right] \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{X}_{it}^{\tau}\tilde{\epsilon}_{it}^{\tau}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{X}_{it}^{\tau}(\tilde{X}_{it}^{\tau})'\left(\hat{\theta}-\theta_0\right)\right] \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{X}_{it}^{\tau}\tilde{\epsilon}_{it}^{\tau} \\
&\quad -\begin{pmatrix} \mu(1-\mu)\int_0^1 \Phi_k(r) H_\nu^2(r)\,dr & 0 \\ 0 & \int_0^1 \Phi_k(r)\,dr\cdot G \end{pmatrix}\begin{pmatrix} \sqrt{nT}(\hat{\theta}_1-\theta_{10}) \\ \sqrt{nT}(\hat{\theta}_2-\theta_{20}) \end{pmatrix}+o_p(1) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{X}_{it}^{\tau}\tilde{\epsilon}_{it}^{\tau}-\begin{pmatrix} \mu(1-\mu)\int_0^1 \Phi_k(r) H_\nu^2(r)\,dr\sqrt{nT}(\hat{\theta}_1-\theta_{10}) \\ \int_0^1 \Phi_k(r)\,dr\cdot G\sqrt{nT}(\hat{\theta}_2-\theta_{20}) \end{pmatrix}+o_p(1).
\end{aligned} \tag{29}
$$

So

$$
\begin{aligned}
& R\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^{\tau}(\tilde{X}_{it}^{\tau})'\right]^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\hat{u}_{n,t} \\
&= [\mathcal{S}_{11}(1)]^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat}_i\cdot Post_t^{\tau}\cdot\epsilon_{it}^{\tau} \\
&\quad -[\mathcal{S}_{11}(1)]^{-1}\int_0^1 \Phi_k(r) H_\nu^2(r)\,dr\times[\mu(1-\mu)]\sqrt{nT}(\hat{\theta}_1-\theta_{10})+o_p(1).
\end{aligned}
$$

Note that

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot Post_t^{\tau}\cdot\epsilon_{it}^{\tau}\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\Phi_k\left(\frac{t}{T}\right)Post_t^{\tau}\right]^{\tau}\cdot\epsilon_{it}\\
&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\Phi_k\left(\frac{t}{T}\right)Post_t^{\tau}\right]^{\tau}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}\right),
\end{aligned}$$

where

$$\begin{aligned}
&\left[\Phi_k\left(\frac{t}{T}\right)Post_t^{\tau}\right]^{\tau}\\
&=\Phi_k\left(\frac{t}{T}\right)Post_t^{\tau}-\left(\frac{1}{T}\sum_{s=1}^{T}\Phi_k\left(\frac{s}{T}\right)Post_s^{\tau}\cdot\tau_D(s)'\right)\left(\frac{1}{T}\sum_{s=1}^{T}\tau_D(s)\,\tau_D(s)'\right)^{-1}\tau_D(t).
\end{aligned}$$

Using Assumption 3.5, we have

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\Phi_k\left(\frac{t}{T}\right)Post_t^{\tau}\right]^{\tau}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}\right)\\
&\to^d \Lambda\int_0^1\left\{\Phi_k(r)\,H_\nu(r)-\left[\int_0^1\Phi_k(s)\,H_\nu(s)\,\tau(s)'\right]\left[\int_0^1\tau(s)\,\tau(s)'\,ds\right]^{-1}\tau(r)\right\}dB(r).
\end{aligned}$$

Combining this with (29) and Lemma 3.2 yields

$$\begin{aligned}
&R\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^{\tau}(\tilde{X}_{it}^{\tau})'\right]^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\hat{u}_{n,t}\\
&\to^d \Lambda\left[\mathcal{S}_{11}(1)\right]^{-1}\int_0^1\left\{\Phi_k(r)\,H_\nu(r)-\left[\int_0^1\Phi_k(s)\,H_\nu(s)\,\tau(s)'\right]\left[\int_0^1\tau(s)\,\tau(s)'\,ds\right]^{-1}\tau(r)\right\}dB(r)\\
&-\Lambda\left[\mathcal{S}_{11}(1)\right]^{-1}\left[\int_0^1\Phi_k(s)\,H_\nu^2(s)\,ds\right]\left[\int_0^1H_\nu^2(s)\,ds\right]^{-1}\int_0^1H_\nu(r)\,dB(r)\\
&=\Lambda\left[\mathcal{S}_{11}(1)\right]^{-1}\int_0^1\Phi_k^H(r)\,dB(r),
\end{aligned}$$

where

$$\begin{aligned}
\Phi_k^H(r)&=\Phi_k(r)\,H_\nu(r)-\left[\int_0^1\Phi_k(s)\,H_\nu(s)\,\tau(s)'\right]\left[\int_0^1\tau(s)\,\tau(s)'\,ds\right]^{-1}\tau(r)\\
&-\left[\int_0^1\Phi_k(s)\,H_\nu^2(s)\,ds\right]\left[\int_0^1H_\nu^2(s)\,ds\right]^{-1}H_\nu(r)\\
&=\left\{\Phi_k(r)-\left[\int_0^1\Phi_k(s)\,H_\nu^2(s)\,ds\right]\left[\int_0^1H_\nu^2(s)\,ds\right]^{-1}\right\}H_\nu(r)\\
&-\left[\int_0^1\Phi_k(s)\,H_\nu(s)\,\tau(s)'\,ds\right]\left[\int_0^1\tau(s)\,\tau(s)'\,ds\right]^{-1}\tau(r).
\end{aligned}$$

It then follows that

$$\hat{V}_R \to^d \Lambda^2 \frac{1}{K} \sum_{k=1}^{K} \left[\int_0^1 \Phi_k^H(r)\, dB(r) \right]^2 \left[\mu(1-\mu) \int_0^1 H_\nu^2(s)\, ds \right]^{-2}.$$

This completes the proof of Lemma 3.3. ■

Proof of Theorem 3.1. Using Lemmas 3.2 and 3.3, we immediately have

$$\begin{aligned}
\mathbb{T} &= \frac{\sqrt{nT}\left(\hat{\theta}_1 - \theta_{10}\right)}{\sqrt{\hat{V}_R}} \\
&\to^d \frac{1}{\mu(1-\mu)} \frac{\int_0^1 H_\nu(r)\, dB(r)}{\int_0^1 H_\nu^2(r)\, dr} \frac{\mu(1-\mu)\int_0^1 H_\nu^2(s)\, ds}{\left\{ \frac{1}{K}\sum_{k=1}^K \left[\int_0^1 \Phi_k^H(r)\, dB(r) \right]^2 \right\}^{1/2}} \\
&= \frac{\int_0^1 H_\nu(r)\, dB(r)}{\left\{ \frac{1}{K}\sum_{k=1}^K \left[\int_0^1 \Phi_k^H(r)\, dB(r) \right]^2 \right\}^{1/2}} := \mathcal{T}_\infty.
\end{aligned}$$

■

Proof of Theorem 5.1. (a) We have

$$\begin{aligned}
\hat{\Lambda}_k &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{Treat}_i \cdot \hat{\epsilon}_{it}^{\tau} \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{Treat}_i \cdot \epsilon_{it}^{\tau} \\
&- \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{Treat}_i \cdot \begin{pmatrix} \widetilde{Treat}_i \cdot Post_t^{\tau} \\ \tilde{Z}_{it}^{\tau} \end{pmatrix}' \left(\hat{\theta} - \theta_0\right).
\end{aligned} \tag{30}$$

Let

$$\left[\Phi_k\left(\frac{t}{T}\right)\right]^{\tau} = \Phi_k\left(\frac{t}{T}\right) - \left[\frac{1}{T}\sum_{s=1}^{T} \Phi_k\left(\frac{s}{T}\right) \cdot \tau_D(s)'\right] \left[\frac{1}{T}\sum_{s=1}^{T} \tau_D(s)\, \tau_D(s)'\right]^{-1} \tau_D(t).$$

Then the first term in (30) satisfies

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{Treat}_i \cdot \epsilon_{it}^{\tau} \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\Phi_k\left(\frac{t}{T}\right)\right]^{\tau} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{Treat}_i \cdot \epsilon_{it} \to^d \Lambda \int \Phi_k^{\tau}(r)\, dB(r).
\end{aligned} \tag{31}$$

The second term in (30) satisfies

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\left(\begin{array}{c}\widetilde{Treat_i}\cdot Post_t^{\tau}\\ \tilde{Z}_{it}^{\tau}\end{array}\right)'\left(\hat{\theta}-\theta_0\right)\\
&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\widetilde{Treat_i})^2 Post_t^{\tau}\left(\hat{\theta}_1-\theta_{10}\right)\\
&+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot(\tilde{Z}_{it}^{\tau})'\left(\hat{\theta}_2-\theta_{20}\right)\\
&=\left[\frac{1}{n}\sum_{i=1}^{n}(\widetilde{Treat_i})^2\right]\left[\frac{1}{T}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)Post_t^{\tau}\right]\sqrt{nT}\left(\hat{\theta}_1-\theta_{10}\right)\\
&+\frac{1}{nT}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\sum_{i=1}^{n}\widetilde{Treat_i}\cdot(\tilde{\mathcal{Z}}_{it}^{\tau})'\sqrt{nT}\left(\hat{\theta}_2-\theta_{20}\right):=I_1+I_2. \qquad (32)
\end{aligned}$$

To obtain an upper bound for I_2, we have, using Assumption 3.3:

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{[Tr]}\frac{1}{n}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot(\tilde{\mathcal{Z}}_{it}^{\tau})'\\
&=\frac{1}{T}\sum_{t=1}^{[Tr]}\frac{1}{n}\sum_{i=1}^{n}Treat_i\cdot(\tilde{\mathcal{Z}}_{it}^{\tau})'=\frac{1}{T}\sum_{t=1}^{[Tr]}\frac{1}{n}\sum_{i=1}^{\mu n}(\tilde{\mathcal{Z}}_{it}^{\tau})'\\
&=(1-\mu)\,\mu\frac{1}{T}\sum_{t=1}^{[Tr]}\left[\frac{1}{n\mu}\sum_{i=1}^{\mu n}(\mathcal{Z}_{it}^{\tau})'-\frac{1}{n(1-\mu)}\sum_{j=n\mu+1}^{n}(\mathcal{Z}_{jt}^{\tau})'\right]\\
&=(1-\mu)\,\mu\left[\frac{1}{T}\sum_{t=1}^{[Tr]}\bar{\mathcal{Z}}_{\cdot,t}^{treat}-\left(\frac{1}{T}\sum_{s=1}^{T}\bar{\mathcal{Z}}_{\cdot,s}^{treat}\tau_D(s)'\right)A_{\tau\tau}^{-1}\frac{1}{T}\sum_{t=1}^{[Tr]}\tau_D(t)\right]\\
&-(1-\mu)\,\mu\left[\frac{1}{T}\sum_{t=1}^{[Tr]}\bar{\mathcal{Z}}_{\cdot,t}^{control}-\left(\frac{1}{T}\sum_{s=1}^{T}\bar{\mathcal{Z}}_{\cdot,s}^{control}\tau_D(s)'\right)A_{\tau\tau}^{-1}\frac{1}{T}\sum_{t=1}^{[Tr]}\tau_D(t)\right]\\
&=o_p(1)
\end{aligned}$$

uniformly in r. Combining this with a proof similar to that of Lemma 9.1, we have

$$\frac{1}{nT}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\sum_{i=1}^{n}\widetilde{Treat_i}\cdot(\tilde{\mathcal{Z}}_{it})'=o_p(1).$$

It then follows that $I_2=o_p(1)$. Therefore, the second term in (30) converges in distribution to

$$\begin{aligned}
&\mu(1-\mu)\int_0^1\Phi_k(r)\,H_\nu(r)\,dr\frac{\Lambda}{\mu(1-\mu)}\frac{\int_0^1 H_\nu(r)\,dB(r)}{\int_0^1 H_\nu^2(r)\,dr}\\
&=\Lambda\frac{\int_0^1\Phi_k(r)\,H_\nu(r)\,dr}{\int_0^1 H_\nu^2(r)\,dr}\int_0^1 H_\nu(r)\,dB(r).
\end{aligned}$$

Combining this with (31) yields

$$\hat{\Lambda}_k \to^d \Lambda \left[\int_0^1 \Phi_k^\tau (r)\, dB(r) - \frac{\int_0^1 \Phi_k(r)\, H_\nu(r)\, dr}{\int_0^1 H_\nu^2(r)\, dr} \int_0^1 H_\nu(r)\, dB(r) \right]$$
$$= \Lambda \int_0^1 \Phi_k^{\mathcal{H}}(r)\, dB(r).$$

Part (a) follows immediately.

(b) Using part (a), we have

$$\tilde{\mathbb{T}} = \frac{\sqrt{nT}(\hat{\theta}_1 - \theta_{10})}{\hat{\sigma}}$$
$$\to^d \frac{1}{\mu(1-\mu)} \frac{\int_0^1 H_\nu(r)\, dB(r)}{\int_0^1 H_\nu^2(r)\, dr} \frac{\mu(1-\mu)\sqrt{\int_0^1 H_\nu^2(s)\, ds}}{\left\{ \frac{1}{K}\sum_{k=1}^K \left[\int_0^1 \Phi_k^{\mathcal{H}}(r)\, dB(r) \right]^2 \right\}^{1/2}}$$
$$= \frac{\int_0^1 H_\nu(r)\, dB(r)}{\left\{ \frac{1}{K}\sum_{k=1}^K \left[\int_0^1 \Phi_k^{\mathcal{H}}(r)\, dB(r) \right]^2 \right\}^{1/2} \sqrt{\int_0^1 H_\nu^2(s)\, ds}} := \tilde{\mathcal{T}}_\infty.$$

■

It follows from (19) that

$$\hat{\sigma}^2 = \hat{\Lambda}^2 \left[\mu(1-\mu)\right]^{-2} \left\{ \frac{1}{T}\sum_{t=1}^T [Post_t^\tau]^2 \right\}^{-1},$$

where

$$\hat{\Lambda}^2 = \frac{1}{K}\sum_{k=1}^K \hat{\Lambda}_k^2 \text{ and } \hat{\Lambda}_k = \frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbf{\Phi}_{k,\mathcal{H},t}\hat{e}_t^\tau.$$

To prove Theorem 6.1, we need to first prove the following lemma, which establishes the asymptotic bias and variance of $\hat{\Lambda}^2$.

Lemma 9.2 *Let Assumptions 3.2 and 6.1 hold. If $K \to \infty$ such that $K/T + T/K^2 \to 0$, then*
(i) $E(\hat{\Lambda}^2 - \Lambda^2) = \left(\frac{K}{T}\right)^2 B + O\left(\frac{1}{T}\right),$
(ii) $var(\hat{\Lambda}^2) = \frac{2\Lambda^4}{K}(1 + o(1)) + O\left(\frac{1}{T}\right).$

Proof of Lemma 9.2. We prove (i) only, as (ii) follows from standard arguments, e.g., Theorem 9 in Hannan (1970, p. 280). By definition, we have

$$\hat{\Lambda}_k = \frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbf{\Phi}_{k,\mathcal{H},t}\hat{e}_t^\tau$$
$$= \frac{1}{\sqrt{T}}\mathbf{\Phi}'_{k,\mathcal{H}} \left[\left(I_T - M_\tau \cdot Post\left(Post' \cdot M_\tau \cdot Post\right)^{-1} Post' \cdot M_\tau \right) M_\tau e \right]$$
$$= \frac{1}{\sqrt{T}}\mathbf{\Phi}'_{k,\mathcal{H}} M_\tau e - \frac{1}{\sqrt{T}}\mathbf{\Phi}'_{k,\mathcal{H}} Post^\tau \cdot \left(T\|Post^\tau\|^2\right)^{-1} (Post^\tau)' e$$
$$= \frac{1}{\sqrt{T}}\mathbf{\Phi}'_{k,\mathcal{H}} \frac{\mathbf{C}_{\mathcal{H}}}{T} e = \frac{1}{\sqrt{T}}\left(\mathbf{\Phi}_k^*\right)' e,$$

where $\mathbf{C}_{\mathcal{H}}$ is defined in (16),

$$\mathbf{\Phi}_k^* = \frac{\mathbf{C}_{\mathcal{H}}}{T}\mathbf{\Phi}_{k,\mathcal{H}} = M_{Post,\tau}\mathbf{\Phi}R^{(k)},$$

and $R^{(k)}$ is the k-th column of $(R_{\mathcal{H}})^{-1}$. Here we have used $\mathbf{C}_{\mathcal{H}} = TM_{Post,\tau}$, where $M_{Post,\tau}$ is defined in (16).

Let $\mathbf{\Phi}^* = (\mathbf{\Phi}_1^*, \mathbf{\Phi}_2^*, ..., \mathbf{\Phi}_K^*) = T^{-1}\mathbf{C}_{\mathcal{H}}\mathbf{\Phi}_{\mathcal{H}}$, where $\mathbf{\Phi}_{\mathcal{H}} = \mathbf{\Phi}R_{\mathcal{H}}^{-1}$. Then

$$\left(\frac{\mathbf{\Phi}^*}{\sqrt{T}}\right)'\frac{\mathbf{\Phi}^*}{\sqrt{T}} = \mathbf{\Phi}_{\mathcal{H}}'\frac{\mathbf{C}_{\mathcal{H}}}{T^2}\mathbf{\Phi}_{\mathcal{H}} = \left(R_{\mathcal{H}}^{-1}\right)'\mathbf{\Phi}'\frac{\mathbf{C}_{\mathcal{H}}}{T^2}\mathbf{\Phi}R_{\mathcal{H}}^{-1} = \left(R_{\mathcal{H}}^{-1}\right)'R_{\mathcal{H}}'R_{\mathcal{H}}R_{\mathcal{H}}^{-1} = I_K.$$

Therefore, $\mathbf{\Phi}_k^*/\sqrt{T}$ is a series of orthonormal basis vectors in $\mathbb{R}^T$. Each column $\mathbf{\Phi}_k^*$ of the matrix $\mathbf{\Phi}^*$ corresponds to the basis function $\Phi_k^*(r)$ defined by

$$\Phi_k^*(r) = \sum_{j=1}^{K}\left[\int_0^1 C_\nu^{\mathcal{H}}(r,s)\,\Phi_j(s)\,ds\right]R_\infty^{(j,k)} = \sum_{j=1}^{K}\Phi_j^{\mathcal{H}}(r)\,R_\infty^{(j,k)}, \tag{33}$$

where $R_\infty^{(j,k)}$ is the (j,k)-th element of R_∞^{-1} and $R_\infty = \lim_{T\to\infty} R_{\mathcal{H}}$ is the upper triangular factor of the Cholesky decomposition of the matrix $\int_0^1 \Phi_F^{\mathcal{H}}(r)\,\Phi_F^{\mathcal{H}}(r)'\,dr$. The second equality in (33) follows from simple calculations using the definition of $C_\nu^{\mathcal{H}}(r,s)$ given in (15).

Let $c_k = \sum_{j=1}^K \tilde{c}_j R_\infty^{(j,k)}$ and $d_k = \sum_{j=1}^K \tilde{d}_j R_\infty^{(j,k)}$. Then $\Phi_k^*(r)$ can be further represented as

$$\begin{aligned}\Phi_k^*(r) &= \sum_{j=1}^{K}\left[\Phi_j(r) - \tau(r)'\tilde{d}_j - 1(r\geq\nu)\,\tilde{c}_j\right]R_\infty^{(j,k)}\\ &= \sum_{j=1}^{K}\Phi_j(r)\,R_\infty^{(j,k)} - \tau(r)'d_k - 1(r\geq\nu)\,c_k\\ &:= \pi_k(r) - 1(r\geq\nu)\,c_k,\end{aligned}$$

where $\pi_k(r) = \sum_{j=1}^K \Phi_j(r)\,R_\infty^{(j,k)} - \tau(r)'d_k$.

Under Assumption 6.1(c), $\pi_k(r)$ is twice continuously differentiable. Under Assumption 6.1(b), the coefficients $\{c_k\}$ satisfy

$$\begin{aligned}\sum_{k=1}^{K}|c_k|^2 &= \sum_{k=1}^{K}\sum_{j_1=1}^{K}\sum_{j_2=1}^{K}\tilde{c}_{j_1}\tilde{c}_{j_2}R_\infty^{(j_1,k)}R_\infty^{(j_2,k)} = \sum_{j_1=1}^{K}\sum_{j_2=1}^{K}\tilde{c}_{j_1}\tilde{c}_{j_2}\sum_{k=1}^{K}R_\infty^{(j_1,k)}R_\infty^{(j_2,k)}\\ &= \tilde{c}'\left\{\int_0^1\Phi_F^{\mathcal{H}}(r)\,\Phi_F^{\mathcal{H}}(r)'\,dr\right\}^{-1}\tilde{c}\\ &= O(\|\tilde{c}\|^2) = O\left(\sum_{k=1}^{K}|\tilde{c}_k|^2\right),\end{aligned} \tag{34}$$

where $\tilde{c} = (\tilde{c}_1, ..., \tilde{c}_K)' \in \mathbb{R}^K$. But

$$\sum_{k=1}^{K} |\tilde{c}_k|^2 = \sum_{k=1}^{K} |P_H \Phi_k|^2 = \sum_{k=1}^{K} \left[\int_0^1 \Phi_k(r) H_\nu(r) dr \right]^2 \left[\int_0^1 H_\nu^2(s) ds \right]^{-2}$$
$$= O(1) \sum_{k=1}^{K} \left(\int_0^1 \Phi_k(r)^2 dr \right) \left(\int_0^1 H_\nu^2(s) ds \right)^{-1}$$
$$= O\left(\int_0^1 \sum_{k=1}^{K} \Phi_k(r)^2 dr \right) = O\left(\int_0^1 \|\Phi_F(r)\|_2^2 dr \right), \tag{35}$$

where the third equality follows from the Cauchy inequality. Similarly, we can show that

$$\sum_{k=1}^{K} \|d_k\|^2 = O\left(\sum_{k=1}^{K} ||\tilde{d}_k||^2 \right) \text{ and } \sum_{k=1}^{K} ||\tilde{d}_k||^2 = O\left(\int_0^1 \|\Phi_F(r)\|_2^2 dr \right). \tag{36}$$

Now

$$E\hat{\Lambda}_k^2 = \frac{1}{T} E\left[e' \mathbf{\Phi}_k^* (\mathbf{\Phi}_k^*)' e \right] = \frac{1}{T} E\left[\sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{\Phi}_{k,t}^* \mathbf{\Phi}_{k,s}^* e_t e_s \right]$$
$$= \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{\Phi}_{k,t}^* \mathbf{\Phi}_{k,s}^* \sigma_{e,t-s}^2 = \sum_{p=-T+1}^{T+1} \omega_{k,T}(p/T) \left(1 - \frac{|p|}{T} \right) \sigma_{e,p}^2,$$

where

$$\omega_{k,T}(p/T) = \frac{1}{T - |p|} \sum_{t=1}^{T} \mathbf{\Phi}_{k,t}^* \mathbf{\Phi}_{k,t-p}^* 1\{1 \le t - p \le T\}.$$

As a result, we have

$$E\hat{\Lambda}^2 = \sum_{p=-T+1}^{T+1} \omega_T^K\left(\frac{p}{S} \right) \left(1 - \frac{|p|}{T} \right) \sigma_{e,p}^2,$$

where $S = TK^{-1}$ is the usual truncation lag parameter and

$$\omega_T^K\left(\frac{p}{S} \right) = \frac{1}{K} \sum_{k=1}^{K} \omega_{k,T}\left(\frac{1}{K} \frac{p}{S} \right).$$

The above representation is in the same format as what we would obtain in the case of kernel LRV estimation.

As $T \to \infty$, we have

$$\omega_{k,T}(\varsigma) = \omega_k(\varsigma) + O\left(\frac{1}{T} \right)$$

for

$$\omega_k(\varsigma) := \frac{1}{1 - |\varsigma|} \int_{\max(0,\varsigma)}^{\min(1+\varsigma,1)} \Phi_k^*(s) \Phi_k^*(s - \varsigma) ds,$$

and for $\tilde{\varsigma} = K\varsigma$,

$$\omega_T^K(\tilde{\varsigma}) \to \frac{1}{K} \sum_{k=1}^{K} \omega_k\left(\frac{1}{K} \tilde{\varsigma} \right).$$

It is easy to show that for each $k = 1, ..., K, \omega_k(\varsigma)$ is an even function, $\omega_k(0) = 1$, and $\int_0^1 \varsigma\omega_k(\varsigma)\, d\varsigma < \infty$.

Observing that $\sum_{p=-\infty}^{\infty} |p|^2 \sigma_{e,p}^2 < \infty$ under Assumption 6.1(a), we have

$$\begin{aligned}
E(\hat{\Lambda}^2 - \Lambda^2) &= \sum_{p=-T+1}^{T+1} \omega_T^K\left(\frac{p}{S}\right)\left(1 - \frac{|p|}{T}\right)\sigma_{e,p}^2 - \sum_{p=-\infty}^{\infty}\sigma_{e,p}^2 \\
&= -\sum_{p=-T+1}^{T+1} \frac{\left[1 - \omega_T^K\left(\frac{p}{S}\right)\right]}{\left(\frac{|p|}{S}\right)^q}\left(\frac{|p|}{S}\right)^q \sigma_{e,p}^2 + O\left(\frac{1}{T}\right) \\
&= -\sum_{p=-T+1}^{T+1} \frac{\left[1 - \frac{1}{K}\sum_{k=1}^{K}\omega_{k,T}\left(\frac{1}{K}\frac{p}{S}\right)\right]}{\left(\frac{|p|}{S}\right)^q}\left(\frac{|p|}{S}\right)^q \sigma_{e,p}^2 + O\left(\frac{1}{T}\right) \\
&= \lim_{(K,S)\to\infty} \frac{\left[1 - \frac{1}{K}\sum_{k=1}^{K}\omega_{k,T}\left(\frac{1}{KS}\right)\right]}{\left(\frac{1}{S}\right)^q}\frac{1}{S^q}\sum_{p=-T+1}^{T+1}|p|^q\sigma_{e,p}^2\left(1 + o\left(1\right)\right) + O\left(\frac{1}{T}\right) \\
&= -\lim_{(K,S)\to\infty}\frac{1}{K^{1+q}}\sum_{k=1}^{K}\frac{\left[1 - \omega_k\left(\frac{1}{KS}\right)\right]}{\left(\frac{1}{KS}\right)^q}\frac{1}{S^q}\sum_{p=-T+1}^{T+1}|p|^q\sigma_{e,p}^2\left(1 + o\left(1\right)\right) + O\left(\frac{1}{T}\right) \\
&= -\left(\frac{K}{T}\right)^q \omega^{(q)}(0)\sum_{p=-\infty}^{\infty}|p|^q\sigma_{e,p}^2\left(1 + o\left(1\right)\right) + O\left(\frac{1}{T}\right),
\end{aligned}$$

where $\omega^{(q)}(0)$ is defined according to

$$\omega^{(q)}(0) = \lim_{(K,S)\to\infty}\frac{1}{K^{1+q}}\sum_{k=1}^{K}\frac{\left[1 - \omega_k\left(\frac{1}{KS}\right)\right]}{\left(\frac{1}{KS}\right)^q}.$$

In addition, $q = 1$ if $\omega^{(1)}(0) \neq 0$, and $q = 2$ otherwise.

We now show that $\omega^{(1)}(0) = 0$. It is easy to see that

$$\omega^{(1)}(0) = \lim_{K\to\infty}\frac{1}{K^2}\sum_{k=1}^{K}\omega_k^{(1)}(0),$$

where

$$\omega_k^{(1)}(0) = \lim_{\varsigma\to 0+}\frac{1 - \omega_k(\varsigma)}{\varsigma}.$$

Denote $\dot{\pi}_k(s) = d\pi_k(s)/ds$. Noting that

$$\begin{aligned}
\Phi_k^*(s - \varsigma) &= \pi_k(s - \varsigma) - c_k 1(s - \varsigma \geq \nu) \\
&= \pi_k(s) - \dot{\pi}_k(s)\varsigma - c_k 1(s \geq \nu) + c_k 1\{\nu \leq s < \nu + \varsigma\} + o(\varsigma) \\
&= \Phi_k^*(s) - \left[\dot{\pi}_k(s)\varsigma - c_k 1\{s \in [\nu, \nu + \varsigma)\}\right] + o(\varsigma),
\end{aligned}$$

as $\varsigma \to 0+$, we have

$$
\begin{aligned}
\omega_k^{(1)}(0) &= \lim_{\varsigma\to 0+} \frac{1 - \frac{1}{1-|\varsigma|}\int_{\max(0,\varsigma)}^{\min(1+\varsigma,1)} \Phi_k^*(s)\,\Phi_k^*(s-\varsigma)\,ds}{\varsigma} \\
&= \lim_{\varsigma\to 0+} \frac{1 - \varsigma - \int_{\varsigma}^{1} \Phi_k^*(s)\,\Phi_k^*(s-\varsigma)\,ds}{\varsigma} \\
&= \lim_{\varsigma\to 0+} \frac{1 - \varsigma - \int_{\varsigma}^{1} [\Phi_k^*(s)]^2\,ds}{\varsigma} + \lim_{\varsigma\to 0+} \frac{1}{\varsigma}\int_{\varsigma}^{1} \Phi_k^*(s)\,[\dot{\pi}_k(s)\,\varsigma - c_k 1\{s\in[\nu,\nu+\varsigma)\}]\,ds \\
&= -1 + \Phi_k^*(0)^2 + \lim_{\varsigma\to 0+} \frac{1}{\varsigma}\int_{\varsigma}^{1} \Phi_k^*(s)\,[\dot{\pi}_k(s)\,\varsigma - c_k 1\{s\in[\nu,\nu+\varsigma)\}]\,ds,
\end{aligned}
$$

where

$$
\begin{aligned}
&\lim_{\varsigma\to 0+} \frac{1}{\varsigma}\int_{\varsigma}^{1} \Phi_k^*(s)\,[\dot{\pi}_k(s)\,\varsigma - c_k 1\{s\in[\nu,\nu+\varsigma)\}]\,ds \\
&= \lim_{\varsigma\to 0+} \frac{\int_0^1 \Phi_k^*(s)\,[\dot{\pi}_k(s)\,\varsigma - c_k 1\{s\in[\nu,\nu+\varsigma)\}]\,ds}{\varsigma} \\
&- \lim_{\varsigma\to 0+} \frac{\int_0^{\varsigma} \Phi_k^*(s)\,[\dot{\pi}_k(s)\,\varsigma - c_k 1\{s\in[\nu,\nu+\varsigma)\}]\,ds}{\varsigma} \\
&= \int_0^1 \Phi_k^*(s)\,\dot{\pi}_k(s)\,ds - c_k \Phi_k^*(\nu)\,.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\omega_k^{(1)}(0) &= -1 + \Phi_k^*(0)^2 + \int_0^1 [\pi_k(s) - c_k 1(s\geq\nu)]\,\dot{\pi}_k(s)\,ds - c_k\Phi_k^*(\nu) \\
&= -1 + \Phi_k^*(0)^2 + \frac{1}{2}\left[\pi_k(1)^2 - \pi_k(0)^2\right] - c_k(\pi_k(1) - \pi_k(\nu)) - c_k\Phi_k^*(\nu) \\
&= -1 + \Phi_k^*(0)^2 + \frac{1}{2}\left[\left\{[\Phi_k^*(1)+c_k]^2 - \Phi_k^*(0)^2\right\}\right] - c_k[\Phi_k^*(1) - \Phi_k^*(\nu)] - c_k\Phi_k^*(\nu) \\
&= -1 + \Phi_k^*(0)^2 + \frac{1}{2}\left[\Phi_k^*(1)^2 - \Phi_k^*(0)^2\right] + \frac{1}{2}c_k^2 \\
&= -1 + \frac{1}{2}\left[\Phi_k^*(1)^2 + \Phi_k^*(0)^2\right] + \frac{1}{2}c_k^2.
\end{aligned}
$$

So,

$$
\omega^{(1)}(0) = \frac{1}{2}\lim_{K\to\infty}\frac{1}{K^2}\sum_{k=1}^{K}\left[\Phi_k^*(1)^2 + \Phi_k^*(0)^2\right] + \frac{1}{2}\lim_{K\to\infty}\frac{1}{K^2}\sum_{k=1}^{K} c_k^2.
$$

Using $\Phi_k^*(r) = \sum_{j=1}^{K}\Phi_j^{\mathcal{H}}(r)\,R_\infty^{(j,k)}$, we have

$$
\begin{aligned}
\sum_{k=1}^{K}\Phi_k^*(0)^2 &= \sum_{k=1}^{K}\sum_{j=1}^{K}\Phi_j^{\mathcal{H}}(0)\,R_\infty^{(j,k)}\sum_{i=1}^{K}\Phi_i^{\mathcal{H}}(0)\,R_\infty^{(i,k)} \\
&= \sum_{i=1}^{K}\sum_{j=1}^{K}\Phi_i^{\mathcal{H}}(0)\,\Phi_j^{\mathcal{H}}(0)\sum_{k=1}^{K}R_\infty^{(i,k)}R_\infty^{(j,k)} \\
&= \Phi_F^{\mathcal{H}}(0)'\left\{\int_0^1 \Phi_F^{\mathcal{H}}(r)\,\Phi_F^{\mathcal{H}}(r)'\,dr\right\}^{-1}\Phi_F^{\mathcal{H}}(0)\,.
\end{aligned}
$$

Similarly,

$$\sum_{k=1}^{K}\Phi_k^*(1)^2=\Phi_F^{\mathcal{H}}(1)'\left\{\int_0^1\Phi_F^{\mathcal{H}}(r)\,\Phi_F^{\mathcal{H}}(r)'\,dr\right\}^{-1}\Phi_F^{\mathcal{H}}(1)\,.$$

Using (36) and Assumption 6.1 (d), we have

$$\begin{aligned}\sum_{k=1}^{K}\Phi_k^*(0)^2&=O\left(\left\|\Phi_F^{\mathcal{H}}(0)\right\|^2\right)=O\left(\sum_{k=1}^{K}\left[\Phi_k(0)-1(0\geq\nu)\,\tilde{c}_k-\tau(0)'\,\tilde{d}_k\right]^2\right)\\&=O\left(\sum_{k=1}^{K}\left[\Phi_k(0)-\tau(0)'\,\tilde{d}_k\right]^2\right)=O\left(\sum_{k=1}^{K}\Phi_k(0)^2\right)+O\left(\sum_{k=1}^{K}\left\|\tilde{d}_k\right\|^2\right)\\&=O\left(\left\|\Phi_F(0)\right\|^2\right)+O\left(\int_0^1\left\|\Phi_F(r)\right\|^2dr\right)=O(K)\,.\end{aligned}$$

Similarly,

$$\sum_{k=1}^{K}\Phi_k^*(1)^2=O\left(\left\|\Phi_F^{\mathcal{H}}(1)\right\|^2\right)=O(K)\,.$$

Therefore,

$$\begin{aligned}\omega^{(1)}(0)&=\frac{1}{2}\lim_{K\to\infty}\frac{1}{K^2}\sum_{k=1}^{K}c_k^2\leq O(1)\cdot\lim_{K\to\infty}\frac{1}{K^2}\sum_{k=1}^{K}\tilde{c}_k^2\\&=O(1)\cdot\lim_{K\to\infty}\frac{1}{K^2}\int_0^1\left\|\Phi_F(r)\right\|^2dr=0.\end{aligned}$$

We proceed to evaluate $\omega^{(2)}(0)$. Letting $\varsigma=1/(KS)$, we have

$$\begin{aligned}\frac{1-\omega_k\left(\frac{1}{KS}\right)}{\left(\frac{1}{KS}\right)^2}&=\frac{1-\omega_k(\varsigma)}{\varsigma^2}=\frac{1}{\varsigma^2}\left[1-\frac{1}{1-\varsigma}\int_\varsigma^1\Phi_k^*(s)\,\Phi_k^*(s-\varsigma)\,ds\right]\\&=\frac{1}{\varsigma^2(1-\varsigma)}\left[1-\varsigma-\int_\varsigma^1\Phi_k^*(s)\,\Phi_k^*(s-\varsigma)\,ds\right].\end{aligned}$$

Using the assumption that $\pi_k(\cdot)$ is twice continuously differentiable, as $\varsigma\to0+$ we have

$$\begin{aligned}\Phi_k^*(s-\varsigma)&=\pi_k(s-\varsigma)-c_k1(s-\varsigma\geq\nu)\\&=\pi_k(s)-\dot{\pi}_k(s)\,\varsigma+\frac{1}{2}\ddot{\pi}_k(s)\,\varsigma^2-c_k1(s\geq\nu)+c_k1\{\nu\leq s<\nu+\varsigma\}+o\left(\varsigma^2\right)\\&=\Phi_k^*(s)-\dot{\pi}_k(s)\,\varsigma+\frac{1}{2}\ddot{\pi}_k(s)\,\varsigma^2+c_k1\{s\in[\nu,\nu+\varsigma)\}+o\left(\varsigma^2\right),\end{aligned}$$

where $\ddot{\pi}_k(s)=d^2\pi_k(s)/ds^2$. So,

$$\begin{aligned}&\frac{1}{\varsigma^2(1-\varsigma)}\left[1-\varsigma-\int_\varsigma^1\Phi_k^*(s)\,\Phi_k^*(s-\varsigma)\,ds\right]\\&=\frac{1}{\varsigma^2}\left[1-\varsigma-\int_\varsigma^1\left[\Phi_k^*(s)\right]^2ds+\int_\varsigma^1\Phi_k^*(s)\,\dot{\pi}_k(s)\,\varsigma ds-c_k\int_0^1\Phi_k^*(s)\,1\{s\in[\nu,\nu+\varsigma)\}\,ds\right](1+o(\varsigma))\\&-\int_\varsigma^1\Phi_k^*(s)\,\frac{1}{2}\ddot{\pi}_k(s)\,ds\,(1+o(\varsigma))+o(1)\,.\end{aligned}$$

In the proof of $\omega^{(1)}(0) = 0$, we have effectively shown that

$$\frac{1}{K}\sum_{k=1}^{K}\frac{1-\varsigma-\int_{\varsigma}^{1}\left[\Phi_k^*(s)\right]^2 ds+\int_{\varsigma}^{1}\Phi_k^*(s)\,\dot{\pi}_k(s)\,\varsigma ds - c_k\int_0^1\Phi_k^*(s)\,1\left\{s\in[\nu,\nu+\varsigma)\right\}ds}{\varsigma}=O(1),$$

and so

$$\begin{aligned}&\frac{1}{K^3}\sum_{k=1}^{K}\frac{1-\varsigma-\int_{\varsigma}^{1}\left[\Phi_k^*(s)\right]^2 ds+\int_{\varsigma}^{1}\Phi_k^*(s)\,\dot{\pi}_k(s)\,\varsigma ds - c_k\int_0^1\Phi_k^*(s)\,1\left\{s\in[\nu,\nu+\varsigma)\right\}ds}{\varsigma^2}\\&=O\left(\frac{1}{K^2\varsigma}\right)=O\left(\frac{KS}{K^2}\right)=O\left(\frac{S}{K}\right)=O\left(\frac{T}{K^2}\right)=o(1),\end{aligned}$$

where the last equality follows from the rate condition in the lemma. As a consequence, we have

$$\begin{aligned}\omega^{(2)}(0)&=-\frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}\sum_{k=1}^{K}\int_0^1\Phi_k^*(s)\,\ddot{\pi}_k(s)\,ds\\&=-\frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}\sum_{k=1}^{K}\int_0^1\left[\pi_k(s)-c_k1(s\geq\nu)\right]\ddot{\pi}_k(s)\,ds\\&=-\frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}\sum_{k=1}^{K}\int_0^1\pi_k(s)\,\ddot{\pi}_k(s)\,ds+\frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}\sum_{k=1}^{K}c_k\left[\dot{\pi}_k(1)-\dot{\pi}_k(\nu)\right].\end{aligned}$$

Using (34), (35), and Assumption 6.1(d), we have

$$\begin{aligned}\left|\sum_{k=1}^{K}c_k\left[\dot{\pi}_k(1)-\dot{\pi}_k(\nu)\right]\right|&\leq\left(\sum_{k=1}^{K}c_k^2\right)^{1/2}\left(\sum_{k=1}^{K}\left[\dot{\pi}_k(1)-\dot{\pi}_k(\nu)\right]^2\right)^{1/2}\\&=O(\sqrt{K})\left(\sum_{k=1}^{K}\left[\dot{\pi}_k(1)-\dot{\pi}_k(\nu)\right]^2\right)^{1/2},\end{aligned}$$

and by the same argument as in (34) we have

$$\begin{aligned}\sum_{k=1}^{K}\left[\dot{\pi}_k(1)-\dot{\pi}_k(\nu)\right]^2&\leq 2\sum_{k=1}^{K}\left\{\left[\sum_{j=1}^{K}\left[\dot{\Phi}_j(1)-\dot{\Phi}_j(\nu)\right]R_{\infty}^{(j,k)}\right]^2+\left\{\left[\dot{\tau}(1)-\dot{\tau}(\nu)\right]'d_k\right\}^2\right\}\\&\leq 2\sum_{k=1}^{K}\left[\sum_{j=1}^{K}\left[\dot{\Phi}_j(1)-\dot{\Phi}_j(\nu)\right]R_{\infty}^{(j,k)}\right]^2+O\left(\sum_{k=1}^{K}\|d_k\|^2\right)\\&=O(1)\left(\sum_{j=1}^{K}\left[\dot{\Phi}_j(1)-\dot{\Phi}_j(\nu)\right]^2\right)+O\left(\sum_{k=1}^{K}\left\|\tilde{d}_k\right\|^2\right)\\&=O\left(K^3\right)+O(K)=O\left(K^3\right),\end{aligned}$$

where we have used (36). The above bounds imply that $\left|\sum_{k=1}^{K}c_k\left[\dot{\pi}_k(1)-\dot{\pi}_k(\nu)\right]\right|=O\left(K^2\right)$. Hence

$$\lim_{K\to\infty}\frac{1}{K^3}\sum_{k=1}^{K}c_k\left[\dot{\pi}_k(1)-\dot{\pi}_k(\nu)\right]=0 \text{ and } \omega^{(2)}(0)=-\frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}\sum_{k=1}^{K}\int_0^1\pi_k(s)\,\ddot{\pi}_k(s)\,ds.$$

It now suffices to compute the above limit. We have

$$\int_0^1 \pi_k(s)\ddot{\pi}_k(s)\,ds = \int_0^1 \pi_k(s)\,d\dot{\pi}_k(s) = \pi_k(s)\dot{\pi}_k(s)|_0^1 - \int_0^1 [\dot{\pi}_k(s)]^2\,ds$$
$$= \pi_k(1)\dot{\pi}_k(1) - \pi_k(0)\dot{\pi}_k(0) - \int_0^1 [\dot{\pi}_k(s)]^2\,ds.$$

Under Assumption 6.1(d), we have

$$\begin{aligned}
&\left|\sum_{k=1}^{K} \pi_k(i)\dot{\pi}_k(i)\right| \\
&= \left|\sum_{k=1}^{K}\left[\left(\sum_{j_1=1}^{K}\Phi_{j_1}(i)R_\infty^{(j_1,k)} - \tau(i)' d_k\right)\left(\sum_{j_2=1}^{K}\dot{\Phi}_{j_2}(i)R_\infty^{(j_2,k)} - \dot{\tau}(i)' d_k\right)\right]\right| \\
&\le \left|\sum_{j_2=1}^{K}\sum_{j_1=1}^{K}\Phi_{j_1}(i)\dot{\Phi}_{j_2}(i)\sum_{k=1}^{K}R_\infty^{(j_1,k)}R_\infty^{(j_2,k)}\right| + \left|\sum_{k=1}^{K} d_k'\tau(i)\dot{\tau}(i)' d_k\right| \\
&+ \left|\sum_{k=1}^{K}\sum_{j_1=1}^{K}\sum_{j_2=1}^{K}\Phi_{j_1}(i)\dot{\tau}(i)'\tilde{d}_{j_2}R_\infty^{(j_1,k)}R_\infty^{(j_2,k)}\right| + \left|\sum_{j_1=1}^{K}\sum_{j_2=1}^{K}\tau(i)'\tilde{d}_{j_1}\dot{\Phi}_{j_2}(i)\sum_{k=1}^{K}R_\infty^{(j_1,k)}R_\infty^{(j_2,k)}\right| \\
&\le \|\Phi_F(i)\|\left\|\dot{\Phi}_F(i)\right\| + O\left(\sum_{k=1}^{K}\left\|\tilde{d}_k\right\|^2\right) + \left(\|\Phi_F(i)\| + \left\|\dot{\Phi}_F(i)\right\|\right)\left(\sum_{k=1}^{K}\left\|\tilde{d}_k\right\|^2\right)^{1/2} \\
&= O\left(K^2\right) + O(K) + \|\Phi_F(i)\|\,O(\sqrt{K}) + \left\|\dot{\Phi}_F(i)\right\| O(\sqrt{K}) = O\left(K^2\right).
\end{aligned}$$

It then follows that for $\tilde{d}_F = (\tilde{d}_1, ..., \tilde{d}_K)'$ we have

$$\begin{aligned}
\omega^{(2)}(0) &= \frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}\sum_{k=1}^{K}\int_0^1 [\dot{\pi}_k(s)]^2\,ds \\
&= \frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}\sum_{k=1}^{K}\int_0^1\left[\sum_{j=1}^{K}\dot{\Phi}_j(s)R_\infty^{(j,k)} - \dot{\tau}(s)' d_k\right]^2 ds \\
&= \frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}\sum_{k=1}^{K}\int_0^1\left[\sum_{j=1}^{K}\dot{\Phi}_j(s)R_\infty^{(j,k)}\right]^2 ds + \frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}\sum_{k=1}^{K} d_k'\left[\int_0^1 \dot{\tau}(s)\dot{\tau}(s)'\,ds\right] d_k \\
&- \lim_{K\to\infty}\frac{1}{K^3}\sum_{j_1=1}^{K}\sum_{j_2=1}^{K}\left[\int_0^1 d_{j_1}'\dot{\tau}(s)\dot{\Phi}_{j_2}(s)\,ds\right]\sum_{k=1}^{K}R_\infty^{(j_1,k)}R_\infty^{(j_2,k)}.
\end{aligned}$$

But

$$\frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}\sum_{k=1}^{K} d_k'\left[\int_0^1 \dot{\tau}(s)\dot{\tau}(s)'\,ds\right] d_k = 0,$$

$$\lim_{K\to\infty}\frac{1}{K^3}\sum_{j_1=1}^{K}\sum_{j_2=1}^{K}\left[\int_0^1 \tilde{d}_{j_1}'\dot{\tau}(s)\dot{\Phi}_{j_2}(s)\,ds\right]\sum_{k=1}^{K}R_\infty^{(j_1,k)}R_\infty^{(j_2,k)}$$
$$=\lim_{K\to\infty}\frac{1}{K^3}\int_0^1 \dot{\Phi}_F(s)'\left[\int_0^1 \Phi_F^{\mathcal{H}}(s)\Phi_F^{\mathcal{H}}(s)\,ds\right]^{-1}\left[\tilde{d}_F\dot{\tau}(s)\right]ds=0,$$

and so

$$\omega^{(2)}(0)=\frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}\sum_{j_1=1}^{K}\sum_{j_2=1}^{K}\int_0^1 \dot{\Phi}_{j_1}(s)\dot{\Phi}_{j_2}(s)\,ds\sum_{k=1}^{\infty}R_\infty^{(j_1,k)}R_\infty^{(j_2,k)}$$
$$=\frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}\int_0^1 \dot{\Phi}_F(s)'\left[\int_0^1 \Phi_F^{\mathcal{H}}(s)\left[\Phi_F^{\mathcal{H}}(s)\right]'ds\right]^{-1}\dot{\Phi}_F(s)\,ds.$$

Combining the above results, we can conclude that

$$E(\hat{\Lambda}^2-\Lambda^2)=-\left(\frac{K}{T}\right)^2\omega^{(2)}(0)\sum_{p=-\infty}^{\infty}|p|^2\sigma_{e,p}^2+o\left(\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right), \tag{37}$$

as desired. ■

Proposition 9.1 *Suppose we use the Fourier basis functions $\Phi_{2j-1}(s)=\sqrt{2}\cos(2\pi js)$ and $\Phi_{2j}(s)=\sqrt{2}\sin(2\pi js)$ for $j=1,...,K/2$ and $\tau(r)$ is a vector of polynomial trend functions. Then $\omega^{(2)}(0)=\pi^2/6$.*

Proof of Proposition 9.1. Letting $m(r):=\left[1(r\geq\nu),\tau(r)'\right]'$, we have

$$\Phi_k^{\mathcal{H}}(r)=\Phi_k(r)-m(r)'\vartheta_k,$$

where

$$\vartheta_k=\left[\int_0^1 m(r)m(r)'dr\right]^{-1}\left[\int_0^1 m(r)\Phi_k(r)\,dr\right].$$

Some simple calculations show that

$$\int_0^1 \Phi_k^{\mathcal{H}}(r)\Phi_j^{\mathcal{H}}(r)\,dr$$
$$=\int_0^1\left[\Phi_k(r)-m(r)'\vartheta_k\right]\left[\Phi_j(r)-m(r)'\vartheta_j\right]$$
$$=1\{k=j\}-\vartheta_j'\int_0^1\Phi_k(r)m(r)\,dr-\vartheta_k'\int_0^1\Phi_j(r)m(r)\,dr+\vartheta_k'\left[\int_0^1 m(r)m(r)'dr\right]\vartheta_j$$
$$=1\{k=j\}-\vartheta_j'\left[\int_0^1 m(r)m(r)'dr\right]\vartheta_k$$
$$=1\{k=j\}-\tilde{\vartheta}_j'\tilde{\vartheta}_k,$$

where

$$\tilde{\vartheta}_k=\left[\int_0^1 m(r)m(r)'dr\right]^{1/2}\vartheta_k=\left[\int_0^1 m(r)m(r)'dr\right]^{-1/2}\left[\int_0^1 m(r)\Phi_k(r)\,dr\right].$$

Next, we evaluate $\int_0^1 m(r)\Phi_k(r)\,dr$. The absolute value of the first element is of the form

$$\left|\int_\nu^1 \sqrt{2}\cos(2\pi kr)\,dr\right| = \sqrt{2}\left|\frac{\sin(2\pi k\nu)}{2\pi k}\right| \le \frac{C}{k} \text{ or}$$
$$\left[\int_\nu^1 \sqrt{2}\sin(2\pi kr)\,dr\right] = \sqrt{2}\left|\frac{1-\cos(2\pi k\nu)}{2\pi k}\right| \le \frac{C}{k}.$$

The absolute value of each of the other elements is of the form

$$\begin{aligned}\left|\int_0^1 \tau(r)\left(\sqrt{2}\cos 2\pi kr\right)dr\right| &= \frac{\sqrt{2}}{2\pi k}\left|\int_0^1 \tau(r)\,d(\sin 2\pi kr)\right| \\ &= \frac{\sqrt{2}}{2\pi k}\left|\int_0^1 \sin(2\pi kr)\,\dot{\tau}(r)\,dr\right| \le \frac{C}{k}\end{aligned}$$

or

$$\begin{aligned}\left|\int_0^1 \tau(r)\left(\sqrt{2}\sin 2\pi kr\right)dr\right| &= \frac{\sqrt{2}}{2\pi k}\left|\int_0^1 \tau(r)\,d(\cos 2\pi kr)\right| \\ &= \frac{\sqrt{2}}{2\pi k}\left|\tau(1)-\tau(0)-\int_0^1 \cos(2\pi kr)\,\dot{\tau}(r)\,dr\right| \le \frac{C}{k}.\end{aligned}$$

In the above, the absolute value and inequality should be understood elementwise. Therefore,

$$\left|\left[\int_0^1 m(r)\,\Phi_k(r)\,dr\right]\right| \le \frac{C}{k} \tag{38}$$

for some constant C.

Let $\tilde{\vartheta} = (\tilde{\vartheta}_1, ..., \tilde{\vartheta}_K)' \in \mathbb{R}^{K\times(d_\tau+1)}$. Then

$$\left[\int_0^1 \Phi_F^{\mathcal{H}}(s)\left[\Phi_F^{\mathcal{H}}(s)\right]' ds\right]^{-1} = \left[I_K - \tilde{\vartheta}\tilde{\vartheta}'\right]^{-1} = I_K + \tilde{\vartheta}(I_{d_\tau+1} - \tilde{\vartheta}'\tilde{\vartheta})^{-1}\tilde{\vartheta}' := I_K + \tilde{\vartheta}^*(\tilde{\vartheta}^*)',$$

where $\tilde{\vartheta}^* = \tilde{\vartheta}(I_{d_\tau+1} - \tilde{\vartheta}'\tilde{\vartheta})^{-1/2}$. In view of (38), we have $(\tilde{\vartheta}_k^*)'\tilde{\vartheta}_k^* \le C/k^2$.

It then follows that

$$\begin{aligned}\omega^{(2)}(0) &= \frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3} tr\left\{\left[\int_0^1 \Phi_F^{\mathcal{H}}(s)\left[\Phi_F^{\mathcal{H}}(s)\right]' ds\right]^{-1}\int_0^1 \dot{\Phi}_F(s)\,\dot{\Phi}_F(s)'\,ds\right\} \\ &= \frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3} tr\left\{\left[I_K + \tilde{\vartheta}^*(\tilde{\vartheta}^*)'\right]\begin{pmatrix} (2\pi)^2 & 0 & 0 & ... & 0 \\ 0 & (2\pi)^2 & 0 & & 0 \\ ... & ... & \ddots & \ddots & ... \\ 0 & 0 & \ddots & [2\pi(K/2)]^2 & 0 \\ 0 & ... & ... & 0 & [2\pi(K/2)]^2 \end{pmatrix}\right\} \\ &= \lim_{K\to\infty}\frac{1}{K^3}\sum_{j=1}^{K/2}(2\pi j)^2 + \frac{1}{2}\lim_{K\to\infty}\frac{1}{K^3}\sum_{j=1}^{K/2}\left[(\tilde{\vartheta}_{2j-1}^*)'\tilde{\vartheta}_{2j-1}^* + (\tilde{\vartheta}_{2j}^*)'\tilde{\vartheta}_{2j}^*\right](2\pi j)^2 \\ &= \lim_{K\to\infty}\frac{1}{K^3}\sum_{j=1}^{K/2}(2\pi j)^2 = \frac{1}{6}\pi^2.\end{aligned}$$

■

Proof of Theorem 6.1 . Part (a). We first establish a moment bound for $\hat{\sigma}^2/\sigma^2_{\text{GLS}}$. Under Assumption 6.1(a), we have $\sqrt{nT}E[(\hat{\theta}_{1,\text{GLS}} - \hat{\theta}_{1,\text{OLS}})]^2 = O\left(1/T\right)$, and so

$$\sigma^2_{\text{GLS}} = \Lambda^2 \left[\mu\left(1-\mu\right)\right]^{-2} \left\{\frac{1}{T}\sum_{t=1}^{T}[Post_t^{\tau}]^2\right\}^{-1} + O\left(\frac{1}{T}\right). \tag{39}$$

Using Lemma 9.2, we have

$$\begin{aligned} E\left(\frac{\hat{\sigma}^2}{\sigma^2_{\text{GLS}}} - 1\right) &= E\left[\frac{\hat{\Lambda}^2}{\Lambda^2}\left(1 + O\left(\frac{1}{T}\right)\right) - 1\right] = \frac{E(\hat{\Lambda}^2 - \Lambda^2)}{\Lambda^2} + O\left(\frac{1}{T}\right) \\ &= \frac{K^2}{T^2}\frac{B}{\Lambda^2} + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right) \\ &= \frac{K^2}{T^2}\bar{B} + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right) \end{aligned}$$

and

$$\begin{aligned} E\left(\frac{\hat{\sigma}^2}{\sigma^2_{\text{GLS}}} - 1\right)^2 &= E\left[\frac{\hat{\Lambda}^2}{\Lambda^2}\left(1 + O\left(\frac{1}{T}\right)\right) - 1\right]^2 \\ &= \frac{2}{K}(1 + o(1)) + O\left(\frac{1}{T}\right). \end{aligned}$$

Then, by applying (39) and (37), we have

$$\begin{aligned} &P\left(\left|\frac{\sqrt{nT}(\hat{\theta}_{1,\text{OLS}} - \theta_{10})}{\hat{\sigma}}\right| \le z\right) = EG\left(\frac{z^2\hat{\sigma}^2}{\sigma^2_{\text{GLS}}}\right) + O\left(\frac{1}{T}\right), \\ &= EG(z^2) + G'\left(z^2\right)E\left(\frac{\hat{\sigma}^2}{\sigma^2_{\text{GLS}}} - 1\right)z^2 + \frac{1}{2}G''(z^2)E\left(\frac{\hat{\sigma}^2}{\sigma^2_{\text{GLS}}} - 1\right)^2 z^4 \\ &+ o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right). \\ &= G(z^2) + \frac{K^2}{T^2}\bar{B}G'\left(z^2\right)z^2 + \frac{1}{K}G''(z^2)z^4 + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right). \end{aligned}$$

Using this, we have

$$\begin{aligned} &P\left(\left|\frac{\sqrt{nT}(\hat{\theta}_{1,\text{OLS}} - \theta_{10})}{\hat{\sigma}}\right| > t_K^{\alpha/2}\right) \\ &= 1 - G((t_K^{\alpha/2})^2) - \frac{K^2\bar{B}}{T^2}G'((t_K^{\alpha/2})^2)(t_K^{\alpha/2})^2 \\ &- \frac{1}{K}G''((t_K^{\alpha/2})^2)(t_K^{\alpha/2})^4 + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right). \end{aligned} \tag{40}$$

On the other hand, we have

$$(t_K^{\alpha/2})^2 = \chi_1^{\alpha} - \frac{1}{K}\frac{G''(\chi_1^{\alpha})}{G'(\chi_1^{\alpha})}(\chi_1^{\alpha})^2 + o\left(\frac{1}{K}\right). \tag{41}$$

See equation (14) in Sun (2011). Combining (40) and (41) yields

$$
\begin{aligned}
&P\left(\left|\frac{\sqrt{nT}(\hat{\theta}_{1,\mathrm{OLS}}-\theta_{10})}{\hat{\sigma}}\right|>t_K^{\alpha/2}\right)\\
&=1-G(\chi_1^{\alpha})+G'\left(\chi_1^{\alpha}\right)\frac{1}{K}\frac{G''(\chi_1^{\alpha})}{G'(\chi_1^{\alpha})}(\chi_1^{\alpha})^2\\
&-\frac{K^2\bar{B}}{T^2}G'(\chi_1^{\alpha})\chi_1^{\alpha}-\frac{1}{K}G''(\chi_1^{\alpha})\left(\chi_1^{\alpha}\right)^2+o\left(\frac{1}{K}\right)+o\left(\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right)\\
&=\alpha-\frac{K^2\bar{B}}{T^2}G'(\chi_1^{\alpha})\chi_1^{\alpha}+o\left(\frac{1}{K}\right)+o\left(\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right).
\end{aligned}
$$

Part (b). Under $H_1(\delta^2)$, we have

$$
\begin{aligned}
&P\left(\left|\frac{\sqrt{nT}(\hat{\theta}_{1,\mathrm{OLS}}-\theta_{10})}{\hat{\sigma}}\right|\leq z|H_1(\delta^2)\right)=EG_{\delta^2}\left(\frac{z^2\hat{\sigma}^2}{\sigma^2}\right)+O(T^{-1}),\\
&=G_{\delta^2}(z^2)+\frac{K^2\bar{B}}{T^2}G'_{\delta^2}\left(z^2\right)z^2+\frac{1}{K}G''_{\delta^2}(z^2)z^4+o\left(\frac{1}{K}\right)+o\left(\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right).
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
&P\left(\left|\frac{\sqrt{nT}(\hat{\theta}_{1,\mathrm{OLS}}-\theta_{10})}{\hat{\sigma}}\right|\leq t_K^{\alpha/2}|H_1(\delta^2)\right)\\
&=G_{\delta^2}((t_K^{\alpha/2})^2)+\frac{K^2\bar{B}}{T^2}G'_{\delta^2}((t_K^{\alpha/2})^2)(t_K^{\alpha/2})^2+\frac{1}{K}G''_{\delta^2}((t_K^{\alpha/2})^2)(t_K^{\alpha/2})^4\\
&+o\left(\frac{1}{K}\right)+o\left(\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right)\\
&=G_{\delta^2}(\chi_1^{\alpha})+\frac{K^2\bar{B}}{T^2}G'_{\delta^2}(\chi_1^{\alpha})\chi_1^{\alpha}+\frac{\delta^2}{2K}G'_{3,\delta^2}(\chi_1^{\alpha})\chi_1^{\alpha}+o(\frac{1}{K})+o\left(\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right),
\end{aligned}
$$

where we have used the result that

$$
G''_{\delta^2}\left(\chi_p^{\alpha}\right)-\frac{G''\left(\chi_1^{\alpha}\right)}{G'\left(\chi_1^{\alpha}\right)}G'_{\delta^2}\left(\chi_p^{\alpha}\right)=\frac{\delta^2}{2\chi_1^{\alpha}}G'_{3,\delta^2}\left(\chi_1^{\alpha}\right),
$$

which follows from simple calculations. For details of the calculation, see the proof of Theorem 5 in Sun (2011). ■

References

[1] Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59, 817–858.

[2] Bertrand, M., Duflo, E., and Mullainathan, S. (2004): "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119, 249–275.

[3] Driscoll, J. and Kraay, A. (1998): "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics* 80, 549–560.

[4] Doukhan, P. (1994): *Mixing. Properties and Example*, Springer, New York.

[5] Hansen, C. (2007): "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when T Is Large." *Journal of Econometrics* 141, 597–620.

[6] Hannan, E. J. (1970): *Multiple Time Series*, Wiley: New York.

[7] Hwang, J. and Sun, Y. (2017): "Asymptotic F and t Tests in an Efficient GMM Setting." *Journal of Econometrics* 198(2), 277–295.

[8] Kiefer, N. and Vogelsang, T. (2002a): "Heteroskedasticity-Autocorrelation Robust Testing Using Bandwidth Equal to Sample Size." *Econometric Theory* 18, 1350–1366.

[9] Kiefer, N. and Vogelsang, T. (2002b): "Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation." *Econometrica* 70, 2093–2095.

[10] Kiefer, N. and Vogelsang, T. (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests." *Econometric Theory* 21, 1130–1164.

[11] Kim, M. S. and Sun, Y. (2013): "Heteroskedasticity and Spatiotemporal Dependence Robust Inference for Linear Panel Models with Fixed Effects." *Journal of Econometrics* 177(1), 85–108.

[12] Lazarus, E., Lewis, D. J., Stock, J. H. and Watson, M. W. (2016): "HAR Inference: Kernel Choice, Size Distortions, and Power Losses." Working paper, Department of Economics, Harvard University.

[13] Phillips, P. C. B. (2005): "HAC Estimation by Automated Regression." *Econometric Theory* 21, 116–142.

[14] Sun, Y. (2011): "Robust Trend Inference with Series Variance Estimator and Testing-optimal Smoothing Parameter." *Journal of Econometrics* 164(2), 345–366.

[15] Sun, Y. (2013): "A Heteroskedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator." *Econometrics Journal* 16, 1–26.

[16] Sun, Y. (2014a): "Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference." *Journal of Econometrics* 178(3), 659–677.

[17] Sun, Y. (2014b): "Fixed-smoothing Asymptotics in a Two-step GMM Framework." *Econometrica* 82(6), 2327–2370.

[18] Sun, Y. (2014c): "Fixed-smoothing Asymptotics and Asymptotic F and t Tests in the Presence of Strong Autocorrelation." *Advances in Econometrics* 33, 23–63.

[19] Sun, Y. and Phillips, P. C. B. (2009): "Bandwidth Choice for Interval Estimation in GMM Regression." Working paper, UCSD.

[20] Sun, Y., Phillips, P. C. B., and Jin, S. (2008): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing." *Econometrica* 76(1), 175–194.

[21] Sun, Y., Phillips, P. C. B., and Jin, S. (2011): "Power Maximization and Size Control in Heteroscedasticity and Autocorrelation Robust Tests with Exponentiated Kernels." *Econometric Theory* 27(6), 1320–1368.

[22] Sun, Y. and Kim, M. S. (2012): "Simple and Powerful GMM Over-identification Tests with Accurate Size." *Journal of Econometrics* 166, 267–281.

[23] Sun, Y. and Kim, M. S. (2015): "Asymptotic F Test in a GMM Framework with Cross Sectional Dependence." *Review of Economics and Statistics* 97(1), 210–223

[24] Sun, Yu (2016): "Fixed-b Inference for Difference-in-Differences Estimation." Working paper, Capital University of Economics and Business, China.

[25] White, Halbert (2001): *Asymptotic Theory for Econometricians,* Academic Press, San Diego, CA.