A Simple and Trustworthy Asymptotic t Test 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 an asymptotically valid t test 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. The proposed t test is based on a special heteroscedasticity and autocorrelation robust (HAR) variance estimator that is tailored to inference problems in the DD setting. The asymptotic distribution of the t test depends on the smoothing parameter K in the HAR variance estimator, and a testing-optimal procedure for choosing K 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 estimator, the t test has more accurate size than the corresponding normal test and is just as powerful as the latter. Compared to the nonstandard test that is designed to reduce the size distortion of the normal test, the proposed t test is just as accurate but much more convenient to use, as the critical values are from the standard t table. Model-based and empirical-data-based Monte Carlo simulations show that the proposed t test works quite well in finite samples.

Keywords: Basis Functions, Difference-in-Differences, Fixed-smoothing Asymptotics, Heteroscedasticity and Autocorrelation Robust, 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 the data 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, BDM hereafter). If one clusters the data 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

*Email: chengliu_eco@whu.edu.cn and yisun@ucsd.edu. Liu gratefully acknowledges partial research support from Grant NSFC 71501144. Sun gratefully acknowledges partial research support from NSF under Grant No. SES-1530592. For constructive comments, we thank Han Hong (the coeditor), three anonymous referees, and an associate editor. Address correspondence to Yixiao Sun, Department of Economics, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA.

but allow the clusters to be temporally dependent. Our approach is in the spirit of Driscoll and Kraay (1998), but we employ a different heteroscedasticity and autocorrelation robust (HAR) 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 measuring the 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 (2017, hereafter SY) 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 average 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 a 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 the critical values are readily available from standard statistical tables and programming environments.

The smoothing parameter K plays a key role in determining the size and power tradeoff of the asymptotic t test. In the literature on LRV estimation and HAR inference, Phillips (2005) proposes to choose K by minimizing the asymptotic mean square error (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 choose K to minimize the type II error of the asymptotic t test while controlling its type I error.

We conduct two sets of simulation experiments. In the first set of experiments, we consider the data that is generated from a theoretical econometric model. These experiments are designed to evaluate the performance of our test relative to other tests under different simulation configurations such as the time-series and cross-sectional sample sizes, the time series and cross-sectional dependence, and the smoothing-parameter choices. More specifically, we compare the performance of a fixed-smoothing test with that of the corresponding asymptotic normal test. Each type of tests actually consists of two tests, reflecting whether a transformation is applied to the Fourier bases or not. 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 under different simulation configurations. Power study under data-driven K-values shows that all tests have similar power properties. In view of its accurate size, competitive power, and its convenience to use, we recommend using the asymptotic t test, especially when cross-sectional dependence may be present.

In the second set of experiments, we follow BDM (2004) and consider the data that is empirically calibrated to the Current Population Survey (CPS). These experiments are designed to evaluate the relative performance of our test in an empirically relevant situation. We find that our test is competitive even relative to a most trustworthy test considered by BDM (2004). This is encouraging, especially given that the latter test exploits additional information embedded in our simulation design while our test does not. Our test, therefore, can lead to more trustworthy inferences in empirical applications, even when the time series sample size is relatively small.

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 linear trend regressions, 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 specifically, 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, this resembles a deterministic mean shift, and the process has energy concentrated at the origin. As a result, the asymptotic variance of the DD estimator depends only on the long-run variance of the regression- error process. This is in contrast to the stationary case in which the asymptotic variance depends on the long-run variance of the product of the regressor process and the regression-error process.

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), Atchadé and Cattaneo (2014) and Sun (2014a) and the references therein. A paper that is closest to this paper is the paper by SY (2017), who considers the fixed-b asymptotic theory for DD regressions. There are a number of theoretical and practical differences between SY (2017) and this paper. First, SY (2017) requires n to be finite while we can allow n to be finite or grow with T, and even larger than T. Second, the asymptotic distribution of SY's test statistic is a non-standard distribution while that of our recommended test statistic is a standard t distribution. As a result, our recommended test is easier to use than SY's test, which requires simulations to obtain the critical values. Third, we have provided a data-driven procedure for choosing our smoothing parameter K while SY (2017) doesn't provide a method to choose her smoothing parameter b. In conclusion, our asymptotic theory accommodates more general cases, and our test is more convenient for practical use.

The rest of the paper is organized as follows. Section 2 presents the basic setting and introduces the DD estimator. Section 3 establishes the fixed-smoothing asymptotics of the t statistic, and Section 4 develops an asymptotically valid t test. Section 5 proposes a data-driven and testing-optimal approach to choose the smoothing parameter K. Section 6 presents a step-by-step summary of our testing procedure and provides some guidance on applying our t test to multi-level data that are quite prevalent in DD regressions. 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

$$Y_{it} = \lambda_t + \tau(t)' \alpha_i + Treat_i \cdot \beta_{10} + Post_t \cdot \beta_{20} + Treat_i \cdot Post_t \cdot \theta_{10} + Z_{it}' \theta_{20} + \epsilon_{it}, \qquad (1)$$

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. The rest elements of $\tau(t)$ take a parametric form such as polynomials. $Treat_i$ 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 $Treat_i = 1\{i \le \mu n\}$ for some $\mu \in (0, 1)$. $Post_t$ is a dummy variable indicating the post-treatment periods. That is, $Post_t = 1\{t \ge \nu T + 1\}$ for some $\nu \in (0, 1)$. For notational convenience, we assume that μn and νT are positive integers[1]. 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 the slope coefficient, we detrend each time series individually. Let

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

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

[1] In an empirical application where the policy intervention takes place at $t^* + 1$, we can set $\nu = t^*/T$. We can set μ similarly.

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

$$\tilde{Y}_{it}^\tau = Y_{it}^\tau - \frac{1}{n}\sum_{j=1}^{n} Y_{jt}^\tau, \tag{2}$$

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

$$\tilde{Y}_{it}^\tau = \widetilde{Treat}_i \cdot Post_t^\tau \cdot \theta_{10} + (\tilde{Z}_{it}^\tau)'\theta_{20} + \tilde{\epsilon}_{it}^\tau. \tag{3}$$

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

The above two transformations remove individual fixed effects, time fixed effects, and individual-specific parametric trends. The order of the two transformations does not matter. We obtain the same equation (3) if we employ cross-sectional demeaning first and then apply individual detrending.

Let

$$X_{it} = \begin{pmatrix} Treat_i \cdot Post_t \\ Z_{it} \end{pmatrix}, \quad \tilde{X}_{it}^\tau = \begin{pmatrix} \widetilde{Treat}_i \cdot Post_t^\tau \\ \tilde{Z}_{it}^\tau \end{pmatrix}, \tag{4}$$

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

$$\hat{\theta} = \left[\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{X}_{it}^\tau (\tilde{X}_{it}^\tau)'\right]^{-1} \left[\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{X}_{it}^\tau \tilde{Y}_{it}^\tau\right]. \tag{5}$$

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 the interactions between individual dummies and 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 can be represented as a difference in two differences.

3 Fixed-Smoothing Asymptotics

To investigate the asymptotic properties of $\hat{\theta}$, we make the following assumption on the trend function.

Assumption 3.1 *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 3.1 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 parametric trend effects, respectively. Note that α_{zi} is a matrix with dimension $d_Z \times d_\tau$. 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

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

where $\bar{\mathcal{Z}}_{\cdot,t} = n^{-1}\sum_{i=1}^{n} \mathcal{Z}_{it}$ is the overall cross-sectional average. We make the following assumptions on $\mathcal{Z}_{it}$.

Assumption 3.2

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

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

Assumption 3.3 $(nT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n} \tilde{\mathcal{Z}}_{it}^{\tau}(\tilde{\mathcal{Z}}_{it}^{\tau})' \to^p G$ *for some positive-definite matrix* G.

Assumption 3.2 is weaker than $\bar{\mathcal{Z}}_{\cdot,t}^{treat} = \bar{\mathcal{Z}}_{\cdot,t}^{control}$ for all t. It 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 3.2 holds. This is similar to the "parallel paths" assumption that is often imposed in a difference-in-differences regression. To make such an assumption more plausible in a DD regression, we may follow a standard empirical practice and redefine the treatment and control groups. We assume that this practice has been followed, if needed, so that Assumption 3.2 holds.

Assumption 3.3 holds if

(i) $T^{-1}\sum_{t=1}^{T}\tilde{\mathcal{Z}}_{it}\tilde{\mathcal{Z}}_{it}' \to^p G$ uniformly over $i = 1, 2, \ldots, n$;

(ii) $T^{-1}\sum_{t=1}^{T}\tilde{\mathcal{Z}}_{it}\tau_D(t)' = o_p(1)$ uniformly over $i = 1, 2, \ldots, n$.

Uniformity over $i = 1, 2, \ldots, n$ can be obtained by using a classical argument. Consider condition (i) as an example. We have

$$\Pr\left(\max_{i\in\{1,2,\ldots,n\}} ||T^{-1}\sum_{t=1}^{T}(\tilde{\mathcal{Z}}_{it}\tilde{\mathcal{Z}}_{it}' - G)|| > \varepsilon\right) \le \sum_{i=1}^{n}\Pr\left(||\sum_{t=1}^{T}(\tilde{\mathcal{Z}}_{it}\tilde{\mathcal{Z}}_{it}' - G)|| > T\varepsilon\right).$$

To evaluate the above upper bound, we use the second part of Lemma 2.1 in Merlevède and Peligrad (2000). We assume that each element of $\tilde{\mathcal{Z}}_{it}\tilde{\mathcal{Z}}_{it}' - G$ is a strictly stationary and strong mixing (α-mixing) sequence. Under some conditions that control the upper-tail quantile function and the strong mixing coefficient uniformly over i, we have

$$E\left(||\sum_{t=1}^{T}(\tilde{\mathcal{Z}}_{it}\tilde{\mathcal{Z}}_{it}' - G)||\right)^4 = O\left(T^2\right) \tag{6}$$

uniformly over $i = 1, 2, \ldots, n$. By the Markov inequality, we have

$$\sum_{i=1}^{n}\Pr\left(||\sum_{t=1}^{T}(\tilde{\mathcal{Z}}_{it}\tilde{\mathcal{Z}}_{it}' - G)|| > T\varepsilon\right) \le O\left(\sum_{i=1}^{n}\frac{T^2}{T^4}\right) = O\left(\frac{n}{T^2}\right).$$

Therefore, condition (i) can hold when $n/T^2 \to 0$. Such a rate condition on n and T accommodates the case that n is larger than T. In fact, using the same arguments as in Doukhan (1994, Sec 1.4.1, Theorem 2 and Remark 2, pp. 25–31), the moment bound in (6) can be strengthened to

$$E\left(||\sum_{t=1}^{T}(\tilde{\mathcal{Z}}_{it}\tilde{\mathcal{Z}}_{it}' - G)||\right)^p = O\left(T^{p/2}\right) \text{ for } p \ge 4.$$

So condition (i) can hold when $n/T^{p/2} \to 0$. When p is large enough, the rate condition $n/T^{p/2} \to 0$ can hold even if n is much larger than T. Our simulation results show that our test performs quite well when n is much larger than T.

To investigate the strength of the signal in $\tilde{X}_{it}^{\tau}$, we write

$$\frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\tilde{X}_{it}^{\tau}(\tilde{X}_{it}^{\tau})' = S := \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \tag{7}$$

where

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

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

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

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 asymptotic properties of S_{11}, S_{21}, and S_{22}.

Lemma 3.1 *Let Assumptions 3.1–3.3 hold. Then*
(*a*) $S_{11} = \mu(1-\mu)\int_0^1 H_{\nu}^2(s)\,ds + O\left(T^{-1}\right),$
(*b*) $S_{21} = o_p(1),$
(*c*) $S_{22} = G + o_p(1).$

Given that $S_{21} = o_p(1)$, Lemma 3.1 shows that the regressor of interest in the serially detrended and cross-sectionally demeaned regression is asymptotically 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 3.2. Without this assumption, S_{21} will not be $o_p(1)$. We leave the case when S_{21} does not vanish asymptotically to future research.

To establish the limiting distributions of $\hat{\theta}$ and the asymptotic variance estimator to be defined later, we maintain the following assumption.

Assumption 3.4 (*a*) $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}\right) \to^d \Lambda B(r)$ *for some* $\Lambda > 0$.
(*b*) $\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i=1}^{n}\tilde{\mathcal{Z}}_{it}^{\tau}\cdot\epsilon_{it} = O_p(1).$

We discuss Assumption 3.4(a) only as similar discussions apply to Assumption 3.4(b).

When n is fixed, Assumption 3.4(a) is a functional central limit theorem (FCLT) for the time series $\{\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}/\sqrt{n}\}$. When n grows with T, $\{\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}/\sqrt{n}\}$ should be regarded as a triangular array, and Assumption 3.4(a) 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 3.4(a) is a high-level assumption. Sufficient conditions for it 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*) $E\left((nT)^{-1}\sum_{t=1}^{[Tr]}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}\right) = 0$ *for all* n *and* T.

(*ii*) *For all* n, $E\left(|n^{-1/2}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}|^{\delta}\right) \leq \Delta < \infty$ *for some* $\delta > 2$.

(*iii*) *For all* n, *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* $\alpha(m) = O\left(m^{-\delta/(\delta-2)+\delta^*}\right)$ *for some* $\delta > 2$ *and* $\delta^* > 0$.

(*iv*) *For all* n, *var*$\left((nT)^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{Treat_i}\cdot\epsilon_{it}\right) > C > 0$ *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

$$
E\left[T^{-1}\sum_{t=1}^{[Tr]}\left(\bar{\epsilon}_{\cdot,t}^{treat}-\bar{\epsilon}_{\cdot,t}^{control}\right)\right]=0. \tag{8}
$$

That is, the condition holds 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.

Note that if the trend function $\tau(t)$ is present but detrending is not applied, then ϵ_{it} effectively contains $\tau(t)'\alpha_i$ as a component. In this case, a component of $\frac{1}{nT}\sum_{t=1}^{[Tr]}\left(\sum_{i=1}^{n}\widetilde{Treat}_i\cdot\epsilon_{it}\right)$ is

$$
\begin{aligned}
&(1-\mu)\,\mu\cdot\frac{1}{T}\sum_{t=1}^{[Tr]}\tau(t)'\left(\frac{1}{n\mu}\sum_{i=1}^{n\mu}\alpha_i-\frac{1}{n\,(1-\mu)}\sum_{i=n\mu+1}^{n}\alpha_i\right)\\
&:=(1-\mu)\,\mu\cdot\frac{1}{T}\sum_{t=1}^{[Tr]}\tau(t)'\cdot\left(\bar{\alpha}^{treat}-\bar{\alpha}^{control}\right).
\end{aligned}
$$

So, unless the average trend effects are the same across the two groups, i.e., $E\left(\bar{\alpha}^{treat}\right)=E\left(\bar{\alpha}^{control}\right)$, Condition 3.1(i) will be violated and the DD estimator will be inconsistent.

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 3.1–3.4 hold. Then*

$$
\sqrt{nT}(\hat{\theta}_1-\theta_{10})\to^d\frac{\Lambda}{\mu\,(1-\mu)}\frac{\int_0^1 H_\nu(r)\,dB(r)}{\int_0^1 H_\nu^2(r)\,dr}\overset{d}{=}\frac{\Lambda}{\mu\,(1-\mu)\sqrt{\int_0^1 H_\nu^2(r)\,dr}}N(0,1). \tag{9}
$$

For Lemma 3.2 to hold, we need Assumption 3.4(a) for only $r=\nu$ and 1 and

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

In this case, (8) needs to hold for only $r = \nu$ and 1. That is, the averages of $\bar{\epsilon}^{treat}_{\cdot,t}$ and $\bar{\epsilon}^{control}_{\cdot,t}$ 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 deterministic trend. We maintain the stronger Assumption 3.4 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 $\epsilon_{it} = Treat_i \times e_t^{(1)} + (1 - Treat_i) \times e_t^{(2)} + \epsilon^*_{it}$ for some sequences $e_t^{(1)}$ and $e_t^{(2)}$ where $\{\epsilon^*_{it}\}$ 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 $n^{-1/2}T^{-1/2}\sum_{t=1}^{[Tr]}(\sum_{i=1}^{n}\widetilde{Treat}_i \cdot \epsilon_{it})$ satisfies an FCLT, we require that $n^{-1}T^{-1/2}\sum_{t=1}^{[Tr]}(\sum_{i=1}^{n}\widetilde{Treat}_i \cdot \epsilon_{it})$ satisfies 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 procedure remains asymptotically valid without any modification.

Lemma 3.2 is similar to what one obtains in a trend regression or cointegrating regression. A common feature of these regressions and the DD regression is that the regressor of interest is a deterministic function whose energy is concentrated at the origin. Because of such energy concentration, the asymptotic variance of the parameter estimator depends only on the long-run variance of the regression-error process. This is in contrast to regressions with a stationary regressor where the asymptotic variance depends on the interaction between the regressor process and the regressor-error process.

Lemma 3.2 shows that

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

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

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

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 inferences about θ_{10}.

Let

$$\hat{\epsilon}^\tau_{it} = \tilde{Y}^\tau_{it} - (\tilde{X}^\tau_{it})'\hat{\theta} \text{ and } \hat{e}_t = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat}_i \cdot \hat{\epsilon}^\tau_{it}.$$

Then Λ^2 can be estimated by

$$\hat{\Lambda}^2 = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}Q_K\left(\frac{t}{T}, \frac{s}{T}\right)\hat{e}_t\hat{e}_s,$$

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 $\hat{\Lambda}^2$, we obtain

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

for

$$\hat{\Lambda}_k = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\hat{e}_t.$$

Thus $\hat{\Lambda}^2$ is a simple average of some "direct" estimators $\hat{\Lambda}_k^2$, 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 discussions of series LRV estimation, see Sun (2013).

The asymptotic variance of $\hat{\theta}_1$ can then be estimated by[2]

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

The corresponding t statistic is

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

To establish the asymptotic distribution of $\mathbb{T}$, we maintain the following assumption on the basis functions $\{\Phi_k(r)\}$.

Assumption 3.5 *The basis functions $\Phi_k(\cdot)$, $k = 1,2,\ldots,K$, are piecewise monotonic and continuously differentiable.*

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

(a)

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

[2] The asymptotic variance can also be estimated using the usual sandwich form that does not exploit the fact that the regressor of interest has energy concentrated at the origin. See Liu and Sun (2007) for a detailed treatment of this case.

jointly with (9), where

$$\Phi_k^{\mathcal{H}}(r) = \Phi_k(r) - (P_H\Phi_k)\cdot H_\nu(r) - \left[\int_0^1 \Phi_k(s)\,\tau(s)'\,ds\right]\left[\int_0^1 \tau(s)\,\tau(s)'\,ds\right]^{-1}\tau(r) \tag{11}$$

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*)

$$\begin{aligned}\mathbb{T} \to^d \mathcal{T}_\infty &:= \frac{\int_0^1 H_\nu(r)\,dB(r)}{\left\{\frac{1}{K}\sum_{k=1}^K\left[\int_0^1 \Phi_k^{\mathcal{H}}(r)\,dB(r)\right]^2\right\}^{1/2}\left(\int_0^1 H_\nu^2(s)\,ds\right)^{1/2}} \\ &\overset{d}{=} \frac{N(0,1)}{\left\{\frac{1}{K}\sum_{k=1}^K\left[\int_0^1 \Phi_k^{\mathcal{H}}(r)\,dB(r)\right]^2\right\}^{1/2}}.\end{aligned} \tag{12}$$

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 $\tau(r)$.

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 asymptotic 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 values 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 without trend and a model with a linear trend, respectively. It is clear that the critical values depend on ν, especially when K is small, 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.

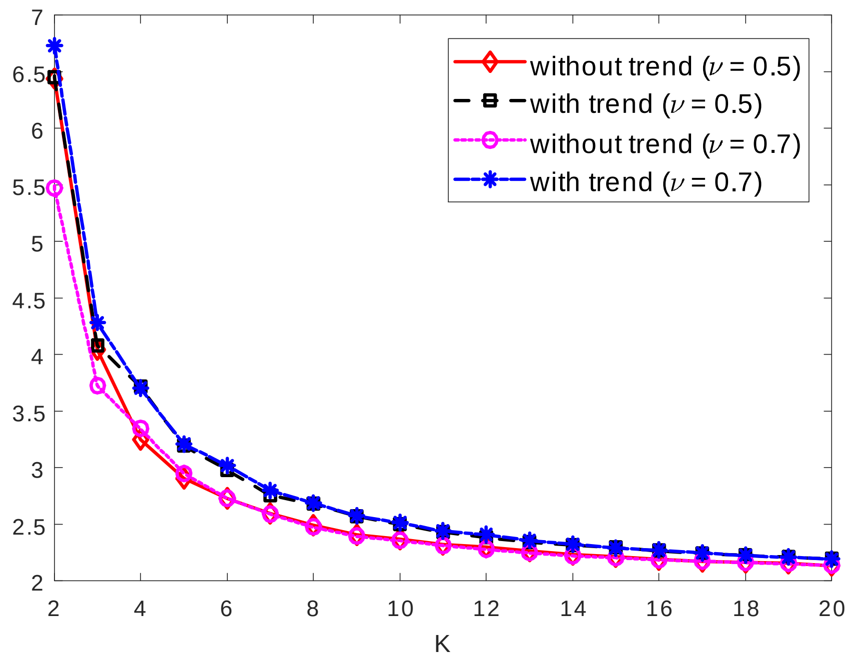

Figure 1: Nonstandard fixed-smoothing critical values for models with and without linear trends and for different values of ν.

4 Asymptotic t Test

4.1 Asymptotic t theory

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

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

and

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

which are all normal. Then

$$\mathcal{T}_\infty = \frac{\eta_0}{\left(\frac{1}{K}\sum_{k=1}^K \eta_k^2\right)^{1/2}}. \tag{13}$$

Since $\int_0^1 H_\nu(r)\,\Phi_k^{\mathcal{H}}(r)\,dr = 0$, we have

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

For normal random variables, zero covariance implies independence. So η_0 and η_k are independent for $k = 1, 2, \ldots, K$. If $\eta_k \sim iid\ N(0,1)$ for $k = 1, 2, \ldots, K$, then $\mathcal{T}_\infty$ follows the standard t distribution with K degrees of freedom.

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

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

where

$$C_\nu^{\mathcal{H}}(r,s) = \delta(r-s) - \frac{H_\nu(r)\,H_\nu(s)}{\int_0^1 H_\nu^2(t)\,dt} - \tau(r)' \left[\int_0^1 \tau(t)\,\tau(t)'\,dt\right]^{-1} \tau(s) \tag{14}$$

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

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

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

$$\int_0^1 \int_0^1 \Phi_{k_1}(r)\, C_\nu^{\mathcal{H}}(r,s)\, \Phi_{k_2}(s)\,drds = 1\{k_1 = k_2\} \text{ for } k_1, k_2 = 1, \ldots, K. \tag{15}$$

Instead of searching for the basis functions that satisfy (15), 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 $\mathbf{C}_{\mathcal{H}}$ be the $T \times T$ matrix whose (i,j)-th element is equal to

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

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

$$\langle \ell_1, \ell_2 \rangle = \ell_1' \mathbf{C}_{\mathcal{H}} \ell_2 / T^2, \tag{16}$$

which makes $\mathbb{R}^T$ a Hilbert space. The discrete analogue of (15) is

$$\langle \mathbf{\Phi}_{k_1}, \mathbf{\Phi}_{k_2} \rangle = 1\{k_1 = k_2\} \text{ for } k_1, k_2 = 1, \ldots, K. \tag{17}$$

Note that (17) is different from the usual orthonormality in the Euclidean sense. In general, the basis vectors $\{\mathbf{\Phi}_k\}$ do not satisfy (17) 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

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

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

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

then $\{\mathbf{\Phi}_{1,\mathcal{H}}, \ldots, \mathbf{\Phi}_{K,\mathcal{H}}\}$ is a set of bases in $\mathbb{R}^T$ that satisfies the conditions in (17).

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

$$\mathbf{\Phi}_{\mathcal{H}} = \mathbf{\Phi} \left(R_{\mathcal{H}} \right)^{-1}.$$

For such a choice of $\mathbf{\Phi}_{\mathcal{H}}$, we have

$$(\mathbf{\Phi}_{\mathcal{H}})'\mathbf{C}_{\mathcal{H}}\mathbf{\Phi}_{\mathcal{H}}/T^2 = \left(R'_{\mathcal{H}}\right)^{-1} \mathbf{\Phi}'\mathbf{C}_{\mathcal{H}}\mathbf{\Phi} \left(R_{\mathcal{H}}\right)^{-1} / T^2 = \left(R'_{\mathcal{H}}\right)^{-1} R'_{\mathcal{H}} R_{\mathcal{H}} \left(R_{\mathcal{H}}\right)^{-1} = I_K,$$

so the conditions in (17) are satisfied.

As $T \to \infty$, $\mathbf{\Phi}'\mathbf{C}_{\mathcal{H}}\mathbf{\Phi}/T^2$ converges to the variance Σ_η of $\eta = (\eta_1, \ldots, \eta_K)'$. This implies that $R_{\mathcal{H}}$ converges to the upper triangular factor of the Cholesky decomposition of Σ_η. As a result, every transformed basis vector is approximately equal to a linear combination of the original basis vectors. The implied basis functions are thus equal to linear combinations of the original basis functions. Therefore, if Assumption 3.5 holds for the original basis functions, it also holds for the transformed basis functions.

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

$$\mathcal{T}_\infty =^d t_K.$$

That is, the t statistic $\mathbb{T}$ constructed based on the transformed basis functions is asymptotically distributed as the standard t distribution with K degrees of freedom.

4.2 Understanding the asymptotic t test

To understand the asymptotic t theory, we abstract away nonessential details in a DD regression and consider the time series regression

$$\mathbb{Y}_t = \tau(t)' \alpha + Post_t \times \beta + e_t, \; t = 1, 2, \ldots, T \tag{18}$$

with β as the parameter of interest, where $\mathbb{Y}_t$ is the outcome of interest and e_t satisfies the FCLT $T^{-1/2}\sum_{t=1}^{[Tr]} e_t \rightarrow^d \Lambda B(r)$. After partialling out the trend component, the time series model becomes $\mathbb{Y}_t^\tau = Post_t^\tau \times \beta + e_t^\tau$. The OLS estimator of β is then

$$\hat{\beta} = \left[\sum_{t=1}^{T} (Post_t^\tau)^2\right]^{-1} \sum_{t=1}^{T} (Post_t^\tau \cdot \mathbb{Y}_t^\tau),$$

and the t statistic for testing the null $\beta = \beta_0$ is

$$\mathbb{T}_\beta = \frac{\sqrt{T}(\hat{\beta} - \beta_0)}{\sqrt{\left[T^{-1}\sum_{t=1}^{T}(Post_t^\tau)^2\right]^{-1} \hat{\Lambda}^2}} = \frac{\sqrt{T}(\hat{\beta} - \beta_0)}{\sqrt{\left[T^{-1}\sum_{t=1}^{T}(Post_t^\tau)^2\right]^{-1} \Lambda^2}} \cdot \sqrt{\frac{\Lambda^2}{\hat{\Lambda}^2}} \tag{19}$$

where

$$\hat{\Lambda}^2 = \frac{1}{K}\sum_{k=1}^{K}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right) \cdot \hat{e}_t^\tau\right]^2 \tag{20}$$

and $\hat{e}_t^\tau = \mathbb{Y}_t^\tau - Post_t^\tau \cdot \hat{\beta}$.

It is not difficult to show that the fixed-K limit distribution of $\mathbb{T}_\beta$ is identical to what is given in (13). Therefore, the time series regression in (18) and the DD regression can be regarded as asymptotically equivalent. To develop the asymptotic t theory for $\mathbb{T}_\beta$, we have to use the Gram-Schmidt orthogonalization as in the DD regression. The reason is that the basis functions $\{\Phi_k^{\mathcal{H}}(r)\}$ are not orthonormal on $L^2[0,1]$ and so η_k is not iid, even if the original basis functions $\{\Phi_k(r)\}$ are. In fact, for the asymptotic t theory, we do not care whether $\{\Phi_k(r)\}$ are orthonormal *per se*. What we care about is whether $\{\Phi_k^{\mathcal{H}}(r)\}$ are orthonormal.

There are some special cases where the Gram-Schmidt orthogonalization is not needed. The first example is Sun (2011) who considers a linear trend regression, which is a special case of the model in (18) and can be obtained by setting $\tau(t) = (1,t)'$ and dropping the regressor $Post_t$. It is not hard to show that the limiting distribution of the t statistic for the linear-trend coefficient can still be represented by (13), but now

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

If we employ the special cosine bases $\{\Phi_k(r) := \sqrt{2}\cos(\pi k r), k = 0,1,\ldots\}$, then $\int_0^1 \Phi_k(s)\tau(s)\,ds = 0$ with $\tau(s) = (1,s)'$ and hence $\Phi_k^{\mathcal{H}}(r) = \Phi_k(r)$. So $\eta_k = \int_0^1 \Phi_k^{\mathcal{H}}(r)\,dB(r) = \int_0^1 \Phi_k(r)\,dB(r)$. Given that these cosine bases are orthonormal on $L^2[0,1]$, we have $\eta_k \sim iid\ N(0,1)$ for $k = 1,2,\ldots,K$. Asymptotic t theory can be then developed without applying the Gram-Schmidt orthogonalization to the special cosine bases. Such a theory in Sun (2011) takes advantage of the nature of the regressor, i.e., the linear trend, and the special property of the carefully crafted cosine bases. If the regression contains other deterministic regressors such as higher-order polynomial trends and the regressor $Post_t$, an asymptotic t theory cannot be developed without employing the Gram-Schmidt orthogonalization

The second example is Sun (2013) who develops the asymptotic t theory in the GMM framework for stationary data. To understand the asymptotic t theory there, we consider the simplest location model, which can serve as the limit experiment of more general models. The location model can be cast as the regression model in (18) with the special regressor $\tau(t) = 1$ but without the regressor $Post_t$. Again the representation in (13) is still valid for the limiting distribution of the t statistic. Simple calculations show that now

$$\Phi_k^{\mathcal{H}}(r) = \Phi_k(r) - \int_0^1 \Phi_k(s)\,ds.$$

If the basis functions $\{\Phi_k(\cdot)\}$ are orthonormal on $L^2[0,1]$ and satisfy $\int_0^1 \Phi_k(s)\,ds = 0$, then $\Phi_k^{\mathcal{H}}(r) = \Phi_k(r)$ and $\eta_k = \int_0^1 \Phi_k^{\mathcal{H}}(r)\,dB(r) = \int_0^1 \Phi_k(r)\,dB(r) \sim iid\ N(0,1)$. To develop the asymptotic t theory, we only need to maintain that $\{\Phi_k(\cdot)\}$ are orthonormal and satisfy $\int_0^1 \Phi_k(s)\,ds = 0$. The commonly used Fourier bases meet these requirements. As in the linear trend regression, we do not need to employ the Gram-Schmidt orthogonalization in order to develop the asymptotic t theory.

The idea of using the Gram-Schmidt orthogonalization to develop the asymptotic t approximations (and F approximations) in series HAR inference is quite general. It can be readily extended to regressions with other types of deterministic regressors. It can also be extended to the cases where the standard FCLT $T^{-1/2}\sum_{t=1}^{[Tr]} e_t \to^d \Lambda B(r)$ does not hold and the limiting process may not have independent increments. For example, if e_t is a near unit root process with local-to-unity parameter c, then we could have: $T^{-1/2} e_{[Tr]} \to^d \Lambda J_c(r)$ where $J_c(r)$ is an OU process whose increments are not independent of each other in general. In such cases, the asymptotic t theory can still be developed, but more sophisticated orthonormalizations are needed. See Sun (2014c) for a study in this direction.

5 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 the 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

$$M_\tau \mathcal{Y}_t = M_\tau \cdot Post_t \cdot \sqrt{n}\mu(1-\mu)\theta_{10} + M_\tau e_t, \tag{21}$$

where

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

$$e_t = \sqrt{n}\mu\,(1-\mu)\left(\frac{1}{n\mu}\sum_{i=1}^{n\mu} \epsilon_{it} - \frac{1}{n\,(1-\mu)}\sum_{i=n\mu+1}^{n} \epsilon_{it}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot \epsilon_{it}, \tag{22}$$

$M_\tau = I_{T\times T} - \boldsymbol{\tau}\left(\boldsymbol{\tau}\boldsymbol{\tau}'\right)^{-1}\boldsymbol{\tau}'$, and $I_{T\times T}$ is a $T\times T$ dimensional identity matrix. 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

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

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

Denote the variance matrix of $e = (e_1, e_2, \ldots, e_T)'$ by Ω. On the basis of (21), we can also estimate θ_1 by the generalized least-squares estimator:

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

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 $E(\hat{\theta}_{1,\text{GLS}} - \theta_{10})(\hat{\theta}_{1,\text{GLS}} - \hat{\theta}_{1,\text{OLS}}) = 0$. In addition, letting

$$\hat{e}^\tau = [I_{T\times T} - M_\tau\cdot Post\cdot\left(Post'\cdot M_\tau\cdot Post\right)^{-1} Post'\cdot M_\tau]M_\tau e$$

be the OLS residual, we can show that

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

Hence $\hat{\theta}_{1,\text{GLS}} - \theta_{10}$ is independent of both $\hat{\theta}_{1,\text{GLS}} - \hat{\theta}_{1,\text{OLS}}$ and $\hat{e}^\tau$. Using the definition of e_t given in (22), we can show that $\hat{e}_t^\tau = \sum_{i=1}^n \widetilde{Treat_i}\cdot\hat{\epsilon}_{it}^\tau/\sqrt{n}$. It then follows that

$$\hat{\sigma}^2 = \frac{1}{K}\sum_{k=1}^{K}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{\Phi}_{k,\mathcal{H},t}\hat{e}_t^\tau\right]^2 [\mu\,(1-\mu)]^{-2}\left\{\frac{1}{T}\sum_{t=1}^{T}[Post_t^\tau]^2\right\}^{-1}, \tag{23}$$

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

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

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

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

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

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

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

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

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 convergence rate of $E[\sqrt{nT}(\hat{\theta}_{1,\text{GLS}}-\hat{\theta}_{1,\text{OLS}})]^2$. To this end, we maintain the following assumption.

Assumption 5.1 *(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 $\Phi_F^{\mathcal{H}}(r) = \left[\Phi_1^{\mathcal{H}}(r), \ldots, \Phi_K^{\mathcal{H}}(r)\right]'$, 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 $\Phi_F(r) = [\Phi_1(r), \ldots, \Phi_K(r)]'$, $\dot{\Phi}_F(i) = \left[\dot{\Phi}_1(r), \ldots, \dot{\Phi}_K(r)\right]$, 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 5.1 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 5.1 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 $\mathbb{T}$.

Theorem 5.1 *Let Assumptions 3.1 and 5.1 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 $\mathbb{T}$ satisfies

$$P(|\mathbb{T}| > t_K^{\alpha/2} | H_0) = \alpha - \frac{K^2 \bar{B}}{T^2} G'(\chi_1^\alpha)\chi_1^\alpha + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right). \tag{24}$$

(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 $\mathbb{T}$ satisfies

$$\begin{aligned} P(|\mathbb{T}| < t_K^{\alpha/2} | H_1(\delta^2)) &= G_{\delta^2}(\chi_1^\alpha) + \frac{K^2\bar{B}}{T^2} G'_{\delta^2}(\chi_1^\alpha)\chi_1^\alpha \\ &\quad + \frac{\delta^2}{2K} G'_{3,\delta^2}(\chi_1^\alpha)\chi_1^\alpha + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right), \end{aligned} \tag{25}$$

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 $j = 1, \ldots, K/2$. 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 here.

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

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

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:

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

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 $e_{\mathrm{I}} \leq \kappa\alpha$. 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

$$K_{\mathrm{opt}} = \left\{ \frac{\delta^2 G'_{3,\delta^2}(\chi_1^\alpha)}{4\bar{B}\left[G'_{\delta^2}(\chi_1^\alpha) - \lambda_{\mathrm{opt}} G'(\chi_1^\alpha)\right]} \right\}^{1/3} T^{2/3}, \tag{26}$$

where

$$\lambda_{\mathrm{opt}} = \begin{cases} 0, & \text{if } \bar{B} > 0 \\ \frac{G'_{\delta^2}(\chi_1^\alpha)}{G'(\chi_1^\alpha)} + \delta^2 \frac{|\bar{B}|^{1/2} G'_{3,\delta^2}(\chi_1^\alpha)[\chi_1^\alpha]^{3/2}[G'(\chi_1^\alpha)]^{1/2}}{4[(\kappa-1)\alpha]^{3/2} T}, & \text{if } \bar{B} \leq 0. \end{cases} \tag{27}$$

The optimal K_{opt} in (26) 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 $\hat{e}_t = \sum_{i=1}^n \widetilde{Treat}_i \cdot \hat{\epsilon}_{it}^\tau / \sqrt{n}$. 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

$$\hat{K}_{\mathrm{opt}} = \begin{cases} \left(\frac{(1-\hat{\rho}_e)^2}{8\omega^{(2)}(0)|\hat{\rho}_e|}\right)^{1/3} \left(\frac{G'_{3,\delta^2}(\chi_1^\alpha)\delta^2}{G'_{\delta^2}(\chi_1^\alpha)}\right)^{1/3} T^{2/3}, & \text{if } \bar{B}^{est} > 0 \\ \left(\frac{(1-\hat{\rho}_e)^2}{2\omega^{(2)}(0)|\hat{\rho}_e|}\right)^{1/2} \left(\frac{(\kappa-1)\alpha}{G'(\chi_1^\alpha)\chi_1^\alpha}\right)^{1/2} T, & \text{if } \bar{B}^{est} \leq 0. \end{cases} \tag{28}$$

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.

6 Testing Procedure and Practical Guidance

6.1 Summary of the Proposed t Test

Our asymptotic t test consists of the following steps:

1. Construct model (1) and estimate the parameter of interest.

   (a) Detrend each time series separately, and then remove the cross-sectional average from each detrended variable as described in (2).

   (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 (2) and (4), respectively. Denote the estimates by $\hat{\theta}_1$ and $\hat{\theta}_2$ and the residual by $\hat{\epsilon}_{it}^{\tau}$.

2. Transform the original basis vectors.

   (a) Let $\boldsymbol{\tau} = (\tau(1), \ldots, \tau(T))' \in \mathbb{R}^{T \times d_\tau}$ and $Post^{\tau} = (Post_1^{\tau}, \ldots, Post_T^{\tau})' \in \mathbb{R}^{T \times 1}$, where $Post_t^{\tau}$ is the detrended "$Post_t$" dummy:

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

   Construct the projection matrix

   $$\mathbf{C}_{\mathcal{H}} = T\left[I_{T\times T} - Post^{\tau} \cdot \left[(Post^{\tau})'\,Post^{\tau}\right]^{-1}(Post^{\tau})' - \boldsymbol{\tau}\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}'\right] := T \cdot M_{Post,\tau}. \tag{29}$$

   (b) Let $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_K) \in \mathbb{R}^{T\times K}$ be the matrix of the original basis vectors, where K is the greatest even number less than $\hat{K}_{\text{opt}}$ given by equation (28). The columns of $\boldsymbol{\Phi}$ are

   $$\boldsymbol{\Phi}_{2j-1} = \left(\sqrt{2}\cos(2j\pi\cdot 1/T), \sqrt{2}\cos(2j\pi\cdot 2/T), \ldots, \sqrt{2}\cos(2j\pi\cdot T/T)\right)', \tag{30}$$

   $$\boldsymbol{\Phi}_{2j} = \left(\sqrt{2}\sin(2j\pi\cdot 1/T), \sqrt{2}\sin(2j\pi\cdot 2/T), \ldots, \sqrt{2}\sin(2j\pi\cdot T/T)\right)', \tag{31}$$

   for $j = 1, 2, \ldots, K/2$.
   Compute the upper triangular factor $R_{\mathcal{H}}$ of the Cholesky decomposition of $\boldsymbol{\Phi}'\mathbf{C}_{\mathcal{H}}\boldsymbol{\Phi}/T^2$ such that $\boldsymbol{\Phi}'\mathbf{C}_{\mathcal{H}}\boldsymbol{\Phi}/T^2 = R_{\mathcal{H}}'R_{\mathcal{H}}$.

   (c) Compute the matrix

   $$\boldsymbol{\Phi}_{\mathcal{H}} = (\boldsymbol{\Phi}_{1,\mathcal{H}}, \ldots, \boldsymbol{\Phi}_{K,\mathcal{H}}) = \boldsymbol{\Phi}\,(R_{\mathcal{H}})^{-1},$$

   where each column of $\boldsymbol{\Phi}_{\mathcal{H}}$ consists of a transformed basis vector.

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

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

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

where

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

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

6.2 Practical Guidance for Multi-level Data

In empirical applications, DD analyses are often applied to multi-level data that consist of individuals from different groups such as states and regions and each individual is observed over a number of periods. The DD regression at the individual level is

$$\begin{aligned} Y_{ig_{(i)}t} &= \lambda_t + \tau(t)'\alpha_{g_{(i)}} + Treat_{g_{(i)}}\cdot\beta_{10} + Post_t\cdot\beta_{20} \\ &\quad + Treat_{g_{(i)}}\cdot Post_t\cdot\theta_{10} + Z'_{ig_{(i)}t}\theta_{20,g_{(i)}t} + \epsilon_{g_{(i)}t} + \epsilon_{ig_{(i)}t}, \end{aligned} \tag{32}$$

where the new subscript $g_{(i)}$ indexes the group that individual i belongs to and $\epsilon_{g_{(i)}t}$ is an additional error component capturing unobserved group/time effects. The model can be rewritten in a two-levels form:

$$Y_{ig_{(i)}t} = Y_{g_{(i)}t} + Z'_{ig_{(i)}t}\theta_{20,g_{(i)}t} + \epsilon_{ig_{(i)}t} \tag{33}$$

and

$$Y_{gt} = \lambda_t + \tau(t)'\alpha_g + Treat_g\cdot\beta_{10} + Post_t\cdot\beta_{20} + Treat_g\cdot Post_t\cdot\theta_{10} + \epsilon_{gt}. \tag{34}$$

The first equation is for the data at the individual level, and the second equation is for the aggregate data at the group level. If we observe Y_{gt} and formally change the index g into i, then the aggregate model in (34) is exactly the same as the model we consider in (1). So our proposed test can be directly applied[3].

The problem is that we do not observe Y_{gt} and have to estimate it. To this end, we can first use the individual-level data for each (g,t) pair and run the OLS regression in (33) to obtain an estimator $\hat{\theta}_{20,gt}$ of $\theta_{20,gt}$ and then estimate Y_{gt} by

$$\hat{Y}_{gt} = \frac{1}{N_{gt}}\sum_{i:g_{(i)}=g}\left(Y_{ig_{(i)}t} - Z'_{ig_{(i)}t}\hat{\theta}_{20,g_{(i)}t}\right),$$

where N_{gt} is the number of individuals in the (g,t) pair. If N_{gt} is reasonably large in the sense that it is much larger than the number of groups and the number of time periods, we can safely ignore the estimation error in $\hat{\theta}_{20,gt}$. In this case, we can proceed as if $\hat{Y}_{gt}$ is the same as Y_{gt}. Our asymptotic t theory continues to hold, and we can follow the testing procedure in the previous subsection to perform the asymptotic t test.

[3] The only exception is that there are additional covariates Z_{it} in (1). In principle, we can add a group and time specific component $Z'_{g_{(i)}t}\theta_{20}$ to the multi-level model in (32). Such a component is then present in the aggregate model in (34). Nevertheless, in empirical applications, such a component is often not included in the multi-level model.

7 Simulation Evidence

7.1 Model-based Simulation

We consider the following data generating process

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

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

$$\epsilon_{it} = \rho \epsilon_{it-1} + e_{it}^{\epsilon},\ t \geq 1 \text{ and } \epsilon_{i0} = 0.$$

While $\{e_{it}^{\epsilon}\}$ is iid over time, there may be 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

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

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

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

where $v_{i_1,i_2,t}$ is iid $N(0,1)$ across i_1, i_2, and t. That is, $e_{it}^{\epsilon} \sim SMA(2)$, 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 $\left(3^2, 10\right)$, $\left(3^2, 100\right)$, $\left(8^2, 10\right)$, $\left(8^2, 50\right)$, $(8^2, 100)$, and $(8^2, 200)$.

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 consider the following tests: the nonstandard fixed-K test based on the sine and cosine basis functions and simulated critical values, the standard (fixed-K) t test as described in Section 4, and the nonstandard fixed-b test developed in SY (2017). For the former two tests, we also consider the corresponding normal tests that employ standard normal critical values. There are five types of tests in total.

The test statistic for the fixed-b test is

$$t_b = \frac{\sqrt{nT}(\hat{\theta}_1 - \theta_{10})}{\sqrt{R'\hat{V}_b R}},$$

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

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

$$\hat{\Omega}_b = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} k\left(\frac{|t-s|}{bT} \right) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{X}_{it}^{\tau} \hat{\epsilon}_{it}^{\tau} \right) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{X}_{is}^{\tau} \hat{\epsilon}_{is}^{\tau} \right),$$

$k(\cdot)$ is a kernel function, and $b \in (0, 1]$ is a smoothing parameter. We employ the same kernel function, namely, the Bartlett kernel, as in SY (2017). Since no data-driven method for choosing b is given in SY (2017), we consider the fixed-b test with $b = 0.01, 0.5$, and 1.

For the standard fixed-smoothing t test, we use the data-driven $\hat{K}_{\text{opt}}$ given in (28), but we make two adjustments. First, we use the truncated LS estimator

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

instead of the original estimator $\hat{\rho}_e$ in computing $\hat{K}_{\text{opt}}$. Second, we truncate $\hat{K}_{\text{opt}}$ to be between 4 and $T/2$, and we round it to the greatest even number less than $\hat{K}_{\text{opt}}$. Therefore, K is always equal to 4 when $T = 10$. Rounding is used to speed up the computation. It has a minimal effect on test performances and is not necessary in practical implementation. We impose the lower bound $\hat{K}_{\text{opt}} \geq 4$ because 4 is the smallest even degree of freedom for Student's t distribution to have a finite variance. We impose the lower bound to avoid extreme power loss. We set κ to be 1.3 in our testing-oriented criterion for choosing K. The size of our proposed t test does not change much when we consider $\kappa = 1.1., 1.2\ , \ldots, 1.5$.

For the nonstandard fixed-K test, we choose K to minimize the mean squared error of the long-run variance estimator. The long-run variance under consideration is the long-run variance Λ^2 of the process $e_t := \sum_{i=1}^{n} \widetilde{Treat}_i \cdot \epsilon_{it}/\sqrt{n}$. When the Fourier basis functions are used, the MSE optimal K is given by

$$K = \left\lfloor \left(\frac{\Lambda^2}{2B^2} \right)^{1/5} T^{4/5} \right\rfloor,$$

where $\lfloor \cdot \rfloor$ is the floor function. See Phillips (2005). Following a common practice, 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 $\hat{e}_t = \sum_{i=1}^{n} \widetilde{Treat}_i \cdot \hat{\epsilon}_{it}^{\tau}/\sqrt{n}$. Let $\hat{\rho}_e$ be the estimated AR coefficient and $\hat{\sigma}_e^2$ be the estimated error variance, then the plug-in estimator of K is

$$\hat{K}_{\text{mse}} = \left\lfloor \left(\frac{9(1-\hat{\rho}_e)^6}{2\pi^4 \hat{\sigma}_e^2 \hat{\rho}_e^2} \right)^{1/4} T^{4/5} \right\rfloor. \tag{35}$$

We also truncate $\hat{K}$ to be between 4 and $T/2$, and round it to the greatest even number less than $\hat{K}_{\text{mse}}$.

Tables 1–3 report simulation results of the five different types of tests when $(n, T) = (3^2, 10)$, $(n, T) = (8^2, 10)$, and $(n, T) = (8^2, 100)$, respectively. First, it's clear that the standard fixed-smoothing t test has quite accurate size in all cases when $T = 100$ and acceptable size when $T = 10$. Second, the fixed-K tests are more accurate than the corresponding standard normal tests in almost all cases, especially when the AR parameter is positive and T is small. In those cases, the data-driven K values are relatively small, and the estimation uncertainty in the asymptotic variance estimator becomes large. This is exactly the scenario where the fixed-K approximations can be more accurate. Third, the asymptotic t test outperforms the nonstandard fixed-K test in almost all cases. The reason is that the asymptotic t test employs the test-optimal K that controls the type I error and minimizes the type II error asymptotically while the nonstandard fixed-K test employs the MSE-optimal K that is not targeted at the type I and II errors. Fourth, comparing the standard t test to the nonstandard fixed-b test, we find that both tests under-reject when ρ is negative and over-reject when ρ are positive. However, the standard t test performs much better than the fixed-b test when $\rho \geq 0$ and $T = 10$. Fifth, comparing the results when $\phi = 0$ with those when $\phi = 0.5$, we can see that the cross-sectional dependence does not affect the size properties of any of the five tests: there is no big difference between the empirical rejection probabilities in the two cases, especially when $T = 100$. Finally, 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 the test performances. Comparing the results in Tables 1 and 2, we find 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 simulation result that cross-sectional dependence does not affect the size properties of all five tests. In essence, each test involves collapsing 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 all five tests. On the other hand, when there is substantial temporal dependence, all five tests become more accurate as T increases as shown in Tables 2 and 3.

Figures 2 and 3 present the size-adjusted power for the five tests when $(n, T) = (3^2, 10)$ and $(n, T) = (8^2, 100)$, respectively. Note that the fixed-K test and the corresponding asymptotic normal test have the same size-adjusted power, as they use the same test statistic. We only report the case with a linear trend and without 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 tests have more or less the same size-adjusted power function. This, coupled with its size accuracy and convenience to use, suggests that we use the proposed t test in empirical applications.

7.2 Empirical-data-based Simulation

In this section, we apply our proposed tests to women's wages constructed from the Current Population Survey (CPS). Following BDM (2004), we extract the variables — weekly earnings (from their fourth interview month), employment status, education, age (between 25 and 50), and state of residence from the Merged Outgoing Rotation Group of the CPS from 1979 to 1999. We use the sample with positive reported weekly earnings with a sample size around 540,000.

The log weekly earnings, denoted as Y_{ist}, are the outcomes of interest for individual i in state s at year t. We employ the following linear model to study the treatment effect of a hypothetical policy intervention:

$$Y_{ist} = \alpha_s + \lambda_t + I_{st}\beta + Z_{ist}'c + \epsilon_{ist}, \tag{36}$$

where α_s and λ_t are state and year fixed effects, Z_{ist} contains individual-level covariates, I_{st} is the policy indicator with $I_{st} = Treat_s \times Post_t$, and ϵ_{ist} is the error term.

We follow BDM (2004) to randomly generate the pseudo-sample and the intervention. We treat the state as the sampling unit and draw an *iid* sample of 50 states with replacement from the real empirical data. Individuals within the same state are either all drawn into the pseudo-sample or none of them is drawn into the pseudo-sample. We then randomly select 25 states, i.e., half of the states in the pseudo-sample and designate them "affected" by the intervention. We designate the rest "unaffected" by the intervention. Finally, we randomly draw a year, say $t^* + 1$, between 1985 to 1995 as the year when the policy intervention takes place. $I_{st} = 1$ only for the treatment states and for the years after year t^*. We perform each candidate test on the pseudo-sample and record the outcome of each test. Repeating the whole process a number of times, we obtain the relative rejection frequency of each test.

Our simulation design mimics the following hypothetical scenario: Hundreds of researchers obtain a simple random sample of 50 states and each analyzes the effects of various laws in the CPS independently. We expect 5% of the researchers to reject the null of no effect if the laws indeed have no effect and each researcher uses an accurate 5% test.

By design, each pseudo-sample retains time series dynamics including the temporal dependence and the trend effect, if any, in individual time series. For individuals in the same state, cross-sectional dependence is also retained. To a great extent, each pseudo-sample represents cross-sectional and time-series dependence in the CPS data. Note that there is no cross-sectional dependence for individuals in different states. This is an empirically plausible assumption, which could also be restrictive in other empirical applications.

Table 4 reports the rejection frequencies for seven different tests, including the tests considered by BDM (2004) and the tests proposed here. The tests are based on different point estimators of β (or different ways of estimating β) and different variance estimators.

The first two tests are based on individual-level data. The first test (t_1), reported in Column 1, is based on the OLS estimator of β in model (36). For the standard error estimation, the test assumes that ϵ_{ist} is *iid* across i, s, t. The test, therefore, does not account for cross-sectional dependence, time series dependence or conditional heteroscedasticity. For this reason, the test is expected to have a large size distortion. This is supported by the simulation result in Table 4. The second test (t_2), reported in Column 2, is based on the same OLS estimator as test t_1 but employs a cluster-robust variance estimator that allows for arbitrary correlation among errors in the same state-year cell. This test accounts for cross-sectional dependence within each state and conditional heteroscedasticity but not time series dependence. As expected, Table 4 shows that the second test is still quite size distorted but less so than the first test.

The rest five tests are based on the aggregate data: we aggregate the data into state-year cells using the same procedure discussed in Subsection 6.2. As in BDM (2004), we assume that the effect of Z_{ist} is a constant. So we can pool all data at the individual level to estimate c, and the estimation error in estimating c can be safely ignored. The model based on the aggregate data is

$$Y_{st} = \alpha_s + \lambda_t + \beta I_{st} + \epsilon_{st}. \tag{37}$$

Let

$$Y_{st}^{\tau} = Y_{st} - \frac{1}{T}\sum_{t=1}^{T} Y_{st},\ \tilde{Y}_{st}^{\tau} = Y_{st}^{\tau} - \frac{1}{S}\sum_{s=1}^{S} Y_{st}^{\tau},$$

where $S = 50$ is the number of states. Define ϵ_{st}^{τ} similarly. Then

$$\tilde{Y}_{st}^{\tau} = \widetilde{Treat}_s \times Post_t^{\tau} \times \beta + \tilde{\epsilon}_{st}^{\tau}.$$

Assume that the pseudo-sample has been sorted so that the first $S/2$ states receive the treatment. The DD estimator $\hat{\beta}_{DD}$ of β satisfies

$$\sqrt{ST}\left(\hat{\beta}_{DD}-\beta\right)=\left[\frac{1}{ST}\sum_{s,t}\left(\widetilde{Treat_s\times Post_t^{\tau}}\right)^2\right]^{-1}\left(\frac{1}{\sqrt{ST}}\sum_{s,t}\widetilde{Treat_s Post_t^{\tau}}\times\tilde{\epsilon}_{st}^{\tau}\right). \tag{38}$$

Depending on how the standard error of the DD estimator is constructed, we obtain different t statistics and tests. The test in Column 3 (t_3) is based on the OLS standard error under the assumption that ϵ_{st} is *iid* across s and t. Since the test ignores the autocorrelation in each time series $\{\epsilon_{st}, t=1,2,\ldots,T\}$, the test does not have accurate size. Table 4 shows that the null rejection probability is 51%, which is much larger than the nominal level of 5%. The test in Column 4 (t_4) assumes that each time series $\{\epsilon_{st}, t=1,2,\ldots,T\}$ follows an independent AR(1) process with possibly different AR parameters. Such a test can be reliable if the AR(1) model is correctly specified. Even if the model is not correctly specified, the test based on t_4 should be more reliable than that based on t_3, as the time series dependence is partially accounted for. The simulation results in Table 4 support this observation.

The test in Column 5 (t_5) computes the standard error under the assumption that ϵ_{st_1} and ϵ_{st_2} are correlated for all t_1 and t_2 and the covariance between ϵ_{st_1} and ϵ_{st_2} is the same for different states. Under this assumption, $cov(\epsilon_{st_1},\,\epsilon_{st_2})$ is estimated by $S^{-1}\sum_{s=1}^{S}\hat{\epsilon}_{st_1}^{\tau}\hat{\epsilon}_{st_2}^{\tau}$ where $\hat{\epsilon}_{st_1}^{\tau}=\tilde{Y}_{st}^{\tau}-\widetilde{Treat_s}\times Post_t^{\tau}\times\hat{\beta}_{DD}$. The test in Column 6 ($t_6$) is based on a different estimation procedure. For each state, we first define and compute two time series averages:

$$\bar{Y}_s^1=\frac{1}{t^*}\sum_{t=1}^{t^*}Y_{st}\text{ and }\bar{Y}_s^2=\frac{1}{T-t^*}\sum_{t=t^*+1}^{T}Y_{st}.$$

Then

$$\bar{Y}_s^2-\bar{Y}_s^1=\bar{\lambda}^2-\bar{\lambda}^1+Treat_s\times\beta+\bar{\epsilon}_s^2-\bar{\epsilon}_s^1,\ s=1,\ldots,S, \tag{39}$$

where $\bar{\lambda}^j$ and $\bar{\epsilon}_s^j$ are defined similarly as $\bar{Y}_s^j$ for $j=1,2$. The OLS estimator of β based on the above cross-sectional regression is the DD estimator. That is, the cross-section OLS estimator is numerically identical to $\hat{\beta}_{DD}$ given in (38). However, the above formulation allows us to employ the usual heteroscedasticity-robust variance estimator, which is the asymptotic variance estimator used in Column 6. While the test in Column 5 assumes that the variances and autocovariances of ϵ_{st} are the same across all s, the test in Column 6 allows them to be different across the treatment states and the control states.

Both tests t_5 and t_6 allow for rich enough error correlation, although both tests assume that the states are independent of each other, an assumption that holds in each pseudo sample. Also, the two tests employ cross-sectional averages to estimate autocovariances, and so they are expected to perform well when S is large. In our simulation, $S=50$ may be regarded as reasonably large. Not surprisingly, Table 4 shows that both t_5 and t_6 are quite accurate and that t_6 is the most trustworthy test among tests t_1-t_6.

The tests in Columns 7 and 8 (t_{LS}^0 and t_{LS}^1) are our proposed t tests with data-driven testing-optimal K. As in our model-based simulations, we set κ to be 1.3. The difference between these two tests lies in the trend functions used. While test t_{LS}^0 assumes that no trend is included in the DD regression, test t_{LS}^1 assumes that a linear trend is included with possibly different trend coefficients for different individuals. Table 4 shows that test t_{LS}^1 is as accurate as test t_6. This is encouraging, as t_6 takes advantage of cross-sectional independence of the states but t_{LS}^1 does not. Also, it is encouraging to know that t_{LS}^1 works well when the time series are relatively short — each time series has only 21 observations.

Table 4 shows that test t_{LS}^0 is not as accurate as tests $t_4 - t_6$, although it is more accurate than tests $t_1 - t_3$. The reason for the inaccuracy of t_{LS}^0 is the trend misspecification. As we discussed before, each pseudo-sample retains the time series trends if they are present in the original empirical sample. Note that the dependent variable is the logarithm of weekly earnings. It is empirically plausible that the dependent variable contains a linear trend so that the weekly earnings increase at a constant rate for each individual, even after controlling for some individual-level covariates. A model that is more compatible with this empirical situation is the model in (32) where a trending component is included. Test t_{LS}^0 ignores the linear trend and thus suffers from some size distortion. This explanation is supported by our model-based simulation results not reported here.

A natural question is why tests t_5 and t_6 are not affected by trend misspecification. The reason is that both tests ignore the time series variation and use only time series averages in constructing the asymptotic variance estimator. Consider test t_6 as an example. If a trend function is allowed so that ϵ_{st} becomes

$$\epsilon_{st} = \tau(t)' \tilde{\alpha}_s + \epsilon_{st,-\tau},$$

where $\epsilon_{st,-\tau}$ contains no trend. Then the cross-sectional regression in (39) becomes

$$\bar{Y}_s^2 - \bar{Y}_s^1 = \underbrace{\bar{\lambda}^2 - \bar{\lambda}^1 + \tau_\Delta' E\tilde{\alpha}_s}_{intercept} + Treat_s \times \beta + \underbrace{\tau_\Delta' (\tilde{\alpha}_s - E\tilde{\alpha}_s) + \bar{\epsilon}_{s,-\tau}^2 - \bar{\epsilon}_{s,-\tau}^1}_{error} \tag{40}$$

where

$$\tau_\Delta = \frac{1}{T - t^*} \sum_{t=t^*+1}^{T} \tau(t) - \frac{1}{t^*} \sum_{t=1}^{t^*} \tau(t)$$

is a constant,

$$\bar{\epsilon}_{s,-\tau}^1 = \frac{1}{t^*} \sum_{t=1}^{t^*} \epsilon_{st,-\tau}, \text{ and } \bar{\epsilon}_{s,-\tau}^2 = \frac{1}{T - t^*} \sum_{t=t^*+1}^{T} \epsilon_{st,-\tau}.$$

The random trend effects[4] have been absorbed into the intercept and the error term. If the states are independent of each other, then it is clear that the OLS estimator of β and the associated heteroscedasticity-robust variance estimator are consistent. This explains why in our simulation experiments t_5 and t_6 have good size properties. The cost of achieving the invariance to the trend effects is that the point estimator of β is only $\sqrt{n}$ consistent when trends are present, even if T is large. The relative efficiency loss can be huge. Also, it is useful to reiterate that t_5 and t_6 have satisfactory size performances only when the trend effects are random, and the groups/states are independent. Both conditions hold in our simulation experiment. However, these two conditions are unlikely to hold in empirical applications.

We have also considered the power of the tests. A true $0.0x$ log point effect is generated by replacing y_{ist} by $y_{ist} + I_{st} * 0.0x$ if state s is chosen to be in the treatment group and $t \geq t^* + 1$. The second row of Table 4 reports the results for $x = 2$, which corresponds to an effect of 2%. Our proposed test t_{LS}^1 is more powerful than test t_6, even though the null rejection probability of t_{LS}^1 is the same as that of t_6. As we discussed above, by pushing the random trend effects into the error term, the point estimator behind t_6 is less efficient than the point estimator behind t_{LS}^1,

[4] In the pseudo-sample, the trend effects are random, as they are not correlated with the treatment dummy, the key regressor of interest.

which includes the random trend as part of the regression and hence reduces the size of the error term.

8 Conclusion

This paper develops an asymptotically valid t test in the DD regression when T is relatively large. The t test employs standard t critical values and is thus easy to use. It is more accurate than the normal test that employs standard normal critical values. The proposed t test has competitive power properties. The cross-sectional sample size n can be fixed or grow with T. Simulations show that the proposed t test works well even when n is comparable to T. Given these attractive properties, we recommend using our proposed t test in place of the normal test 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 test. This extension is straightforward. Second, while the paper considers only panel data, it is easy to see that the proposed t test 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.

Table 1: Empirical size of different 5% tests in DD regression with sample size $n = 9, T = 10$, and data-driven choice of K.

	Orthonormal series test				Kernel fixed-b test		
	Transformed Bases		Fourier Bases		b-value		
	t	$\mathcal{N}$	$\mathcal{T}_\infty$	$\mathcal{N}$	0.02	0.5	1
			$\tau = 1, \quad \phi = 0$				
$\rho = -0.6$	0.015	0.075	0.060	0.141	0.015	0.014	0.013
$\rho = -0.3$	0.023	0.094	0.066	0.144	0.043	0.031	0.029
$\rho = 0$	0.034	0.122	0.080	0.169	0.102	0.063	0.060
$\rho = 0.3$	0.054	0.176	0.119	0.222	0.204	0.118	0.112
$\rho = 0.6$	0.099	0.263	0.185	0.307	0.348	0.201	0.191
$\rho = 0.9$	0.174	0.387	0.279	0.424	0.490	0.290	0.282
			$\tau(t) = 1, \quad \phi = 0.5$				
$\rho = -0.6$	0.018	0.077	0.062	0.141	0.016	0.016	0.013
$\rho = -0.3$	0.024	0.098	0.067	0.149	0.045	0.034	0.033
$\rho = 0$	0.031	0.123	0.082	0.173	0.104	0.065	0.064
$\rho = 0.3$	0.048	0.170	0.117	0.218	0.202	0.120	0.114
$\rho = 0.6$	0.083	0.243	0.173	0.291	0.323	0.191	0.176
$\rho = 0.9$	0.166	0.378	0.268	0.413	0.484	0.277	0.268
			$\tau(t) = (1, t)', \quad \phi = 0$				
$\rho = -0.6$	0.018	0.088	0.083	0.228	0.080	0.040	0.038
$\rho = -0.3$	0.029	0.108	0.088	0.223	0.127	0.060	0.057
$\rho = 0$	0.036	0.127	0.095	0.220	0.168	0.079	0.075
$\rho = 0.3$	0.039	0.134	0.099	0.223	0.193	0.089	0.087
$\rho = 0.6$	0.036	0.136	0.100	0.227	0.216	0.091	0.090
$\rho = 0.9$	0.030	0.122	0.088	0.207	0.207	0.078	0.078
			$\tau = (1, t)', \quad \phi = 0.5$				
$\rho = -0.6$	0.021	0.097	0.089	0.249	0.083	0.042	0.039
$\rho = -0.3$	0.031	0.116	0.095	0.232	0.129	0.065	0.065
$\rho = 0$	0.036	0.128	0.096	0.226	0.170	0.080	0.077
$\rho = 0.3$	0.037	0.132	0.097	0.226	0.199	0.088	0.087
$\rho = 0.6$	0.035	0.130	0.094	0.222	0.213	0.086	0.085
$\rho = 0.9$	0.028	0.119	0.085	0.203	0.205	0.075	0.077

Note: The t test, denoted by "t", is based on t critical values. The normal tests, denoted by "$\mathcal{N}$", are based on standard normal critical values. The nonstandard fixed-K test, denoted by "$\mathcal{T}_\infty$", is based on simulated nonstandard critical values. K is chosen to be 4 for orthonormal series tests.

Table 2: Empirical size of different 5% tests in DD regression with sample size $n = 64, T = 10$, and data-driven choice of K.

	Orthonormal series test				Kernel fixed-b test		
	Transformed Bases		Fourier Bases		b-value		
	t	$\mathcal{N}$	$\mathcal{T}_\infty$	$\mathcal{N}$	0.02	0.5	1
			$\tau = 1, \quad \phi = 0$				
$\rho = -0.6$	0.018	0.075	0.063	0.139	0.016	0.017	0.016
$\rho = -0.3$	0.026	0.096	0.069	0.143	0.043	0.034	0.033
$\rho = 0$	0.034	0.120	0.081	0.165	0.100	0.063	0.064
$\rho = 0.3$	0.052	0.172	0.116	0.221	0.199	0.115	0.111
$\rho = 0.6$	0.094	0.261	0.184	0.309	0.345	0.201	0.193
$\rho = 0.9$	0.170	0.376	0.272	0.411	0.486	0.279	0.275
			$\tau(t) = 1, \quad \phi = 0.5$				
$\rho = -0.6$	0.017	0.079	0.065	0.148	0.017	0.014	0.012
$\rho = -0.3$	0.025	0.100	0.070	0.152	0.044	0.034	0.032
$\rho = 0$	0.035	0.126	0.085	0.171	0.105	0.065	0.063
$\rho = 0.3$	0.051	0.174	0.117	0.219	0.200	0.117	0.115
$\rho = 0.6$	0.086	0.246	0.171	0.300	0.330	0.189	0.178
$\rho = 0.9$	0.159	0.373	0.264	0.411	0.485	0.277	0.269
			$\tau(t) = (1, t)', \quad \phi = 0$				
$\rho = -0.6$	0.020	0.089	0.085	0.232	0.077	0.042	0.040
$\rho = -0.3$	0.031	0.111	0.093	0.222	0.124	0.059	0.060
$\rho = 0$	0.036	0.125	0.096	0.222	0.166	0.078	0.077
$\rho = 0.3$	0.037	0.129	0.096	0.221	0.193	0.086	0.084
$\rho = 0.6$	0.032	0.127	0.087	0.219	0.204	0.080	0.079
$\rho = 0.9$	0.026	0.119	0.082	0.199	0.200	0.074	0.074
			$\tau = (1, t)', \quad \phi = 0.5$				
$\rho = -0.6$	0.023	0.094	0.093	0.247	0.079	0.040	0.041
$\rho = -0.3$	0.031	0.117	0.099	0.230	0.127	0.060	0.061
$\rho = 0$	0.038	0.127	0.101	0.229	0.163	0.081	0.082
$\rho = 0.3$	0.040	0.136	0.100	0.230	0.199	0.091	0.091
$\rho = 0.6$	0.036	0.136	0.097	0.222	0.211	0.090	0.090
$\rho = 0.9$	0.030	0.120	0.086	0.199	0.198	0.074	0.072

Note: The t test, denoted by "t", is based on t critical values. The normal tests, denoted by "$\mathcal{N}$", are based on standard normal critical values. The nonstandard fixed-K test, denoted by "$\mathcal{T}_\infty$", is based on simulated nonstandard critical values. K is chosen to be 4 for orthonormal series tests.

Table 3: Empirical size of different 5% tests in DD regression with sample size $n = 64, T = 100$, and data-driven choice of K.

	Orthonormal series test				Kernel fixed-b test				
	Transformed Bases		Fourier Bases		b-value			$\bar{K}_{\text{mse}}$	$\bar{K}_{\text{test}}$
	t_1	$\mathcal{N}_1$	$\mathcal{T}_{1\infty}$	$\mathcal{N}_1$	0.02	0.5	1		
				$\tau = 1, \quad \phi = 0$					
$\rho = -0.6$	0.036	0.053	0.042	0.053	0.014	0.032	0.032	37.85	26.02
$\rho = -0.3$	0.038	0.055	0.042	0.051	0.027	0.042	0.041	42.54	28.59
$\rho = 0$	0.049	0.061	0.052	0.062	0.050	0.048	0.048	48.60	43.88
$\rho = 0.3$	0.062	0.081	0.067	0.083	0.102	0.054	0.056	28.74	29.73
$\rho = 0.6$	0.055	0.101	0.074	0.109	0.207	0.070	0.070	13.45	12.22
$\rho = 0.9$	0.068	0.192	0.120	0.232	0.538	0.161	0.150	5.04	4.52
				$\tau(t) = 1, \quad \phi = 0.5$					
$\rho = -0.6$	0.033	0.052	0.038	0.049	0.012	0.030	0.030	37.84	26.02
$\rho = -0.3$	0.035	0.054	0.038	0.048	0.022	0.038	0.038	42.45	28.55
$\rho = 0$	0.044	0.059	0.049	0.058	0.048	0.045	0.045	48.565	43.59
$\rho = 0.3$	0.057	0.079	0.063	0.082	0.098	0.055	0.054	29.01	30.07
$\rho = 0.6$	0.052	0.104	0.072	0.112	0.218	0.070	0.068	13.57	12.35
$\rho = 0.6$	0.063	0.204	0.120	0.239	0.544	0.164	0.157	5.07	4.54
				$\tau(t) = (1, t)', \quad \phi = 0$					
$\rho = -0.6$	0.035	0.055	0.049	0.063	0.013	0.035	0.033	37.82	26.02
$\rho = -0.3$	0.035	0.054	0.042	0.053	0.028	0.044	0.043	42.33	28.43
$\rho = 0$	0.043	0.056	0.048	0.057	0.052	0.050	0.049	48.71	43.37
$\rho = 0.3$	0.055	0.074	0.061	0.079	0.098	0.057	0.056	29.83	30.96
$\rho = 0.6$	0.045	0.088	0.060	0.102	0.205	0.067	0.068	14.12	12.93
$\rho = 0.9$	0.029	0.119	0.069	0.173	0.438	0.091	0.094	5.59	4.88
				$\tau = (1, t)', \quad \phi = 0.5$					
$\rho = -0.6$	0.037	0.058	0.051	0.065	0.017	0.037	0.036	37.81	26.02
$\rho = -0.3$	0.039	0.058	0.046	0.058	0.030	0.047	0.046	42.25	28.40
$\rho = 0$	0.047	0.062	0.053	0.065	0.057	0.055	0.053	48.69	43.10
$\rho = 0.3$	0.062	0.081	0.066	0.087	0.111	0.062	0.061	30.09	31.31
$\rho = 0.6$	0.047	0.095	0.066	0.110	0.208	0.073	0.071	14.25	13.06
$\rho = 0.9$	0.032	0.126	0.073	0.186	0.439	0.099	0.096	5.62	4.92

Note: The t test, denoted by "t", is based on t critical values. The normal tests, denoted by "$\mathcal{N}$", are based on standard normal critical values. The nonstandard fixed-K test, denoted by "$\mathcal{T}_\infty$", is based on simulated nonstandard critical values. $\bar{K}_{\text{mse}}$ and $\bar{K}_{\text{opt}}$ are the averages of the MSE-optimal K and the test-optimal K developed in this paper.

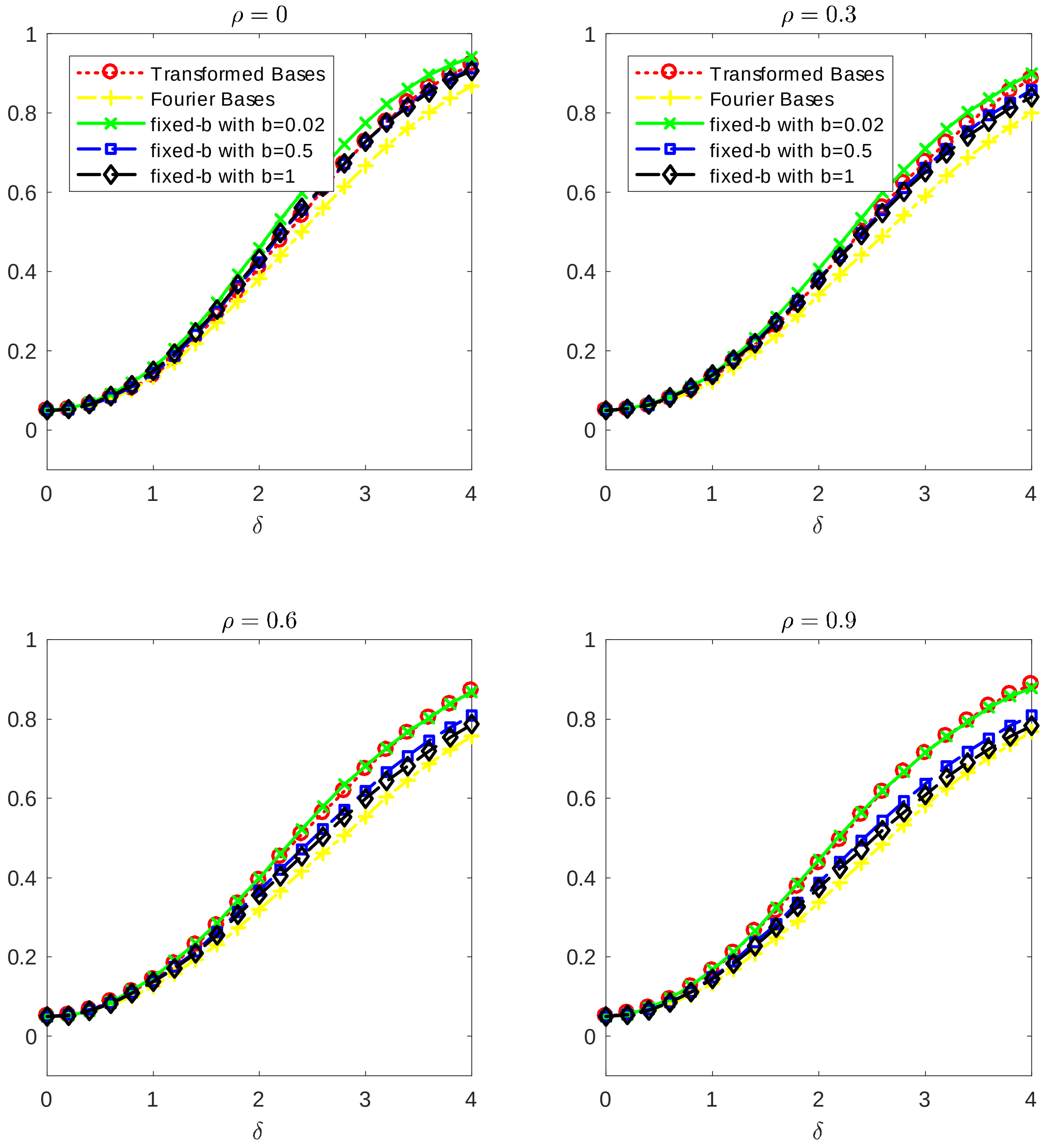

Figure 2: Size-adjusted power of the tests with different variance estimators and basis functions for $n = 9$, $T = 10$ in the presence of linear trends but no cross sectional dependence.

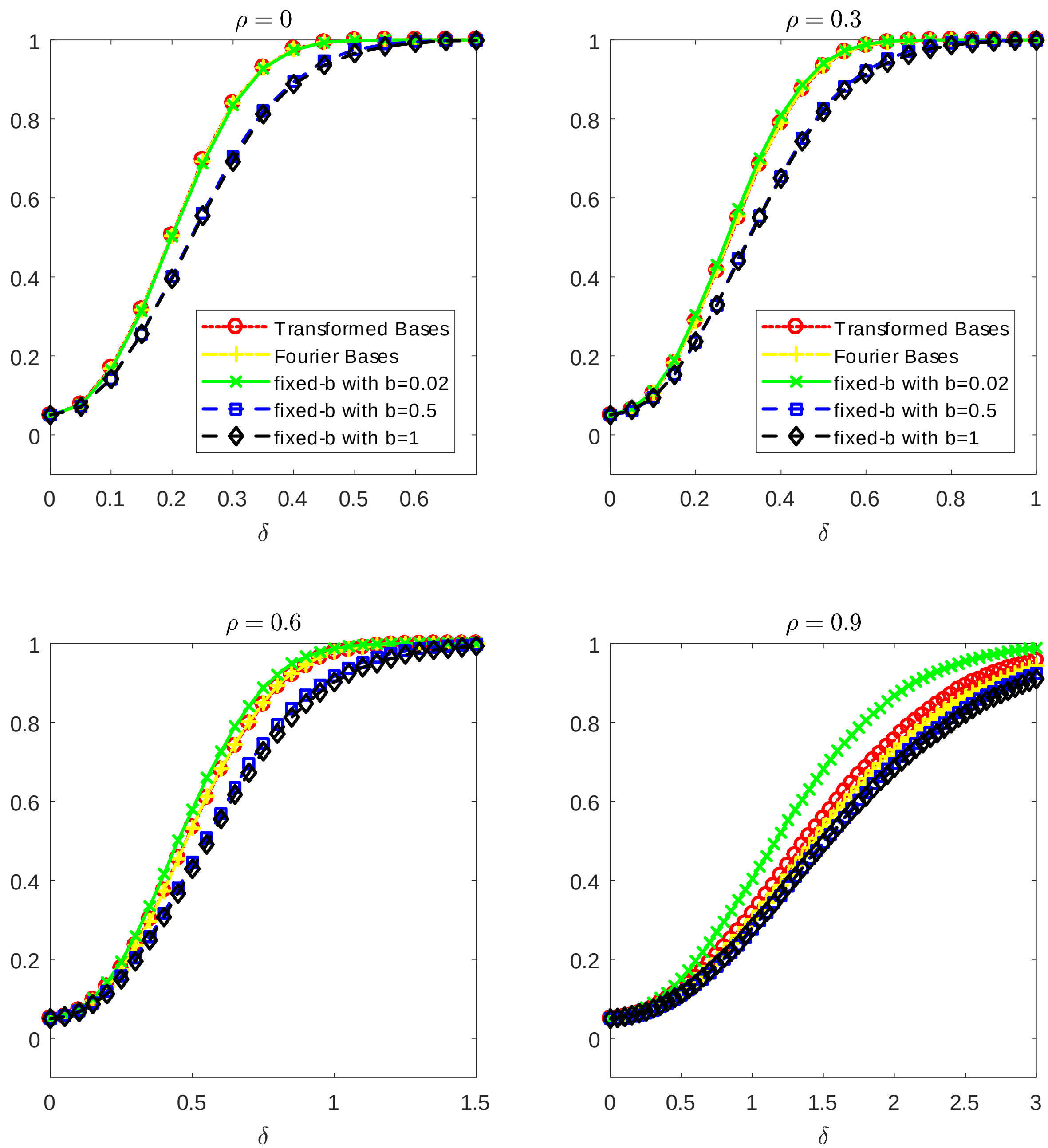

Figure 3: Size-adjusted power of the tests with different variance estimators and basis functions for $n = 64$, $T = 100$ in the presence of linear trends but no cross sectional dependence.

Table 4: Rejection rates of different 5% tests based on CPS data for years 1979-1999.

	t_1	t_2	t_3	t_4	t_5	t_6	t_{LS}^0	t_{LS}^1
No effect	0.665	0.400	0.455	0.215	0.070	0.055	0.290	0.055
2% effect	0.830	0.730	0.660	0.510	0.350	0.285	0.535	0.455

Note: Tests t_1 and t_2 are based on the individual-level data. The other tests are based on the aggregate data with 50 pseudo-states and 21 years. Tests t_{LS}^0 and t_{LS}^1 are our proposed t tests when the DD regression contains no trend and a linear trend, respectively.

9 Appendix of Proofs

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

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

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

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

where the last line follows from some simple calculations.

Note that

$$
\begin{aligned}
Z_{it}^{\tau} &= Z_{it} - \left(\sum_{s=1}^{T} Z_{is}\tau(s)'\right)\left(\sum_{s=1}^{T}\tau(s)\tau(s)'\right)^{-1}\tau(t) \\
&= \lambda_{zt} + \alpha_{zi}\tau(t) + \mathcal{Z}_{it} \\
&- \left(\sum_{s=1}^{T}\lambda_{zs}\tau(s)'\right)\left(\sum_{s=1}^{T}\tau(s)\tau(s)'\right)^{-1}\tau(t) \\
&- \left(\sum_{s=1}^{T}\alpha_{zi}\tau(s)\tau(s)'\right)\left(\sum_{s=1}^{T}\tau(s)\tau(s)'\right)^{-1}\tau(t) \\
&- \left(\sum_{s=1}^{T}\mathcal{Z}_{is}\tau(s)'\right)\left(\sum_{s=1}^{T}\tau(s)\tau(s)'\right)^{-1}\tau(t) = \lambda_{zt}^{\tau} + \mathcal{Z}_{it}^{\tau}
\end{aligned}
\tag{41}
$$

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

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

where

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

In the above expression for S_{21}, 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

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

where we have used

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

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 $T^{-1}\sum_{t=1}^{T}Post_t^{\tau}\cdot(\bar{\mathcal{Z}}_{\cdot,t}^{control})^{\tau}$. It then follows from Assumption 3.2 that $S_{21}=o_p\left(1\right)$.

Part (c). Using (41) and Assumption 3.3, we have

$$
S_{22}=\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\tilde{Z}_{it}^{\tau}(\tilde{Z}_{it}^{\tau})'=\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathcal{Z}}_{it}^{\tau}\left(\tilde{\mathcal{Z}}_{it}^{\tau}\right)'=G+o_p\left(1\right).
$$

■

Proof of Lemma 3.2. Using Lemma 3.1 and Assumption 3.4(b), we have

$$
\begin{aligned}
&\sqrt{nT}\left(\hat{\theta}_1-\theta_{10}\right) &(42)\\
&=\left[\frac{1}{nT}\sum_{i,t}(\widetilde{Treat_i})^2\cdot(Post_t^{\tau})^2\right]^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Post_t^{\tau}\cdot\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}^{\tau}\right)+o_p(1)\\
&=\frac{1}{\mu(1-\mu)\int_0^1 H_\nu^2(r)\,dr}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Post_t^{\tau}\cdot\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}^{\tau}\right]+o_p(1).
\end{aligned}
$$

By Assumption 3.4(a), we have

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

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. ■

Proof of Theorem 3.1. (a) We have

$$
\begin{aligned}
\hat{\Lambda}_k&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\hat{\epsilon}_{it}^{\tau}\\
&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}^{\tau}\\
&-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\begin{pmatrix}\widetilde{Treat_i}\cdot Post_t^{\tau}\\ \tilde{Z}_{it}^{\tau}\end{pmatrix}'\left(\hat{\theta}-\theta_0\right). \qquad (43)
\end{aligned}
$$

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 (43) satisfies

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}^{\tau}$$
$$=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\Phi_k\left(\frac{t}{T}\right)\right]^{\tau}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\epsilon_{it}\to^{d}\Lambda\int\Phi_k^{\tau}(r)\,dB(r)\,. \tag{44}$$

The second term in (43) satisfies

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot\begin{pmatrix}\widetilde{Treat_i}\cdot Post_t^{\tau}\\ \tilde{Z}_{it}^{\tau}\end{pmatrix}'\left(\hat{\theta}-\theta_0\right)$$
$$=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\widetilde{Treat_i})^2 Post_t^{\tau}\left(\hat{\theta}_1-\theta_{10}\right)$$
$$+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Treat_i}\cdot(\tilde{Z}_{it}^{\tau})'\left(\hat{\theta}_2-\theta_{20}\right)$$
$$=\left[\frac{1}{n}\sum_{i=1}^{n}(\widetilde{Treat_i})^2\right]\left[\frac{1}{T}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)Post_t^{\tau}\right]\sqrt{nT}\left(\hat{\theta}_1-\theta_{10}\right)$$
$$+\frac{1}{nT}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\sum_{i=1}^{n}\widetilde{Treat_i}\cdot(\tilde{\mathcal{Z}}_{it}^{\tau})'\sqrt{nT}\left(\hat{\theta}_2-\theta_{20}\right):=I_1+I_2. \tag{45}$$

Under Assumptions 3.1–3.4, we can show that $\sqrt{nT}\left(\hat{\theta}_2-\theta_{20}\right)=O_p(1)$. To obtain an upper bound for I_2, we have, using Assumption 3.2:

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

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

Let $S^*(r) = \frac{1}{nT}\sum_{t=1}^{[Tr]}\sum_{i=1}^{n}\widetilde{Treat}_i \cdot (\tilde{\mathcal{Z}}_{it})'$, then we have

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\frac{1}{n}\sum_{i=1}^{n}\widetilde{Treat}_i \cdot (\tilde{\mathcal{Z}}_{it})' \\
&= \sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\left[S^*\left(\frac{t}{T}\right) - S^*\left(\frac{t-1}{T}\right)\right] \\
&= \sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)S^*\left(\frac{t}{T}\right) - \sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)S^*\left(\frac{t-1}{T}\right) \\
&= \sum_{t=1}^{T-1}\left[\Phi_k\left(\frac{t}{T}\right) - \Phi_k\left(\frac{t+1}{T}\right)\right]S^*\left(\frac{t}{T}\right) + \Phi_k(1)\, S^*(1).
\end{aligned}$$

Under the piecewise monotonicity condition in Assumption 3.5, we have, for some finite κ, we can partition the set $\{1, 2, \ldots, T-1\}$ into $\tilde{\kappa}$ maximal non-overlapping subsets $\cup_{j=1}^{\tilde{\kappa}}\mathcal{I}_j$ such that $\Phi_k(t/T)$ is monotonic on each $\mathcal{I}_j := \{\mathcal{I}_{jL}, \ldots, \mathcal{I}_{jU}\}$. Now

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

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 $[\mathcal{I}_{jL}, \mathcal{I}_{jU}]$. Therefore, 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 (43) 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 (44) 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

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

■

It follows from (23) 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 5.1, we need to first prove the following lemma, which establishes the asymptotic bias and variance of $\hat{\Lambda}^2$.

Lemma 9.1 *Let Assumptions 3.1 and 5.1 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 9.1. We prove (i) only, as (ii) follows from standard arguments, e.g., Theorem 9 in Hannan (1970, p. 280). By definition, we have

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

where $\mathbf{C}_{\mathcal{H}}$ is defined in (29),

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

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 (29).

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

$$\Phi_k^*(r) = \sum_{j=1}^{K}\left[\int_0^1 C_\nu^{\mathcal{H}}(r,s)\,\Phi_j(s)\,ds\right]R_\infty^{(j,k)} = \sum_{j=1}^{K}\Phi_j^{\mathcal{H}}(r)\,R_\infty^{(j,k)}, \tag{46}$$

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 (46) follows from simple calculations using the definition of $C_\nu^{\mathcal{H}}(r,s)$ given in (14).

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 5.1(c), $\pi_k(r)$ is twice continuously differentiable. Under Assumption 5.1(b), the coefficients $\{c_k\}$ satisfy

$$\begin{aligned}\sum_{k=1}^K |c_k|^2 &= \sum_{k=1}^K\sum_{j_1=1}^K\sum_{j_2=1}^K \tilde{c}_{j_1}\tilde{c}_{j_2}R_\infty^{(j_1,k)}R_\infty^{(j_2,k)} = \sum_{j_1=1}^K\sum_{j_2=1}^K \tilde{c}_{j_1}\tilde{c}_{j_2}\sum_{k=1}^K R_\infty^{(j_1,k)}R_\infty^{(j_2,k)}\\ &= \tilde{c}'\left\{\int_0^1 \Phi_F^{\mathcal{H}}(r)\,\Phi_F^{\mathcal{H}}(r)'\,dr\right\}^{-1}\tilde{c}\\ &= O(\|\tilde{c}\|^2) = O\left(\sum_{k=1}^K|\tilde{c}_k|^2\right),\end{aligned} \tag{47}$$

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

$$\begin{aligned}
\sum_{k=1}^{K} |\tilde{c}_k|^2 = \sum_{k=1}^{K} |P_H \Phi_k|^2 &= \sum_{k=1}^{K} \left[\int_0^1 \Phi_k(r) H_\nu(r)\, dr\right]^2 \left[\int_0^1 H_\nu^2(s)\, ds\right]^{-2} \\
&= O(1) \sum_{k=1}^{K} \left(\int_0^1 \Phi_k(r)^2\, dr\right) \left(\int_0^1 H_\nu^2(s)\, ds\right)^{-1} \\
&= O\left(\int_0^1 \sum_{k=1}^{K} \Phi_k(r)^2\, dr\right) = O\left(\int_0^1 \|\Phi_F(r)\|_2^2\, dr\right), \qquad (48)
\end{aligned}$$

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

$$\sum_{k=1}^{K} \|d_k\|^2 = O\left(\sum_{k=1}^{K} ||\tilde{d}_k||^2\right) \text{ and } \sum_{k=1}^{K} ||\tilde{d}_k||^2 = O\left(\int_0^1 \|\Phi_F(r)\|_2^2\, dr\right). \qquad (49)$$

Now

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

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 $k = 1, \ldots, K, \omega_k(\varsigma)$ 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 5.1(a), we have

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

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

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

where

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

Therefore,

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

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 (49) and Assumption 5.1 (d), we have

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

Similarly,

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

Therefore,

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

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 $\varsigma \to 0+$ we have

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

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

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

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

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

and so

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

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

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

Using (47), (48), and Assumption 5.1(d), we have

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

and by the same argument as in (47) we have

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

where we have used (49). 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 5.1(d), we have

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

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

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

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

and so

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

Combining the above results, we can conclude that

$$
E(\hat{\Lambda}^2-\Lambda^2)=-\left(\frac{K}{T}\right)^2\omega^{(2)}(0)\sum_{p=-\infty}^{\infty}|p|^2\,\sigma_{e,p}^2+o\left(\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right), \tag{50}
$$

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* $j=1,\ldots,K/2$ *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

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

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,

$$\left|\left[\int_0^1 m(r)\,\Phi_k(r)\,dr\right]\right| \le \frac{C}{k} \tag{51}$$

for some constant C.

Let $\tilde{\vartheta} = (\tilde{\vartheta}_1, \ldots, \tilde{\vartheta}_K)' \in \mathbb{R}^{K\times(d_\tau+1)}$. 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 (51), we have $(\tilde{\vartheta}_k^*)'\tilde{\vartheta}_k^* \le C/k^2$.

It then follows that

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

■

Proof of Theorem 5.1 . Part (a). We first establish a moment bound for $\hat{\sigma}^2/\sigma^2_{\text{GLS}}$. Under Assumption 5.1(a), we have $\sqrt{nT}E[(\hat{\theta}_{1,\text{GLS}} - \hat{\theta}_{1,\text{OLS}})]^2 = O(1/T)$, and so

$$\sigma^2_{\text{GLS}} = \Lambda^2 \left[\mu\left(1-\mu\right)\right]^{-2} \left\{\frac{1}{T}\sum_{t=1}^{T}[Post_t^{\tau}]^2\right\}^{-1} + O\left(\frac{1}{T}\right). \tag{52}$$

Using Lemma 9.1, 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 (52) and (50), 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

$$\begin{aligned}
&P\left(\left|\frac{\sqrt{nT}(\hat{\theta}_{1,\text{OLS}} - \theta_{10})}{\hat{\sigma}}\right| > t_K^{\alpha/2}\right) \\
&= 1 - G((t_K^{\alpha/2})^2) - \frac{K^2\bar{B}}{T^2}G'((t_K^{\alpha/2})^2)(t_K^{\alpha/2})^2 \\
&- \frac{1}{K}G''((t_K^{\alpha/2})^2)(t_K^{\alpha/2})^4 + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right).
\end{aligned} \tag{53}$$

On the other hand, we have

$$(t_K^{\alpha/2})^2 = \chi_1^{\alpha} - \frac{1}{K}\frac{G''(\chi_1^{\alpha})}{G'(\chi_1^{\alpha})}(\chi_1^{\alpha})^2 + o\left(\frac{1}{K}\right). \tag{54}$$

See equation (14) in Sun (2011). Combining (53) and (54) yields

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

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

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

Therefore, we have

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

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] Atchadé, Y. F. and Cattaneo, M. D. (2014): "A martingale decomposition for quadratic forms of Markov chains (with applications)." *Stochastic Processes and their Application* 124, 646-677.

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

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

[5] Doukhan, P. (1994): *Mixing: Properties and Examples*. New York: Springer-Verlag.

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

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

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

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

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

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

[12] 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.

[13] 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.

[14] Liu, C. and Sun, Y. (2017): "Simple and Trustworthy Asymptotic t Tests in Difference-in-Differences Regressions", Working paper, Department of Economics, UC San Diego.

[15] Merlevède, F. and Peligrad, M. (2000): "The Functional Central Limit Theorem under the Strong Mixing Condition." *The Annals of Probability* 28(3), 1336–1352.

[16] Müller, U. K. (2007): "A Theory of Robust Long-run Variance Estimation." *Journal of Econometrics* 141(2), 1331–1352.

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

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

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

[20] 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.

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

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

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

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

[25] 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.

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

[27] 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

[28] Sun, Yu (2017): "Fixed-b Inference for Difference-in-Differences Estimation." Working paper, Capital University of Economics and Business, China.

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