cemmap

centre for microdata methods and practice

Inference on a semiparametric model with global power law and local nonparametric trends

Jiti Gao
Oliver Linton
Bin Peng

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP05/18

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Inference on a Semiparametric Model with Global Power Law and Local Nonparametric Trends[1]

Jiti Gao*, Oliver Linton⋆[2] and Bin Peng†

*Monash University, ⋆University of Cambridge,
and †University of Bath

July, 2017

Abstract

This paper studies a model with both a parametric global trend and a nonparametric local trend. This model may be of interest in a number of applications in economics, finance, ecology, and geology. The model nests the parametric global trend model considered in Phillips (2007) and Robinson (2012), and the nonparametric local trend model. We first propose two hypothesis tests to detect whether either of the special cases are appropriate. For the case where both null hypotheses are rejected, we propose an estimation method to capture both aspects of the time trend. We establish consistency and some distribution theory in the presence of a large sample. Moreover, we examine the proposed hypothesis tests and estimation methods through both simulated and real data examples. Finally, we discuss some potential extensions and issues when modelling time effects.

Keywords: Global Mean Sea Level; Nonparametric Kernel Estimation; Nonstationarity

JEL classification: C14, C22, Q54

---

[1]The first author would like to thank the Australian Research Council Discovery Grants Program for its support under Grant numbers: DP150101012 & DP170104421.

[2]*Corresponding Author*: Oliver Linton, Faculty of Economics, University of Cambridge, Cambridge CB3 9DD, U.K. Email: obl20@cam.ac.uk

1 Introduction

Time trends have been widely studied and used for more than a century (e.g., Jones, 1943; Anderson, 1971; Hamilton, 2017; Andrews and McDermott, 1995; Phillips, 2001, 2005, 2007, 2009), see Mills and Patterson (2015) for an historical review. There is no doubt that time trends exist in many data sets from different fields, so that how to mimic time effects always plays a crucial role in data-driven science (e.g., economics, finance, ecology, geology, etc.). In some applications, like climate modelling, the trend is the object of interest. In other applications, like some in macroeconomics, interest focusses on the fluctuations about the trend, which is why so many applied works start from detrending the data (e.g., Greene, 2005; Feng and Serletis, 2008). Either way, it is important to have a good methodology for dealing with the trend. There are several general approaches to trend modelling that have widespread appeal for practitioners, these include:

1. using a deterministic global trend under a parametric setting (cf., Chapter 3 of Anderson (1971)). For example, production economists usually incorporate time trends by simply adding a linear term t and/or a quadratic term t^2 to so-called translog production/cost functions in order to capture time effects (e.g., Greene, 2005, Eq. 10; Feng and Serletis, 2008, Eq. 13 and 19, and so on); or

2. using a local deterministic trend under the nonparametric setting (cf., Robinson, 1997; Chen et al., 2012*b*; Dong and Linton, 2016). For example, Engle and Rangel (2008) and Hafner and Linton (2010) use such nonparametric trends to capture slowly varying long run components of volatility. The Hodrick-Prescott filter widely deployed in macroeconomics is best interpreted as fitting such a trend model to the level of the series (Phillips and Jin, 2015); or

3. using a stochastic trend driven by a unit root or random walk process (cf., Harvey, 1989; Greene, 2002).

We are concerned with deterministic trend models, i.e., the first two cases considered above. Not much work has been done to examine the correct functional form in the parametric global trend model, with linear or quadratic being the dominant choices. On the other hand, the nonparametric trend literature confines its attention to the case where the trend is bounded as the sample size increases, which puts some limits on its applicability. In our empirical study we will consider the global mean sea level (GMSL) data, which is plotted below in Figure 1. The plot looks like having a strong linear time trend, but how to defend (or deny) this conjecture against nonlinear alternatives by using a proper statistical tool has not been fully resolved yet.

Power trends have been studied by Phillips (2007) and Robinson (2012) under parametric frameworks respectively, where the traditional least squares method remains valid due to the

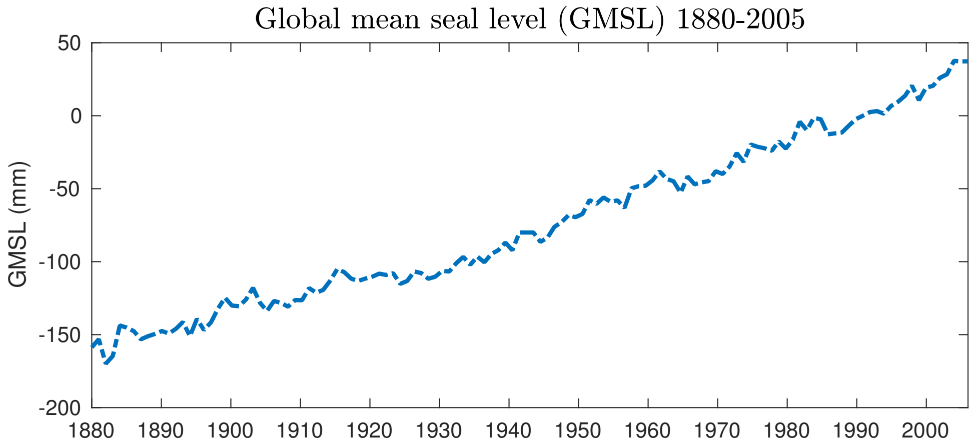

Figure 1: Global mean sea level

parametric nature of their models. The corresponding rates of convergence and asymptotic normalities are established therein. Inspired by these two works, we consider the following model

$$y_t = g(\tau_t)t^{\theta_0} + \varepsilon_t, \tag{1.1}$$

where $\tau_t = t/T$ with $t = 1, \ldots, T$, ε_t is a stationary mixing error process, $g(\cdot)$ is an unknown but smooth function, and θ_0 is an unknown parameter defined on a compact set Θ with $\theta_0 \geqslant 0$. The slowly varying component g can capture nonlinear trend of a quite varied nature, so long as it is bounded and smoothly varying, whereas the global trend part t^{θ_0} allows the outcome variable to increase without bound as the horizon lengthens. The error term ε_t is allowed to be weakly dependent and can represent short term "cyclical" behaviour, which we do not model or estimate. Our model nests the parametric global trend models considered in Phillips (2007) and Robinson (2012) and the nonparametric local trend model that underpins a lot of statistical trend fitting. In this study, we are interested in estimating θ_0 and $g(\cdot)$ from a time series dataset on y_t.

We comment briefly on some related literature. Sornette (2003) proposes deterministic trend and cusp models for modelling stock market crashes. A markedly different approach is provided by unobserved components models from the state space literature; see Harvey (1989) for a comprehensive overview. In these models, the trend is stochastic in nature. It is hard to compare this approach with ours in theoretical terms, since they are nonnested. The pure random walk model implies linear growth in both mean and variance, so by itself is not well suited to describe the flexible trend we propose. From a practical point of view, the two methods offer alternative ways to flexibly estimate the trend behaviour of a time series. In the unobserved components model, the flexibility comes through small stochastic innovations in the trend and the cycle. Our model in contrast owes its flexibility to the nonparametric nature of the deterministic component functions. Dahlhaus (1997) introduces locally stationary process, which combines deterministic local trends with stochastic variation, see also Giraitis et al. (2014) who consider a time-varying coefficient model with stochastic variation.

The structure of this paper is as follows. In Section 2 we propose two hypothesis tests for evaluating the nested parametric and nonparametric models. In Section 3 we propose estimators of both trend components and derive their consistency and limiting distributions. Some simulation studies are implemented in Section 4 to examine the proposed tests and estimation methods. In Section 5 we apply our methodology to the global mean sea level (GMSL) data. Section 6 discusses some potential extensions and issues; Section 7 concludes. Mathematical proofs of the main theories are given in Appendix A. Finally, in Appendix B, we provide the omitted proofs, and some extensions to include (1) theoretical development supporting some of our discussions given in Section 6, (2) another modelling issue of studying power trend, and (3) corresponding simulation studies.

Before proceeding to Section 2, it is convenient to introduce some notations that will be used throughout this paper. $\to_P$ denotes converging in probability; $\to_D$ denotes converging in distribution; $\lfloor a \rfloor$ means the largest integer not exceeding a; $K(\cdot)$ and h represent a symmetric kernel function and a corresponding bandwidth of the kernel method, respectively; moreover, $K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$.

2 Two Pre-Testing Issues

Sections 2 and 3 together provide the asymptotic results of the paper. To be precise, the main testing and estimation steps are as follows.

1. We first consider two hypothesis tests:

$$\text{(a). Parametric test: } \begin{cases} H_0: \ \theta_0 = 0 \\ H_1: \ \theta_0 > 0 \end{cases}; \tag{2.1}$$

$$\text{(b). Nonparametric test: } \begin{cases} H_0^*: \ g(\tau) \text{ is a constant function} \\ H_1^*: \ g(\tau) \text{ is a non-constant function} \end{cases}. \tag{2.2}$$

2. If we fail to reject either of these null hypotheses, everything goes back to some well studied models of the literature. For example, failure to reject "$H_0: \ \theta_0 = 0$" gives a model $y_t = g(\tau_t) + \varepsilon_t$, which is a special case of Robinson (1997), Dong and Linton (2016) and so forth; and failure to reject "$H_0^*: g(\tau)$ is a constant function" leads to $y_t = \beta_0 t^{\theta_0} + \varepsilon_t$, which has been studied in Phillips (2007) and Robinson (2012).

3. If both null hypotheses get rejected, we move on to Section 3. We point out the failure of some intuitive methods in Section 3.1, and provide consistent estimators of g and θ_0 in

Section 3.2.

We now make some assumptions to facilitate the derivation throughout the paper.

Assumption 1:

1. *(a)* $0 \leqslant \theta_0 \in \Theta$, *and* Θ *is a compact set defined on* $\mathbb{R}$.

   (b) $g(\cdot)$ *is second order differentiable on* $[0,1]$, *and satisfies that* $0 < A_1 \leqslant \inf_{u\in[0,1]} |g(u)| \leqslant \sup_{u\in[0,1]} |g(u)| \leqslant A_2 < \infty$ *and* $\sup_{(\theta,u)\in\Theta\times[h,1]} \left|\frac{d[u^{\theta+\theta_0}g(u)]}{du}\right| \leqslant A_3 < \infty$ *for the same* h *defined in Assumption 1.4 below, where* A_1, A_2 *and* A_3 *are positive constants;*

2. $\{\varepsilon_t \mid t = 1,\ldots,T\}$ *is a strictly stationary and* α*-mixing error process with mixing coefficients* $\{\alpha(i) \mid i = 1,2,\ldots\}$ *such that* $\sum_{i=1}^{\infty}[\alpha(i)]^{\frac{\delta}{2+\delta}} < \infty$ *for some* $\delta > 0$, *where* $\alpha(i) = \sup_j \sup_{A\in\mathcal{F}_{-\infty}^{j},\, B\in\mathcal{F}_{j+i}^{\infty}} |\Pr(A\cap B) - \Pr(A)\Pr(B)|$ *and* $\mathcal{F}_j^k$ *is the* σ*-filed generated by* $\{\varepsilon_t \mid j \leqslant t \leqslant k\}$. *Moreover,* $E[\varepsilon_1] = 0$, $E|\varepsilon_1|^2 = \sigma_\varepsilon^2$ *and* $E|\varepsilon_1|^{2+\delta/2} < \infty$ *for the same* δ.

3. *Let* $K(\cdot)$ *be symmetric and defined on* $[-1,1]$. *Assume further that* $K^{(1)}(u)$ *is uniformly bounded on* $[-1,1]$, $\int_{-1}^{1} K(u)du = 1$ *and* $\int_{-1}^{1} |u|K(u)du < \infty$.

4. *For the bandwidth* h, *suppose that* $h = O(T^{-\nu})$ *for some* $0 < \nu < 1$.

Under a parametric setting, Robinson (2012) allows for $\theta_0 > -\frac{1}{2}$ with $\theta_0 \neq 0$, but, for our nonparametric model, we have to impose a stronger restriction in Assumption 1.1.a, so that the kernel method remains valid for the denominator of (3.1) provided below. In the same spirit, Assumption 1.1.b imposes some conditions on $g(\cdot)$ to ensure that the kernel method works for the numerator of (3.1). Assumptions 1.2-1.4 are standard in the literature (cf. Fan and Yao, 2003).

2.1 Parametric Test

If g were known, it would be easy to obtain the Gaussian likelihood as follows:

$$Q_T(\theta) = \sum_{t=1}^{T} \left(y_t - g(\tau_t)t^{\theta}\right)^2,$$

which yields a score function $S_T(\theta) = \frac{\partial Q_T(\theta)}{\partial\theta} = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - g(\tau_t)t^{\theta}\right) g(\tau_t)t^{\theta}\ln t$. Thus, under the null, it reduces to $S_T(0) = \frac{1}{T}\sum_{t=1}^{T}(y_t - g(\tau_t))\, g(\tau_t)\ln t$.

In practice, since g is unknown, we take the estimate $\widehat{S}_T = \frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{g}(\tau_t))\,\widehat{g}(\tau_t)\ln t$, where $\widehat{g}(u) = \frac{\sum_{t=1}^{T} K_h(u-\tau_t)y_t}{\sum_{t=1}^{T} K_h(u-\tau_t)}$. By the development similar to (A.17) of Wang and Xia (2009), it is easy to obtain that under the null

$$\sup_{u\in[0,1]} |\widehat{g}(u) - g(u)| = O_P\left(\frac{\sqrt{\ln T}}{\sqrt{Th}}\right) + O_P(h). \tag{2.3}$$

Notice that using the full sample to construct the test will get two leading terms to cancel with each other (see (B.2) for more details), so that further difficulties will arise when deriving the asymptotic distribution. In order to avoid this problem, we use the even numbered observations to estimate $g(\cdot)$ and evaluate the score function using the odd numbered observations below. Thus, the final version of the score function is

$$\widehat{S}_T = \frac{1}{T_{odd}} \sum_{t \text{ odd}} (y_t - \widehat{g}(\tau_t))\, \widehat{g}(\tau_t) \ln t, \tag{2.4}$$

where T_{odd} stands for the total number of odd numbered observations, and

$$\widehat{g}(u) = \frac{\sum_{t \text{ even}} K_h(u-\tau_t) y_t}{\sum_{t \text{ even}} K_h(u-\tau_t)}. \tag{2.5}$$

Based on the above discussions, we derive a formal hypothesis test, which is described in the next theorem.

Theorem 2.1. *Let Assumption 1 hold.*

1. *Suppose that (2.1), (2.3) holds under the null. In addition, suppose that ε_t is i.i.d. across t. As $T \to \infty$,*

$$\widehat{LM} = \frac{\frac{1}{2\sqrt{T}} \sum_{t\ odd} (y_t - \widehat{g}(\tau_t))\, \widehat{g}(\tau_t) \ln t}{\left\{\widetilde{\sigma}_\varepsilon^2 \cdot \frac{1}{T}\sum_{t=1}^T \left[\widetilde{g}(\tau_t)\ln t\right]^2\right\}^{1/2}} \to_D N(0,1), \tag{2.6}$$

   where $\widetilde{\sigma}_\varepsilon^2 = \frac{1}{T}\sum_{t=1}^T (y_t - \widetilde{g}(\tau_t))^2 \to_P \sigma_\varepsilon^2$, and $\widetilde{g}$ is defined in the same way as (2.5) but utilizes the full sample.

2. *Suppose that $\theta_0 > 0$. As $T \to \infty$, $\widehat{LM} \to \infty$.*

For the sake of readability, we leave a generalized version of the parametric test (i.e., $H_0 : \theta_0 = a$ vs. $H_1 : \theta_0 > a$) with the corresponding discussions in the Appendix B of the paper. To ensure the test also works for the case where ε_t is not i.i.d. over t, certain development as in Andrews (1991) is required. It may lead to another research paper, so we do not pursue it further in order not to deviate from our main goal.

2.2 Nonparametric Test

In this subsection, we consider the nonparametric test (2.2). Notice that, under H_0^*, we have a parametric model of the form

$$y_t = \beta_0\, t^{\theta_0} + \varepsilon_t,$$

where the unknown parameters (β_0, θ_0) can be estimated by the nonlinear least squares estimation method:

$$(\widehat{\beta}, \widehat{\theta}) = \arg\min_{(\beta,\theta)} \sum_{t=1}^{T} \left(y_t - \beta\, t^{\theta}\right)^2. \tag{2.7}$$

By Theorems 1 and 2 of Robinson (2012), we have

$$\widehat{\theta} - \theta_0 = O_P(T^{\chi-\theta_0-\frac{1}{2}}) \quad \text{and} \quad \widehat{\beta} - \beta_0 = O_P((\ln T)T^{\chi-\theta_0-\frac{1}{2}}) \tag{2.8}$$

for any given sufficiently small $\chi > 0$ under minor restrictions.

By (2.8) and building on Fan and Li (1996) and Li (1999), we propose a nonparametric test of the form

$$\widehat{L} = \max_{h \in \mathcal{H}} L(h) \quad \text{with} \quad L(h) = \frac{\sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} K\left(\frac{\tau_t - \tau_s}{h}\right) \widehat{e}_s\, \widehat{e}_t}{\sqrt{\sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} K^2\left(\frac{\tau_t - \tau_s}{h}\right) \widehat{e}_s^2\, \widehat{e}_t^2}}, \tag{2.9}$$

where $\mathcal{H} = \{h = h_{max} a^k : h \geqslant h_{min}, k = 0.1, 2, \ldots\}$ with $0 < h_{min} < h_{max}$ and $0 < a < 1$, and $\widehat{e}_t = y_t - \widehat{\beta}\, t^{\widehat{\theta}}$.

Moreover, the associated critical values can be drawn by the following bootstrap procedure.

1. For $t = 1, \ldots, T$, generate $y_t^* = \widehat{\beta}\, t^{\widehat{\theta}} + \widehat{e}_t u_t$, where u_t's are sampled randomly from $N(0, 1)$.

2. Use the data set $\{y_t^* \mid t = 1, \ldots, T\}$ to implement (2.7) in order to obtain $(\widetilde{\beta}, \widetilde{\theta})$, and compute the statistic L^* that is obtained by replacing y_t and $(\widehat{\beta}, \widehat{\theta})$ with y_t^* and $(\widetilde{\beta}, \widetilde{\theta})$, respectively, in (2.9).

3. Repeat the above steps J times to produce J versions of L^* denoted by L_j^* for $j = 1, \ldots, J$. Use $\{L_1^*, \ldots, L_J^*\}$ to construct the empirical bootstrap distribution function, that is, $F^*(w) = \frac{1}{J}\sum_{j=1}^{J} 1(L_j^* \leqslant w)$. Further use the empirical bootstrap distribution function to estimate the asymptotic critical value, l_α.

Theorem 2.2. *Let Assumption 1 hold, and suppose that $\theta_0 > 0$.*

1. *For the nonparametric test (2.2), (2.8) holds under the null.*

2. *In addition, for $\mathcal{H}$ of (2.9), let $c_0[\ln(\ln T)]^{-1} = h_{max} > h_{min} \geqslant T^{-\gamma} > 0$ with some constants c_0 and γ such that $0 < \gamma < \frac{1}{3}$. Then we have $\lim_{T\to\infty} \Pr(\widehat{L} > l_\alpha) = \alpha$ for the above procedure.*

The first result of Theorem 2.2 follows from Robinson (2012) straight away. The second result of Theorem 2.2 follows from the development similar to Fan and Li (1996), Li (1999) and Gao and Hawthorne (2006). The same principle of this nonparametric test has also been employed in Su et al. (2015) to study a panel data model. As the alternative hypothesis of (2.2) does not specify a clear function form for $g(\cdot)$, we do not further investigate the limit of (2.9) under the alternative. Instead, we consider different functional forms of $g(\cdot)$ in the simulation study of Section 4.

3 Estimation Method and Theory

We now consider estimating (1.1) for the case in which $\theta_0 > 0$ and $g(\cdot)$ is a non-constant function. For $\forall(\theta, u)$, the kernel based OLS estimator of $g(u)$ is intuitively expressed as follows:

$$\widehat{g}(u, \theta) = \left[\sum_{t=1}^{T} t^{2\theta} K_h\left(u - \tau_t\right)\right]^{-1} \sum_{t=1}^{T} t^{\theta} y_t K_h\left(u - \tau_t\right), \tag{3.1}$$

where $K_h\left(x - u\right) = \frac{1}{h} K\left(\frac{x-u}{h}\right)$, and $K(\cdot)$ and h have been defined in Assumption 1. Then the key question becomes how to recover θ_0. Once we have obtained a consistent estimator for θ_0, we need only to plug it in (3.1) to estimate $g(u)$.

3.1 Failure of Some Intuitive OLS Methods

We first explain why two very intuitive OLS methods fail when encountering time trends with time-varying coefficient.

By the traditional profile method (cf., Robinson, 2012; Dong et al., 2016), the first objective function is defined as follows:

$$Q_T(\theta) = \sum_{t=1}^{T} \left(y_t - t^{\theta} \widehat{g}(\tau_t, \theta)\right)^2, \tag{3.2}$$

where $\widehat{g}(u, \theta)$ is denoted in (3.1). According to Lemma 3.1 below, one finding is that

$$t^{\theta} \widehat{g}(\tau_t, \theta) = t^{\theta} t^{\theta_0 - \theta} g(\tau_t)(1 + o_P(1)) = t^{\theta_0} g(\tau_t)(1 + o_P(1)),$$

where θ disappears from the leading term and only exists in the residual. Thus, it would be difficult to recover θ_0 from (3.2), as the limit of $Q_T(\theta)$ is not in the form of $Q(\theta - \theta_0)$ with $Q(w)$ being a continuous function and having a unique local minimum at $w = 0$.

Alternatively, one may follow Section 6 of Phillips (2007) to define an objective function for any given u as

$$Q_T(\alpha | u) = \sum_{t=1}^{n} \left(y_t - \beta\, t^{\theta}\right)^2 K_h\left(\tau_t - u\right), \tag{3.3}$$

where $\alpha = (\beta, \theta)$. Thus, the corresponding estimator is obtained by

$$\widehat{\alpha}(u) = \left(\widehat{\beta}(u),\ \widehat{\theta}(u)\right) = \operatorname*{argmin}_{\alpha} Q_T(\alpha | u). \tag{3.4}$$

Building on (3.4), the estimator of θ_0 is finally defined as $\widehat{\theta} = \int_0^1 \widehat{\theta}(u)\psi(u)du$, where $\psi(\cdot)$ serves as a weight function.

Note that, in order to minimize $Q_T(\alpha|u)$, the following two equations must hold:

$$\frac{\partial Q_T(\alpha|u)}{\partial \beta}\Big|_{\alpha=\widehat{\alpha}(u)} = 0 \quad \text{and} \quad \frac{\partial Q_T(\alpha|u)}{\partial \theta}\Big|_{\alpha=\widehat{\alpha}(u)} = 0.$$

Simple algebra shows that $\frac{\partial Q_T(\alpha|u)}{\partial \beta}\Big|_{\alpha=\widehat{\alpha}(u)} = 0$ yields

$$\widehat{\beta}(u) = \left[\sum_{t=1}^{T} t^{2\widehat{\theta}(u)} K_h(u - \tau_t)\right]^{-1} \sum_{t=1}^{T} t^{\widehat{\theta}(u)} y_t K_h(u - \tau_t),$$

which has the same form as (3.1), and indicates that the leading term of $Q_T(\widehat{\alpha}|u)$ is independent of $\widehat{\theta}(u)$ by the same discussions under (3.2). In other words, we can find different θ's belonging to Θ (say, $\widehat{\theta}_1(u)$ and $\widehat{\theta}_2(u)$) to ensure $Q_T(\widehat{\alpha}_1(u)|u)$ and $Q_T(\widehat{\alpha}_2(u)|u)$ are asymptotically equivalent, where $\widehat{\alpha}_1(u) = (\widehat{\beta}(u), \widehat{\theta}_1(u))$ and $\widehat{\alpha}_2(u) = (\widehat{\beta}(u), \widehat{\theta}_2(u))$. This concludes why the second approach fails.

We will further examine the above two methods in the simulation study of Section 4.

3.2 Consistent Estimation

In order to establish a consistent estimator of θ_0, we firstly state the next lemma.

Lemma 3.1. *Let Assumption 1 hold. In addition, suppose that*

1. $B_T(\theta_0)$ *represents a subset of* Θ *centred at* θ_0 *with radius* $\frac{M}{\ln T}$*, where* M *is a positive constant;*

2. $B_{\epsilon_1}(h) = [(1+\epsilon_1)h,\ 1]$*, where* ϵ_1 *is a sufficiently small positive constant.*

Then, for $\widehat{g}(u,\theta)$ *defined by (3.1), as* $T \to \infty$*,*

$$\sup_{(\theta,u)\in B_T(\theta_0)\times B_{\epsilon_1}(h)} \left|\widehat{g}(u,\theta) - (uT)^{\theta_0-\theta} g(u)\right| = O_P\left(\frac{\sqrt{\ln T}}{T^{\frac{1}{2}+\theta_0} h^{\frac{1}{2}+2\theta_0}}\right) + O(h).$$

Compared with some similar results in the literature (e.g., Vogt, 2012, Eq. 16; Chen et al., 2012*b*, Eq. B.10), one main difference is that we have to take the power term θ into consideration while deriving the rate of uniform convergence, which is the main reason why we have to introduce ϵ_1 in the above lemma. The constant ϵ_1 controls the minimum value that u can take, and in this sense serves the same purpose as C_1 of Theorem 4.2 of Vogt (2012). The slow rate $O(h)$ has also been achieved in Wang and Xia (2009, Eq. A.17-A.19), where the uniform convergence of another kernel based method is studied. It is noteworthy that if we truncate the interval $[(1+\epsilon_1)h, 1]$ to $[(1+\epsilon_1)h, 1-h]$ as in Vogt (2012, Eq. 16), we can replace h with h^2 after imposing extra restrictions on $g(\cdot)$. Moreover, compared to Theorem 2 of Robinson (2012) (i.e., the second term of (2.8) of Section 2), we actually do not need to introduce a term T^{χ} in the above asymptotic results, which in a sense improves the rate of convergence slightly regardless of the terms caused by the nonparametric method.

In addition, Lemma 3.1 indicates that $\widehat{g}(u,\theta)$ with $\theta \in B_T(\theta_0)$ is a consistent estimator of $g(u)$ subject to a constant term $(uT)^{\theta_0-\theta}$, which is not guaranteed to be 1 if θ is very close to (or on) the boundary of $B_T(\theta_0)$. Below, we are going to show that $\widehat{\theta}$ defined by (3.6) indeed falls in $B_T(\theta_0)$ with probability approaching one in Theorem 3.1, and further deal with the unknown constant in Theorem 3.2.

Finally, Lemma 3.1 suggests that constructing an objective function in logarithmic form may asymptotically converge to a continuous function having a unique minimum at $\theta = \theta_0$. We define the objective function

$$R_T(\theta) = \left\{ \lambda_T \cdot \ln \left[\frac{1}{T} \sum_{t=\lfloor Th \rfloor +1}^{T} \tau_t^{2\theta} \widehat{g}(\tau_t, \theta) \right]^2 \right\}^2, \tag{3.5}$$

where we let $\lambda_T = \frac{1}{\ln T}$ for notational simplicity, and $\widehat{g}(\cdot,\cdot)$ is defined in (3.1).

Remark 3.1.

1. *Note that the number of observations lying between $\lfloor Th \rfloor$ and $\lfloor T(1+\epsilon_1)h \rfloor$ is limited and negligible, as ϵ_1 is an arbitrary small positive constant. Thus, with some abuse of notation, we define (3.5) by using observations from $\lfloor Th \rfloor + 1, \ldots, T$ throughout this paper.*

2. *The term $\tau_t^{2\theta}$ serves the purpose of solving a technical issue when recovering the normalizer of Theorem 3.3. A short explanation is that without the term $\tau_t^{2\theta}$, $\frac{1}{T}\sum_{t=\lfloor Th \rfloor +1}^{T} \frac{\partial \widehat{g}(\tau_t,\theta_0)}{\partial \theta}$ will yield a term $\frac{1}{T}\sum_{t=\lfloor Th \rfloor +1}^{T} \tau_t^{-2\theta_0}$ in the denominator. Intuitively, one may think that $\frac{1}{T}\sum_{t=\lfloor Th \rfloor +1}^{T} \tau_t^{-2\theta_0}$ converges to $\int_0^1 u^{-2\theta_0} du$, however, it is not the case given the assumption on θ_0. Let alone the fact that $\int_0^1 u^{-2\theta_0} du$ does not exist in general, because $\int_0^1 u^{-2\theta_0} du < \infty$ does not hold for $1 - 2\theta_0 < 0$.*

According to (3.5), the estimator of θ_0 is given by

$$\widehat{\theta} = \operatorname*{argmin}_{\theta \in \Theta} R_T(\theta), \tag{3.6}$$

and we summarize the corresponding asymptotic results by the next theorem.

Theorem 3.1. *Suppose that Assumption 1 holds. Then, as $T \to \infty$,*

1. $\widehat{\theta} - \theta_0 = O_P(\lambda_T)$, *where λ_T is defined in (3.5);*

2. $\sup_{u \in [(1+\epsilon_1)h,\ 1]} \left| \widehat{g}(u,\widehat{\theta}) - (uT)^{\theta_0 - \widehat{\theta}} \cdot g(u) \right| = O_P\left(\frac{\sqrt{\ln T}}{T^{\frac{1}{2}+\theta_0} h^{\frac{1}{2}+2\theta_0}} \right) + O(h)$, *where ϵ_1 is the same one as denoted in Lemma 3.1.*

Remark 3.2. *Due to taking the logarithm in (3.5), we can only achieve a slow rate of convergence (i.e., λ_T) for $\widehat{\theta}$. Compared to the parametric setting (Robinson, 2012), the slow rate is caused by the nonparametric nature of (1.1). A similar phenomenon has also been observed in Pesaran and Yang (2016) (cf., discussions under their Eq. 80), even though it is not directly related to our model.*

We now briefly explain the key difference between the fixed coefficient power trend and the time-varying coefficient power trend by using a simple parametric model even without an error term, say $y_t = \tau_t^{\theta_0}$. Simple calculation shows

$$
\begin{aligned}
Q_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T}(y_t - \tau_t^{\theta})^2 = \frac{1}{T}\sum_{t=1}^{T}\tau_t^{2\theta_0} - \frac{2}{T}\sum_{t=1}^{T}\tau_t^{\theta_0+\theta} + \frac{1}{T}\sum_{t=1}^{T}\tau_t^{2\theta} \\
&= \left(\int_0^1 u^{2\theta_0}du - 2\int_0^1 u^{\theta_0+\theta}du + \int_0^1 u^{2\theta}du\right)\cdot(1+o(1)) \\
&= \left(\frac{1}{2\theta_0+1} - \frac{2}{\theta_0+\theta+1} + \frac{1}{2\theta+1}\right)\cdot(1+o(1)) \\
&= \frac{2(\theta_0-\theta)^2}{(2\theta_0+1)(\theta_0+\theta+1)(2\theta+1)}\cdot(1+o(1))
\end{aligned}
$$

under minor restrictions on θ, where the third equality follows from the definition of Riemann integral. Note the parameter θ will not exist in the power any more as T diverges. In other words, it does not require taking logarithm to obtain an objective function having a unique minimum at $\theta = \theta_0$ asymptotically. However, this is not the case any more for models with time-varying power trend.

Remark 3.3. *It is easy to see that the rate of convergence of the second result of Theorem 3.1 will reach the minimum value when $h = O\left(T^{-\frac{1+2\theta_0}{3+4\theta_0}}\cdot(\ln T)^{-\frac{1}{3+4\theta_0}}\right)$. We will further examine this finding, and explain how to select the "optimal" bandwidth practically in the simulation study of Section 4.*

Before proceeding further, we take a careful look at the estimation of $g(\cdot)$, and explain the identification issue of g mentioned under Lemma 3.1. Consider the following distance between (θ, g) and (θ^*, f)

$$
D_T\{(\theta,g),(\theta^*,f)\} = \sum_{t=1}^{T}\left\{g(\tau_t)t^{\theta} - f(\tau_t)t^{\theta^*}\right\}^2 = \sum_{t=1}^{T}\left\{T^{\theta}g(\tau_t)\tau_t^{\theta} - T^{\theta^*}f(\tau_t)\tau_t^{\theta^*}\right\}^2.
$$

Based on Theorem 3.1, we let $\theta = \theta^* + \frac{M}{\ln T}$ with M being a constant. Then we can write

$$
\begin{aligned}
D_T\{(\theta,g),(\theta^*,f)\} &= \sum_{t=1}^{T}\left\{T^{\theta^*}Mg(\tau_t)\tau_t^{\theta} - T^{\theta^*}f(\tau_t)\tau_t^{\theta^*}\right\}^2 \\
&= T^{2\theta^*}\sum_{t=1}^{T}\tau_t^{2\theta^*}\left\{Mg(\tau_t)\tau_t^{M/\ln T} - f(\tau_t)\right\}^2,
\end{aligned}
$$

so any sequence $f_T(u) = Mg(u)u^{M/\ln T} \simeq Mg(u)$ will set this objective function exactly zero. This identification issue is purely due to the slow rate of convergence obtained by (1) of Theorem 3.1. At this stage, how to achieve a faster rate to overcome this problem remains unclear to us.

In order to identify the unknown constant, we let $|g(1)| = 1$ in the rest of this paper. For those $g(\cdot)$'s not satisfying $|g(1)| = 1$, we are essentially recovering a rescaled version of $g(u)$ below, i.e., $\mathrm{g}(u) = g(u)/|g(1)|$ given $g(1) \neq 0$. See Su and Jin (2012) and Dong and Linton (2016) for similar settings on the functional component.

We further make the following assumption.

Assumption 2: *Let $\gamma(t-s) = E[\varepsilon_t \varepsilon_s]$ satisfy that $(Th)\, C(K;Th) \to 0$ as $T \to \infty$, where $C(K;u) = \int_{-1}^1 \int_{-1}^1 K(x)K(y)\, |\gamma((x-y)\,u)|\, I[x \neq y]\, dxdy$.*

Assumption 2 is used to establish the existence of the contribution from the long-run covariance in such kernel estimation. It can easily be verified if we impose $|\gamma(u)| \leqslant A\rho^{|u|}$ for $|u| \geqslant 1$, $0 < \rho < 1$ and $0 < A < \infty$. Another example can be found in Gao and Anh (1999, p. 41).

Before stating the next theorem, we define for $\forall u \in (0,1)$

$$\widehat{\eta}_T = \frac{1}{T}\sum_{t=\lfloor Th \rfloor+1}^{T} \tau_t^{2\widehat{\theta}} \widetilde{g}(\tau_t), \quad \widetilde{g}(u) = (uT)^{-\log_T |\widehat{g}(1,\widehat{\theta})|} \widehat{g}(u,\widehat{\theta}),$$

$$\widehat{\Sigma} = \widehat{\sigma}_\varepsilon^2 \int_{-1}^1 K^2(x)dx, \quad \widehat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum_{t=\lfloor Th \rfloor+1}^{T} \left(y_t - t^{\widehat{\theta}}\, \widehat{g}(\tau_t,\widehat{\theta})\right)^2,$$

$$\kappa_{1T}(\widehat{\theta},u) = |\widehat{g}(1,\widehat{\theta})|^{-1} \cdot \left(\sum_{t=1}^T t^{2\widehat{\theta}} K_h(u-\tau_t)\right)^{-1} \sum_{t=1}^T t^{\widehat{\theta}+\theta_0} g(\tau_t) K_h(u-\tau_t) - g(u). \tag{3.7}$$

Theorem 3.2. *Let Assumptions 1 and 2 hold.*

(1). For $\forall u \in (0,1)$, as $T \to \infty$,

$$T^{\theta_0+\frac{1}{2}} h^{\frac{1}{2}} \cdot \frac{u^{\widehat{\theta}}}{\widehat{\eta}_T \sqrt{\widehat{\Sigma}}} \left(|\widehat{g}(1,\widehat{\theta})|^{-1} \cdot \widehat{g}(u,\widehat{\theta}) - g(u) - \kappa_{1T}(\widehat{\theta},u)\right) \to_D N(0,1),$$

where $\kappa_{1T}(\widehat{\theta},u) = O_P(h)$.

Suppose further that, for $\forall u \in (0,1)$, $\sup_{\theta\in\Theta} \left| \frac{d^2[w^{\theta+\theta_0} g(w)]}{dw^2}\Big|_{w=u} \right| < \infty$, and $h = O(T^{-\nu})$ with $0 < \nu \leqslant 1 - \frac{2+\theta_0}{2.5+2\theta_0}$.

(2). Then $\kappa_{1T}(\widehat{\theta},u)$ will achieve a fast rate, i.e., $\kappa_{1T}(\widehat{\theta},u) = O_P(h^2)$.

Due to the nonparametric nature of our model, the rate of convergence and the normality on the estimate of the coefficient function cannot be established using Theorem 8.1 of Wooldridge (1994) as in the proof Theorem 6.3 of Phillips (2007). The profile method under nonparametric framework employed in this paper allows us to avoid bringing a term diverging at a rate of $\ln T$ to slow down the rate of convergence (see Theorem 6.3 of Phillips (2007) for details).

The fact that $\lim_{T\to\infty} |\widehat{\eta}_T| = \left|\int_0^1 u^{2\theta_0} g(u)du\right| > 0$ is verified by (A.11) and (A.12). The bias term $\kappa_{1T}(\widehat{\theta}, u)$ is due to the use of the kernel method, and the extra conditions in the body of Theorem 3.2 make certain that $\kappa_{1T}(\widehat{\theta}, u)$ will have the usual order $O_P(h^2)$ as in the literature of nonparametric regression (see Chen et al., 2012*b*; Vogt, 2012, for example). Without these restrictions, the slow rate (i.e., $O_P(h)$) applies.

Having established the above results, we are now ready to consider the asymptotic distribution of $\widehat{\theta}$. By definition of (3.6) and Mean Value Theorem,

$$0 = \frac{\partial R_T(\theta)}{\partial \theta}\Big|_{\theta=\widehat{\theta}} = \frac{\partial R_T(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} + \frac{\partial^2 R_T(\theta)}{\partial \theta^2}\Big|_{\theta=\widetilde{\theta}}(\widehat{\theta} - \theta_0), \tag{3.8}$$

where $\widetilde{\theta}$ lies between $\widehat{\theta}$ and θ_0; and $\frac{\partial R_T(\theta)}{\partial \theta}$ and $\frac{\partial^2 R_T(\theta)}{\partial \theta^2}$ are provided in (A.1) of Appendix A of the paper. The following theorem holds, and it associated proof is provided in Appendix A below.

Theorem 3.3. *Suppose that Assumption 1 holds. As $T \to \infty$,*

(1). $(\ln T)(\widehat{\theta} - \theta_0) \to_P \ln\left|\int_0^1 u^{2\theta_0} g(u)du\right|$;

Given that $\left|\int_0^1 u^{2\theta_0} g(u)du\right| \neq 1$,

(2). $\frac{\ln T}{\ln|\widehat{\eta}_T|}(\widehat{\theta} - \theta_0) \to_P 1$, *where η_T has been defined in (3.7).*

For Theorem 3.3, we make some comments in the next remark.

Remark 3.4. *Theorem 3.3 shows that the limit of* $(\ln T)(\widehat{\theta} - \theta)$ *is in fact a constant rather than a distribution in this paper. Without the terms A_1, A_3 and A_5 in the proof of Theorem 3.3, the right hand side of (A.25) would lead to an asymptotic normality as in Theorem 6.3 of Phillips (2007) and Theorem 3 of Robinson (2012). However, these terms cannot be removed using a bias correction procedure for our nonparametric model, so we state Theorem 3.3 as it is.*

We will further examine Theorem 3.3 in the simulation study below.

4 Numerical Studies

We next conduct some simulation studies to examine the asymptotic results established in Sections 2 and 3. For better presentation, we report some selected results below and leave extra simulation results in the Appendix B of this paper.

4.1 Parametric Test

To examine the hypothesis test provided in Section 2.1, the data generating process (DGP) is $y_t = g(\tau_t)t^{\theta_0} + \varepsilon_t$, where $\varepsilon_t \sim i.i.d.\ N(0,1)$. We consider the following cases under different sample sizes in order to evaluate the size and power of (2.6).

- Case 1 – Size: $\theta_0 = 0$

  Case 1.1: $g(w) = \exp(w)$; Case 1.2: $g(w) = w^2 + 1$

- Case 2 – Power: $\theta_0 = 0.3,\ 0.5,\ 0.7$

  Case 2.1: $g(w) = \exp(w)$; Case 2.2: $g(w) = w^2 + 1$

For each generated data set, we calculate $\widehat{LM}$ of (2.6), and let $\alpha_{LM} = 1(\widehat{LM} > 1.6449)$ (i.e., rejecting the null at 95% significant level), where $1(\cdot)$ is an indicator function. After J replications, we calculate the simple average $\bar{\alpha}_{LM} = \frac{1}{J}\sum_{j=1}^{J}\alpha_{LM,j}$, where $\alpha_{LM,j}$ stands for the value of α_{LM} at the j^{th} replication. Below we choose $J = 1000$. In view of (2.3), the bandwidth is set to $h = \left(\frac{\ln T}{T}\right)^{1/3}$, which is the "optimal" one under the null subject to an unknown constant. We plot the value of $\bar{\alpha}_{LM}$ (i.e., rejection rate) at different sample sizes in Figures 2 and 3 instead of reporting them in tables.

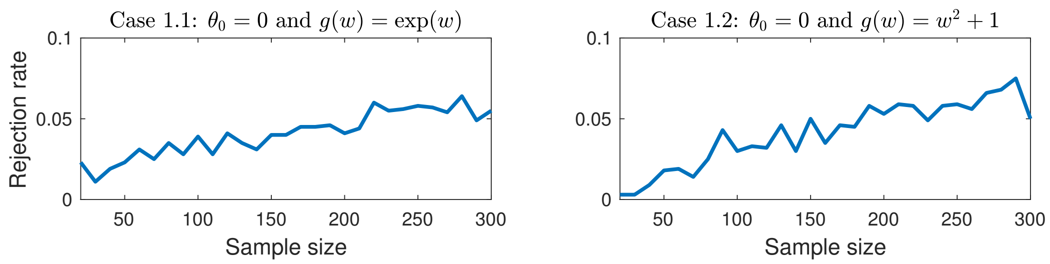

Figure 2: Parametric test: Case 1 – Size

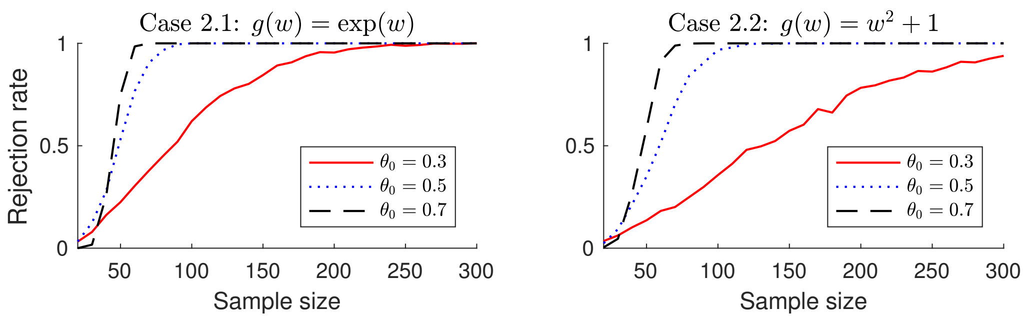

Figure 3: Parametric test: Case 2 – Power

According to Figures 2 and 3, the proposed parametric test (2.6) in general has good finite sample performance. In addition, Figure 3 suggests that as θ_0 gets far away from the null, the power of (2.6) tends to get improved. It should be expected, because when θ_0 is closer to 0, we would need more data to distinguish θ_0 and 0.

4.2 Nonparametric Test

In this subsection, we study the nonparametric test proposed in Section 2.2. It is worth to mention that the principle of this nonparametric test is in fact not new and has been well studied in the literature, so interested readers can refer to the previous studies (e.g., Fan and Li, 1996; Li, 1999; Gao and Hawthorne, 2006; Su et al., 2015) for more detailed and systematic simulation studies on the finite sample performance of this type of test.

The main DGP is still $y_t = g(\tau_t)t^{\theta_0} + \varepsilon_t$, where $\varepsilon_t \sim i.i.d.\ N(0,1)$. In order to examine the size and power, we consider the following cases.

- Case 1 – Size: $g(w) \equiv 1$ and $\theta_0 = 0.5,\ 1$
- Case 2 – Power: $\theta_0 = 0.5,\ 1$

$$\text{Case 2.1: } g(w) = \exp(w); \quad \text{Case 2.2: } g(w) = w^2 + 1$$

For each generated data set, we calculate the statistic value by (2.9), and 95% critical values by Theorem 2.2 based on 299 bootstrap replications.[3] Similar to the above subsection, if we reject the null at 95% significant level for the j^{th} data set, we then record $\alpha_{L,j} = 1$, otherwise $\alpha_{L,j} = 0$. After J replications, we calculate the simple average $\bar{\alpha}_L = \frac{1}{J}\sum_{j=1}^{J}\alpha_{L,j}$. Again, we choose $J = 1000$, and plot the values of $\bar{\alpha}_L$ at different sample sizes in Figures 4 and 5 below.

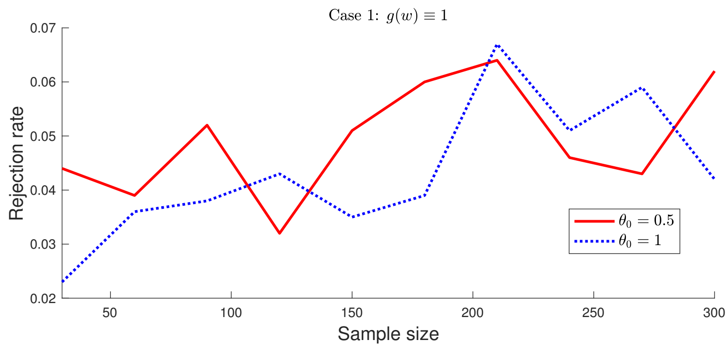

Figure 4: Nonparametric test: Case 1 – Size

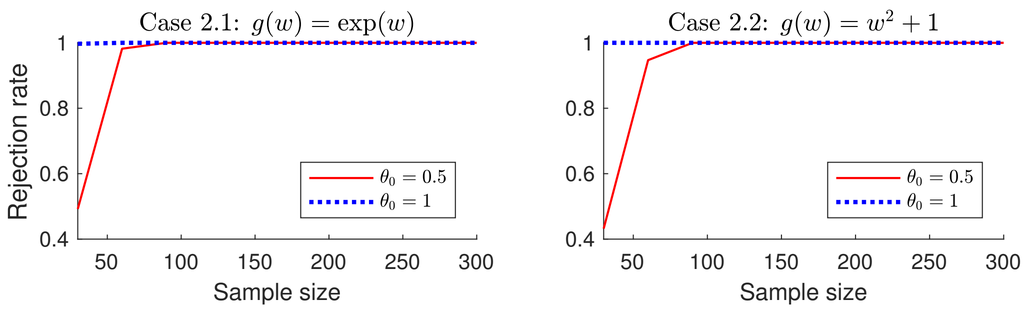

Figure 5: Nonparametric test: Case 2 – Power

[3]Here, we follow exactly the same procedure of (9) of Gao and Hawthorne (2006).

The size of the nonparametric test is still as good as expected by Figure 4, while, according to Figure 5, the power of the nonparametric test is much better than what we see from the parametric test.

4.3 Evaluation of the Estimates

In this subsection, we examine the asymptotic results provided in Section 2.3. Building on Remark 3.3, we firstly explain how to implement our nonparametric method while taking bandwidth selection into account.

Remark 4.1. *Bandwidth selection:*

1. *Provide an initial bandwidth (say $h_0 = T^{-1/3}$);*

2. *For k^{th} ($k \geqslant 1$) iteration, use h_{k-1} obtained from $(k-1)^{th}$ iteration to calculate $\widehat{\theta}_k$. Stop iteration, if $|\widehat{\theta}_k - \widehat{\theta}_{k-1}| \leqslant \epsilon$, where ϵ is a sufficiently small positive number and serves as a stopping criteria. Otherwise, update the bandwidth by $h_k = T^{-\frac{1+2\widehat{\theta}_k}{3+4\widehat{\theta}_k}} \cdot (\ln T)^{-\frac{1}{3+4\widehat{\theta}_k}}$ and proceed to $(k+1)^{th}$ iteration.*

According to Remark 3.3, the above bandwidth selection procedure yields an "optimal" one up to an unknown constant. Unfortunately, how to identify this constant remains unclear.

Specifically, the DGP is $y_t = g(\tau_t)t^{\theta_0} + \varepsilon_t$, where $\theta_0 = 0.8$, $\varepsilon_t \sim i.i.d.\ N(0,1)$, and $g(u) = 3(u-1)^2 + 1$. For each generated data set, we first estimate θ_0 and g by our nonparametric method proposed in this paper (referred to as NM hereafter). More specifically, we recover θ_0 by (3.6), and estimate $g(\tau_t)$ for $t = \lfloor Th \rfloor + 1, \ldots, T$ by $\widetilde{g}(u) = (uT)^{-\log_T |\widehat{g}(1,\widehat{\theta})|}\widehat{g}(u,\widehat{\theta})$ based on the second result of Lemma 3.1. In addition, we calculate $\frac{\ln T}{\ln |\widehat{\eta}_T|}(\widehat{\theta} - \theta_0) - 1$ in order to further examine Theorem 3.3.

For each generated data, we record three squared errors:

1. $\mathrm{se}_g = \dfrac{1}{T - \lfloor Th \rfloor} \displaystyle\sum_{t=\lfloor Th \rfloor + 1}^{T} (\widetilde{g}(\tau_t) - g(\tau_t))^2$;

2. $\mathrm{se}_\theta = (\widehat{\theta} - \theta_0)^2$;

3. $\mathrm{se}_\theta^* = \left(\frac{\ln T}{\ln |\widehat{\eta}_T|}(\widehat{\theta} - \theta_0) - 1\right)^2$.

We repeat the above procedure J times, and calculate three root mean squared errors (RMSE) by $\mathrm{RMSE}_\theta = \left(\frac{1}{J}\sum_{j=1}^J \mathrm{se}_{\theta,j}\right)^{1/2}$, $\mathrm{RMSE}_\theta^* = \left(\frac{1}{J}\sum_{j=1}^J \mathrm{se}_{\theta,j}^*\right)^{1/2}$ and $\mathrm{RMSE}_g = \left(\frac{1}{J}\sum_{j=1}^J \mathrm{se}_{g,j}\right)^{1/2}$, where $\mathrm{se}_{\theta,j}$, $\mathrm{se}_{\theta,j}^*$ and $\mathrm{se}_{g,j}$ stand for the values of se_θ, se_θ^* and se_g obtained from j^{th} replication, respectively.

For the purpose of comparison, we also recover θ_0 by minimizing (3.2) and (3.3) respectively, and estimate $g(\tau_t)$ for $t = \lfloor Th \rfloor + 1, \ldots, T$ by (3.1) with corresponding estimates of θ_0. In order to put all methods on equal footing, we change (3.2) and (3.3) respectively to

$$Q_T(\theta) = \sum_{t=\lfloor Th \rfloor+1}^{T} \left(y_t - t^{\theta}\widehat{g}(\tau_t, \theta)\right)^2, \tag{4.1}$$

$$Q_T(\alpha|u) = \sum_{t=\lfloor Th \rfloor+1}^{T} \left(y_t - \beta t^{\theta}\right)^2 K_h\left(\tau_t - u\right) \quad \text{with} \quad \alpha = (\beta, \theta) \tag{4.2}$$

in the simulation study. For (4.2), we implement $\left(\widehat{\beta}(u),\ \widehat{\theta}(u)\right) = \text{argmin}_{\alpha} Q_T(\alpha|u)$ to obtain $\{\widehat{\theta}(\tau_t)\ | t = \lfloor Th \rfloor + 1, \ldots, T\}$, and then without losing generality take simple average to calculate $\widehat{\theta} = \dfrac{1}{T - \lfloor Th \rfloor} \sum_{t=\lfloor Th \rfloor+1}^{T} \widehat{\theta}(\tau_t)$. We refer to these two methods as OLS1 and OLS2, respectively, and report their RMSEs in the same way as we defined above.

Below, we set $J = 1000$, $T = 200,\ 500,\ 1000$ and $h = h_{opt},\ T^{-1/3},\ T^{-1/5},\ T^{-1/8}$, where "$h_{opt}$" is referred to as the one selected by the procedure of Remark 4.1. For the other methods, we adopt the same combinations of the bandwidth and sample size but exclude $h = h_{\text{opt}}$. The results are reported in Table 1.

Table 1: *Simulation Results for Section 3*

		RMSE_g			RMSE_θ			RMSE^*_θ		
	$h\backslash T$	200	500	1000	200	500	1000	200	500	1000
NM	h_{opt}	0.0226	0.0112	0.0063	0.1114	0.0960	0.0870	0.1126	0.0962	0.0860
	$T^{-1/3}$	0.0638	0.0443	0.0322	0.1164	0.0983	0.0883	0.0880	0.0830	0.0780
	$T^{-1/5}$	0.1293	0.1209	0.1100	0.1369	0.1114	0.0975	0.0182	0.0258	0.0308
	$T^{-1/8}$	0.1112	0.1387	0.1474	0.1687	0.1349	0.1168	0.0377	0.0341	0.0296
OLS1	$T^{-1/3}$	4.8626	7.0742	9.2029	0.3000	0.3000	0.3000			
	$T^{-1/5}$	4.5374	6.5958	8.6375	0.3000	0.3000	0.3000			
	$T^{-1/8}$	4.4457	6.2569	8.0620	0.3000	0.3000	0.3000			
OLS2	$T^{-1/3}$	4.1406	6.2579	7.8744	0.2740	0.2819	0.2786			
	$T^{-1/5}$	3.6473	5.6527	7.7828	0.2661	0.2780	0.2861			
	$T^{-1/8}$	3.0284	5.1373	7.0172	0.2414	0.2725	0.2819			

As expected, both OLS1 and OLS2 perform rather poorly, and NM method with h_{opt} in general provides relatively good estimates in terms of RMSE_g and RMSE_θ. On the other hand, h_{opt} does not yield the best estimate in terms of RMSE^*_θ, but the difference only happens at the second or third decimal, so negligible.

5 Empirical Study

We now provide a case study by investigating the global mean seal level (GMSL). The data is collected from CSIRO (http://www.cmar.csiro.au/sealevel/index.html), and is recorded in millimetres originally. As shown in Figure 1, the range of raw data covering years 1880 to 2005 is from -169.9 to 37.6, and has a strong time trend. Note that although our model (1.1) and the model of Robinson (2012) (i.e., (5.1) below) are defined on $t = 1, \ldots, T$, both models in fact have $y_0 = 0$ if we allow for $t = 0$. Therefore, we shift the data set vertically to let y_0 (i.e., the value of year 1880) be 0 for better fit.

We first implement the two hypothesis tests of Section 2. The detailed testing procedures are identical to Section 4, so we do not repeat them again for conciseness. Table 2 below summarizes the statistic values of two tests and the corresponding decisions at 95% significant level.

Table 2: Two tests

	Statistic Value	Decision
Parametric Test	4.74	Reject
Nonparametric Test	2.44	Reject

Based on Table 2, we have enough evidences to move on to consider (1.1) for the case where $\theta_0 > 0$ and g is a non-constant function. Hereafter, we always refer to our nonparametric method as NM. We select the bandwidth (referred to as h_{opt}) by the procedure given in Remark 4.1. In order to check the sensitivity of our nonparametric approach, we use two more bandwidths $h_L = h_{opt} - 0.03$ and $h_R = h_{opt} + 0.03$ to implement the nonparametric regression below.

For the purpose of comparison, we also consider a parametric setting of Robinson (2012) (referred to as Para-R hereafter) of the form:

$$y_t = \sum_{j=1}^{d} \beta_j t^{\theta_{0,j}} + \varepsilon_t, \tag{5.1}$$

and estimate $\theta_0 = (\theta_{0,1}, \ldots, \theta_{0,d})'$ and $\beta_0 = (\beta_1, \ldots, \beta_d)'$ of (5.1) by the approach of Robinson (2012). It is noteworthy that how to choose the value of d is still an open question. However, in our study, we always get a warning from Matlab saying "*Matrix is close to singular or badly scaled*" when $d \geqslant 2$. Therefore, we set $d = 1$ throughout this study, which essentially gives the model of Phillips (2007).

We report the estimation results of both methods in Table 3, and plot the estimated g_0 under three choices of bandwidth in Figure 6.[4] It is clear that the estimation results of θ_0 and g_0 are quite stable with respect to the choice of bandwidth.

[4]The 95% confidence interval is drawn under the choice of h_{opt} by using Theorem 3.2 and ignoring the bias term.

Table 3: Estimation Results for Section 4

		θ_0	β_0
NM	$h_{opt} = 0.1020$	0.8533	–
	$h_L = 0.0720$	0.8534	–
	$h_R = 0.1320$	0.8530	–
Para-R	–	1.0000	0.4676

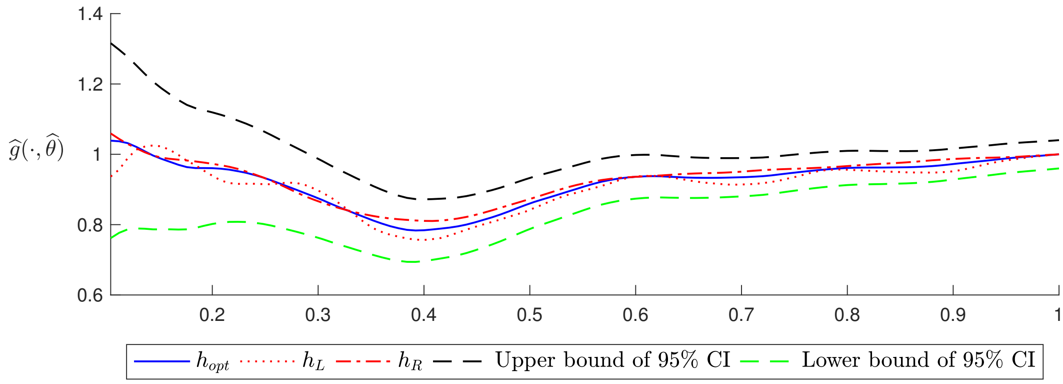

Figure 6: Estimation of g_0 (i.e., $\widehat{g}(\cdot, \widehat{\theta})$)

To further compare the performance of two methods, we finally plot the estimation residuals from both methods in Figure 7 below. It is easy to see that the residual terms of NM indeed are smaller than those of Para-R, which should be expected. As we have rejected the nonparametric test in the beginning of this section, so the model (1.1) potentially can fit the data set in a better fashion.

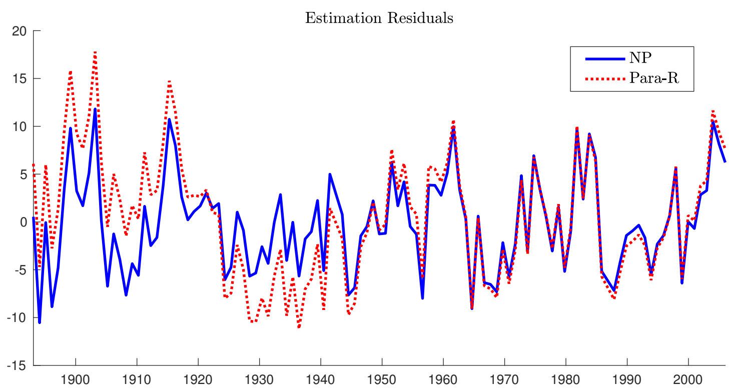

Figure 7: Estimation Residuals

6 Extensions with Discussion

Below, we discuss some potential extensions with the corresponding issues.

Extension 1: Building on Robinson (2012), one intuitive extension might be

$$y_t = \sum_{j=1}^{d} g_j(\tau_t) t^{\theta_{0,j}} + \varepsilon_t \quad \text{with} \quad t = 1, \ldots, T, \tag{6.1}$$

where $g_j(\cdot)$ for $j = 1, \ldots, d$ are unknown functions, and $\theta_0 = (\theta_{0,1}, \ldots, \theta_{0,d})'$ is defined on a compact set $\Theta \subset \mathbb{R}^d$ and satisfies certain restrictions.

However, model (6.1) suffers from an identification issue. To make the explanation clearer and simpler, we now suppose θ_0 is known. Then, for $\forall u \in (0,1)$, the kernel based OLS estimator of $G(u) = (g_1(u), \ldots, g_d(u))'$ is

$$\widehat{G}(u) = \left(\sum_{t=1}^{T} z_t z_t' K_h(u - \tau_t) \right)^{-1} \sum_{t=1}^{T} z_t y_t K_h(u - \tau_t), \tag{6.2}$$

where $z_t = (t^{\theta_{0,1}}, \ldots, t^{\theta_{0,d}})'$. For (6.2), we normalize the matrix in the inverse as follows:

$$D_{\theta_0}^{-1} \sum_{t=1}^{T} z_t z_t' K_h(u - \tau_t) D_{\theta_0}^{-1}, \tag{6.3}$$

where $D_{\theta_0} = \mathrm{diag}\{T^{1/2+\theta_{0,1}}, \ldots, T^{1/2+\theta_{0,d}}\}$. For the $(i,j)^{th}$ element of (6.3) with $1 \leqslant i, j \leqslant d$, we can show that

$$\frac{1}{Th} \sum_{t=1}^{T} \tau_t^{\theta_{0,i}+\theta_{0,j}} K\left(\frac{u - \tau_t}{h} \right) = u^{\theta_{0,i}+\theta_{0,j}} (1 + o(1)) \tag{6.4}$$

after going through a procedure similar to those for Lemma A.2 of this paper. (6.4) indicates that (6.3) can be rewritten as

$$D_{\theta_0}^{-1} \sum_{t=1}^{T} x_t x_t' K_h(u - \tau_t) D_{\theta_0}^{-1} = (u^{\theta_{0,1}}, \ldots, u^{\theta_{0,d}})'(u^{\theta_{0,1}}, \ldots, u^{\theta_{0,d}})(1 + o(1)), \tag{6.5}$$

which is obviously not invertible, i.e., (6.2) is not well defined.

Compared to Robinson (2012), the problem is due to the nonparametric nature of (6.1). The parametric case does not have the kernel function in (6.2), and yields

$$\frac{1}{T} \sum_{t=1}^{T} \tau_t^{\theta_{0,i}+\theta_{0,j}} = \int_0^1 u^{\theta_{0,i}+\theta_{0,j}} du \cdot (1 + o(1)) = \frac{1}{\theta_{0,i} + \theta_{0,j} + 1} \cdot (1 + o(1)). \tag{6.6}$$

Thereby, the limit of $D_{\theta_0}^{-1} \sum_{t=1}^{T} z_t z_t' D_{\theta_0}^{-1}$ is a Cauchy matrix, and is invertible under certain restrictions. Then all the discussions given in Remark 3.2 apply.

Extension 2: One may include some explanatory variables and then consider a generalized trending model of the form:

$$y_t = f(x_t, \tau_t) + g(\tau_t) t^{\theta_0} + \varepsilon_t, \tag{6.7}$$

where x_t is a $d \times 1$ vector including all the observable regressors, $f(\cdot, \cdot)$ is an unknown function, and the other variables are defined in the same way as (1.1). It is worthy pointing out that (1.1) is equivalent to (1) of Vogt (2012) including a time trend, and also nests the following model as special cases:

$$y_t = f(x_t) + g(\tau_t)t^{\theta_0} + \varepsilon_t, \tag{6.8}$$

where (6.8) is similar to (1) of Gao and Hawthorne (2006) and (1.1) of Dong and Linton (2016) but replacing weak trends with strong ones.

However, there are some issues when recovering f. For example, (1) Vogt (2012) argues that $f(x_t, \tau_t)$ suffers the curse of dimensionality, so one can decompose $f(x_t, \tau_t)$ to an additive form $f(x_t, \tau_t) = \sum_{j=1}^{d} f_j(x_{t,j}, \tau_t)$ with $x_t = (x_{t,1}, \ldots, x_{t,d})'$ in order to bypass this issue, which is exactly what Dong and Linton (2016) do in their paper while sieve estimation technique being employed; (2) Phillips et al. (2017) point out that the usual asymptotic methods and limit theory of kernel estimation break down when $f(x_t, \tau_t)$ has a linear form of $f(x_t, \tau_t) = x_t' f(\tau_t)$ with x_t being an integrated process; and so forth. We will leave detailed analysis of $f(\cdot, \cdot)$ to future studies, but we would like to point out that, under some restrictions on $f(\cdot, \cdot)$ and $\{x_t \mid t = 1, \ldots, T\}$, the main results of this paper may still hold after certain modifications on the assumptions and the proofs. A formal statement is given in Appendix B of this paper for the sake of presentation.

7 Conclusions

In summary, this paper provides the practitioner from a variety of fields with a new nonparametric trending method to exam/capture/remove time effects. We firstly study two hypothesis tests. Then we consider the case where both of the null hypotheses get rejected. The consistent estimators and their corresponding asymptotic properties are established in the paper. Moreover, we examine the proposed hypothesis tests, estimation methods through both simulated and real data examples. Finally, we discuss some extensions with corresponding potential issues in the end of this paper, which may guide our future research. Some extra results and simulations are given in Appendix B of this paper. We assume smoothness on g, but it may be possible to extend the methodology to consider a finite number of trend breaks or discontinuities in g, see Delgado and Hidalgo (2000). Likewise the global trend may be subject to some breaks, Bai and Perron (1998).

Appendix A

In this appendix, we firstly introduce some notations and necessary lemmas, before we complete the proofs of the main theorems. It is worthy mentioning that the proof of Theorem 2.1 is relatively straight forward, after we establish Theorem 3.1 to Theorem 3.3. Thus, we leave it in the Appendix B of this paper, although it is the first asymptotic result in the main text. The proof of Theorem 2.2 follows from the development similar to Fan and Li (1996), Li (1999) and Gao and Hawthorne (2006), thus omitted.

Recall that we have denoted $\Lambda_{T,h}(u,\theta) = \sum_{t=1}^{T} t^{2\theta} K_h(u-\tau_t)$ in Theorem 3.2 for notational simplicity. Simple calculation shows that

$$\frac{\partial \widehat{g}(u,\theta)}{\partial \theta} = -2\Lambda_{T,h}^{-2}(u,\theta)\left[\sum_{t=1}^{T}\sum_{s=1}^{T}(t\sqrt{s})^{2\theta} y_s K_h(u-\tau_t)K_h(u-\tau_s)\ln t\right]$$
$$+\Lambda_{T,h}^{-1}(u,\theta)\left[\sum_{t=1}^{T} t^{\theta} y_t K_h(u-\tau_t)\ln t\right] ;$$

$$\frac{\partial^2 \widehat{g}(u,\theta)}{\partial \theta^2} = 8\Lambda_{T,h}^{-3}(u,\theta)\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}(ts\sqrt{r})^{2\theta} y_r K_h(u-\tau_t)K_h(u-\tau_s)K_h(u-\tau_r)(\ln t)(\ln s)\right]$$
$$-4\Lambda_{T,h}^{-2}(u,\theta)\left[\sum_{t=1}^{T}\sum_{s=1}^{T}(t\sqrt{s})^{2\theta} y_s K_h(u-\tau_t)K_h(u-\tau_s)(\ln t)\ln(t\sqrt{s})\right]$$
$$-2\Lambda_{T,h}^{-2}(u,\theta)\left[\sum_{t=1}^{T}\sum_{s=1}^{T}(t\sqrt{s})^{2\theta} y_s K_h(u-\tau_t)K_h(u-\tau_s)(\ln t)(\ln s)\right]$$
$$+\Lambda_{T,h}^{-1}(u,\theta)\left[\sum_{t=1}^{T} t^{\theta} y_t K_h(u-\tau_t)(\ln t)^2\right] ;$$

$$\frac{\partial R_T(\theta)}{\partial \theta} = 4\lambda_T^2\left\{\ln\left[\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)\right]^2\right\}\cdot\left[\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)\right]^{-1}$$
$$\cdot\left\{\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\frac{\partial\widehat{g}(\tau_t,\theta)}{\partial\theta}+\frac{2}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)\ln\tau_t\right\} ;$$

$$\frac{\partial^2 R_T(\theta)}{\partial \theta^2} = -4\lambda_T^2\left\{\ln\left[\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)\right]^2\right\}\cdot\left[\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)\right]^{-2}$$
$$\cdot\left\{\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\frac{\partial\widehat{g}(\tau_t,\theta)}{\partial\theta}+\frac{2}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)\ln\tau_t\right\}^2$$
$$+4\lambda_T^2\left\{\ln\left[\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)\right]^2\right\}\cdot\left[\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)\right]^{-1}$$
$$\cdot\left\{\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\frac{\partial^2\widehat{g}(\tau_t,\theta)}{\partial^2\theta}+\frac{4}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\frac{\partial\widehat{g}(\tau_t,\theta)}{\partial\theta}\ln\tau_t+\frac{4}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)(\ln\tau_t)^2\right\}$$
$$+8\lambda_T^2\left[\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)\right]^{-2}\cdot\left\{\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\frac{\partial\widehat{g}(\tau_t,\theta)}{\partial\theta}+\frac{2}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)\ln\tau_t\right\}^2 . \tag{A.1}$$

Lemma A.1.

1. *Let* $\{X_t, t \geqslant 1\}$ *be a zero-mean* α*-mixing process satisfying* $\Pr(|X_t| \leqslant b) = 1$ *for all* $t \geqslant 1$. *Then for each integer* $q \in [1, \frac{n}{2}]$ *and each* $\epsilon > 0$, *we have*

$$\Pr\left(\left|\sum_{t=1}^{T} X_t\right| > n\epsilon\right) \leqslant 4\exp\left(-8^{-1}\epsilon^2 q[v(q)]^{-2}\right) + 22\left(1 + 4b\epsilon^{-1}\right)^{1/2} q\alpha\left(\lfloor T/(2q)\rfloor\right),$$

where $v^2(q) = \frac{2}{p^2}\sigma^2(q) + \frac{b\epsilon}{2}$ *with* $p = \frac{T}{2q}$ *and*

$$\begin{aligned}\sigma^2(q) = \max_{1\leqslant j\leqslant 2q-1} & E\{(\lfloor jp\rfloor + 1 - jp)X_{\lfloor jp\rfloor+1} + X_{\lfloor jp\rfloor+2} + \cdots + X_{\lfloor (j+1)p\rfloor} \\ & + ((j+1)p - \lfloor (j+1)p\rfloor)X_{\lfloor (j+1)p\rfloor+1}\}^2;\end{aligned}$$

2. $\frac{1}{T}\sum_{t=1}^{T}\ln t = \ln T - 1 + o(1)$, *as* $T \to \infty$.

Proof of Lemma A.1:

(1). The detailed proof can been seen in Bosq (1998), thus omitted here.

(2). Write

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\ln t = \frac{1}{T}\sum_{t=1}^{T}(\ln\tau_t + \ln T) &= \int_0^1 (\ln u)du + o(1) + \ln T \\ &= u(\ln u)\big|_0^1 - \int_0^1 ud(\ln u) + o(1) + \ln T \\ &= -1 + o(1) + \ln T,\end{aligned}$$

where the second equality follows from the definition of Riemann integral.

The proof is now completed. ∎

Lemma A.2. *Let Assumption 1 hold. As* $T \to \infty$,

1. $\sup_{u\in[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}{\tau_t}^{\theta}\varepsilon_t K_h(u-\tau_t)\right| = O_P\left(\sqrt{\frac{\ln T}{Th}}\right)$ *for* $\forall\theta\in\Theta$;

2. $\sup_{(\theta,u)\in\Theta\times[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}{\tau_t}^{\theta}\varepsilon_t K_h(u-\tau_t)\right| = O_P\left(\sqrt{\frac{\ln T}{Th}}\right)$;

3. $\sup_{(\theta,u)\in\Theta\times[0,1]}\left|\frac{1}{T}\sum_{t=1}^{n}{\tau_t}^{\theta}(\ln\tau_t)\varepsilon_t K_h(u-\tau_t)\right| = O_P\left(\frac{(\ln T)^{\frac{3}{2}}}{\sqrt{Th}}\right)$;

4. $\sup_{(\theta,u)\in\Theta\times[h,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta+\theta_0}g(\tau_t)K_h(\tau_t-u) - \widetilde{c}u^{\theta+\theta_0}g(u)\right| = O(h)$, *where*

$$\widetilde{c} = \begin{cases} 1, & u\in[h,1-h] \\ \int_{-1}^{c}K(w)dw, & u = 1-ch\in(1-h,1] \quad (i.e.,\ c\in[0,1)) \end{cases};$$

5. $\sup_{(\theta,u)\in\Theta\times[(1+\epsilon_1)h,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t^{2\theta}K_h(\tau_t-u)-\widetilde{c}u^{2\theta}\right|=O(h^{2c^*})$, *where* $\widetilde{c}$ *and* c^* *have been defined in (4) of this lemma and Assumption 1.1.a respectively, and* ϵ_1 *is a sufficiently small positive constant;*

6. $\sup_{\theta\in U(\theta_0)}|v_T(\theta)-v(\theta)|=o(1)$, *where* $U(\theta_0)$ *is a sufficiently small compact set that* θ_0 *belongs to,* $v_T(\theta)=\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta_0+\theta}g(\tau_t)$ *and* $v(\theta)=\int_0^1 u^{\theta_0+\theta}g(u)du$.

Proof of Lemma A.2:

(1). Let $l(T)$ be any positive function satisfying that $l(T)\to\infty$ as $T\to\infty$. By the arguments same as (B.10) and (B.11) of Chen et al. (2012*b*), it suffices to prove that for $\forall\theta\in\Theta$

$$\sup_{u\in[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t{}^{\theta}\varepsilon_t K_h(u-\tau_t)\right|=o_P\left(l(T)\sqrt{\frac{\ln T}{Th}}\right).$$

In order to do so, we cover $[0,1]$ by finite number of subintervals $\{B_i\}$ that are centred at b_i and of length $\delta_T=o(h^2)$. Denote U_T as the number of such subintervals, which immediately gives that $U_T=O(\delta_T^{-1})$. Write

$$\begin{aligned}
&\sup_{u\in[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta}K_h(u-\tau_t)\varepsilon_t\right|\\
\leqslant&\max_{1\leqslant i\leqslant U_T}\sup_{u\in B_i}\left|\frac{1}{Th}\sum_{t=1}^{T}\tau_t^{\theta}K\left(\frac{u-\tau_t}{h}\right)\varepsilon_t-\frac{1}{Th}\sum_{t=1}^{T}\tau_t^{\theta}K\left(\frac{b_i-\tau_t}{h}\right)\varepsilon_t\right|\\
&+\max_{1\leqslant i\leqslant U_T}\left|\frac{1}{Th}\sum_{t=1}^{T}\tau_t^{\theta}K\left(\frac{b_i-\tau_t}{h}\right)\varepsilon_t\right|\\
:=&\,\Pi_{1T}+\Pi_{2T},
\end{aligned}$$

where the definitions of Π_{1T} and Π_{2T} should be obvious.

Below, we take $\delta_T=[l(T)]^{1+\nu}\cdot\sqrt{\frac{\ln T}{Th}}\cdot h^2$ for a sufficiently small $\nu>0$.

For Π_{1T}, write

$$\begin{aligned}
\Pi_{1T}&=\max_{1\leqslant i\leqslant U_T}\sup_{u\in B_i}\left|\frac{1}{Th}\sum_{t=1}^{T}\tau_t^{\theta}K\left(\frac{u-\tau_t}{h}\right)\varepsilon_t-\frac{1}{Th}\sum_{t=1}^{T}\tau_t^{\theta}K\left(\frac{b_i-\tau_t}{h}\right)\varepsilon_t\right|\\
&=\max_{1\leqslant i\leqslant U_T}\sup_{u\in B_i}\left|\frac{1}{Th}\sum_{t=1}^{T}\tau_t^{\theta}\cdot\frac{u-b_i}{h}\cdot K^{(1)}(u^*)\varepsilon_t\right|\\
&\leqslant O(1)\max_{1\leqslant i\leqslant U_T}\sup_{u\in B_i}\frac{\delta_T}{h^2}\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta}|\varepsilon_t|\\
&=O_P(1)\frac{\delta_T}{h^2}\cdot\int_0^1 u^{\theta}du\cdot E|\varepsilon_t|=O_P\left([l(T)]^{1+\nu}\sqrt{\frac{\ln T}{Th}}\right)=o_P\left(l(T)\sqrt{\frac{\ln T}{Th}}\right),
\end{aligned}$$

where u^* lies between $\frac{u-\tau_t}{h}$ and $\frac{b_i-\tau_t}{h}$; the second equality follows from Mean Value Theorem; the third equality follows from the definition of Riemann integral; and the fourth equality follows from the construction of δ_T.

For Π_{2T}, we use truncation technique, so for the same $\nu>0$ above denote

$$\widetilde{\varepsilon}_t = \varepsilon_t \cdot \mathbb{1}(|\varepsilon_t| \leqslant T^{1/\nu} l(T)) \quad \text{and} \quad \widetilde{\varepsilon}_t^c = \varepsilon_t - \widetilde{\varepsilon}_t,$$

where $\mathbb{1}(\cdot)$ is the indicator function.

Thus, we obtain that

$$\begin{aligned}
\Pi_{2T} \leqslant & \max_{1\leqslant i\leqslant U_T} \left| \frac{1}{Th} \sum_{t=1}^{T} \tau_t^{\theta} K\left(\frac{b_i - \tau_t}{h}\right) \widetilde{\varepsilon}_t \right| + \max_{1\leqslant i\leqslant U_T} \left| \frac{1}{Th} \sum_{t=1}^{T} \tau_t^{\theta} K\left(\frac{b_i - \tau_t}{h}\right) \widetilde{\varepsilon}_t^c \right| \\
:= & \ \Pi_{2T,1} + \Pi_{2T,2},
\end{aligned}$$

where the definitions of $\Pi_{2T,1}$ and $\Pi_{2T,2}$ should be obvious.

For $\Pi_{2T,1}$, observe that

$$\left| \frac{1}{Th} \cdot \tau_t^{\theta} K\left(\frac{b_i - \tau_t}{h}\right) \widetilde{\varepsilon}_t \right| \leqslant O(1) T^{1/\nu - 1} l(T) h^{-1} = O(1)\xi,$$

where $\xi = T^{1/\nu-1} l(T) h^{-1}$.

Then, for any $\epsilon > 0$, letting $l(\cdot)$ satisfying

$$l(T) \to \infty \quad \text{and} \quad \frac{T^{1-2/\nu} h}{[l(T)]^4 \cdot \ln T} \to \infty,$$

and applying Lemma A.1 with

$$q = \frac{T}{2p}, \quad p = \frac{1}{\epsilon [l(T)]^2} \sqrt{\frac{T^{1-2/\nu} h}{\ln T}}, \quad \epsilon_1 = \epsilon T^{-1} l(T) \sqrt{\frac{\ln T}{Th}}, \quad \text{and} \quad \frac{2\sigma^2(q)}{p^2} + \frac{\xi \epsilon_1}{2} \leqslant \frac{O(1)}{T^2 hp},$$

we have

$$\begin{aligned}
& \Pr\left(\Pi_{2T,1} > T\epsilon_1\right) = \Pr\left(\Pi_{2T,1} > \epsilon l(T) \sqrt{\frac{\ln T}{Th}}\right) \\
\leqslant & \ O(1)\delta_T^{-1} \exp\left(- \frac{\epsilon^2 [l(T)]^2 q \frac{\ln T}{T^3 h}}{\frac{O(1)}{T^2 hp}} \right) + O(1)\delta_T^{-1} \left(1 + \frac{4\xi}{\epsilon_1}\right)^{1/2} q\alpha(\lfloor T/(2q) \rfloor) \\
\leqslant & \ O(1)\delta_T^{-1} \exp\left(-O(1)\epsilon^2 [l(T)]^2 \ln T\right) + O(1)\delta_T^{-1} \left(1 + \frac{4\xi}{\epsilon_1}\right)^{1/2} q\alpha(\lfloor T/(2q) \rfloor).
\end{aligned}$$

By exactly the same arguments as those for (B.16) of Chen et al. (2012*b*), we immediately obtain that $\Pi_{2T,1} = o_P\left(l(T)\sqrt{\frac{\ln T}{Th}}\right)$.

For $\Pi_{2T,2}$, write

$$\begin{aligned}
& \Pr\left(\Pi_{2T,2} \geqslant \epsilon l(T) \sqrt{\frac{\ln T}{Th}}\right) \\
= & \Pr\left(\max_{1\leqslant i\leqslant U_T} \left| \frac{1}{Th} \sum_{t=1}^{T} \tau_t^{\theta} K\left(\frac{b_i - \tau_t}{h}\right) \widetilde{\varepsilon}_t^c \right| \geqslant \epsilon l(T) \sqrt{\frac{\ln T}{Th}} \right) \\
\leqslant & \Pr\left(\max_{1\leqslant i\leqslant U_T} \max_{1\leqslant t\leqslant T} \left| \frac{1}{h} \tau_t^{\theta} K\left(\frac{b_i - \tau_t}{h}\right) \widetilde{\varepsilon}_t^c \right| \geqslant \epsilon l(T) \sqrt{\frac{\ln T}{Th}} \right) \\
\leqslant & \Pr\left(\max_{1\leqslant i\leqslant U_T} K\left(\frac{b_i - \tau_t}{h}\right) \max_{1\leqslant t\leqslant T} \left| \tau_t^{\theta} \widetilde{\varepsilon}_t^c \right| \geqslant \epsilon l(T) \sqrt{\frac{h \ln T}{T}} \right)
\end{aligned}$$

$$\leqslant \Pr\left(\max_{1\leqslant t\leqslant T}|\widetilde{\varepsilon}_t^c| \geqslant \epsilon l(T)\sqrt{\frac{h\ln T}{T}}\cdot\frac{1}{\max_{1\leqslant i\leqslant U_T}K\left(\frac{b_i-\tau_t}{h}\right)\max_{1\leqslant t\leqslant T}\left|\tau_t^{\theta}\right|}\right)$$

$$\leqslant \Pr\left(\max_{1\leqslant t\leqslant T}|\widetilde{\varepsilon}_t^c| \geqslant 0\right)$$

$$\leqslant \sum_{t=1}^{T}\Pr\left(|\widetilde{\varepsilon}_t^c|\geqslant 0\right) \leqslant \sum_{t=1}^{T}\Pr\left(|\varepsilon_t|\geqslant T^{1/\nu}l(T)\right)\leqslant \sum_{t=1}^{T}\frac{E|\varepsilon_t|^{\nu}}{[l(T)]^{\nu}T}$$

$$= O(1)\frac{1}{[l(T)]^{\nu}} = o(1).$$

Based on the analysis of $\Pi_{2T,1}$ and $\Pi_{2T,2}$, we have $\Pi_{2T} = o_P\left(l(T)\sqrt{\frac{\ln T}{Th}}\right)$. In connection with the analysis of Π_{1T}, the proof is completed.

(2). We now use Lemma A2 of Newey and Powell (2003) to show the second result of this lemma. It suffices to show that

$$\sup_{(\theta,u)\in\Theta\times[0,1]}\frac{\sqrt{Th}}{l(T)\sqrt{\ln T}}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t{}^{\theta}\varepsilon_t K_h(u-\tau_t)\right| = o_P(1),$$

where $l(T)$ is an arbitrary positive function satisfying that $l(T)\to\infty$ as $T\to\infty$.

Step 1: $\Theta\times[0,1]$ is a compact subspace of $\mathbb{R}^2$ with Euclidean norm, which verifies condition (i) of Lemma A2 of Newey and Powell (2003).

Step 2: For $\forall\theta\in\Theta$, $\sup_{u\in[0,1]}\frac{\sqrt{Th}}{l(T)\sqrt{\ln T}}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t{}^{\theta}\varepsilon_t K_h(u-\tau_t)\right| = o_P(1)$ holds by results (1) of this lemma. Thus, we immediately obtain that for $\forall(\theta,u)\in\Theta\times[0,1]$

$$\frac{\sqrt{Th}}{l(T)\sqrt{\ln T}}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t{}^{\theta}\varepsilon_t K_h(u-\tau_t)\right| = o_P(1)$$

Step 3: Condition (iii) of Lemma A2 of Newey and Powell (2003) holds apparently in this case. Therefore, we conclude that the second result of this lemma holds.

(3). The procedure of proof is the same as (1) and (2) of this lemma, so omitted.

(4). Divide $\Theta\times[h,1]$ into the following two subsets:

$$\begin{cases} \textit{Case 1:}\quad (\theta,u)\in\Theta\times[h,1-h] \\ \textit{Case 2:}\quad (\theta,u)\in\Theta\times(1-h,1],\ \text{i.e., } (\theta,c)\in\Theta\times[0,1) \text{ with } u=1-ch. \end{cases}$$

For *Case 1*, write

$$\sup_{(\theta,u)\in\Theta\times[h,1-h]}\left|\frac{1}{Th}\sum_{t=1}^{T}\tau_t^{\theta+\theta_0}g(\tau_t)K\left(\frac{\tau_t-u}{h}\right)-u^{\theta+\theta_0}g(u)\right|$$

$$= \sup_{(\theta,u)\in\Theta\times[h,1-h]}\left|\frac{1}{h}\int_0^1 w^{\theta+\theta_0}g(w)K\left(\frac{w-u}{h}\right)dw + O\left(\frac{1}{Th}\right)-u^{\theta+\theta_0}g(u)\right|$$

$$= \sup_{(\theta,u)\in\Theta\times[h,1-h]}\left|\int_{-u/h}^{(1-u)/h} m_1(u+wh)K(w)dw + O\left(\frac{1}{Th}\right)-u^{\theta+\theta_0}g(u)\right|$$

$$
\begin{aligned}
&= \sup_{(\theta,u)\in\Theta\times[h,1-h]} \left| \int_{-1}^{1} \left(m_1(u) + m_1^{(1)}(\tilde{u})wh \right) K(w)dw + O\left(\frac{1}{Th}\right) - u^{\theta+\theta_0} g(u) \right| \\
&= \sup_{(\theta,u)\in\Theta\times[h,1-h]} \left| \int_{-1}^{1} m_1^{(1)}(\tilde{u})whK(w)dw + O\left(\frac{1}{Th}\right) \right| \\
&= O(h) + O\left(\frac{1}{Th}\right) = O(h),
\end{aligned}
$$

where $\tilde{u}$ lies between u and $u+wh$; $m_1(u) = u^{\theta+\theta_0} g(u)$; the first equality follows from the definition of Riemann integral; the third equality follows from Taylor expansion and the fact that $K(w)$ is defined on $[-1,1]$; the fifth equality follows from Assumption 1.1.b; and the sixth equality follows from Assumption 1.4.

For *Case 2*, $(\theta,u) \in \Theta \times (1-h,1]$ is equivalent to $(\theta,c) \in \Theta \times [0,1)$ with $u = 1-ch$. Notice that for u^* lies between u and $u+wh$ where $w \in [-1,c]$, we have

$$
1 - 2h \leqslant u - h \leqslant u^* \leqslant u + ch = 1, \tag{A.2}
$$

where the equality follows from the construction of u. Thus, write

$$
\begin{aligned}
&\sup_{(\theta,c)\in\Theta\times[0,1)} \left| \frac{1}{Th}\sum_{t=1}^{T} \tau_t^{\theta+\theta_0} g(\tau_t) K\left(\frac{\tau_t - u}{h}\right) - u^{\theta+\theta_0} g(u) \int_{-1}^{c} K(w)dw \right| \\
&= \sup_{(\theta,c)\in\Theta\times[0,1)} \left| \int_{-u/h}^{(1-u)/h} m_1(u+wh)K(w)dw + O\left(\frac{1}{Th}\right) - u^{\theta+\theta_0} g(u) \int_{-1}^{c} K(w)dw \right| \\
&= \sup_{(\theta,c)\in\Theta\times[0,1)} \left| \int_{-1}^{c} \left(m_1(u) + m_1^{(1)}(\tilde{u})wh \right) K(w)dw + O\left(\frac{1}{Th}\right) - u^{\theta+\theta_0} g(u) \int_{-1}^{c} K(w)dw \right| \\
&= \sup_{(\theta,c)\in\Theta\times[0,1)} \left| \int_{-1}^{c} m_1^{(1)}(\tilde{u})whK(w)dw + O\left(\frac{1}{Th}\right) \right| \\
&= O(h) + O\left(\frac{1}{Th}\right) = O(h),
\end{aligned}
$$

where $\tilde{u}$ lies between u and $u+wh$; $m_1(w) = w^{\theta+\theta_0} g(w)$; the first equality follows from the definition of Riemann integral; the second equality follows from Taylor expansion and the construction of $u = 1 - ch$; the fourth equality follows from (A.2) and Assumption 1.1.b; and the fifth equality follows from Assumption 1.4.

Based on the above analysis, the result follows.

(5). Similar to result (4) of this lemma, divide $\Theta \times [(1+\epsilon_1)h, 1]$ into the following two subsets:

$$
\begin{cases}
\textit{Case 1:} \quad (\theta,u) \in \Theta \times [(1+\epsilon_1)h, 1-h] \\
\textit{Case 2:} \quad (\theta,u) \in \Theta \times (1-h,1], \text{ i.e., } (\theta,c) \in \Theta \times [0,1) \text{ with } u = 1-ch.
\end{cases}
$$

Before considering *Case 1*, note that for u^* lying between u and $u+wh$ with $u \in [(1+\epsilon_1)h, 1-h]$ and $w \in [-1,1]$, we have

$$
\epsilon_1 h \leqslant (1+\epsilon_1 - 1)h \leqslant u - h \leqslant u^* \leqslant u + h \leqslant 1. \tag{A.3}
$$

Thus,

$$\sup_{(\theta,u)\in\Theta\times[(1+\epsilon_1)h,1]}|(u^*)^{2\theta-1}h| = \sup_{\theta\in\Theta}(\epsilon_1 h)^{2c^*-1}h = O(h^{2c^*}), \tag{A.4}$$

where c^* has been defined in Assumption 1.1.a.

Then we are able to write

$$\begin{aligned}
&\sup_{(\theta,u)\in\Theta\times[(1+\epsilon_1)h,1-h]}\left|\frac{1}{Th}\sum_{t=1}^{T}\tau_t^{2\theta}K\left(\frac{\tau_t-u}{h}\right)-u^{2\theta}\right| \\
&= \sup_{(\theta,u)\in\Theta\times[(1+\epsilon_1)h,1-h]}\left|\frac{1}{h}\int_0^1 w^{2\theta}K\left(\frac{w-u}{h}\right)dw + O\left(\frac{1}{Th}\right)-u^{2\theta}\right| \\
&= \sup_{(\theta,u)\in\Theta\times[(1+\epsilon_1)h,1-h]}\left|\int_{-u/h}^{(1-u)/h}(u+wh)^{2\theta}K(w)dw + O\left(\frac{1}{Th}\right)-u^{2\theta}\right| \\
&= \sup_{(\theta,u)\in\Theta\times[(1+\epsilon_1)h,1-h]}\left|\int_{-1}^{1}\left(u^{2\theta}+2\theta\widetilde{u}^{2\theta-1}wh\right)K(w)dw + O\left(\frac{1}{Th}\right)-u^{2\theta}\right| \\
&= \sup_{(\theta,u)\in\Theta\times[(1+\epsilon_1)h,1-h]}\left|\int_{-1}^{1}2\theta\widetilde{u}^{2\theta-1}whK(w)dw + O\left(\frac{1}{Th}\right)\right| \\
&= O(h^{2c^*}) + O\left(\frac{1}{Th}\right) = O(h^{2c^*}),
\end{aligned}$$

where $\widetilde{u}$ lies between u and $u+wh$; the first equality follows from the definition of Riemann integral; the third equality follows from Mean Value Theorem and the fact that $K(w)$ is defined on $[-1,1]$; and the fifth equality follows from (A.4).

For *Case 2*, write

$$\begin{aligned}
&\sup_{(\theta,c)\in\Theta\times[0,1)}\left|\frac{1}{Th}\sum_{t=1}^{T}\tau_t^{2\theta}K\left(\frac{\tau_t-u}{h}\right)-u^{2\theta}\int_{-1}^{c}K(w)dw\right| \\
&= \sup_{(\theta,c)\in\Theta\times[0,1)}\left|\int_{-u/h}^{(1-u)/h}w^{2\theta}K(w)dw + O\left(\frac{1}{Th}\right)-u^{2\theta}\int_{-1}^{c}K(w)dw\right| \\
&= \sup_{(\theta,c)\in\Theta\times[0,1)}\left|\int_{-1}^{c}\left(u^{2\theta}+2\theta\widetilde{u}^{2\theta-1}wh\right)K(w)dw + O\left(\frac{1}{Th}\right)-u^{2\theta}\int_{-1}^{c}K(w)dw\right| \\
&= \sup_{(\theta,c)\in\Theta\times[0,1)}\left|\int_{-1}^{c}2\theta\widetilde{u}^{2\theta-1}whK(w)dw + O\left(\frac{1}{Th}\right)\right| \\
&= O(h) + O\left(\frac{1}{Th}\right) = O(h),
\end{aligned}$$

where $\widetilde{u}$ lies between u and $u+wh$; the first equality follows from the definition of Riemann integral; the second equality follows from Taylor expansion and the construction of $u = 1-ch$; the fourth equality follows from (A.2); and the fifth equality follows from Assumption 1.4.

Therefore, the result follows immediately.

(6). We now consider the sixth result of this lemma.

Step 1: $U(\theta_0)$ is a compact subspace of $\mathbb{R}$ with Euclidean norm, which verifies condition (i) of Lemma A2 of Newey and Powell (2003).

Step 2: For $\forall\theta\in U(\theta_0)$, it is easy to know $v_T(\theta)-v(\theta) = o(1)$ by the definition of Riemann integral.

Step 3: Note that by keeping using integration by parts, it is easy to know $\int_0^1(\ln u)^4du < \infty$. We now verify the continuity of $v(\theta)$, and write

$$
\begin{aligned}
|v(\theta_1) - v(\theta_2)| &= \left| \int_0^1 (u^{\theta_0+\theta_1} - u^{\theta_0+\theta_2}) g(u) du \right| = \left| (\theta_1 - \theta_2) \cdot \int_0^1 u^{\theta^*} g(u)(\ln u) du \right| \\
&\leqslant |\theta_1 - \theta_2| \left\{ \int_0^1 u^{2\theta^*} du \cdot \int_0^1 g^2(u)(\ln u)^2 du \right\}^{1/2} \\
&= |\theta_1 - \theta_2| \left\{ \frac{1}{2\theta^* + 1} u^{2\theta^*+1} \Big|_0^1 \right\}^{1/2} \left\{ \int_0^1 g^2(u)(\ln u)^2 du \right\}^{1/2} \\
&= |\theta_1 - \theta_2| \left\{ \frac{1}{2\theta^* + 1} u^{2\theta^*+1} \Big|_0^1 \right\}^{1/2} \left\{ \int_0^1 g^4(u) du \cdot \int_0^1 (\ln u)^4 du \right\}^{1/4} \\
&= O(|\theta_1 - \theta_2|), \qquad (A.5)
\end{aligned}
$$

where θ^* lies between $\theta_0 + \theta_1$ and $\theta_0 + \theta_2$; the second equality follows from Mean Value Theorem; the first inequality follows from Cauchy Schwarz inequality; the fifth equality follows from Assumption 1.1.b and the fact that we point out in the beginning of *Step 3*. In connection with *Step 2*, we obtain $|v_T(\theta_1) - v_T(\theta_2)| \leqslant O(1)|\theta_1 - \theta_2|$, which verifies condition (iii) of Lemma A2 of Newey and Powell (2003).

Then the proof is completed. ■

Proof of Lemma 3.1:

(1). Write

$$
\begin{aligned}
&\sup_{(\theta,u)\in B_T(\theta_0)\times B_{\epsilon_1}(h)} \left| \widehat{g}(u,\theta) - (uT)^{\theta_0-\theta} g(u) \right| \\
\leqslant &\sup_{(\theta,u)\in B_T(\theta_0)\times B_{\epsilon_1}(h)} \frac{1}{T^\theta} \left| \left(\frac{1}{T} \sum_{t=1}^T \tau_t^{2\theta} K_h(u-\tau_t) \right)^{-1} \frac{1}{T} \sum_{t=1}^T \tau_t^{\theta} \varepsilon_t K_h(u-\tau_t) \right| \\
&+ \sup_{(\theta,u)\in B_T(\theta_0)\times B_{\epsilon_1}(h)} T^{\theta_0-\theta} \left| \left(\frac{1}{T} \sum_{t=1}^T \tau_t^{2\theta} K_h(u-\tau_t) \right)^{-1} \frac{1}{T} \sum_{t=1}^T \tau_t^{\theta+\theta_0} g(\tau_t) K_h(u-\tau_t) - (uT)^{\theta_0-\theta} g(u) \right| \\
:= &A_1 + A_2,
\end{aligned}
$$

where the definitions of A_1 and A_2 should be obvious.

Firstly, note that one simple fact is

$$
\sup_{\theta\in B_T(\theta_0)} \left(\frac{1}{h} \right)^{\theta-\theta_0} \leqslant \sup_{\theta\in B_T(\theta_0)} T^{|\theta-\theta_0|} = O(1). \qquad (A.6)
$$

We then consider A_1 to A_3 one by one. Start from A_1.

$$
\begin{aligned}
A_1 &= O_P\left(\sqrt{\frac{\ln T}{Th}} \right) \sup_{(\theta,u)\in B_T(\theta_0)\times[(1+\epsilon_1)h,1]} T^{-\theta} \left([u^{2\theta} + O(h)]^{-1} \right) \\
&\leqslant O_P\left(\sqrt{\frac{\ln T}{Th}} \right) \left\{ \sup_{\theta\in B_T(\theta_0)} h^{-\theta} \right\} \left\{ \sup_{\theta\in B_T(\theta_0)} (Th)^{-\theta} \right\} \\
&= O_P\left(\sqrt{\frac{\ln T}{Th}} \right) T^{-\theta_0} h^{-2\theta_0} \left\{ \sup_{\theta\in B_T(\theta_0)} h^{\theta_0-\theta} \right\} \left\{ \sup_{\theta\in B_T(\theta_0)} (Th)^{\theta_0-\theta} \right\} \\
&= O\left(\frac{\sqrt{\ln T}}{T^{\frac{1}{2}+\theta_0} h^{\frac{1}{2}+2\theta_0}} \right), \qquad (A.7)
\end{aligned}
$$

where the first equality follows from (2) and (5) of Lemma A.2; the first inequality follows from Assumption 1.1.a; and the third equality follows from (A.6).

By (4) of Lemma A.2, write

$$A_2 = O(h) \sup_{\theta \in B_T(\theta_0)} T^{\theta_0 - \theta} = O(h). \tag{A.8}$$

By (A.7) and (A.8), the proof is now complete. ■

As explained in Remark 3.1, the number of observations lying between $\lfloor Th \rfloor$ and $\lfloor T(1+\epsilon_1)h \rfloor$ is limited and negligible, as ϵ_1 is an arbitrary small positive constant. Thus, with a bit abuse notation, we define (3.5) by using observations from $\lfloor Th \rfloor + 1, \ldots, T$ throughout the following proofs.

Proof of Theorem 3.1:

(1). Note that $R_T(\theta) = \lambda_T^2 \cdot R_T^*(\theta)$, where $R_T^*(\theta) = \left\{ \ln \left[\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^{T} \tau_t^{2\theta} \widehat{g}(\tau_t, \theta) \right]^2 \right\}^2$. As λ_T is independent of θ, we simply focus on $R_T^*(\theta)$ below. More specifically, we show that for any given $\epsilon > 0$, there exists a sufficiently large positive constant C such that

$$\liminf_T \Pr \left\{ R_T^*(\theta_0 + \lambda_T C) > R_T^*(\theta_0) \right\} \geqslant 1 - \epsilon, \tag{A.9}$$

$$\liminf_T \Pr \left\{ R_T^*(\theta_0 - \lambda_T C) > R_T^*(\theta_0) \right\} \geqslant 1 - \epsilon. \tag{A.10}$$

Both (A.9) and (A.10) holding true implies with probability at least $1 - \epsilon$ that there exists a local minimum in the interval $U_T(\theta_0) = [\theta_0 - \lambda_T C, \theta_0 + \lambda_T C]$. Hence, there exists a local minimizer such that $\widehat{\theta} - \theta_0 = O_P(\lambda_T)$. The above argument is in line with the same spirit of the proofs of Theorem 1 of Fan and Li (2001) and Lemma A.1 of Wang and Xia (2009).

Write

$$\begin{aligned}
& R_T^*(\theta) - R_T^*(\theta_0) \\
= & \left\{ \ln \left[\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^{T} \tau_t^{2\theta} \widehat{g}(\tau_t, \theta) \right]^2 \right\}^2 - \left\{ \ln \left[\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^{T} \tau_t^{2\theta_0} \widehat{g}(\tau_t, \theta_0) \right]^2 \right\}^2 \\
= & \left\{ \ln \left[\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^{T} \tau_t^{2\theta} \left[(\tau_t T)^{\theta_0 - \theta} g(\tau_t) + o_P(1) \right] \right]^2 \right\}^2 \\
& - \left\{ \ln \left[\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^{T} \tau_t^{2\theta_0} \left[g(\tau_t) + o_P(1) \right] \right]^2 \right\}^2 \\
= & \left\{ 2(\theta_0 - \theta) \ln T + \ln \left[\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^{T} \tau_t^{\theta_0 + \theta} g(\tau_t) + o_P(1) \right]^2 \right\}^2 \\
& - \left\{ \ln \left[\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^{T} \tau_t^{2\theta_0} g(\tau_t) + o_P(1) \right]^2 \right\}^2 \\
= & \; 4(\theta_0 - \theta)^2 (\ln T)^2 + 2(\theta_0 - \theta)(\ln T) \cdot \ln \left[\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^{T} \tau_t^{\theta_0 + \theta} g(\tau_t) + o_P(1) \right]^2
\end{aligned}$$

$$
\begin{aligned}
&+\left\{\ln\left[\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{\theta_0+\theta}g(\tau_t)+o_P(1)\right]^2\right\}^2-\left\{\ln\left[\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta_0}g(\tau_t)+o_P(1)\right]^2\right\}^2\\
&:=4B_{1T}(\theta)+2B_{2T}(\theta)+B_{3T}(\theta)-B_{4T}(\theta_0),
\end{aligned}
$$

where the definitions of $B_{1T}(\theta)$, $B_{2T}(\theta)$, $B_{3T}(\theta)$ and $B_{4T}(\theta_0)$ should be obvious; and the second equality follows from Lemma 3.1.

Notice that, for $\left|\int_h^1 u^{\theta_0+\theta}g(u)du\right|^2$, the following two expressions hold uniformly in $\theta\in\Theta$:

$$
\begin{aligned}
\left|\int_h^1 u^{\theta_0+\theta}g(u)du\right|^2 &\geqslant A_1^2\left|\int_h^1 u^{\theta_0+\theta}du\right|^2=A_1^2\left(\frac{1}{\theta_0+\theta+1}u^{\theta_0+\theta+1}\Big|_h^1\right)^2\\
&\geqslant\frac{1}{2\sup_{\theta\in\Theta}(\theta_0+\theta+1)}A_1^2\geqslant K_1>0,
\end{aligned}\tag{A.11}
$$

and

$$
\begin{aligned}
\left|\int_h^1 u^{\theta_0+\theta}g(u)du\right|^2 &\leqslant\int_0^1 u^{2(\theta_0+\theta)}du\int_0^1 g^2(u)du\leqslant A_2^2\int_0^1 u^{2(\theta_0+\theta)}du\\
&=\frac{A_2^2\cdot u^{2(\theta_0+\theta)+1}\big|_0^1}{2(\theta_0+\theta)+1}\leqslant\frac{1}{2\inf_{\theta\in\Theta}(\theta_0+\theta)+1}2A_2^2\leqslant K_2<\infty,
\end{aligned}\tag{A.12}
$$

where A_1 and A_2 are defined in Lemma 1.1.b; K_1 and K_2 indeed exist and are two finite positive constants due to the compactness of Θ and Assumption 1.1.a.

Thus, it is easy to know that $B_{2T}(\theta)=O_P(|\theta_0-\theta|\cdot(\ln T))$. Similarly, we can show that $B_{3T}(\theta)=O_P(1)$ uniformly in θ. $B_{4T}(\theta_0)$ is independent of θ, so ignored.

Based on the above analysis, we immediately obtain that for $\theta=\theta_0\pm\lambda_T C$

$$
R_T^*(\theta)-R_T^*(\theta_0)=4C^2\pm 2C\cdot O_P(1)+O_P(1),
$$

which immediately indicates that (A.9) and (A.10) hold true with sufficiently large C.

The proof of the first result is now completed.

(2). By Lemma 3.1, the second result follows similarly. ■

Proof of Theorem 3.2:

In order to establish the normality of $g(u)$ for $\forall u\in(0,1)$, write

$$
\begin{aligned}
&|\widehat{g}(1,\widehat{\theta})|^{-1}\cdot\widehat{g}(u,\widehat{\theta})-g(u)\\
=\;&|\widehat{g}(1,\widehat{\theta})|^{-1}\cdot\left(\sum_{t=1}^{T}t^{2\widehat{\theta}}K_h(u-\tau_t)\right)^{-1}\sum_{t=1}^{T}t^{\widehat{\theta}+\theta_0}g(\tau_t)K_h(u-\tau_t)-g(u)\\
&+|\widehat{g}(1,\widehat{\theta})|^{-1}\cdot\left(\sum_{t=1}^{T}t^{2\widehat{\theta}}K_h(u-\tau_t)\right)^{-1}\sum_{t=1}^{T}t^{\widehat{\theta}}\varepsilon_t K_h(u-\tau_t)\\
:=\;&A_1+A_2,
\end{aligned}
$$

where the definitions of A_1 and A_2 should be obvious.

Note that by (4) and (5) of Lemma A.2, it is easy to know that $A_1 = O_P(h) + O_P\left(\frac{1}{Th}\right)$. After imposing the conditions in the body of this theorem, it is easy to show that A_1 will have a faster rate $O_P(h^2)$.

We now focus on the normalized version of $\frac{1}{T}\sum_{t=1}^{T} t^{\widehat{\theta}} \varepsilon_t K_h(u-\tau_t)$ and write

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} \tau_t^{\widehat{\theta}} \varepsilon_t K_h(u-\tau_t) \\
=& \frac{1}{T}\sum_{t=1}^{T} \tau_t^{\theta_0} \varepsilon_t K_h(u-\tau_t) + \frac{1}{T}\sum_{t=1}^{T}\left(\tau_t^{\widehat{\theta}} - \tau_t^{\theta_0}\right)\varepsilon_t K_h(u-\tau_t) \\
:=& B_1 + B_2.
\end{aligned}$$

By (3) of Lemme A.2, we know that

$$\begin{aligned}
B_2 &= \frac{1}{T}\sum_{t=1}^{T}(\widehat{\theta}-\theta_0)\tau_t^{\theta^*}(\ln \tau_t)\varepsilon_t K_h(u-\tau_t) \\
&= (\widehat{\theta}-\theta_0)\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta^*}(\ln \tau_t)\varepsilon_t K_h(u-\tau_t) \\
&= O_P\left((\widehat{\theta}-\theta_0)\cdot\frac{(\ln T)^{\frac{3}{2}}}{\sqrt{Th}}\right).
\end{aligned}$$

Also, by standard argument of time series analysis (e.g., proof of Theorem 2 of Cai (2007)), we can prove that

$$\sqrt{Th}B_1 \to_D N(0,\Sigma^*),$$

where $\Sigma^* = \lim_{T\to\infty}\frac{1}{Th}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_t^{\theta_0}\tau_s^{\theta_0}K\left(\frac{w-\tau_t}{h}\right)K\left(\frac{w-\tau_s}{h}\right)E[\varepsilon_t\varepsilon_s]$.

Further note that we have

$$\begin{aligned}
&\frac{1}{Th}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_t^{\theta_0}\tau_s^{\theta_0}K\left(\frac{u-\tau_t}{h}\right)K\left(\frac{u-\tau_s}{h}\right)E[\varepsilon_t\varepsilon_s] \\
=& \frac{1}{Th}\sum_{t=1}^{T}t^{2\theta_0}K^2\left(\frac{u-\tau_t}{h}\right)E[\varepsilon_1^2] + \frac{1}{Th}\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T}\tau_t^{\theta_0}\tau_s^{\theta_0}K\left(\frac{u-\tau_t}{h}\right)K\left(\frac{u-\tau_s}{h}\right)E[\varepsilon_t\varepsilon_s] \\
:=& V_{1T} + V_{2T}.
\end{aligned} \tag{A.13}$$

By Assumption 2, we then have as $T\to\infty$

$$V_{1T} = (1+o(1))\sigma_\varepsilon^2 u^{2\theta_0}\int_{\frac{w-1}{h}}^{\frac{w}{h}} K^2(w)dw, \tag{A.14}$$

$$\begin{aligned}
|V_{2T}| &= \frac{T}{h}\left|\int_0^1\int_0^1 v_1^{\theta_0}v_2^{\theta_0}K\left(\frac{u-v_1}{h}\right)K\left(\frac{u-v_2}{h}\right)\gamma((v_2-v_1)T)\,dv_1\,dv_2\right| \\
&= (1+o(1))(Th)\left|\int_{\frac{u-1}{h}}^{\frac{u}{h}}\int_{\frac{u-1}{h}}^{\frac{u}{h}}(u-xh)^{\theta_0}(u-yh)^{\theta_0}K(x)K(y)\gamma((x-y)Th)\,dy\,dx\right| \\
&\leqslant (1+o(1))\int_{\frac{u-1}{h}}^{\frac{u}{h}}\int_{\frac{u-1}{h}}^{\frac{u}{h}}|u-xh|^{\theta_0}|u-yh|^{\theta_0}K(x)K(y)\,(Th)\,|\gamma((x-y)\,Th)|\,I[x\neq y]\,dydx
\end{aligned}$$

$$= o(1). \tag{A.15}$$

It is easy to show that $V_{1T} = (1 + o(1))\, \sigma_\varepsilon^2 u^{2\theta_0} \int_{-1}^1 K^2(w)dw$. By the identical development for the term I_2 on page 182 of Cai (2007), we have $V_{2T} = o(1)$.

Further note that

$$|\widehat{g}(1, \widehat{\theta})| = T^{\theta_0 - \widehat{\theta}} \to_P \left| \int_0^1 u^{2\theta_0} g(u)du \right|^{-1}, \tag{A.16}$$

where the last step follows from the first result of Theorem 3.3 immediately (the details are temporarily omitted for now).

Also, simple calculation yields

$$\begin{aligned} \widehat{\eta}_T &= \frac{1}{T} \sum_{t=\lfloor Th \rfloor+1}^{T} \tau_t^{2\widehat{\theta}} g(\tau_t) + \frac{1}{T} \sum_{t=\lfloor Th \rfloor+1}^{T} \tau_t^{2\widehat{\theta}} \left(\widetilde{g}(\tau_t) - g(\tau_t)\right) \\ &= \frac{1}{T} \sum_{t=\lfloor Th \rfloor+1}^{T} \tau_t^{2\widehat{\theta}} g(\tau_t) + o_P(1) = \int_0^1 u^{2\theta_0} g(u)du + o_P(1), \end{aligned} \tag{A.17}$$

where $\widetilde{g}$ has been defined in the body of this theorem, and the last equality follows from similar development of (A.18).

Thus, the result follows based on the above analyses. ■

Remark A.1. *Recall that we have defined $v_T(\cdot)$ and $v(\cdot)$ in (6) of Lemma A.2, so write*

$$\left| v_T(\widehat{\theta}) - v(\theta_0) \right| \leqslant \left| v_T(\widehat{\theta}) - v(\widehat{\theta}) \right| + \left| v(\widehat{\theta}) - v(\theta_0) \right| = o_P(1),$$

where $\left| v_T(\widehat{\theta}) - v(\widehat{\theta}) \right| = o_P(1)$ follows from (6) of Lemma A.2, and $\left| v(\widehat{\theta}) - v(\theta_0) \right| = o_P(1)$ follows from (A.5).

By Theorem 3.1, we have $|\widehat{\theta} - \theta| \ln T = O_P(1)$. Then the next limit indeed exists:

$$\phi_1 = \operatorname*{plim}_{T\to\infty} T^{\theta_0 - \widetilde{\theta}} \cdot \frac{1}{T} \sum_{t=\lfloor Th \rfloor+1}^{T} \tau_t^{\theta_0 + \widetilde{\theta}} g(\tau_t) = \widetilde{\alpha}_0 \int_0^1 u^{2\theta_0} g(u)du, \tag{A.18}$$

where $\widetilde{\theta}$ is defined in (3.8), and lies between $\widehat{\theta}$ and θ_0; and $\widetilde{\alpha}_0 = \operatorname*{plim}_{T\to\infty} T^{\theta_0 - \widetilde{\theta}}$.

Similarly, the next two limits exist:

$$\phi_2 = \operatorname*{plim}_{T\to\infty} T^{\theta_0 - \widetilde{\theta}} \cdot \frac{1}{T} \sum_{t=\lfloor Th \rfloor+1}^{T} \tau_t^{\theta_0 + \widetilde{\theta}} g(\tau_t) \ln \tau_t = \widetilde{\alpha}_0 \int_0^1 u^{2\theta_0} g(u)(\ln u)du, \tag{A.19}$$

$$\phi_3 = \operatorname*{plim}_{T\to\infty} T^{\theta_0 - \widetilde{\theta}} \cdot \frac{1}{T} \sum_{t=\lfloor Th \rfloor+1}^{T} \tau_t^{\theta_0 + \widetilde{\theta}} g(\tau_t)(\ln \tau_t)^2 = \widetilde{\alpha}_0 \int_0^1 u^{2\theta_0} g(u)(\ln u)^2 du, . \tag{A.20}$$

With (A.18) to (A.20) in hand, we are now ready to provide the next lemma.

Lemma A.3. *Under Assumption 1, as $T \to \infty$,*

1. $\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\frac{\partial^2\widehat{g}(\tau_t,\theta)}{\partial\theta^2}\Big|_{\theta=\widetilde{\theta}} = (\ln T)^2\phi_1 + 2(\ln T)\phi_2 + \phi_3 + o_P(1)$,

2. $\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\frac{\partial\widehat{g}(\tau_t,\theta)}{\partial\theta}\Big|_{\theta=\widetilde{\theta}} = -(\ln T)\phi_1 - \phi_2 + o_P(1)$,

3. $\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\frac{\partial\widehat{g}(\tau_t,\theta)}{\partial\theta}\ln\tau_t\Big|_{\theta=\widetilde{\theta}} = -(\ln T)\phi_2 - \phi_3 + o_P(1)$,

4. $\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)\big|_{\theta=\widetilde{\theta}} = \phi_1 + o_P(1)$,

5. $\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)\ln\tau_t\big|_{\theta=\widetilde{\theta}} = \phi_2 + o_P(1)$,

6. $\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)(\ln\tau_t)^2\big|_{\theta=\widetilde{\theta}} = \phi_3 + o_P(1)$,

7. $\frac{\partial^2 R_T(\theta)}{\partial\theta^2}\Big|_{\theta=\widetilde{\theta}} = 8 + o_P(1)$,

where ϕ_1 to ϕ_3 are defined by (A.18) to (A.20) respectively; and $\widetilde{\theta}$ is defined in (3.8).

Proof of Lemma A.3:

(1). Recall that we have defined $\frac{\partial^2\widehat{g}(u,\theta)}{\partial\theta^2}$ and $\Lambda_{T,h}(u,\theta)$ in the beginning of Appendix A. Write

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\frac{\partial^2\widehat{g}(\tau_t,\theta)}{\partial\theta^2}\Big|_{\theta=\widetilde{\theta}}\\
=\ &\frac{8}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\widetilde{\theta}}\Lambda_{T,h}^{-3}(\tau_t,\widetilde{\theta})\left[\sum_{u=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}(us\sqrt{r})^{2\widetilde{\theta}}y_rK_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)K_h(\tau_t-\tau_r)(\ln u)(\ln s)\right]\\
&-\frac{4}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\widetilde{\theta}}\Lambda_{T,h}^{-2}(\tau_t,\widetilde{\theta})\left[\sum_{r=1}^{T}\sum_{s=1}^{T}(r\sqrt{s})^{2\widetilde{\theta}}y_sK_h(\tau_t-\tau_r)K_h(\tau_t-\tau_s)(\ln r)\ln(r\sqrt{s})\right]\\
&-\frac{2}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\widetilde{\theta}}\Lambda_{T,h}^{-2}(\tau_t,\widetilde{\theta})\left[\sum_{r=1}^{T}\sum_{s=1}^{T}(r\sqrt{s})^{2\widetilde{\theta}}y_sK_h(\tau_t-\tau_r)K_h(\tau_t-\tau_s)(\ln r)(\ln s)\right]\\
&+\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\widetilde{\theta}}\Lambda_{T,h}^{-1}(\tau_t,\widetilde{\theta})\left[\sum_{s=1}^{T}s^{\widetilde{\theta}}y_sK_h(\tau_t-\tau_s)(\ln s)^2\right]\\
:=\ &8A_1 - 4A_2 - 2A_3 + A_4,
\end{aligned}
$$

where the definitions of A_1 to A_4 should be obvious.

We now consider A_1 to A_4 one by one. Firstly, further decompose A_1 as follows:

$$
A_1 = \frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\widetilde{\theta}}T^{-6\widetilde{\theta}-3}\left[\frac{1}{T}\sum_{s=1}^{T}\tau_s^{2\widetilde{\theta}}K_h(\tau_t-\tau_s)\right]^{-3}
$$

$$
\begin{aligned}
&\cdot T^{5\widetilde{\theta}+\theta_0+3}\left[\frac{1}{T^3}\sum_{u=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}(\tau_u\tau_s)^{2\widetilde{\theta}}\tau_r^{\widetilde{\theta}+\theta_0}g(\tau_r)K_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)K_h(\tau_t-\tau_r)(\ln u)(\ln s)\right]\\
&+\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\widetilde{\theta}}T^{-6\widetilde{\theta}-3}\left[\frac{1}{T}\sum_{s=1}^{T}\tau_s^{2\widetilde{\theta}}K_h(\tau_t-\tau_s)\right]^{-3}\\
&\cdot T^{5\widetilde{\theta}+3}\left[\frac{1}{T^3}\sum_{u=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}(\tau_u\tau_s)^{2\widetilde{\theta}}\tau_r^{\widetilde{\theta}}\varepsilon_rK_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)K_h(\tau_t-\tau_r)(\ln u)(\ln s)\right]\\
:=\;& A_{11}+A_{12},
\end{aligned}
$$

where the definitions of A_{11} and A_{12} should be clear.

For A_{11}, write

$$
\begin{aligned}
A_{11}=\;&\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\widetilde{\theta}}T^{-6\widetilde{\theta}-3}\left[\frac{1}{T}\sum_{s=1}^{T}\tau_s^{2\widetilde{\theta}}K_h(\tau_t-\tau_s)\right]^{-3}\\
&\cdot T^{5\widetilde{\theta}+\theta_0+3}\left[\frac{1}{T^3}\sum_{u=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}(\tau_u\tau_s)^{2\widetilde{\theta}}\tau_r^{\widetilde{\theta}+\theta_0}g(\tau_r)K_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)K_h(\tau_t-\tau_r)(\ln u)(\ln s)\right]\\
=\;&T^{\theta_0-\widetilde{\theta}}(1+O_P(h^{2c^*}))\cdot\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\widetilde{\theta}}\tau_t^{-6\widetilde{\theta}}\\
&\cdot\left[\frac{1}{T}\sum_{u=1}^{T}\tau_u^{2\widetilde{\theta}}(\ln u)K_h(\tau_t-\tau_u)\right]^2\left[\frac{1}{T}\sum_{u=1}^{T}\tau_u^{\widetilde{\theta}+\theta_0}g(\tau_u)K_h(\tau_t-\tau_u)\right]\\
=\;&T^{\theta_0-\widetilde{\theta}}(1+O_P(h^{2c^*}))\cdot\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{\theta_0-3\widetilde{\theta}}g(\tau_t)\left[\frac{1}{T}\sum_{u=1}^{T}\tau_u^{2\widetilde{\theta}}(\ln\tau_u+\ln T)K_h(\tau_t-\tau_u)\right]^2\\
=\;&T^{\theta_0-\widetilde{\theta}}(\ln T)^2(1+O_P(h^{2c^*}))\cdot\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{\theta_0-3\widetilde{\theta}}g(\tau_t)\left[\frac{1}{T}\sum_{u=1}^{T}\tau_u^{2\widetilde{\theta}}K_h(\tau_t-\tau_u)\right]^2\\
&+2T^{\theta_0-\widetilde{\theta}}(\ln T)(1+O_P(h^{2c^*}))\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{\theta_0-3\widetilde{\theta}}g(\tau_t)\\
&\cdot\left[\frac{1}{T}\sum_{u=1}^{T}\tau_u^{2\widetilde{\theta}}(\ln\tau_u)K_h(\tau_t-\tau_u)\right]\left[\frac{1}{T}\sum_{u=1}^{T}\tau_u^{2\widetilde{\theta}}K_h(\tau_t-\tau_u)\right]\\
&+T^{\theta_0-\widetilde{\theta}}(1+O_P(h^{2c^*}))\cdot\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{\theta_0-3\widetilde{\theta}}g(\tau_t)\left[\frac{1}{T}\sum_{u=1}^{T}\tau_u^{2\widetilde{\theta}}(\ln\tau_u)K_h(\tau_t-\tau_u)\right]^2\\
=\;&T^{\theta_0-\widetilde{\theta}}(\ln T)^2(1+O_P(h^{2c^*}))\cdot\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{\theta_0+\widetilde{\theta}}g(\tau_t)\\
&+2T^{\theta_0-\widetilde{\theta}}(\ln T)(1+O_P(h^{2c^*}))\cdot\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{\theta_0+\widetilde{\theta}}g(\tau_t)\ln\tau_t\\
&+T^{\theta_0-\widetilde{\theta}}(1+O_P(h^{2c^*}))\cdot\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{\theta_0+\widetilde{\theta}}g(\tau_t)(\ln\tau_t)^2\\
=\;&(\ln T)^2\phi_1+2(\ln T)\phi_2+\phi_3+o_P(1),
\end{aligned}
\tag{A.21}
$$

where the second, third and fifth equalities follow from (4) and (5) of Lemma A.2; and the last equality follows from (A.18) to (A.20) and the definition of Riemann integral.

Similar to the analysis of A_{11}, we have

$$
\begin{aligned}
A_{12} &= O_P(1)T^{-\widetilde{\theta}}(\ln T)^2 \cdot \frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\widetilde{\theta}}\left[\frac{1}{T}\sum_{s=1}^{T}\tau_s^{2\widetilde{\theta}}K_h(\tau_t-\tau_s)\right]^{-3} \\
&\quad \cdot\left[\frac{1}{T^3}\sum_{u=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}(\tau_u\tau_s)^{2\widetilde{\theta}}\tau_r^{\widetilde{\theta}}\varepsilon_r K_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)K_h(\tau_t-\tau_r)\right] \\
&= O_P(1)T^{-\theta_0}T^{\theta_0-\widetilde{\theta}}(\ln T)^2 \cdot \frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\left[\frac{1}{T}\sum_{r=1}^{T}\tau_r^{\widetilde{\theta}}\varepsilon_r K_h(\tau_t-\tau_r)\right] \\
&= O_P\left(T^{-\theta_0}(\ln T)^2\frac{\sqrt{\ln T}}{\sqrt{Th}}\right) = O_P\left(\frac{1}{T^{\theta_0}}\cdot\frac{(\ln T)^{5/2}}{\sqrt{Th}}\right),
\end{aligned}
$$

where the second equality follows from (5) of Lemma A.2; and the third equality follows from (2) of Lemma A.2 and Theorem 3.1.

Based on the development of A_{11} and A_{12}, we immediately obtain that

$$
A_1 = (\ln T)^2\phi_1 + 2(\ln T)\phi_2 + \phi_3 + o_P(1).
$$

Similarly, we have

$$
\begin{aligned}
A_2 &= \frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T} t^{2\widetilde{\theta}}T^{-4\widetilde{\theta}-2}\left[\frac{1}{T}\sum_{s=1}^{T}\tau_s^{2\widetilde{\theta}}K_h(\tau_t-\tau_s)\right]^{-2} \\
&\quad \cdot T^{3\widetilde{\theta}+\theta_0+2}\left[\frac{1}{T^2}\sum_{r=1}^{T}\sum_{s=1}^{T}\tau_r^{2\widetilde{\theta}}\tau_s^{\widetilde{\theta}}y_s K_h(\tau_t-\tau_r)K_h(\tau_t-\tau_s)(\ln r)\left(\ln r+\frac{1}{2}\ln s\right)\right] \\
&= \frac{3}{2}\left[(\ln T)^2\phi_1 + 2(\ln T)\phi_2 + \phi_3\right] + o_P(1), \\
A_3 &= (\ln T)^2\phi_1 + 2(\ln T)\phi_2 + \phi_3 + o_P(1), \\
A_4 &= (\ln T)^2\phi_1 + 2(\ln T)\phi_2 + \phi_3 + o_P(1).
\end{aligned}
$$

Based on the above, the second result of this lemma holds.

(2). We now consider $\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\frac{\partial\widehat{g}(\tau_t,\theta)}{\partial\theta}\big|_{\theta=\widetilde{\theta}}$ and write

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\frac{\partial\widehat{g}(\tau_t,\theta)}{\partial\theta}\Big|_{\theta=\widetilde{\theta}} \\
=\ &\frac{-2}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\widetilde{\theta}}\left[\sum_{u=1}^{T}u^{2\widetilde{\theta}}K_h(\tau_t-\tau_u)\right]^{-2}\left[\sum_{u=1}^{T}\sum_{s=1}^{T}(u\sqrt{s})^{2\widetilde{\theta}}y_s K_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)\ln u\right] \\
&+\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\widetilde{\theta}}\left[\sum_{u=1}^{T}u^{2\widetilde{\theta}}K_h(\tau_t-\tau_u)\right]^{-1}\left[\sum_{u=1}^{T}u^{\widetilde{\theta}}y_u K_h(\tau_t-\tau_u)\ln u\right] \\
:=\ &-2A_1 + A_2,
\end{aligned}
$$

where the definitions of A_1 and A_2 should be obvious.

For A_1, write

$$\begin{aligned}
A_1 &= (1+O_P(h^{2c^*}))T^{-4\widetilde{\theta}-2}\cdot\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{-2\widetilde{\theta}}\left[\sum_{u=1}^{T}\sum_{s=1}^{T}(u\sqrt{s})^{2\widetilde{\theta}}g(\tau_s)s^{\theta_0}K_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)\ln u\right]\\
&\quad+(1+O_P(h^{2c^*}))T^{-4\widetilde{\theta}-2}\cdot\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{-2\widetilde{\theta}}\left[\sum_{u=1}^{T}\sum_{s=1}^{T}(u\sqrt{s})^{2\widetilde{\theta}}\varepsilon_s K_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)\ln u\right]\\
&= \frac{(1+O_P(h^{2c^*}))T^{\theta_0-\widetilde{\theta}}(\ln T)}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{-2\widetilde{\theta}}\left[\frac{1}{T^2}\sum_{u=1}^{T}\sum_{s=1}^{T}\tau_u^{2\widetilde{\theta}}\tau_s^{\widetilde{\theta}+\theta_0}g(\tau_s)K_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)\right]\\
&\quad+\frac{(1+O_P(h^{2c^*}))T^{\theta_0-\widetilde{\theta}}(\ln T)}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{-2\widetilde{\theta}}\left[\frac{1}{T^2}\sum_{u=1}^{T}\sum_{s=1}^{T}(\tau_u\sqrt{\tau_s})^{2\widetilde{\theta}}\varepsilon_s K_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)\right]\\
&\quad+\frac{(1+O_P(h^{2c^*}))T^{\theta_0-\widetilde{\theta}}}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{-2\widetilde{\theta}}\cdot\left[\frac{1}{T^2}\sum_{u=1}^{T}\sum_{s=1}^{T}\tau_u^{2\widetilde{\theta}}\tau_s^{\widetilde{\theta}+\theta_0}g(\tau_s)K_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)\ln\tau_u\right]\\
&\quad+\frac{(1+O_P(h^{2c^*}))T^{\theta_0-\widetilde{\theta}}}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{-2\widetilde{\theta}}\cdot\left[\frac{1}{T^2}\sum_{u=1}^{T}\sum_{s=1}^{T}(\tau_u\sqrt{\tau_s})^{2\widetilde{\theta}}\varepsilon_s K_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)\ln\tau_u\right]\\
&= (\ln T)\phi_1+\phi_2+o_P(1),
\end{aligned}$$

where the first equality follows from (5) of Lemma A.2; and the third equality follows the development similar to (A.21).

Similarly, we can show that

$$A_2 = (\ln T)\phi_1+\phi_2+o_P(1).$$

Based on the above, the third result of this lemma holds.

(3). We now consider $\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\frac{\partial\widehat{g}(\tau_t,\theta)}{\partial\theta}\ln\tau_t\big|_{\theta=\widetilde{\theta}}$ and write

$$\begin{aligned}
&\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\frac{\partial\widehat{g}(\tau_t,\theta)}{\partial\theta}\ln\tau_t\Big|_{\theta=\widetilde{\theta}}\\
&= \frac{-2}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}(\ln\tau_t)\tau_t^{2\widetilde{\theta}}\left[\sum_{u=1}^{T}u^{2\widetilde{\theta}}K_h(\tau_t-\tau_u)\right]^{-2}\left[\sum_{u=1}^{T}\sum_{s=1}^{T}(u\sqrt{s})^{2\widetilde{\theta}}y_sK_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)\ln u\right]\\
&\quad+\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}(\ln\tau_t)\tau_t^{2\widetilde{\theta}}\left[\sum_{u=1}^{T}u^{2\widetilde{\theta}}K_h(\tau_t-\tau_u)\right]^{-1}\left[\sum_{u=1}^{T}u^{\widetilde{\theta}}y_uK_h(\tau_t-\tau_u)\ln u\right]\\
&:= -2A_1+A_2,
\end{aligned}$$

where the definitions of A_1 and A_2 should be obvious.

For A_1, write

$$\begin{aligned}
A_1 &= (1+O_P(h^{2c^*}))T^{-4\widetilde{\theta}-2}\cdot\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}(\ln\tau_t)\tau_t^{-2\widetilde{\theta}}\left[\sum_{u=1}^{T}\sum_{s=1}^{T}(u\sqrt{s})^{2\widetilde{\theta}}g(\tau_s)s^{\theta_0}K_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)\ln u\right]\\
&\quad+(1+O_P(h^{2c^*}))T^{-4\widetilde{\theta}-2}\cdot\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}(\ln\tau_t)\tau_t^{-2\widetilde{\theta}}\left[\sum_{u=1}^{T}\sum_{s=1}^{T}(u\sqrt{s})^{2\widetilde{\theta}}\varepsilon_sK_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)\ln u\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{(1+O_P(h^{2c^*}))T^{\theta_0-\widetilde{\theta}}(\ln T)}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}(\ln \tau_t)\tau_t^{-2\widetilde{\theta}}\left[\frac{1}{T^2}\sum_{u=1}^{T}\sum_{s=1}^{T}\tau_u^{2\widetilde{\theta}}\tau_s^{\widetilde{\theta}+\theta_0}g(\tau_s)K_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)\right]\\
&\quad+\frac{(1+O_P(h^{2c^*}))T^{\theta_0-\widetilde{\theta}}(\ln T)}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}(\ln \tau_t)\tau_t^{-2\widetilde{\theta}}\left[\frac{1}{T^2}\sum_{u=1}^{T}\sum_{s=1}^{T}(\tau_u\sqrt{\tau_s})^{2\widetilde{\theta}}\varepsilon_sK_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)\right]\\
&\quad+\frac{(1+O_P(h^{2c^*}))T^{\theta_0-\widetilde{\theta}}}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}(\ln \tau_t)\tau_t^{-2\widetilde{\theta}}\cdot\left[\frac{1}{T^2}\sum_{u=1}^{T}\sum_{s=1}^{T}\tau_u^{2\widetilde{\theta}}\tau_s^{\widetilde{\theta}+\theta_0}g(\tau_s)K_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)\ln\tau_u\right]\\
&\quad+\frac{(1+O_P(h^{2c^*}))T^{\theta_0-\widetilde{\theta}}}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}(\ln \tau_t)\tau_t^{-2\widetilde{\theta}}\cdot\left[\frac{1}{T^2}\sum_{u=1}^{T}\sum_{s=1}^{T}(\tau_u\sqrt{\tau_s})^{2\widetilde{\theta}}\varepsilon_sK_h(\tau_t-\tau_u)K_h(\tau_t-\tau_s)\ln\tau_u\right]\\
&= (\ln T)\phi_2+\phi_3+o_P(1),
\end{aligned}
$$

where the first equality follows from (5) of Lemma A.2; and the third equality follows the development similar to (A.21).

Similarly, we can show that

$$A_2 = (\ln T)\phi_2+\phi_3+o_P(1).$$

Based on the above, the third result of this lemma holds.

(4)-(6). Similar to results (2)-(3) of this lemma,

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)\Big|_{\theta=\widetilde{\theta}} = \phi_1+o_P(1),\\
&\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)\ln\tau_t\Big|_{\theta=\widetilde{\theta}} = \phi_2+o_P(1),\\
&\frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T}\tau_t^{2\theta}\widehat{g}(\tau_t,\theta)(\ln\tau_t)^2\Big|_{\theta=\widetilde{\theta}} = \phi_3+o_P(1).
\end{aligned}
$$

(7). By (1)-(6) of this lemma, simple calculation immediately gives

$$\frac{\partial^2 R_T(\theta)}{\partial\theta^2}\Big|_{\theta=\widetilde{\theta}} = 8+o_P(1).$$

The proof is now completed. ■

Before proving Theorem 3.3, we denote some variables for notational simplicity and provide some discussions.

$$
\begin{aligned}
&\Sigma = \lim_{T\to\infty}\sum_{t=1}^{T}\sum_{s=1}^{T}E[V_tV_s],\quad V_t = V_{1t}+V_{2t},\quad V_{1t} = -\frac{1}{T^{3/2}}\sum_{u=\lfloor Th\rfloor+1}^{T}\tau_u^{\theta_0}\varepsilon_uK_h(\tau_u-\tau_t),\\
&V_{2t} = \frac{1}{T^{3/2}\ln T}\sum_{v=\lfloor Th\rfloor+1}^{T}\tau_v^{\theta_0}(\ln\tau_v)\varepsilon_tK_h(\tau_v-\tau_t). \qquad (A.22)
\end{aligned}
$$

Remark A.2. *We now verify the existence of* Σ*. Simple algebra shows that* $\frac{\ln\tau_t}{\ln T} = -(1-\frac{\ln t}{\ln T})$*, so* V_{2t} *is a rescaled version of* V_{1t}*. Thus, we just focus on* $\sum_{t=1}^{T}\sum_{s=1}^{T}E[V_{1t}V_{1s}]$ *for the purpose of demonstration.*

Note that it is easy to obtain

$$\int_h^1 K_h(w-u)dw = \begin{cases} \int_{-c}^1 K(w)dw, & u = h + ch \in [0,h) \quad (i.e.,\ c \in [0,1)) \\ 1, & u \in [2h, 1-h] \\ \int_{-1}^c K(w)dw, & u = 1 - ch \in (1-h, 1] \quad (i.e.,\ c \in [0,1)) \end{cases}, \tag{A.23}$$

which indicates $0 \leqslant \sup_{u\in[0,1]} \int_h^1 K_h(w-u)dw \leqslant 1$. *Thus, for* $\sum_{t=1}^T \sum_{s=1}^T E[V_{1t}V_{1s}]$, *we have*

$$\begin{aligned} \sum_{t=1}^T \sum_{s=1}^T E[V_{1t}V_{1s}] &= \frac{1}{T^3} \sum_{s_1=1}^T \sum_{s_2=1}^T \sum_{t_1=\lfloor Th\rfloor+1}^T \sum_{t_2=\lfloor Th\rfloor+1}^T E[\varepsilon_{s_1}\varepsilon_{s_2}]\tau_{s_1}^{\theta_0}\tau_{s_2}^{\theta_0} K_h(\tau_{t_1}-\tau_{s_1})K_h(\tau_{t_2}-\tau_{s_2}) \\ &= \frac{1}{T}\sum_{s_1=1}^T \sum_{s_2=1}^T E[\varepsilon_{s_1}\varepsilon_{s_2}]\tau_{s_1}^{\theta_0}\tau_{s_2}^{\theta_0} \int_h^1 K_h(w-\tau_{s_1})dw \int_h^1 K_h(w-\tau_{s_2})dw + o(1), \end{aligned}$$

where the second the equality follows from the definition of Riemann integral, and the right hand side converges by (A.23) and standard argument of time series analysis.

Proof of Theorem 3.3:

(1). Write

$$\begin{aligned} &\left\{ \frac{1}{T}\sum_{u=\lfloor Th\rfloor+1}^T \tau_u^{2\theta} \frac{\partial \widehat{g}(\tau_u,\theta)}{\partial\theta} + \frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^T \tau_u^{2\theta}\widehat{g}(\tau_u,\theta)\ln\tau_u \right\}\Big|_{\theta=\theta_0} \\ =& -\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^T \tau_u^{2\theta_0}\left[\sum_{t=1}^T t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-2}\left[\sum_{t=1}^T\sum_{s=1}^T (t\sqrt{s})^{2\theta_0} y_s K_h(\tau_u-\tau_t)K_h(\tau_u-\tau_s)\ln t\right] \\ &+\frac{1}{T}\sum_{u=\lfloor Th\rfloor+1}^T \tau_u^{2\theta_0}\left[\sum_{t=1}^T t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-1}\left[\sum_{t=1}^T t^{\theta_0}y_t K_h(\tau_u-\tau_t)\ln t\right] \\ &+\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^T (\ln\tau_u)\tau_u^{2\theta_0}\left[\sum_{t=1}^T t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-1}\left[\sum_{t=1}^T t^{\theta_0}y_t K_h(\tau_u-\tau_t)\right] \\ =& -\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^T \tau_u^{2\theta_0}\left[\sum_{t=1}^T t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-2}\left[\sum_{t=1}^T\sum_{s=1}^T (t\sqrt{s})^{2\theta_0} s^{\theta_0} g(\tau_s) K_h(\tau_u-\tau_t)K_h(\tau_u-\tau_s)\ln t\right] \\ &-\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^T \tau_u^{2\theta_0}\left[\sum_{t=1}^T t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-2}\left[\sum_{t=1}^T\sum_{s=1}^T (t\sqrt{s})^{2\theta_0} \varepsilon_s K_h(\tau_u-\tau_t)K_h(\tau_u-\tau_s)\ln t\right] \\ &+\frac{1}{T}\sum_{u=\lfloor Th\rfloor+1}^T \tau_u^{2\theta_0}\left[\sum_{t=1}^T t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-1}\left[\sum_{t=1}^T t^{2\theta_0}g(\tau_t) K_h(\tau_u-\tau_t)\ln t\right] \\ &+\frac{1}{T}\sum_{u=\lfloor Th\rfloor+1}^T \tau_u^{2\theta_0}\left[\sum_{t=1}^T t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-1}\left[\sum_{t=1}^T t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\ln t\right] \\ &+\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^T (\ln\tau_u)\tau_u^{2\theta_0}\left[\sum_{t=1}^T t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-1}\left[\sum_{t=1}^T t^{2\theta_0}g(\tau_t) K_h(\tau_u-\tau_t)\right] \\ &+\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^T (\ln\tau_u)\tau_u^{2\theta_0}\left[\sum_{t=1}^T t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-1}\left[\sum_{t=1}^T t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\right] \end{aligned}$$

$$:= -2A_1 - 2A_2 + A_3 + A_4 + 2A_5 + 2A_6, \tag{A.24}$$

where the definitions of A_1 to A_6 should be obvious.

Focus on $\frac{T^{\theta_0+\frac{1}{2}}}{\ln T}(-2A_2 + A_4 + 2A_6)$ first. By repeatedly using Lemma A.2 as we have done in the proof of Lemma A.3, we are able to write

$$\begin{aligned}
&\frac{T^{\theta_0+\frac{1}{2}}}{\ln T}(-2A_2 + A_4 + 2A_6)\\
&= -\frac{T^{\theta_0+\frac{1}{2}}}{\ln T}\cdot\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}\tau_u^{2\theta_0}\left[\sum_{t=1}^{T}t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-2}\left[\sum_{t=1}^{T}\sum_{s=1}^{T}(t\sqrt{s})^{2\theta_0}\varepsilon_s K_h(\tau_u-\tau_t)K_h(\tau_u-\tau_s)\ln t\right]\\
&\quad+\frac{T^{\theta_0+\frac{1}{2}}}{\ln T}\cdot\frac{1}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}\tau_u^{2\theta_0}\left[\sum_{t=1}^{T}t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-1}\left[\sum_{t=1}^{T}t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\ln t\right]\\
&\quad+\frac{T^{\theta_0+\frac{1}{2}}}{\ln T}\cdot\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}(\ln\tau_u)\tau_u^{2\theta_0}\left[\sum_{t=1}^{T}t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-1}\left[\sum_{t=1}^{T}t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\right]\\
&= -\frac{T^{\frac{1}{2}}}{\ln T}\cdot\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}\tau_u^{2\theta_0}\left[\frac{1}{T}\sum_{t=1}^{T}\tau_t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-2}\left[\frac{1}{T}\sum_{t=1}^{T}\tau_t^{2\theta_0}K_h(\tau_u-\tau_t)\ln t\right]\\
&\quad\times\left[\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\right]\\
&\quad+\frac{T^{\frac{1}{2}}}{\ln T}\cdot\frac{1}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}\tau_u^{2\theta_0}\left[\frac{1}{T}\sum_{t=1}^{T}\tau_t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\ln t\right]\\
&\quad+\frac{T^{\frac{1}{2}}}{\ln T}\cdot\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}(\ln\tau_u)\tau_u^{2\theta_0}\left[\frac{1}{T}\sum_{t=1}^{T}\tau_t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\right]\\
&= -(1+o(1))\cdot\frac{T^{\frac{1}{2}}}{\ln T}\cdot\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}\tau_u^{-2\theta_0}\left[\tau_u^{2\theta_0}\ln\tau_u+\tau_u^{2\theta_0}\ln T\right]\left[\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\right]\\
&\quad+(1+o(1))\cdot\frac{T^{\frac{1}{2}}}{\ln T}\cdot\frac{1}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}\left[\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)(\ln\tau_t+\ln T)\right]\\
&\quad+(1+o(1))\cdot\frac{T^{\frac{1}{2}}}{\ln T}\cdot\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}(\ln\tau_u)\left[\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\right]\\
&= -(1+o(1))\cdot\frac{T^{\frac{1}{2}}}{\ln T}\cdot\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}\left[\ln\tau_u+\ln T\right]\left[\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\right]\\
&\quad+(1+o(1))\cdot\frac{T^{\frac{1}{2}}}{\ln T}\cdot\frac{1}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}\left[\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)(\ln\tau_t+\ln T)\right]\\
&\quad+(1+o(1))\cdot\frac{T^{\frac{1}{2}}}{\ln T}\cdot\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}(\ln\tau_u)\left[\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\right]\\
&= (1+o_P(1))\cdot\frac{1}{T^{3/2}}\sum_{u=\lfloor Th\rfloor+1}^{T}\sum_{t=1}^{T}\left\{-2\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)+\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\right\}\\
&\quad+(1+o_P(1))\cdot\frac{1}{T^{3/2}\ln T}\sum_{u=\lfloor Th\rfloor+1}^{T}\sum_{t=1}^{T}\tau_t^{\theta_0}(\ln\tau_t)\varepsilon_t K_h(\tau_u-\tau_t) = (1+o_P(1))\cdot\sum_{t=1}^{T}V_t,
\end{aligned} \tag{A.25}$$

where $V_t = V_{1t} + V_{2t}$,

$$V_{1t} = -\frac{1}{T^{3/2}} \sum_{u=\lfloor Th \rfloor+1}^{T} \tau_t^{\theta_0} \varepsilon_t K_h(\tau_u - \tau_t),$$

$$V_{2t} = \frac{1}{T^{3/2}\ln T} \sum_{u=\lfloor Th \rfloor+1}^{T} \tau_t^{\theta_0} (\ln \tau_t) \varepsilon_t K_h(\tau_u - \tau_t). \tag{A.26}$$

We then can use the large block and small block technique (e.g., Lemma A of Chen et al. (2012*a*)) to show that $\sum_{t=1}^{T} V_t \to_D N(0, \Sigma)$, where Σ has been defined in (A.22).

Thus, we know that

$$-2A_2 + A_4 + 2A_6 = O_P\left(\frac{\ln T}{T^{\theta_0+\frac{1}{2}}}\right). \tag{A.27}$$

To further simplify the notation, letting $\xi_T = \frac{1}{T}\sum_{t=\lfloor Th \rfloor+1}^{T} \tau_t^{2\theta_0} \widehat{g}(\tau_t, \theta_0)$, it is easy to know that

$$\xi_T \to_P \int_0^1 u^{2\theta_0} g(u) du \tag{A.28}$$

by the development of (A.18). Thus, rearranging (3.8) using the decomposition (A.24) gives

$$\begin{aligned} &\left[\frac{\partial^2 R_T(\theta)}{\partial \theta^2}\Big|_{\theta=\widetilde{\theta}}\right]^{-1} \left\{\frac{-4\lambda_T^2 \cdot \ln \xi_T^2}{\xi_T} \cdot (\ln T)(-2A_2 + A_4 + 2A_6)\right\} \\ &= (\ln T)\left\{(\widehat{\theta} - \theta_0) - \left[\frac{\partial^2 R_T(\theta)}{\partial \theta^2}\Big|_{\theta=\widetilde{\theta}}\right]^{-1} \frac{4\lambda_T^2 \cdot \ln \xi_T^2}{\xi_T}(2A_1 - A_3 - 2A_5)\right\}. \end{aligned} \tag{A.29}$$

Note that (A.27) and (7) of Lemma A.3 together imply

$$\left[\frac{\partial^2 R_T(\theta)}{\partial \theta^2}\Big|_{\theta=\widetilde{\theta}}\right]^{-1} \left\{\frac{-4\lambda_T^2 \cdot \ln \xi_T^2}{\xi_T} \cdot (\ln T)(-2A_2 + A_4 + 2A_6)\right\} = O_P\left(\frac{1}{T^{\theta_0+\frac{1}{2}}}\right).$$

Thus, we can further simplify (A.29) to obtain

$$\begin{aligned} (\ln T)(\widehat{\theta} - \theta_0) &= (\ln T)\left[\frac{\partial^2 R_T(\theta)}{\partial \theta^2}\Big|_{\theta=\widetilde{\theta}}\right]^{-1} \frac{4\lambda_T^2 \cdot \ln \xi_T^2}{\xi_T}(2A_1 - A_3 - 2A_5) + O_P\left(\frac{1}{T^{\theta_0+\frac{1}{2}}}\right) \\ &= \lambda_T \frac{\ln|\xi_T|}{\xi_T}(2A_1 - A_3 - 2A_5) + O_P\left(\frac{1}{T^{\theta_0+\frac{1}{2}}}\right). \end{aligned} \tag{A.30}$$

Below we just need to focus on A_1, A_3 and A_5. Start from A_1.

$$\begin{aligned} A_1 &= \frac{1}{T}\sum_{u=\lfloor Th \rfloor+1}^{T} \tau_u^{2\theta_0} \left[\sum_{t=1}^{T} \tau_t^{2\theta_0} K_h(\tau_u - \tau_t)\right]^{-2} \\ &\quad \cdot \left[\sum_{t=1}^{T}\sum_{s=1}^{T} \tau_t^{2\theta_0}\tau_s^{2\theta_0} g(\tau_s) K_h(\tau_u - \tau_t) K_h(\tau_u - \tau_s)(\ln \tau_t + \ln T)\right] \\ &= (\ln T)\cdot \frac{1}{T}\sum_{u=\lfloor Th \rfloor+1}^{T} \tau_u^{2\theta_0} \left[\sum_{t=1}^{T} \tau_t^{2\theta_0} K_h(\tau_u - \tau_t)\right]^{-2} \left[\sum_{t=1}^{T}\sum_{s=1}^{T} \tau_t^{2\theta_0}\tau_s^{2\theta_0} g(\tau_s) K_h(\tau_u - \tau_t) K_h(\tau_u - \tau_s)\right] \\ &\quad + \frac{1}{T}\sum_{u=\lfloor Th \rfloor+1}^{T} \tau_u^{2\theta_0} \left[\sum_{t=1}^{T} \tau_t^{2\theta_0} K_h(\tau_u - \tau_t)\right]^{-2} \left[\sum_{t=1}^{T}\sum_{s=1}^{T} \tau_t^{2\theta_0}\tau_s^{2\theta_0} g(\tau_s) K_h(\tau_u - \tau_t) K_h(\tau_u - \tau_s)\ln \tau_t\right] \end{aligned}$$

$$:= A_{11} + A_{12}.$$

For A_{11}, we have

$$\begin{aligned}
A_{11} &= (\ln T) \cdot \left(1 + O(h^{2c^*})\right) \cdot \frac{1}{T} \sum_{u=\lfloor Th \rfloor+1}^{T} \tau_u^{2\theta_0} \tau_u^{-4\theta_0} \tau_u^{2\theta_0} \tau_u^{2\theta_0} g(\tau_u) \\
&= (\ln T) \cdot \left(1 + O(h^{2c^*})\right) \cdot \frac{1}{T} \sum_{u=\lfloor Th \rfloor+1}^{T} \tau_u^{2\theta_0} g(\tau_u) \\
&= (\ln T) \cdot \left(1 + O(h^{2c^*})\right) \cdot \left(\int_h^1 u^{2\theta_0} g(u) du + O\left(\frac{1}{Th}\right)\right) \\
&= (\ln T) \int_0^1 g(u) du + o(1),
\end{aligned}$$

where the first equality follows from (4) and (5) of Lemma A.2; and the third equality follows from the definition of Riemann integral.

Similarly,

$$\begin{aligned}
A_{12} &= \left(1 + O(h^{2c^*})\right) \cdot \frac{1}{T} \sum_{u=\lfloor Th \rfloor+1}^{T} \tau_u^{2\theta_0} \tau_u^{-4\theta_0} \tau_u^{2\theta_0} \tau_u^{2\theta_0} g(\tau_u)(\ln \tau_u) \\
&= \left(1 + O(h^{2c^*})\right) \cdot \left(\int_h^1 u^{2\theta_0} g(u)(\ln u) du + O\left(\frac{\ln(1/h)}{T}\right)\right) \\
&= \int_0^1 u^{2\theta_0} g(u)(\ln u) du + o(1),
\end{aligned}$$

where the second equality follows from the definition of Riemann integral.

Therefore,

$$A_1 = (\ln T) \int_0^1 u^{2\theta_0} g(u) du + \int_0^1 u^{2\theta_0} g(u)(\ln u) du + o(1).$$

Similarly, we can show that

$$\begin{aligned}
A_3 &= (\ln T) \int_0^1 u^{2\theta_0} g(u) du + \int_0^1 u^{2\theta_0} g(u)(\ln u) du + o(1), \\
A_5 &= \int_0^1 u^{2\theta_0} g(u)(\ln u) du + o(1).
\end{aligned}$$

By the analyses of A_1, A_3 and A_5, we obtain that

$$2A_1 - A_3 - 2A_5 = (\ln T) \int_0^1 u^{2\theta_0} g(u) du \cdot (1 + O_P(\lambda_T)). \tag{A.31}$$

In connection with (A.30) and (A.28), we can conclude that

$$(\ln T)(\widehat{\theta} - \theta_0) = \frac{\ln |\xi_T|}{\xi_T} \int_0^1 u^{2\theta_0} g(u) du + O_P(\lambda_T) = \ln \left| \int_0^1 u^{2\theta_0} g(u) du \right| + o_P(1),$$

where the existence of $\ln \left| \int_0^1 u^{2\theta_0} g(u) du \right|$ has been verified by (A.11) and (A.12) already.

Thus, the proof of the first result of this theorem is now complete.

(2). The second result follows from (A.17) straight away. ■

Appendix B

We first provide the omitted proof for Theorem 2.1.

Proof of Theorem 2.1:

(1). The proof of (2.3) follows from the standard arguments, so omitted. We take a further look at (2.4) at first, and write

$$\begin{aligned}
\widehat{S}_T &= -\frac{1}{T_{odd}}\sum_{t \text{ odd}}\left[-\varepsilon_t+\widehat{g}(\tau_t)-g(\tau_t)\right]\cdot\left[\widehat{g}(\tau_t)-g(\tau_t)+g(\tau_t)\right]\ln t \\
&= \frac{1}{T_{odd}}\sum_{t \text{ odd}}\varepsilon_t g(\tau_t)\ln t+\frac{1}{T_{odd}}\sum_{t \text{ odd}}\varepsilon_t\cdot\left[\widehat{g}(\tau_t)-g(\tau_t)\right]\ln t \\
&\quad -\frac{1}{T_{odd}}\sum_{t \text{ odd}}\left[\widehat{g}(\tau_t)-g(\tau_t)\right]g(\tau_t)\ln t-\frac{1}{T_{odd}}\sum_{t \text{ odd}}\left[\widehat{g}(\tau_t)-g(\tau_t)\right]^2\ln t \\
&:= S_{T,1}+S_{T,2}-S_{T,3}-S_{T,4},
\end{aligned} \tag{B.1}$$

where the definitions of $S_{T,1}$ to $S_{T,4}$ should be obvious. Since it is easy to show that $S_{T,2}=o_P(S_{T,1})$ and $S_{T,4}=o_P(S_{T,1})$, we just focus on $S_{T,1}-S_{T,3}$ as follows:

$$\begin{aligned}
S_{T,1}-S_{T,3} &= \frac{1}{T_{odd}}\sum_{t \text{ odd}}\varepsilon_t g(\tau_t)\ln t-\frac{1}{T_{odd}}\sum_{t \text{ odd}}\left[\widehat{g}(\tau_t)-g(\tau_t)\right]g(\tau_t)\ln t \\
&= \frac{1}{T_{odd}}\sum_{t \text{ odd}}\varepsilon_t g(\tau_t)\ln t-\frac{1}{T_{odd}}\sum_{t \text{ odd}}\frac{\sum_{s \text{ even}}K_h(\tau_t-\tau_s)\varepsilon_s}{\sum_{s \text{ even}}K_h(\tau_t-\tau_s)}g(\tau_t)\ln t \\
&\quad -\frac{1}{T_{odd}}\sum_{t \text{ odd}}\left[\frac{\sum_{s \text{ even}}K_h(\tau_t-\tau_s)g(\tau_s)}{\sum_{s \text{ even}}K_h(\tau_t-\tau_s)}-g(\tau_t)\right]g(\tau_t)\ln t \\
&= \frac{1}{T_{odd}}\sum_{t \text{ odd}}\varepsilon_t g(\tau_t)\ln t-\frac{T_{even}}{T_{odd}}\cdot\frac{1}{T_{even}}\sum_{t \text{ even}}\varepsilon_t\sum_{s \text{ odd}}\frac{K_h(\tau_t-\tau_s)}{\sum_{j \text{ even}}K_h(\tau_j-\tau_s)}g(\tau_s)\ln s \\
&\quad +o_P(1) \\
&= \frac{1}{T_{odd}}\sum_{t \text{ odd}}\varepsilon_t g(\tau_t)\ln t-\frac{1+o_P(1)}{T_{even}}\sum_{t \text{ even}}\varepsilon_t g(\tau_t)\ln t+o_P(1) \\
&= \frac{2}{T}\sum_{t \text{ odd}}\varepsilon_t g(\tau_t)\ln t-\frac{2+o_P(1)}{T}\sum_{t \text{ even}}\varepsilon_t g(\tau_t)\ln t+o_P(1),
\end{aligned} \tag{B.2}$$

where the fourth equality follows from

$$g(\tau_t)\ln t-\sum_{s \text{ odd}}\frac{K_h(\tau_t-\tau_s)}{\sum_{j \text{ even}}K_h(\tau_j-\tau_s)}g(\tau_s)\ln s=o_P(1)$$

uniformly in t by the proof similar to those given for Theorem 3.3 of the main text.

Based on (B.2) and the assumptions in the body of this theorem, the result follows immediately. Then the proof is complete.

(2). We now consider what happens under the alternative hypothesis, i.e., $\theta_0>0$. For $\forall u\in(0,1)$, we have

$$|\widehat{g}(u)| = \left|\frac{\sum_{t=1}^T K_h(u-\tau_t)y_t}{\sum_{t=1}^T K_h(u-\tau_t)}\right| = \left|\frac{\sum_{t=1}^T K_h(u-\tau_t)g(\tau_t)t^{\theta_0}}{\sum_{t=1}^T K_h(u-\tau_t)}+\frac{\sum_{t=1}^T K_h(u-\tau_t)\varepsilon_t}{\sum_{t=1}^T K_h(u-\tau_t)}\right|$$

$$= \left| T^{\theta_0} \cdot \frac{\sum_{t=1}^T K_h(u-\tau_t) g(\tau_t)\tau_t^{\theta_0}}{\sum_{t=1}^T K_h(u-\tau_t)} \right| + o_P(1) = T^{\theta_0} \cdot \left(u^{\theta_0} |g(u)| + o_P(1) \right) + o_P(1)$$
$$\to_P \infty. \tag{B.3}$$

In connection with (B.1), it is easy to see that $S_{T,4}$ is the true leading term due to the involvement of the quadratic term. Then by the definition of (2.6) and under the alternative hypothesis, $\widehat{LM} \to \infty$ as $T \to \infty$. ■

A Generalized Parametric Test with Discussions

We now discuss if a more generalized version of (2.1) can be achieved. To be precise, the test is specified as follows:

$$H_0: \ \theta_0 = a \quad \text{vs.} \quad H_1: \ \theta_0 > a, \tag{B.4}$$

where a is a positive constant. Under the null, the estimator of g reduces to a special case of (3.1), i.e.,

$$\widehat{g}(u, \theta_0) = \left[\sum_{t=1}^T t^{2\theta_0} K_h(u-\tau_t) \right]^{-1} \sum_{t=1}^T t^{\theta_0} y_t K_h(u-\tau_t). \tag{B.5}$$

In order to avoid $\left[\sum_{t=1}^T t^{2\theta_0} K_h(u-\tau_t)\right]^{-1}$ blowing up the rate of convergence in the sup norm below, we further restrict u to the set $[c, 1-h]$ and suppose that $\sup_{(\theta,u)\in\Theta\times[c,1-h]} \left| \frac{d^2[u^{\theta+\theta_0} g(u)]}{du^2} \right| < \infty$, where $c \in (0,1)$ is a fixed constant. Then by the proof of Lemma 3.1, a faster rate convergence for $\widehat{g}(u,\theta_0)$ can be achieved as follows:

$$\sup_{u\in[c,1-h]} |\widehat{g}(u,\theta_0) - g(u)| = O_P\left(\frac{\sqrt{\ln T}}{T^{\frac{1}{2}+\theta_0} h^{\frac{1}{2}}} \right) + O(h^2). \tag{B.6}$$

Note that in this case, the score function becomes

$$S_T(\theta_0) = \frac{1}{T} \sum_{t=1}^T \left(y_t - g(\tau_t) t^{\theta_0} \right) g(\tau_t) t^{\theta_0} \ln t.$$

In order to use (B.6), we denote $B_h = \{t \mid \lfloor cT \rfloor \leqslant t \leqslant \lfloor (1-h)T \rfloor\}$ and let T^* be the cardinality of B_h. Then consider the following approximation for the score function in practice:

$$\widehat{S}_T = \frac{1}{T^*/2} \sum_{t \text{ odd}\in B_h} \left(y_t - \widehat{g}(\tau_t) t^{\theta_0} \right) \widehat{g}(\tau_t) t^{\theta_0} \ln t, \tag{B.7}$$

where

$$\widehat{g}(u) = \left[\sum_{t \text{ even}\in B_h} t^{2\theta_0} K_h(u-\tau_t) \right]^{-1} \sum_{t \text{ even}\in B_h} t^{\theta_0} y_t K_h(u-\tau_t). \tag{B.8}$$

Based on the above, write

$$\widehat{S}_T = -\frac{1}{T^*/2} \sum_{t \text{ odd}\in B_h} \left[-\varepsilon_t + \widehat{g}(\tau_t) t^{\theta_0} - g(\tau_t) t^{\theta_0} \right] \cdot \left[\widehat{g}(\tau_t) - g(\tau_t) + g(\tau_t) \right] t^{\theta_0} \ln t$$

$$
\begin{aligned}
&= \frac{2}{T^*} \sum_{t \text{ odd} \in B_h} \varepsilon_t g(\tau_t) t^{\theta_0} \ln t + \frac{2}{T^*} \sum_{t \text{ odd} \in B_h} \varepsilon_t \cdot [\widehat{g}(\tau_t) - g(\tau_t)]\, t^{\theta_0} \ln t \\
&\quad - \frac{2}{T^*} \sum_{t \text{ odd} \in B_h} [\widehat{g}(\tau_t) - g(\tau_t)]\, g(\tau_t) t^{2\theta_0} \ln t - \frac{2}{T^*} \sum_{t \text{ odd} \in B_h} [\widehat{g}(\tau_t) - g(\tau_t)]^2\, t^{2\theta_0} \ln t \\
&:= S_{T,1} + S_{T,2} - S_{T,3} - S_{T,4}. \qquad \text{(B.9)}
\end{aligned}
$$

Similar to the proof of Theorem 2.1, it is easy to show that $\sqrt{T^*} S_{T,2}$ and $\sqrt{T^*} S_{T,4}$ are negligible, and $S_{T,1} - S_{T,3}$ can be rewritten as

$$
S_{T,1} - S_{T,3} = \frac{2}{T^*} \sum_{t \text{ odd} \in B_h} \varepsilon_t g(\tau_t) t^{\theta_0} \ln t - \frac{2 + o_P(1)}{T^*} \sum_{t \text{ even} \in B_h} \varepsilon_t g(\tau_t) t^{\theta_0} \ln t + o_P(1) \qquad \text{(B.10)}
$$

provided $h^2 T^{2\theta_0} \ln T \to 0$ in view of (B.6) and the development of (B.2).

Therefore, we are able to state the next result.

Corollary B.1. *Let the conditions of Theorem 2.1 hold. Suppose further that $h^2 T^{2\theta_0} \ln T \to 0$ and $\sup_{(\theta,u) \in \Theta \times [c,1-h]} \left| \frac{d^2 [u^{\theta+\theta_0} g(u)]}{du^2} \right| < \infty$.*

1. *Under the null, as $T \to \infty$,*

$$
\widehat{LM} = \frac{\frac{\sqrt{T^*}}{2} \widehat{S}_T}{\left\{ \widehat{\sigma}_\varepsilon^2 \cdot \frac{1}{T^*} \sum_{t \in B_h} \left[\widehat{g}(\tau_t) t^{\theta_0} \ln t \right]^2 \right\}^{1/2}} \to_D N(0,1), \qquad \text{(B.11)}
$$

where $\widehat{\sigma}_\varepsilon^2 = \frac{1}{T^*} \sum_{t \in B_h} (y_t - \widehat{g}(\tau_t))^2 \to_P \sigma_\varepsilon^2$, and $\widehat{S}_T$ and $\widehat{g}(u)$ have been defined in (B.7) and (B.8) respectively.

2. *Under the alternative, as $T \to \infty$, $\widehat{LM} \to \infty$.*

The proof of the second result of the above corollary follows from almost an identical procedure as (B.3), thus omitted.

Remark B.1.

1. *The condition $h^2 T^{2\theta_0} \ln T \to 0$ implies θ_0 cannot be greater than or equal to 1.*

2. *Note that $\{\theta \mid \widehat{LM}(\theta) < z_\alpha\}$ gives the $1 - \alpha$ confidence interval for θ_0, where α stands for the significant level and z_α presents the corresponding critical value.*

Below, we implement some simulation study to back up our arguments. Particularly, the DGP is $y_t = \exp(\tau_t) t^{\theta_0} + \varepsilon_t$, where the variables are generated in exactly the same way as Case 1.1 of Section 4.1. We choose the value of θ_0 from $\{0.2,\ 0.4,\ 0.6,\ 0.8,\ 1.0\}$. The bandwidth is set to $h = (\frac{\ln T}{T})^{7/10}$, and we let $c = 0.3$ without losing generality. As the above development requires $h^2 T^{2\theta_0} \ln T \to 0$, so we would expect that the size of the test will go wrong when $\theta_0 \geqslant 0.7$. For simplicity, we report the size based on 1000 replications in Figure 8, which is sufficient to explain our argument on the requirement of $\theta_0 < 0.7$. The power test can be done as in Section 4.1, so we do not pursue it further.

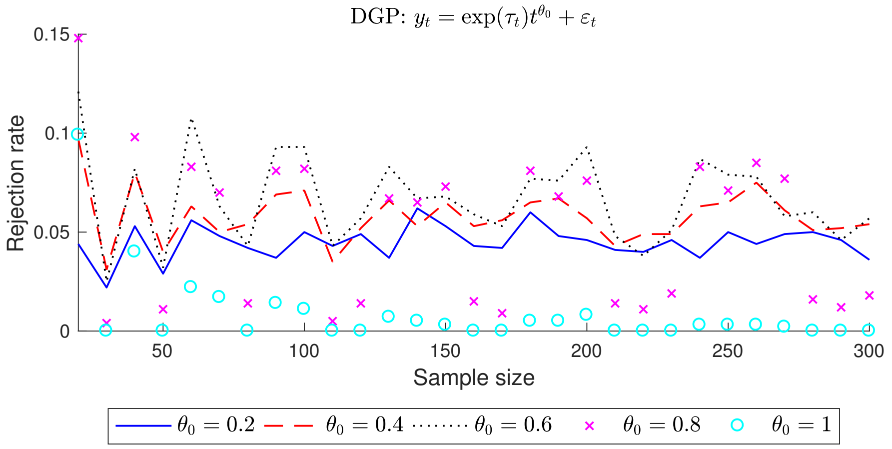

Figure 8: Size of (B.11) at 5% Significant Level

As expected, while $\theta_0 = 0.2,\ 0.4,\ 0.6$, the size of (B.11) is reasonably well (i.e., moving around 5%). However, when $\theta_0 = 0.8,\ 1$, the size of the test totally goes wrong, which confirms our argument on the requirement of $h^2 T^{2\theta_0} \ln T \to 0$.

Extension 2 of Section 5

We now formalize the statement made in Extension 2 of Section 5. Consider a general trending model of the form:

$$y_t = f(x_t, \tau_t) + g(\tau_t)t^{\theta_0} + \varepsilon_t. \tag{B.12}$$

Assumption 3:

Suppose that $f(\cdot,\cdot)$ and $\{x_t \mid t = 1, \ldots, T\}$ satisfy one of the following three cases:

1. *$\{x_t \mid t = 1, \ldots, T\}$ is a strictly stationary and α-mixing error process with a density function $p(w)$. Moreover, $\sup_{(w,u)\in\mathbb{R}^d\times[0,1]} p(w)\dfrac{\partial f(w,u)}{\partial u} < \infty$ and $E\big[\sup_{u\in[0,1]} |f(x_1,u)|\big]$; or*
2. *$\{x_t \mid t = 1, \ldots, T\}$ is a locally stationary process.[5] Let $f(\cdot,\cdot)$ be uniformly bounded and satisfy that $|f(x_1,u) - f(x_2,u)| \leqslant A_4\|x_1 - x_2\|$ for $\forall u \in [0,1]$, where A_4 is a positive constant; or*
3. *(a) Let $f(\cdot,\cdot)$ be uniformly bounded, and $x_t = x_{t-1} + w_t$ for $t \geqslant 1$ and $\|x_0\| = O_P(1)$;*

   (b) Let $w_t = \sum_{j=0}^{\infty}\psi_j\epsilon_{t-j}$, where $\sum_{j=0}^{\infty} j\|\psi_j\| < \infty$ and $\psi := \sum_{j=0}^{\infty}\psi_j \neq 0$;

[5]We adopt the following definition for a locally stationary process (cf., Vogt, 2012; Dong and Linton, 2016):

Definition 7.1. *The process $\{x_t \mid t = 1, \ldots, T\}$ is locally stationary if for each rescaled time point $u \in [0,1]$ there exists an associated process $\{x_t(u) \mid t = 1, \ldots, T\}$ with the following two properties:*

(a) $\{x_t(u) \mid t = 1, \ldots, T\}$is strictly stationary with density $f_u(w)$;

(b) It holds that $\|x_t - x_t(u)\|_r \leqslant \left(|\tau_t - u| + T^{-1}\right) U_t(u)$ a.s., where $\tau_t = t/T$, $\{U_t(u)\}$ is a process of positive variables satisfying $E|U_t(u)|^\rho < C$ for some $\rho \geqslant 1$ and $C < \infty$ independent of u, t, and T. Moreover, $\|\cdot\|_r$ denotes an arbitrary norm on $\mathbb{R}^d$.

(c) *Let $\{\epsilon_j \mid -\infty < j < \infty\}$ be a scalar sequence of i.i.d. random variables having an absolutely continuous distribution with respect to the Lebesgue measure and satisfying $E[\epsilon_1] = 0_{d\times 1}$, $E[\epsilon_1\epsilon_1'] = I_d$, $E\|\epsilon_1\|^q < \infty$ for some $q > 2$. The characteristic function of ϵ_1 is integrable.*

Corollary B.2. *Let $\widehat{\theta}$ and $\widehat{g}(u,\theta)$ be those defined in Section 2 of the paper. Under Assumptions 1 and 3, suppose further that $h = O(T^{-\nu})$ with ν being a positive constant and satisfying $0 < \nu < \frac{1}{2}$. As $T \to \infty$,*

1. $\sup_{(\theta,u)\in\Theta\times[0,1]} \left|\frac{1}{T}\sum_{t=1}^{T} \tau_t^{\theta} f(x_t,\tau_t) K_h(u-\tau_t)\right| = O_P(1)$;
2. $\widehat{\theta} - \theta_0 = O_P(\lambda_T)$, *where λ_T is defined in (3.5);*
3. $\sup_{u\in[h,1]} \left|\widehat{g}(u,\widehat{\theta}) - (uT)^{\theta_0-\widehat{\theta}} \cdot g(u)\right| = o_P(1)$.

Proof of Corollary B.2:

(1). First, we point out one simple fact below:

$$\int_{-u/h}^{(1-u)/h} K(w)dw = \begin{cases} 1, & u \in [h, 1-h] \\ \int_{-1}^{c} K(w)dw, & u = 1 - ch \text{ with } c \in [0,1) \\ \int_{-c}^{1} K(w)dw, & u = ch \text{ with } c \in [0,1) \end{cases}.$$

Therefore, it is easy to know that

$$\sup_{u\in[0,1]} = \int_{-u/h}^{(1-u)/h} K(w)dw = O(1). \tag{B.13}$$

Case 1: Under Assumption 3.1, we have

$$\begin{aligned}
&E\left[\sup_{(\theta,u)\in\Theta\times[0,1]} \left|\frac{1}{T}\sum_{t=1}^{T} \tau_t^{\theta} f(x_t,\tau_t) K_h(u-\tau_t)\right|\right] \\
&\leqslant \int \sup_{(\theta,u)\in\Theta\times[0,1]} \frac{1}{T}\sum_{t=1}^{T} \tau_t^{\theta} |f(w,\tau_t)| K_h(u-\tau_t) p(w)dw \\
&\leqslant O(1) \int \sup_{u\in[0,1]} \frac{1}{T}\sum_{t=1}^{T} |f(w,\tau_t)| K_h(u-\tau_t) p(w)dw \\
&= O(1) \int \sup_{u\in[0,1]} \int_0^1 |f(w_1,w_2)| \cdot p(w_1) K_h(u-w_2) dw_2 dw_1 \\
&= O(1) \int \sup_{u\in[0,1]} \int_{-u/h}^{(1-u)/h} |f(w_1, u+w_2h)| \cdot p(w_1) K(w_2) dw_2 dw_1 \\
&= O(1) \int \sup_{u\in[0,1]} |f(w_1,u)| \int_{-u/h}^{(1-u)/h} K(w_2)dw_2 \cdot p(w_1)dw_1 \\
&\leqslant O(1) \int \sup_{u\in[0,1]} |f(w,u)| p(w)dw = O(1),
\end{aligned}$$

where the second inequality follows from the fact that $0 \leqslant \tau^{\theta} < 1$ uniformly; the first equality follows from the definition of Riemann integral; the third and fourth equalities follows from Assumption 3.1.; the third inequality follows from (B.13).

Therefore, $\sup_{(\theta,u)\in\Theta\times[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta}f(x_t,\tau_t)K_h(u-\tau_t)\right| = O_P(1)$ under Assumption 3.a.

Case 2: Let Assumption 3.2 hold. Note that by the definition of a locally stationary process, it is easy to know that $U_t(u) = O_P(1)$ uniformly in t and u.

Write

$$\begin{aligned}
&\sup_{(\theta,u)\in\Theta\times[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta}f(x_t,\tau_t)K_h(u-\tau_t)\right| \\
\leqslant\ &\sup_{(\theta,u)\in\Theta\times[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta}\left(f(x_t,\tau_t)-f(x_t(\tau_t),\tau_t)\right)K_h(u-\tau_t)\right| \\
&+\sup_{(\theta,u)\in\Theta\times[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta}f(x_t(\tau_t),\tau_t)K_h(u-\tau_t)\right| \\
:=\ &A_1+A_2,
\end{aligned}$$

where the definitions of A_1 and A_2 should be obvious.

For A_1, we have

$$\begin{aligned}
A_1 &= \sup_{(\theta,u)\in\Theta\times[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta}\left(f(x_t,\tau_t)-f(x_t(\tau_t),\tau_t)\right)K_h(u-\tau_t)\right| \\
&\leqslant O(1)\sup_{(\theta,u)\in\Theta\times[0,1]}\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta}\|x_t-x_t(\tau_t)\|K_h(u-\tau_t) \\
&\leqslant O(1)\sup_{(\theta,u)\in\Theta\times[0,1]}\frac{1}{T^2}\sum_{t=1}^{T}\tau_t^{\theta}U_t(\tau_t)K_h(u-\tau_t) \\
&\leqslant O(1)\frac{1}{T^2h}\sum_{t=1}^{T}U_t(\tau_t)\leqslant O_P(1)\frac{1}{Th}.
\end{aligned}$$

where the first inequality follows from Assumption 3.2; the second inequality follows from the definition of locally stationary process; and the fourth inequality follows from the fact (i.e., $U_t(\tau_t) = O_P(1)$) that we point out in the beginning of *Case 2*.

For A_2, it is easy to obtain that

$$\begin{aligned}
E[A_2] &= E\left[\sup_{(\theta,u)\in\Theta\times[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta}f(x_t(\tau_t),\tau_t)K_h(u-\tau_t)\right|\right] \\
&\leqslant \sup_{(\theta,u)\in\Theta\times[0,1]}\frac{1}{T}\sum_{t=1}^{T}K_h(u-\tau_t) = O(1)\sup_{u\in[0,1]}\frac{1}{h}\int_0^1 K_h(u-w)dw \\
&= O(1)\sup_{u\in[0,1]}\int_{-u/h}^{(1-u)/h}K(w)dw = O(1),
\end{aligned}$$

where the first inequality follows from Assumption 3.2; and the second equality follows from the definition of Riemann integral; and the fourth equality follows from (B.13).

Based on the analyses of A_1 and A_2, we have $\sup_{(\theta,u)\in\Theta\times[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta}f(x_t,\tau_t)K_h(u-\tau_t)\right| = O_P(1)$.

Case 3: Let Assumption 3.3 hold. Construct a ν_T satisfying that $\nu_T\to\infty$ and $\nu_T/(Th)\to 0$. By Lemma C.5 of Dong et al. (2016), we know that, for sufficiently large t, $x_t/\sqrt{t}$ has a pdf function $\phi_t(w)$, which is uniformly bounded in both t and w.

$$
\begin{aligned}
& E\left[\sup_{(\theta,u)\in\Theta\times[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta} f(x_t,\tau_t)K_h(u-\tau_t)\right|\right] \\
= & E\left[\sup_{(\theta,u)\in\Theta\times[0,1]}\left|\frac{1}{T}\sum_{t=1}^{\nu_T}\tau_t^{\theta} f(x_t,\tau_t)K_h(u-\tau_t)\right|\right] \\
& + E\left[\sup_{(\theta,u)\in\Theta\times[0,1]}\left|\frac{1}{T}\sum_{t=\nu_T+1}^{T}\tau_t^{\theta} f(x_t,\tau_t)K_h(u-\tau_t)\right|\right] \\
= & O(1)\frac{\nu_T}{Th} + E\left[\sup_{(\theta,u)\in\Theta\times[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta} f\left(\sqrt{t}\cdot\frac{x_t}{\sqrt{t}},\tau_t\right)K_h(u-\tau_t)\right|\right] \\
\leqslant & O(1)\frac{\nu_T}{Th} + \frac{1}{T}\sum_{t=\nu_T+1}^{T}\int \sup_{(\theta,u)\in\Theta\times[0,1]}\tau_t^{\theta}|f(\sqrt{t}w,\tau_t)|K_h(u-\tau_t)\phi_t(w)dw \\
\leqslant & O(1)\frac{\nu_T}{Th} + \sup_{u\in[0,1]}\frac{1}{T}\sum_{t=1}^{T}K_h(u-\tau_t)\int\phi_t(w)dw = O(1),
\end{aligned}
$$

where the second inequality follows from Assumption 3.3; and the last equality follows from (B.13) and the fact that $\phi_t(w)$ is a density function.

Thus, we have $\sup_{(\theta,u)\in\Theta\times[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta} f(x_t,\tau_t)K_h(u-\tau_t)\right| = O_P(1)$, so the proof of the first result is complete.

Based on the first result of this corollary, the second and third results can be verified by exactly the same procedure as documented in the Appendix A of the this paper. ■

Another Potential Issue

We now explain the failure of a sieve based OLS method based on model (1.1). Still consider $y_t = g(\tau_t)t^{\theta_0} + \varepsilon_t$. Further assume θ_0 is known. Following Newey (1997), we can expand $g(\cdot)$ by power series on certain support as follows:

$$
\begin{aligned}
T^{-\theta_0}y_t &= T^{-\theta_0}\sum_{i=0}^{k-1}c_i\tau_t^i t^{\theta_0} + T^{-\theta_0}\sum_{i=k}^{\infty}c_i\tau_t^i t^{\theta_0} + T^{-\theta_0}\varepsilon_t \\
&= \sum_{i=0}^{k-1}c_i\tau_t^{i+\theta_0} + \sum_{i=k}^{\infty}c_i\tau_t^{i+\theta_0} + T^{-\theta_0}\varepsilon_t.
\end{aligned}
$$

In view of (6.6), it is easy to obtain

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}(\tau_t^{\theta_0},\tau_t^{\theta_0+1},\ldots,\tau_t^{\theta_0+k-1})(\tau_t^{\theta_0},\tau_t^{\theta_0+1},\ldots,\tau_t^{\theta_0+k-1})' \\
= & \left\{\frac{1}{2\theta_0+i+j+1}\right\}_{k\times k}\cdot(1+o(1)) \qquad \text{(B.14)}
\end{aligned}
$$

for $0\leqslant i,j\leqslant k-1$ under proper restrictions on k and T. Thus, as k diverges, the right hand side of (B.14) is asymptotically singular, which indicates that the sieve based OLS method also does not work for model (1.1). Certainly, the choice of basis functions plays an important role when implementing the sieve based OLS method. However, it is not clear to us which series can solve the ill-posed problem at this stage.

Numerical Studies

We now use simulation to examine Corollary B.2 and the potential issue mentioned above together.

Specifically, we adopt the following DGPs:

$$\begin{aligned} &\textit{DGP 1:} \quad y_t = f(x_t, \tau_t) + g(\tau_t)t^{\theta_0} + \varepsilon_t \quad \text{with} \quad g(u) = 3(u-1)^2 + 1, \\ &\textit{DGP 2:} \quad y_t = f(x_t, \tau_t) + g(\tau_t)t^{\theta_0} + \varepsilon_t \quad \text{with} \quad g(u) = 3|u-1|^{0.7} + 1. \end{aligned} \tag{B.15}$$

The $f(\cdot,\cdot)$ and $\{x_t\}$ are generated as follows:

- *Case 1 (Stationary)*: x_t follows an AR(1) process $x_t = 0.5x_{t-1} + v_t$, and $f(x,u) = \sum_{j=1}^d |x_j| + 5\sin(u\cdot\pi)$ with $x = (x_1,\ldots,x_d)'$;
- *Case 2 (Nonstationary)*: x_t follows an integrated process $x_t = x_{t-1} + v_t$, and $f(x,u) = \exp\left\{-(\sum_{j=1}^d x_j)^2\right\} + 5\sin(u\cdot\pi)$ with $x = (x_1,\ldots,x_d)'$.

In both cases, $x_0 \sim N(0_{d\times 1}, I_d)$ and $v_t \sim i.i.d.\ N(0_{d\times 1}, I_d)$. We set $d = 1$ without losing generality. The other variables are generated in exactly the same way as Section 3 of this paper.

We first implement NM, FOLS1 and FOLS2 methods documented in the main text to DGP 1 under both Cases 1 and 2, and report results in Tables 4 and 5 below.

Table 4: *(DGP1, Case 1)*

		RMSE$_\theta$			RMSE$_g$		
	$h\backslash T$	200	500	1000	200	500	1000
NM	$T^{-2/5}$	0.1004	0.0914	0.0846	0.1020	0.0534	0.0326
	$T^{-1/3}$	0.1057	0.0938	0.0859	0.0383	0.0215	0.0157
	$T^{-1/5}$	0.1263	0.1069	0.0951	0.0869	0.0946	0.0928
	$T^{-1/8}$	0.1581	0.1304	0.1144	0.0976	0.1283	0.1396
FOLS1	$T^{-2/5}$	0.3000	0.3000	0.3000	5.4512	7.4493	9.4524
	$T^{-1/3}$	0.3000	0.3000	0.3000	5.3382	7.3816	9.4220
	$T^{-1/5}$	0.3000	0.3000	0.3000	4.9488	6.8721	8.8398
	$T^{-1/8}$	0.3000	0.3000	0.3000	4.8027	6.4911	8.2334
FOLS2	$T^{-2/5}$	0.2746	0.2688	0.2676	4.6967	6.0574	7.4868
	$T^{-1/3}$	0.2773	0.2808	0.2794	4.6713	6.4921	8.1174
	$T^{-1/5}$	0.2628	0.2790	0.2865	3.9134	5.9406	7.9951
	$T^{-1/8}$	0.2431	0.2728	0.2827	3.3330	5.3493	7.2129

Table 5: *(DGP1, Case 2)*

		RMSE$_\theta$			RMSE$_g$		
	$h\backslash T$	200	500	1000	200	500	1000
NM	$T^{-2/5}$	0.1025	0.0923	0.0851	0.0948	0.0482	0.0286
	$T^{-1/3}$	0.1078	0.0947	0.0864	0.0382	0.0229	0.0175
	$T^{-1/5}$	0.1284	0.1078	0.0956	0.0873	0.0951	0.0933
	$T^{-1/8}$	0.1603	0.1314	0.1149	0.0976	0.1284	0.1397
FOLS1	$T^{-2/5}$	0.3000	0.3000	0.3000	5.3572	7.3848	9.4045
	$T^{-1/3}$	0.3000	0.3000	0.3000	5.2526	7.3231	9.3785
	$T^{-1/5}$	0.3000	0.3000	0.3000	4.8719	6.8208	8.8019
	$T^{-1/8}$	0.3000	0.3000	0.3000	4.7293	6.4433	8.1987
FOLS2	$T^{-2/5}$	0.2733	0.2670	0.2670	4.5755	5.9264	7.4130
	$T^{-1/3}$	0.2762	0.2808	0.2791	4.5590	6.4374	8.0573
	$T^{-1/5}$	0.2652	0.2783	0.2866	3.9096	5.8606	7.9608
	$T^{-1/8}$	0.2416	0.2722	0.2822	3.2417	5.2840	7.1504

As can be seen, the procedure of recovering θ_0 and g_0 is not affected by $f(\cdot,\cdot)$ and $\{x_t \,|t=1,\ldots,T\}$ too much, which indicates that one can implement our procedure to detrend the data set in a better fashion practically. However, when FOLS1 and FOLS2 get employed, huge biases arise. Thus, detrending the data set by a proper econometric tool indeed matters.

Below we focus on DGPs 1 and 2 under Case 1 only in order to examine the issue raised by sieve estimation technique. Apart from our proposed method, we also use sieve based OLS method (referred to as SOLS). In particular, we use power series $\{1, u, u^2, \ldots\}$ to approximate $g(u)$ in our simulation study (cf., Newey (1997)). Specifically, the new objective function is

$$Q_T(\theta) = \sum_{t=1}^{T}\left(y_t - t^\theta \widehat{g}_k(\tau_t,\theta)\right)^2, \tag{B.16}$$

where $\widehat{g}_k(\tau_t,\theta) = z_t'\widehat{C}(\theta)$, $z_t = (1, \tau_t^1, \ldots, \tau_t^{k-1})'$, and $\widehat{C}(\theta) = \left(\sum_{t=1}^T \left[t^\theta z_t\right]\cdot\left[t^\theta z_t\right]'\right)^{-1}\sum_{t=1}^T \left[t^\theta z_t\right] y_t$. In order to demonstrate our arguments under (B.14), we set the truncation parameter to $k = 2,\ 3,\ 5,\ 10,\ 15$. For the purpose of comparison, we set the bandwidth to $h = 1/k$ when implementing our method.[6] In Table 6, it is not surprising to see the best estimate comes from SOLS method with $k = 3$, as this choice of power series perfectly fits the DGP 1. However, when we increase the value of truncation parameter, the matrix in the inverse is getting closer to singular as explained under (B.14), which is also confirmed by Matlab over simulation study as we always receive warnings saying "*Matrix is close to singular or badly scaled*".

[6]The setting of $h = 1/k$ is indeed reasonable. As for a nonparametric model $y_t = g(x_t) + e_t$ with $t = 1, \ldots, T$, it is easy to see that the leading terms of the rates of convergence are $\sqrt{\frac{k^d}{T}}$ and $\frac{1}{\sqrt{Th^d}}$ for the sieve based method and the kernel based method, respectively, under certain restrictions, where k is the truncation parameter, h is the bandwidth, and d is the dimension of x_t. For more details, see Chen (2007) and Gao (2007) for excellent reviews of nonparametric regression.

Table 6: *(DGP 1, Case 1)*

		RMSE$_\theta$			RMSE$_g$		
		$T=200$	$T=500$	$T=1000$	$T=200$	$T=500$	$T=1000$
NM	$h=1/2$	0.1548	0.1378	0.1263	0.1019	0.1183	0.1263
	$h=1/3$	0.1240	0.1121	0.1034	0.0822	0.1122	0.1254
	$h=1/5$	0.1083	0.0985	0.0911	0.0295	0.0492	0.0659
	$h=1/10$	0.1015	0.0927	0.0859	0.0826	0.0281	0.0157
	$h=1/15$	0.1003	0.0918	0.0850	0.1046	0.0432	0.0198
SOLS	$k=2$	0.3000	0.3000	0.3000	4.7499	6.4232	8.0886
	$k=3$	0.0161	0.0057	0.0030	0.1035	0.0365	0.0178
	$k=5$	0.0599	0.0199	0.0094	0.6624	0.2456	0.1313
	$k=10$	0.2407	0.2124	0.1998	1.0880	1.2357	1.3105
	$k=15$	0.3242	0.3167	0.1237	1.2188	1.4768	0.9684

Although the power series may work well with a relatively small truncation parameter when $g(\cdot)$ is a certain polynomial function, it may not even work well for the case where the powers of polynomial functions are not integers, which is confirmed by the simulation study for DGP 2. In Table 7, we see that the results of SOLS generally perform worse than our proposed method, which indicates that the choice of the basis functions indeed matters. However, at this stage, it is not clear which particular class of basis functions can potentially solve the problem discussed under (B.14), while the finite sample results are generally good.

Table 7: *(DGP 2, Case 1)*

		RMSE$_\theta$			RMSE$_g$		
		$T=200$	$T=500$	$T=1000$	$T=200$	$T=500$	$T=1000$
NM	$h=1/2$	0.0723	0.0650	0.0598	0.9221	0.9424	0.9522
	$h=1/3$	0.0435	0.0408	0.0381	0.8646	0.8981	0.9133
	$h=1/5$	0.0312	0.0301	0.0284	0.6699	0.7209	0.7436
	$h=1/10$	0.0269	0.0265	0.0252	0.4152	0.4806	0.5090
	$h=1/15$	0.0262	0.0259	0.0247	0.2948	0.3642	0.3943
SOLS	$k=2$	0.1879	0.1876	0.1877	1.2468	1.4045	1.5010
	$k=3$	0.2131	0.2192	0.2219	4.5228	6.0737	7.5396
	$k=5$	0.1866	0.1719	0.1656	3.8811	4.2123	4.6908
	$k=10$	0.2736	0.2853	0.2002	1.6633	2.0399	1.8726
	$k=15$	0.2679	0.2573	0.2002	1.6132	1.9826	1.8734

References

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, John Wiley and Sons, INC.

Andrews, D. W. K. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**(3), 817–858.

Andrews, D. W. K. and McDermott, C. J. (1995), 'Nonlinear econometric models with deterministically trending variables', *The Review of Economic Studies* **62**(3), 343–360.

Bai, J. and Perron, P. (1998), 'Estimating and testing linear models with multiple structural changes', *Econometrica* **66**(1), 47–78.

Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction, second edition. In: Lecture Notes in Statistics*, Springer-Verlag.

Cai, Z. (2007), 'Trending time-varying coefficient time series models with serially correlated errors', *Journal of Econometrics* **136**(1), 163–188.

Chen, J., Gao, J. and Li, D. (2012*a*), 'A new diagnostic test for cross-section uncorrelatedness in nonparametric panel data models', *Econometric Theory* **28**(5), 1144–1163.

Chen, J., Gao, J. and Li, D. (2012*b*), 'Semiparametric trending panel data models with cross-sectional dependence', *Journal of Econometrics* **171**(1), 71–85.

Chen, X. (2007), 'Large sample sieve estimation of semi-nonparametric models', *Handbook of Econometrics* pp. 5549–5632.

Dahlhaus, R. (1997), 'Fitting time series models to nonstationary processes', *The Annals of Statistics* **25**(1), 1–37.

Delgado, M. A. and Hidalgo, J. (2000), 'Nonparametric inference on structural breaks', *Journal of Econometrics* **96**(1), 113–144.

Dong, C., Gao, J. and Tjøstheim, D. (2016), 'Estimation for single-index and partially linear single-index nonstationary time series models', *The Annals of Statistics* **44**(1), 425–453.

Dong, C. and Linton, O. (2016), Additive nonparametric models with time variable and both stationary and nonstationary regressors. https://ssrn.com/abstract=2847681.

Engle, R. F. and Rangel, J. G. (2008), 'The spline-garch model for low-frequency volatility and its global macroeconomic causes', *The Review of Financial Studies* **21**(3), 1187–1222.

Fan, J. and Li, R. (2001), 'Variable selection via nonconcave penalized likelihood and its oracle properties', *Journal of the American Statistical Association* **96**(456), 1348–1360.

Fan, J. and Yao, Q. (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer-Verlag.

Fan, Y. and Li, Q. (1996), 'Consistent model specification tests: omitted variables, parametric and semiparametric functional forms', *Econometrica* **64**(4), 865–890.

Feng, G. and Serletis, A. (2008), 'Productivity trends in U.S. manufacturing: Evidence from the nq and aim cost functions', *Journal of Econometrics* **142**(1), 281–311.

Gao, J. (2007), *Nonlinear Time Series: Semiparametric and Nonparametric Methods*, Chapman and Hall/CRC.

Gao, J. and Anh, V. V. (1999), 'Semiparametric regression under long-range dependent errors', *Journal of Statistical Planning and Inference* **80**, 37–57.

Gao, J. and Hawthorne, K. (2006), 'Semiparametric estimation and testing of the trend of temperature series', *Econometrics Journal* **9**, 332–355.

Giraitis, L., Kapetanios, G. and Yates, T. (2014), 'Inference on stochastic time-varying coefficient models', *Journal of Econometrics* **179**(1), 46–65.

Greene, W. (2002), *Econometric Analysis (Fifth ed.)*, New Jersey: Prentice Hall.

Greene, W. (2005), 'Reconsidering heterogeneity in panel data estimators of the stochastic frontier model', *Journal of Econometrics* **126**, 269–303.

Hafner, C. M. and Linton, O. (2010), 'Efficient estimation of a multivariate multiplicative volatility model', *Journal of Econometrics* **159**(1), 55–73.

Hamilton, J. D. (2017), Why you should never use the hodrick-prescott filter. https://www.nber.org/papers/w23429.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Presss.

Jones, H. L. (1943), 'Fitting polynomial trends to seasonal data by the method of least squares', *Journal of the American Statistical Association* **38**, 453–465.

Li, Q. (1999), 'Consistent model specification tests for time series econometric models', *Journal of Econometrics* **92**(1), 101–147.

Newey, W. K. (1997), 'Convergence rates and asymptotic normality for series estimators', *Journal of Econometrics* **97**, 147–168.

Newey, W. K. and Powell, J. L. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**(5), 1565–1578.

Pesaran, M. H. and Yang, C. F. (2016), Econometric analysis of production networks with dominant units. Working Paper No. 16-25.

Phillips, P. C. B. (2001), 'Trending time series and macroeconomic activity: Some present and future challenges', *Journal of Econometrics* **100**(1), 21–27.

Phillips, P. C. B. (2005), 'Challenges of trending time series econometrics', *Mathematics and Computers in Simulation* **68**(5), 401–416.

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

Phillips, P. C. B. (2009), 'The mysteries of trend', *Macroeconomic Review* **9**(2), 82–89.

Phillips, P. C. B. and Jin, S. (2015), Business cycles, trend elimination, and the HP filter. Cowles Foundation Discussion Paper No. 2005.

Phillips, P. C. B., Li, D. and Gao, J. (2017), 'Estimating smooth structural change in cointegration models', *Journal of Econometrics* **196**(1), 180–195.

Robinson, P. M. (1997), 'Large-sample inference for nonparametric regression with dependent errors', *The Annals of Statistics* **25**(5), 2054–2083.

Robinson, P. M. (2012), 'Inference on power law spatial trends', *Bernoulli* **18**(2), 644–677.

Sornette, D. (2003), *Why Stock Markets Crash*, Princeton University Press.

Su, L. and Jin, S. (2012), 'Sieve estimation of panel data models with cross section dependence', *Journal of Econometrics* **169**(1), 34–47.

Su, L., Jin, S. and Zhang, Y. (2015), 'Specification test for panel data models with interactive fixed effects', *Journal of Econometrics* **186**(1), 222–244.

Vogt, M. (2012), 'Nonparametric regression for locally stationary time series', *The Annals of Statistics* **40**(5), 2601–2633.

Wang, H. and Xia, Y. (2009), 'Shrinkage estimation of the varying coefficient', *Journal of the American Statistical Association* **104**(486), 747–757.

Wooldridge, J. M. (1994), 'Estimation and inference for dependent processes', *Handbook of Econometrics* pp. 2639–2738.