



*Citation for published version:*

Gao, J, Linton, O & Peng, B 2020, 'Inference on a Semiparametric Model with Global Power Law and Local Nonparametric Trends', *Econometric Theory*, vol. 36, no. 2, pp. 223-249.  
<https://doi.org/10.1017/S0266466619000094>

*DOI:*

[10.1017/S0266466619000094](https://doi.org/10.1017/S0266466619000094)

*Publication date:*

2020

*Document Version*

Other version

[Link to publication](#)

## University of Bath

### Alternative formats

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**Online Supplementary Appendix to  
“Inference on a Semiparametric Model with  
Global Power Law and Local Nonparametric Trends”**

JITI GAO\*, OLIVER LINTON\* AND BIN PENG†

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

In this online supplementary file, we provide two empirical studies, preliminary lemmas with the associated development, omitted proofs of the main text, another modelling issue of studying power trend, and some extra simulation studies.

## Appendix B

### B.1 Empirical Study

We provide two case studies in this section. Firstly, we focus on the global mean sea level (GMSL). Then we move on to investigate the U.S. GDP data.

#### B.1.1 Global Mean Sea Level

The data is collected from CSIRO<sup>1</sup>, and is recorded in millimetres originally. As shown in Figure B.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., (B.1) below) are defined on  $t = 1, \dots, T$ , both models in fact have  $y_0 = 0$  if  $t = 0$  is permitted. Therefore, we shift the data set vertically to let  $y_0$  (i.e., the value of year 1880) be 0 for better fit.

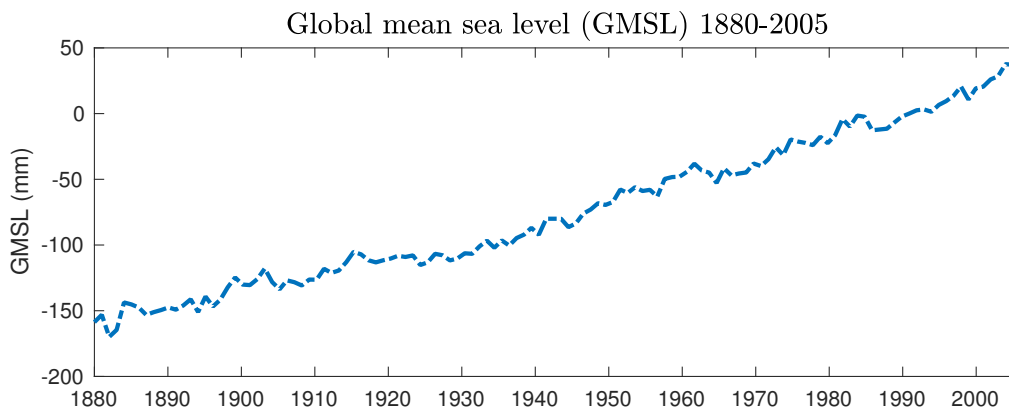


Figure B.1: Global Mean Sea Level

<sup>1</sup><http://www.cmar.csiro.au/sealevel/index.html>

We first implement the two hypothesis tests of Section 3. The detailed testing procedures are identical to the simulation section, so we do not repeat them again for conciseness. Table B.1 below summarizes the statistic values of two tests and the corresponding decisions at 95% significant level.<sup>2</sup>

Table B.1: Results of Two Tests

	Statistic Value	Decision
Testing $\theta_0$	3.57	Reject
Testing $g(\cdot)$	2.44	Reject

Based on Table B.1, we have enough evidence to move on to consider model (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 the simulation section. In order to check the sensitivity of our nonparametric approach, we use two more bandwidths  $h_{left} = h_{opt} - 0.03$  and  $h_{right} = h_{opt} + 0.03$  to implement the nonparametric regression below.

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

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

and estimate  $\theta_0 = (\theta_{0,1}, \dots, \theta_{0,d})'$  and  $\beta_0 = (\beta_1, \dots, \beta_d)'$  of (B.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  $d \geq 2$ . Therefore, we set  $d = 1$  throughout this study, which essentially gives a model of Phillips (2007).

We report the estimation results of both methods in Table B.2, and plot the estimated  $g_0$  under three choices of bandwidth in Figure B.2. It is clear that the estimation results of  $\theta_0$  and  $g_0$  are quite stable with respect to the choice of bandwidth.

Table B.2: Estimation Results for Section 4

	$h$	$\theta_0$	$\beta_0$
NM ( $h_{opt}$ )	0.1666	0.8527	–
NM ( $h_{left}$ )	0.1366	0.8529	–
NM ( $h_{right}$ )	0.1966	0.8521	–
Para-R	–	1.0000	0.4676

By plotting the estimation residuals for  $t = [Th] + 1, \dots, T$  in Figure B.3, it is easy to see that the residuals of NM indeed are smaller than those of Para-R.

<sup>2</sup>Using the odd numbered observations to estimate  $g(\cdot)$  and evaluating the score function with the even numbered observations gives the statistic value 2.54. Either way, we reject the null hypothesis.

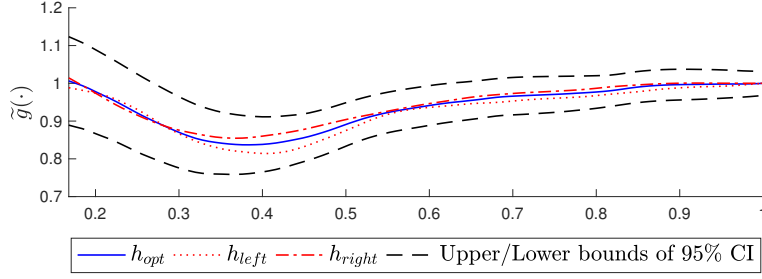


Figure B.2: Estimation of  $g_0$

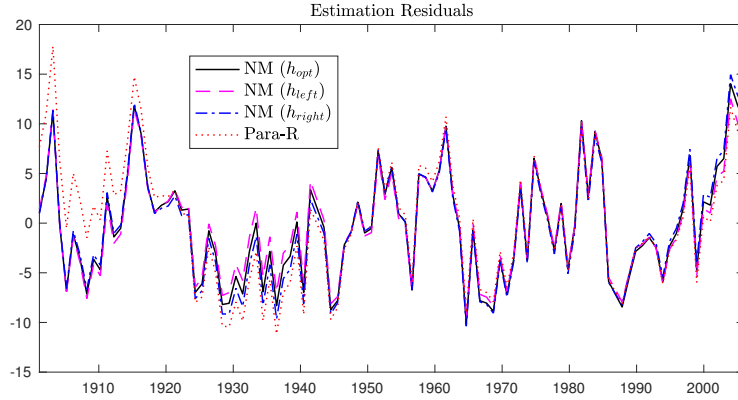


Figure B.3: Estimation Residuals

Finally, we take a look at the out-sample root mean squared errors (OSRMSE) of both methods, and they are specifically calculated as follows.

$$OSRMSE = \sqrt{\frac{1}{5} \sum_{j=0}^4 (y_{T_j} - \hat{y}_{T_j})^2},$$

where  $T_j = T - j$ , and  $\hat{y}_{T_j}$  is obtained by using sample  $\{y_t \mid t = 1, \dots, T_j - 1\}$  for both methods. As how to calculate  $\hat{y}_{T_j}$  is obvious for Para-R, we omit the details. Below we explain how to obtain  $\hat{y}_{T_j}$  using the NM method. Specifically, the objective function is specified as follows.

$$R_{T_j}(\theta) = \left\{ \lambda_{T_j} \cdot \ln \left[ \frac{1}{T_j} \sum_{t=\lfloor T_j h \rfloor + 1}^{T_j - 1} \left( \frac{t}{T_j} \right)^{2\theta} \hat{g}_{T_j}(\tau_t, \theta) \right]^2 \right\}^2,$$

where  $\hat{g}_{T_j}(u, \theta) = \left[ \sum_{t=1}^{T_j - 1} t^{2\theta} K_h(u - \tau_t) \right]^{-1} \sum_{t=1}^{T_j - 1} t^\theta y_t K_h(u - \tau_t)$ . Thus,  $\hat{y}_{T_j} = T_j^{\hat{\theta}_{T_j}} \hat{g}_{T_j}(1, \hat{\theta})$ , where  $\hat{\theta}_{T_j} = \text{argmin}_\theta R_{T_j}(\theta)$ . We summarise the results in the next table. In this case, Para-R slightly outperforms the NM method.

Table B.3: Out-Sample Root Mean Squared Errors

NM	Para-R
11.86	9.28

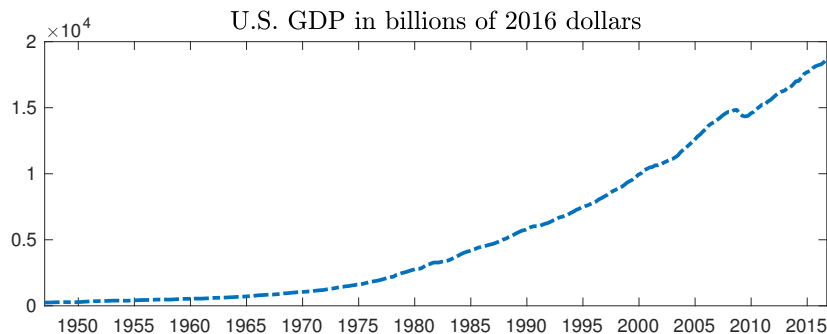


Figure B.4: U.S. GDP Data (1947 Q1 - 2016 Q3)

### B.1.2 U.S. GDP

We now provide a case study by investigating U.S. GDP data, which are collected from the Bureau of Economic Analysis, U.S. Department of Commerce<sup>3</sup>, and are recorded in billions of 2016 U.S. dollars. As shown in Figure B.4, the range of raw data covering 1947 Q1 to 2016 Q3 is from 243 to 18,675, and has a strong nonlinear time trend.

We repeat the testing and estimation procedures as we do for the GMSL. Table B.4 below summarizes the statistic values of two tests and the corresponding decisions at the 95% significance level.<sup>4</sup>

Table B.4: Results of Two Tests

	Statistic Value	Decision
Testing $\theta_0$	2.16	Reject
Testing $g(\cdot)$	26.24	Reject

We report the estimation results in Table B.5, and plot the estimated  $g_0$  under three choices of bandwidth in Figure B.5.

The estimation residuals for  $t = [Th] + 1, \dots, T$  (also considered as detrended series) are plotted in Figure B.6. It is easy to see that the residuals of NM are indeed smaller than those of Para-R, and both methods reveal the trending heteroskedasticity in the residuals. Moreover, if we consider the above procedure as a detrending process, fluctuations about the trend are the true focus. It is then interesting to see that

<sup>3</sup><https://bea.gov/national>

<sup>4</sup>Using the odd numbered observations to estimate  $g(\cdot)$  and evaluating the score function with the even numbered observations gives the statistic value 2.46. Still, we reject the null hypothesis.

Table B.5: Estimation Results

	$h$	$\theta_0$	$\beta_0$
NM ( $h_{opt}$ )	0.1001	1.4653	–
NM ( $h_{left}$ )	0.0701	1.4650	–
NM ( $h_{right}$ )	0.1301	1.4656	–
Para-R	–	2.5844	0.0092

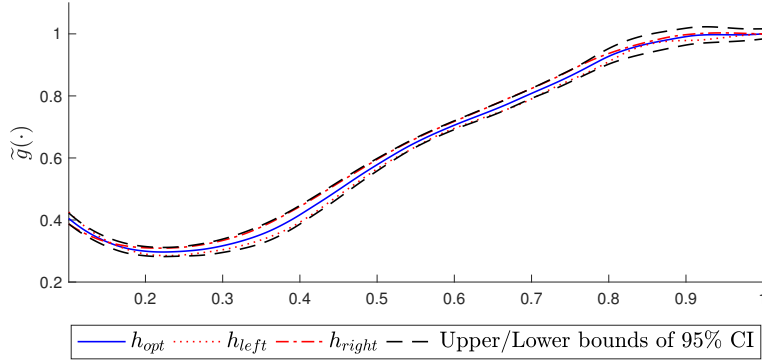


Figure B.5: Estimation of  $g_0$

both methods clearly reveal (1) Early 1980s recession<sup>5</sup>, (2) Recession of the early 1990s<sup>6</sup>, (3) Stock market downturn of 2002<sup>7</sup>, and (4) Global financial crisis<sup>8</sup> (GFS) in the history of the U.S. For the first three, both methods agree with each other well in terms of starting and ending date, but Para-R suggests that the GFS is still going on during 2014–2016, which is contradictory to the economic prevailing climate of these three years of the U.S. (Maria and Wen, 2015).

Finally, we summarise the results of OSRMSE in the next table, and in this case, NM outperforms Para-R.

Table B.6: Out-Sample Root Mean Squared Errors

NM	Para-R
594	671

<sup>5</sup>The early 1980s recession describes the severe global economic recession affecting much of the developed world in the late 1970s and early 1980s.

<sup>6</sup>The recession of the early 1990s describes the period of economic downturn affecting much of the world in the late 1980s and early 1990s.

<sup>7</sup>In 2001, stock prices took a sharp downturn in stock markets across the U.S., Canada, Asia, and Europe.

<sup>8</sup>It began in 2007 with a crisis in the subprime mortgage market in the U.S., and developed into a full-blown international banking crisis in 2008. The crisis was followed by a global economic downturn, the Great Recession.

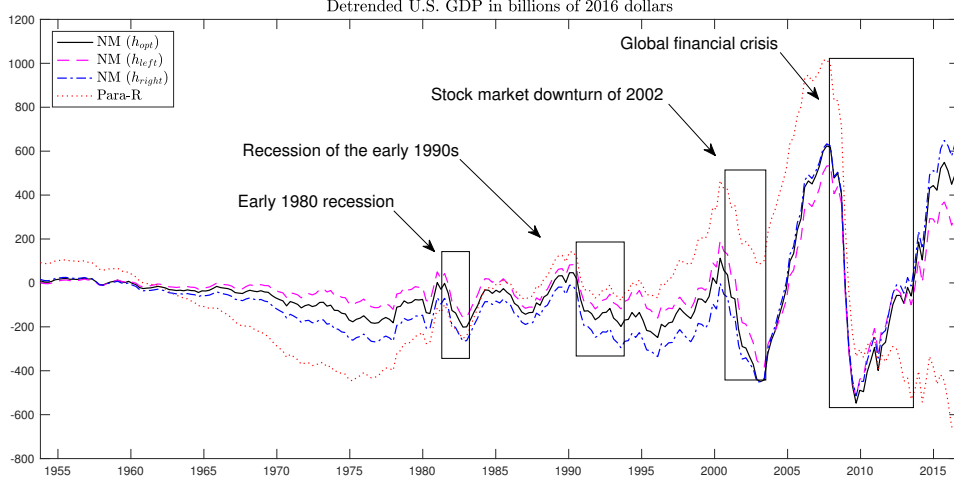


Figure B.6: Detrended U.S. GDP

## B.2 Proofs

This subsection includes preliminary lemmas with the associated development and omitted proofs of the main text. Before proceeding further, we prepare some notations for later use. Let  $\Lambda_{T,h}(u, \theta) = \sum_{t=1}^T t^{2\theta} K_h(u - \tau_t)$ . Simple calculation shows that

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

$$\begin{aligned}
& \cdot \left\{ \frac{1}{T} \sum_{t=[Th]+1}^T \tau_t^{2\theta} \frac{\partial \hat{g}(\tau_t, \theta)}{\partial \theta} + \frac{2}{T} \sum_{t=[Th]+1}^T \tau_t^{2\theta} \hat{g}(\tau_t, \theta) \ln \tau_t \right\} ; \\
& \frac{\partial^2 R_T(\theta)}{\partial \theta^2} = -4\lambda_T^2 \left\{ \ln \left[ \frac{1}{T} \sum_{t=[Th]+1}^T \tau_t^{2\theta} \hat{g}(\tau_t, \theta) \right]^2 \right\} \cdot \left[ \frac{1}{T} \sum_{t=[Th]+1}^T \tau_t^{2\theta} \hat{g}(\tau_t, \theta) \right]^{-2} \\
& \cdot \left\{ \frac{1}{T} \sum_{t=[Th]+1}^T \tau_t^{2\theta} \frac{\partial \hat{g}(\tau_t, \theta)}{\partial \theta} + \frac{2}{T} \sum_{t=[Th]+1}^T \tau_t^{2\theta} \hat{g}(\tau_t, \theta) \ln \tau_t \right\}^2 \\
& + 4\lambda_T^2 \left\{ \ln \left[ \frac{1}{T} \sum_{t=[Th]+1}^T \tau_t^{2\theta} \hat{g}(\tau_t, \theta) \right]^2 \right\} \cdot \left[ \frac{1}{T} \sum_{t=[Th]+1}^T \tau_t^{2\theta} \hat{g}(\tau_t, \theta) \right]^{-1} \\
& \cdot \left\{ \frac{1}{T} \sum_{t=[Th]+1}^T \tau_t^{2\theta} \frac{\partial^2 \hat{g}(\tau_t, \theta)}{\partial^2 \theta} + \frac{4}{T} \sum_{t=[Th]+1}^T \tau_t^{2\theta} \frac{\partial \hat{g}(\tau_t, \theta)}{\partial \theta} \ln \tau_t + \frac{4}{T} \sum_{t=[Th]+1}^T \tau_t^{2\theta} \hat{g}(\tau_t, \theta) (\ln \tau_t)^2 \right\} \\
& + 8\lambda_T^2 \left[ \frac{1}{T} \sum_{t=[Th]+1}^T \tau_t^{2\theta} \hat{g}(\tau_t, \theta) \right]^{-2} \cdot \left\{ \frac{1}{T} \sum_{t=[Th]+1}^T \tau_t^{2\theta} \frac{\partial \hat{g}(\tau_t, \theta)}{\partial \theta} + \frac{2}{T} \sum_{t=[Th]+1}^T \tau_t^{2\theta} \hat{g}(\tau_t, \theta) \ln \tau_t \right\}^2. \quad (\text{B.2})
\end{aligned}$$

### B.2.1 Preliminary Lemmas with Associated Proofs

Recall that we consider the case where  $\theta_0 > 0$  and  $g$  is a non-constant function in Section 3, and we will not repeat this again in the following development.

#### Lemma B.1.

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

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

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

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

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

**Lemma B.2.** Let Assumption 1 hold, and define

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

As  $T \rightarrow \infty$ ,



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.  $\sup_{(\theta, u) \in \Theta \times [h,1]} \left| \frac{1}{T} \sum_{t=1}^T \tau_t^{\theta+\theta_0} g(\tau_t) K_h(\tau_t - u) - \tilde{c} u^{\theta+\theta_0} g(u) \right| = O(h)$ ;
5.  $\sup_{(\theta, u) \in \Theta \times B_{\varepsilon_1}(h)} \left| \frac{1}{T} \sum_{t=1}^T \tau_t^{2\theta} K_h(\tau_t - u) - \tilde{c} u^{2\theta} \right| = O_P(1) h^{\min\{2c^*, 1\}}$ , where  $B_{\varepsilon_1}(h)$  has been defined in Lemma 4.1, and  $c^* = \min_{\{\theta \in \Theta\}} \theta > 0$ ;
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  $\theta_0$  belongs to,  $v_T(\theta) = \frac{1}{T} \sum_{t=1}^T \tau_t^{\theta_0+\theta} g(\tau_t)$  and  $v(\theta) = \int_0^1 u^{\theta_0+\theta} g(u) du$ .

**Lemma B.3.** Under Assumption 1, as  $T \rightarrow \infty$ ,

1.  $\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{2\theta} \frac{\partial^2 \hat{g}(\tau_t, \theta)}{\partial \theta^2} \Big|_{\theta=\tilde{\theta}} = (\ln T)^2 \phi_1 + 2(\ln T) \phi_2 + \phi_3 + o_P(1)$ ,
2.  $\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{2\theta} \frac{\partial \hat{g}(\tau_t, \theta)}{\partial \theta} \Big|_{\theta=\tilde{\theta}} = -(\ln T) \phi_1 - \phi_2 + o_P(1)$ ,
3.  $\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{2\theta} \frac{\partial \hat{g}(\tau_t, \theta)}{\partial \theta} \ln \tau_t \Big|_{\theta=\tilde{\theta}} = -(\ln T) \phi_2 - \phi_3 + o_P(1)$ ,
4.  $\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{2\theta} \hat{g}(\tau_t, \theta) \Big|_{\theta=\tilde{\theta}} = \phi_1 + o_P(1)$ ,
5.  $\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{2\theta} \hat{g}(\tau_t, \theta) \ln \tau_t \Big|_{\theta=\tilde{\theta}} = \phi_2 + o_P(1)$ ,
6.  $\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{2\theta} \hat{g}(\tau_t, \theta) (\ln \tau_t)^2 \Big|_{\theta=\tilde{\theta}} = \phi_3 + o_P(1)$ ,
7.  $\frac{\partial^2 R_T(\theta)}{\partial \theta^2} \Big|_{\theta=\tilde{\theta}} = 8 + o_P(1)$ ,

where  $\phi_1$  to  $\phi_3$  are defined by (B.8) to (B.10) respectively; and  $\tilde{\theta}$  is defined in (4.9) of the main text.

**Proof of Lemma B.1:**

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

(2). Write

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

where the second equality follows from the definition of the Riemann integral. The proof is complete.  $\blacksquare$

**Proof of Lemma B.2:**

(1). Let  $l(T)$  be any positive function satisfying that  $l(T) \rightarrow \infty$  as  $T \rightarrow \infty$ . By the same arguments as (B.10) and (B.11) of Chen et al. (2012), 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 a finite number of subintervals  $\{B_i\}$  that are centred at  $b_i$  and of length  $\kappa_T = o(h^2)$ . Denote  $U_T$  as the number of such subintervals, which immediately gives  $U_T = O(\kappa_T^{-1})$ . Below, we take  $\kappa_T = [l(T)]^{1-v} \cdot \sqrt{\frac{\ln T}{Th}} \cdot h^2$  for a sufficiently large  $v$ , which satisfies  $v \leq 2 + \delta/2$  and  $\delta$  is defined in Assumption 1.2. Write

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

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

For  $\Pi_{1T}$ ,

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

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

For  $\Pi_{2T}$ , we use a truncation technique, so for the same  $v$  above denote  $\tilde{\varepsilon}_t = \varepsilon_t \cdot I[|\varepsilon_t| \leq T^{1/v} l(T)]$  and  $\tilde{\varepsilon}_t^c = \varepsilon_t - \tilde{\varepsilon}_t$ , where  $I[\cdot]$  is the indicator function. Thus, we obtain that

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

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

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

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

$$\begin{aligned}
&\leq \Pr \left( \max_{1 \leq i \leq U_T} \max_{1 \leq t \leq T} \frac{1}{h} \tau_t^\theta K \left( \frac{b_i - \tau_t}{h} \right) |\tilde{\varepsilon}_t^c| \geq \epsilon l(T) \sqrt{\frac{\ln T}{Th}} \right) \\
&= \Pr \left( \max_{1 \leq i \leq U_T} K \left( \frac{b_i - \tau_t}{h} \right) \max_{1 \leq t \leq T} \tau_t^\theta |\tilde{\varepsilon}_t^c| \geq \epsilon l(T) \sqrt{\frac{h \ln T}{T}} \right) \\
&\leq \Pr \left( \max_{1 \leq t \leq T} |\tilde{\varepsilon}_t^c| \geq 0 \right) \leq \sum_{t=1}^T \Pr \left( |\varepsilon_t| > T^{1/v} l(T) \right) \\
&\leq \sum_{t=1}^T \frac{E|\varepsilon_t|^v}{[l(T)]^v T} = O(1) \frac{1}{[l(T)]^v} = o(1),
\end{aligned}$$

where the third equality follows from the existence of  $E|\varepsilon_t|^v$  due to  $v \leq 2 + \delta/2$  and Assumption 1.2.

For  $\Pi_{2T,1}$ , write

$$\begin{aligned}
\Pi_{2T,1} &= \max_{1 \leq i \leq U_T} \left| \frac{1}{Th} \sum_{t=1}^T \tau_t^\theta K \left( \frac{b_i - \tau_t}{h} \right) (\tilde{\varepsilon}_t - E[\tilde{\varepsilon}_t] + E[\tilde{\varepsilon}_t]) \right| \\
&\leq \max_{1 \leq i \leq U_T} \left| \frac{1}{Th} \sum_{t=1}^T \tau_t^\theta K \left( \frac{b_i - \tau_t}{h} \right) (\tilde{\varepsilon}_t - E[\tilde{\varepsilon}_t]) \right| + \max_{1 \leq i \leq U_T} \left| \frac{1}{Th} \sum_{t=1}^T \tau_t^\theta K \left( \frac{b_i - \tau_t}{h} \right) E[\varepsilon_t - \tilde{\varepsilon}_t^c] \right| \\
&= \max_{1 \leq i \leq U_T} \left| \frac{1}{Th} \sum_{t=1}^T \tau_t^\theta K \left( \frac{b_i - \tau_t}{h} \right) (\tilde{\varepsilon}_t - E[\tilde{\varepsilon}_t]) \right| + \max_{1 \leq i \leq U_T} \left| \frac{1}{Th} \sum_{t=1}^T \tau_t^\theta K \left( \frac{b_i - \tau_t}{h} \right) E[\tilde{\varepsilon}_t^c] \right| \\
&:= \Pi_{2T,11} + \Pi_{2T,12}.
\end{aligned}$$

By the proof given for  $\Pi_{2T,2}$ , we know that  $\Pi_{2T,12} = o\left(l(T) \sqrt{\frac{\ln T}{Th}}\right)$ . Thus, we focus on  $\Pi_{2T,11}$ . Observe that

$$\left| \frac{1}{Th} \cdot \tau_t^\theta K \left( \frac{b_i - \tau_t}{h} \right) (\tilde{\varepsilon}_t - E[\tilde{\varepsilon}_t]) \right| \leq O(1) T^{1/v-1} l(T) h^{-1} = O(1) \xi,$$

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

Then, for any  $\epsilon > 0$ , letting  $l(\cdot)$  and  $v$  satisfy  $l(T) \rightarrow \infty$  and  $\frac{T^{1-2/v} h}{[l(T)]^4 \ln T} \rightarrow \infty$  and applying Lemma B.1 with

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

we have

$$\begin{aligned}
\Pr(\Pi_{2T,11} > T \epsilon_1) &= \Pr \left( \Pi_{2T,1} > \epsilon l(T) \sqrt{\frac{\ln T}{Th}} \right) \\
&\leq O(1) \kappa_T^{-1} \exp \left( -\frac{\epsilon^2 [l(T)]^2 q \frac{\ln T}{T^3 h}}{\frac{O(1)}{T^2 h p}} \right) + O(1) \kappa_T^{-1} \left( 1 + \frac{4\xi}{\epsilon_1} \right)^{1/2} q \alpha(\lfloor T/(2q) \rfloor) \\
&\leq O(1) \kappa_T^{-1} \exp(-O(1) \epsilon^2 [l(T)]^2 \ln T) + O(1) \kappa_T^{-1} \left( 1 + \frac{4\xi}{\epsilon_1} \right)^{1/2} q \alpha(\lfloor T/(2q) \rfloor).
\end{aligned}$$

By the same arguments under (B.16) of Chen et al. (2012), we obtain  $\Pi_{2T,11} = o_P\left(l(T) \sqrt{\frac{\ln T}{Th}}\right)$ .

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

(2). As in the first result of this lemma, it suffices to show that

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

where  $l(T)$  is an arbitrary positive function satisfying that  $l(T) \rightarrow \infty$  as  $T \rightarrow \infty$ . Below, we use Lemma A2 of Newey and Powell (2003) to prove this result.

*Step 1:*  $\Theta \times [0, 1]$  is a compact subspace of  $\mathbb{R}^2$  with the 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 result (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.

By *Step 1-Step 3*, the second result of this lemma holds.

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

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

$$\begin{cases} \text{Case 1: } (\theta, u) \in \Theta \times [h, 1-h]; \\ \text{Case 2: } (\theta, u) \in \Theta \times (1-h, 1]. \end{cases}$$

For *Case 1*, write

$$\begin{aligned} & \sup_{(\theta, u) \in \Theta \times [h, 1-h]} \left| \frac{1}{Th} \sum_{t=1}^T \tau_t^{\theta+\theta_0} g(\tau_t) K\left(\frac{\tau_t - u}{h}\right) - u^{\theta+\theta_0} g(u) \right| \\ &= \sup_{(\theta, u) \in \Theta \times [h, 1-h]} \left| \frac{1}{h} \int_0^1 w^{\theta+\theta_0} g(w) K\left(\frac{w-u}{h}\right) dw + O\left(\frac{1}{Th}\right) - u^{\theta+\theta_0} g(u) \right| \\ &= \sup_{(\theta, u) \in \Theta \times [h, 1-h]} \left| \int_{-u/h}^{(1-u)/h} m_1(u+wh) K(w) dw + O\left(\frac{1}{Th}\right) - u^{\theta+\theta_0} g(u) \right| \\ &= \sup_{(\theta, u) \in \Theta \times [h, 1-h]} \left| \int_{-1}^1 \left( m_1(u) + m_1^{(1)}(\tilde{u})wh \right) K(w) dw + O\left(\frac{1}{Th}\right) - u^{\theta+\theta_0} g(u) \right| \\ &= \sup_{(\theta, u) \in \Theta \times [h, 1-h]} \left| \int_{-1}^1 m_1^{(1)}(\tilde{u})wh K(w) dw + O\left(\frac{1}{Th}\right) \right| \\ &= O(h) + O\left(\frac{1}{Th}\right) = O(h), \end{aligned}$$

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 the Riemann integral; the third equality follows from the Taylor expansion and the fact that  $K(w)$  is defined on  $[-1, 1]$ ; the fifth equality follows from Assumption 1.1; 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  $u = 1 - ch$ . Before proceeding further, note that for  $u^*$  lying between  $u$  and  $u + wh$  with  $w \in [-1, c]$ , we have

$$1 - 2h \leq u - h \leq u^* \leq u + ch = 1. \quad (\text{B.3})$$

Thus, we can write

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

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 the Riemann integral; the second equality follows from the Taylor expansion and the construction of  $u = 1 - ch$ ; the fourth equality follows from (B.3) and Assumption 1.1; 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  $B_{\epsilon_1}(h)$  into the following two subsets:

$$\begin{cases} \text{Case 1: } B_1(h) \equiv [(1 + \epsilon_1)h, 1 - h]; \\ \text{Case 2: } B_2(h) \equiv (1 - h, 1]. \end{cases}$$

Before considering *Case 1*, note that for  $u^*$  lying between  $u$  and  $u + wh$  with  $u \in B_1(h)$  with  $w \in [-1, 1]$ , we have

$$\epsilon_1 h \leq (1 + \epsilon_1)h - h \leq u - h \leq u^* \leq u + h \leq 1. \quad (\text{B.4})$$

Thus,

$$\begin{aligned} \sup_{(\theta, u) \in \Theta \times B_1(h)} |(u^*)^{2\theta - 1} h| &= \begin{cases} \sup_{\theta \in \Theta} (\epsilon_1 h)^{2c^* - 1} h = O(h^{2c^*}), & \text{for } \theta \in \Theta \cap (0, \frac{1}{2}) \\ h, & \text{for } \theta \in \Theta \cap [\frac{1}{2}, \infty) \end{cases} \\ &= O_P(1) h^{\min\{2c^*, 1\}}, \end{aligned} \quad (\text{B.5})$$

where  $c^* = \min_{\theta \in \Theta} \theta$  and  $c^* > 0$ . Then we are able to write

$$\begin{aligned} & \sup_{(\theta, u) \in \Theta \times B_1(h)} \left| \frac{1}{Th} \sum_{t=1}^T \tau_t^{2\theta} K\left(\frac{\tau_t - u}{h}\right) - u^{2\theta} \right| \\ &= \sup_{(\theta, u) \in \Theta \times B_1(h)} \left| \frac{1}{h} \int_0^1 w^{2\theta} K\left(\frac{w - u}{h}\right) dw + O\left(\frac{1}{Th}\right) - u^{2\theta} \right| \end{aligned}$$

$$\begin{aligned}
&= \sup_{(\theta, u) \in \Theta \times B_1(h)} \left| \int_{-u/h}^{(1-u)/h} (u + wh)^{2\theta} K(w) dw + O\left(\frac{1}{Th}\right) - u^{2\theta} \right| \\
&= \sup_{(\theta, u) \in \Theta \times B_1(h)} \left| \int_{-1}^1 (u^{2\theta} + 2\theta \tilde{u}^{2\theta-1} wh) K(w) dw + O\left(\frac{1}{Th}\right) - u^{2\theta} \right| \\
&= \sup_{(\theta, u) \in \Theta \times B_1(h)} \left| \int_{-1}^1 2\theta \tilde{u}^{2\theta-1} wh K(w) dw + O\left(\frac{1}{Th}\right) \right| \\
&= O(h^{\min\{2c^*, 1\}}) + O\left(\frac{1}{Th}\right) = O(h^{\min\{2c^*, 1\}}),
\end{aligned}$$

where  $\tilde{u}$  lies between  $u$  and  $u + wh$ ; the first equality follows from the definition of the Riemann integral; the third equality follows from the Mean Value Theorem and the fact that  $K(w)$  is defined on  $[-1, 1]$ ; and the fifth equality follows from (B.5).

Again,  $(\theta, u) \in \Theta \times B_2(h)$  is equivalent to  $(\theta, c) \in \Theta \times [0, 1]$  with  $u = 1 - ch$ . For *Case 2*, write

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

where  $\tilde{u}$  lies between  $u$  and  $u + wh$ ; the first equality follows from the definition of the Riemann integral; the second equality follows from Taylor expansion and the construction of  $u = 1 - ch$ ; the fourth equality follows from (B.3); and the fifth equality follows from Assumption 1.4.

Therefore, the result follows.

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

*Step 2:* Note that it is easy to know  $\int_0^1 (\ln u)^4 du < \infty$  using integration by parts. We now verify the continuity of  $v(\theta)$ .

$$\begin{aligned}
|v(\theta_1) - v(\theta_2)| &= \left| \int_0^1 (u^{\theta_0 + \theta_1} - u^{\theta_0 + \theta_2}) g(u) du \right| = \left| (\theta_1 - \theta_2) \cdot \int_0^1 u^{\theta^*} g(u) (\ln u) du \right| \\
&\leq |\theta_1 - \theta_2| \left\{ \int_0^1 u^{2\theta^*} du \cdot \int_0^1 g^2(u) (\ln u)^2 du \right\}^{1/2} \\
&= |\theta_1 - \theta_2| \left\{ \frac{1}{2\theta^* + 1} u^{2\theta^* + 1} \Big|_0^1 \right\}^{1/2} \left\{ \int_0^1 g^2(u) (\ln u)^2 du \right\}^{1/2} \\
&= |\theta_1 - \theta_2| \left\{ \frac{1}{2\theta^* + 1} u^{2\theta^* + 1} \Big|_0^1 \right\}^{1/2} \left\{ \int_0^1 g^4(u) du \cdot \int_0^1 (\ln u)^4 du \right\}^{1/4} \\
&= O(|\theta_1 - \theta_2|), \tag{B.6}
\end{aligned}$$

where  $\theta^*$  lies between  $\theta_0 + \theta_1$  and  $\theta_0 + \theta_2$ ; the second equality follows from the Mean Value Theorem;

the first inequality follows from the Cauchy Schwarz inequality; the fifth equality follows from Assumption 1.1 and the fact that we point out in the beginning of this step. In connection with *Step 1*, we obtain  $|v_T(\theta_1) - v_T(\theta_2)| \leq O(1)|\theta_1 - \theta_2|$ .

Recall that  $U(\theta_0)$  is a compact subspace of  $\mathbb{R}$  with the Euclidean norm. By *Step 1-Step 2* and a proof similar to Lemma A2 of Newey and Powell (2003), the result follows immediately.  $\blacksquare$

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

$$\left|v_T(\widehat{\theta}) - v(\theta_0)\right| \leq \left|v_T(\widehat{\theta}) - v(\widehat{\theta})\right| + \left|v(\widehat{\theta}) - v(\theta_0)\right| = o_P(1), \quad (\text{B.7})$$

where  $\left|v_T(\widehat{\theta}) - v(\widehat{\theta})\right| = o_P(1)$  follows from (6) of Lemma B.2, and  $\left|v(\widehat{\theta}) - v(\theta_0)\right| = o_P(1)$  follows from (B.6). In addition, by Theorem 4.2, we have  $|\widehat{\theta} - \theta| \ln T = O_P(1)$ . Thus, we know the next limit exists:

$$\phi_1 = \text{plim}_{T \rightarrow \infty} T^{\theta_0 - \tilde{\theta}} \cdot \frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{\theta_0 + \tilde{\theta}} g(\tau_t) = \tilde{\alpha}_0 \int_0^1 u^{2\theta_0} g(u) du, \quad (\text{B.8})$$

where  $\tilde{\theta}$  is defined in (4.9), and  $\tilde{\alpha}_0 = \text{plim}_{T \rightarrow \infty} T^{\theta_0 - \tilde{\theta}}$ .

Similarly, the next two limits exist:

$$\phi_2 = \text{plim}_{T \rightarrow \infty} T^{\theta_0 - \tilde{\theta}} \cdot \frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{\theta_0 + \tilde{\theta}} g(\tau_t) \ln \tau_t = \tilde{\alpha}_0 \int_0^1 u^{2\theta_0} g(u) (\ln u) du, \quad (\text{B.9})$$

$$\phi_3 = \text{plim}_{T \rightarrow \infty} T^{\theta_0 - \tilde{\theta}} \cdot \frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{\theta_0 + \tilde{\theta}} g(\tau_t) (\ln \tau_t)^2 = \tilde{\alpha}_0 \int_0^1 u^{2\theta_0} g(u) (\ln u)^2 du. \quad (\text{B.10})$$

With (B.8) to (B.10) in hand, we are now ready to prove the next lemma.

### Proof of Lemma B.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 this supplementary file. Write

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

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:

$$\begin{aligned}
A_1 &= \frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{2\tilde{\theta}} T^{-6\tilde{\theta}-3} \left[ \frac{1}{T} \sum_{s=1}^T \tau_s^{2\tilde{\theta}} K_h(\tau_t - \tau_s) \right]^{-3} \\
&\quad \cdot T^{5\tilde{\theta} + \theta_0 + 3} \left[ \frac{1}{T^3} \sum_{u=1}^T \sum_{s=1}^T \sum_{r=1}^T (\tau_u \tau_s)^{2\tilde{\theta}} \tau_r^{\tilde{\theta} + \theta_0} g(\tau_r) K_h(\tau_t - \tau_u) K_h(\tau_t - \tau_s) K_h(\tau_t - \tau_r) (\ln u) (\ln s) \right] \\
&\quad + \frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{2\tilde{\theta}} T^{-6\tilde{\theta}-3} \left[ \frac{1}{T} \sum_{s=1}^T \tau_s^{2\tilde{\theta}} K_h(\tau_t - \tau_s) \right]^{-3} \\
&\quad \cdot T^{5\tilde{\theta} + 3} \left[ \frac{1}{T^3} \sum_{u=1}^T \sum_{s=1}^T \sum_{r=1}^T (\tau_u \tau_s)^{2\tilde{\theta}} \tau_r^{\tilde{\theta}} \varepsilon_r K_h(\tau_t - \tau_u) K_h(\tau_t - \tau_s) K_h(\tau_t - \tau_r) (\ln u) (\ln s) \right] \\
&:= A_{11} + A_{12},
\end{aligned}$$

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\tilde{\theta}} T^{-6\tilde{\theta}-3} \left[ \frac{1}{T} \sum_{s=1}^T \tau_s^{2\tilde{\theta}} K_h(\tau_t - \tau_s) \right]^{-3} \\
&\quad \cdot T^{5\tilde{\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\tilde{\theta}} \tau_r^{\tilde{\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 - \tilde{\theta}} (1 + o_P(1)) \cdot \frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{2\tilde{\theta}} \tau_t^{-6\tilde{\theta}} \\
&\quad \cdot \left[ \frac{1}{T} \sum_{u=1}^T \tau_u^{2\tilde{\theta}} (\ln u) K_h(\tau_t - \tau_u) \right]^2 \left[ \frac{1}{T} \sum_{u=1}^T \tau_u^{\tilde{\theta} + \theta_0} g(\tau_u) K_h(\tau_t - \tau_u) \right] \\
&= T^{\theta_0 - \tilde{\theta}} (1 + o_P(1)) \cdot \frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{\theta_0 - 3\tilde{\theta}} g(\tau_t) \left[ \frac{1}{T} \sum_{u=1}^T \tau_u^{2\tilde{\theta}} (\ln \tau_u + \ln T) K_h(\tau_t - \tau_u) \right]^2 \\
&= T^{\theta_0 - \tilde{\theta}} (\ln T)^2 (1 + o_P(1)) \frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{\theta_0 - 3\tilde{\theta}} g(\tau_t) \left[ \frac{1}{T} \sum_{u=1}^T \tau_u^{2\tilde{\theta}} K_h(\tau_t - \tau_u) \right]^2 \\
&\quad + 2T^{\theta_0 - \tilde{\theta}} (\ln T) (1 + o_P(1)) \frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{\theta_0 - 3\tilde{\theta}} g(\tau_t) \\
&\quad \cdot \left[ \frac{1}{T} \sum_{u=1}^T \tau_u^{2\tilde{\theta}} (\ln \tau_u) K_h(\tau_t - \tau_u) \right] \left[ \frac{1}{T} \sum_{u=1}^T \tau_u^{2\tilde{\theta}} K_h(\tau_t - \tau_u) \right] \\
&\quad + T^{\theta_0 - \tilde{\theta}} (1 + o_P(1)) \frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{\theta_0 - 3\tilde{\theta}} g(\tau_t) \left[ \frac{1}{T} \sum_{u=1}^T \tau_u^{2\tilde{\theta}} (\ln \tau_u) K_h(\tau_t - \tau_u) \right]^2 \\
&= T^{\theta_0 - \tilde{\theta}} (\ln T)^2 (1 + o_P(1)) \frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T \tau_t^{\theta_0 + \tilde{\theta}} g(\tau_t)
\end{aligned}$$



$$\begin{aligned}
& +2T^{\theta_0-\tilde{\theta}}(\ln T)(1+o_P(1))\frac{1}{T}\sum_{t=\lfloor Th \rfloor+1}^T \tau_t^{\theta_0+\tilde{\theta}}g(\tau_t)\ln \tau_t \\
& +T^{\theta_0-\tilde{\theta}}(1+o_P(1))\frac{1}{T}\sum_{t=\lfloor Th \rfloor+1}^T \tau_t^{\theta_0+\tilde{\theta}}g(\tau_t)(\ln \tau_t)^2 \\
& = (\ln T)^2\phi_1 + 2(\ln T)\phi_2 + \phi_3 + o_P(1),
\end{aligned} \tag{B.11}$$

where the second, third and fifth equalities follow from (4) and (5) of Lemma B.2; and the last equality follows from (B.8) to (B.10) and the definition of the Riemann integral.

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

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

where the second equality follows from (5) of Lemma B.2; and the third equality follows from (2) of Lemma B.2 and Theorem 4.2.

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

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

Based on the above development, simple calculation yields the first result of this lemma.

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

$$\frac{1}{T}\sum_{t=\lfloor Th \rfloor+1}^T \tau_t^{2\theta} \frac{\partial \hat{g}(\tau_t, \theta)}{\partial \theta} \Big|_{\theta=\tilde{\theta}}$$

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

where the definitions of  $A_1$  and  $A_2$  should be obvious.

For  $A_1$ , write

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

where the first equality follows from (5) of Lemma B.2; and the third equality follows the development similar to (B.11). Similarly, we can show that  $A_2 = (\ln T) \phi_1 + \phi_2 + o_P(1)$ . Based on the above development, simple calculation yields the second result of this lemma.

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

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

where the definitions of  $A_1$  and  $A_2$  should be obvious.

For  $A_1$ , write

$$\begin{aligned}
A_1 &= (1 + o_P(1))T^{-4\tilde{\theta}-2} \frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T (\ln \tau_t) \frac{\sum_{u=1}^T \sum_{s=1}^T (u\sqrt{s})^{2\tilde{\theta}} g(\tau_s) s^{\theta_0} K_h(\tau_t - \tau_u) K_h(\tau_t - \tau_s) \ln u}{\tau_t^{2\tilde{\theta}}} \\
&\quad + (1 + o_P(1))T^{-4\tilde{\theta}-2} \frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^T (\ln \tau_t) \frac{\sum_{u=1}^T \sum_{s=1}^T (u\sqrt{s})^{2\tilde{\theta}} \varepsilon_s K_h(\tau_t - \tau_u) K_h(\tau_t - \tau_s) \ln u}{\tau_t^{2\tilde{\theta}}} \\
&= (\ln T)\phi_2 + \phi_3 + o_P(1),
\end{aligned}$$

where the first equality follows from (5) of Lemma B.2; and the second equality follows the development similar to (B.11). Similarly, we can show that  $A_2 = (\ln T)\phi_2 + \phi_3 + o_P(1)$ . Based on the above development, simple calculation yields the third result of this lemma.

(4)-(6). Similar to the proofs given for (2)-(3) of this lemma, (4)-(6) of this lemma follow.

(7). By (1)-(6) of this lemma, simple calculation immediately gives  $\left. \frac{\partial^2 R_T(\theta)}{\partial \theta^2} \right|_{\theta=\tilde{\theta}} = 8 + o_P(1)$ . The proof is now complete.  $\blacksquare$

## B.2.2 Proofs of Section 4

### Proof of Lemma 4.1:

(1). For notational simplicity, let  $\mathbb{B} := B_T(\theta_0) \times B_{\varepsilon_1}(h)$ . Write

$$\begin{aligned}
&\sup_{(\theta, u) \in \mathbb{B}} |\hat{g}(u, \theta) - (uT)^{\theta_0 - \theta} g(u)| \\
&\leq \sup_{(\theta, u) \in \mathbb{B}} \frac{1}{T^\theta} \left| \left( \frac{1}{T} \sum_{t=1}^T \tau_t^{2\theta} K_h(u - \tau_t) \right)^{-1} \frac{1}{T} \sum_{t=1}^T \tau_t^\theta \varepsilon_t K_h(u - \tau_t) \right| \\
&\quad + \sup_{(\theta, u) \in \mathbb{B}} T^{\theta_0 - \theta} \left| \left( \frac{1}{T} \sum_{t=1}^T \tau_t^{2\theta} K_h(u - \tau_t) \right)^{-1} \frac{1}{T} \sum_{t=1}^T \tau_t^{\theta + \theta_0} g(\tau_t) K_h(u - \tau_t) - (uT)^{\theta_0 - \theta} g(u) \right| \\
&:= A_1 + A_2,
\end{aligned}$$

where the definitions of  $A_1$  and  $A_2$  should be obvious.

Firstly, note that two simple facts are

$$\sup_{\theta \in B_T(\theta_0)} \left( \frac{1}{h} \right)^{\theta - \theta_0} \leq \sup_{\theta \in B_T(\theta_0)} T^{|\theta - \theta_0|} = O(1) \quad \text{and} \quad h^{\frac{1}{\ln T}} = O(T^{-\nu})^{\frac{1}{\ln T}} = O(1). \quad (\text{B.12})$$

We then consider  $A_1$  and  $A_2$  respectively. Start from  $A_1$ .

$$\begin{aligned}
A_1 &= O_P \left( \sqrt{\frac{\ln T}{Th}} \right) \sup_{(\theta, u) \in \mathbb{B}} T^{-\theta} u^{-2\theta} \leq O_P \left( \sqrt{\frac{\ln T}{Th}} \right) \left\{ \sup_{\theta \in B_T(\theta_0)} h^{-2\theta} \right\} \left\{ \sup_{\theta \in B_T(\theta_0)} T^{-\theta} \right\} \\
&= O_P \left( \sqrt{\frac{\ln T}{Th}} \right) T^{-\theta_0} h^{-2\theta_0} \left\{ \sup_{\theta \in B_T(\theta_0)} h^{2\theta_0 - 2\theta} \right\} \left\{ \sup_{\theta \in B_T(\theta_0)} T^{\theta_0 - \theta} \right\} = O \left( \frac{\sqrt{\ln T}}{T^{\frac{1}{2} + \theta_0} h^{\frac{1}{2} + 2\theta_0}} \right), \quad (\text{B.13})
\end{aligned}$$

where the first equality follows from (2) and (5) of Lemma B.2; and the third equality follows from (B.12).

For  $A_2$ , write

$$A_2 = \sup_{(\theta, u) \in \mathbb{B}} T^{\theta_0 - \theta} \left| u^{-2\theta} (1 + O_P(h^{\min\{2b_0, 1\}})) \cdot u^{\theta_0 + \theta} g(u) (1 + O_P(h)) - (uT)^{\theta_0 - \theta} g(u) \right|$$

$$\begin{aligned}
&= O_P(1) \sup_{\theta \in B_T(\theta_0)} T^{\theta_0 - \theta} h^{\min\{2b_0, 1\}} = O_P(1) \sup_{\theta \in B_T(\theta_0)} T^{\theta_0 - \theta} h^{\min\{2(\theta_0 - \frac{M}{\ln T}), 1\}} \\
&= O\left(h^{\min\{2\theta_0, 1\}}\right),
\end{aligned} \tag{B.14}$$

where  $b_0 = \min\{\theta \mid \theta \in B_T(\theta_0)\} = \theta_0 - \frac{M}{\ln T}$ ; the first equality follows from (4) and (5) of Lemma B.2; and the fourth equality follows from (B.12).

Based on the development of  $A_1$  and  $A_2$ , the proof is complete.  $\blacksquare$

### B.2.3 Proofs of Section 3

It is worthy mentioning that the proof of Theorem 3.1 is relatively straightforward after establishing Theorem 4.2 to Theorem 4.4, though Theorem 3.1 is the first asymptotic result of the main text.

#### Proof of Theorem 3.1:

(1). By the development similar to (A.19) of Wang and Xia (2009), it is easy to obtain that under the null

$$\sup_{u \in [0, 1]} |\hat{g}(u) - g(u)| = O_P\left(\frac{\sqrt{\ln T}}{\sqrt{Th}}\right) + O_P(h). \tag{B.15}$$

We then take a further look at (3.3), and write

$$\begin{aligned}
S_T &= -\frac{1}{T/2} \sum_{t \text{ odd}} [-\varepsilon_t + \hat{g}(\tau_t) - g(\tau_t)] \cdot [\hat{g}(\tau_t) - g(\tau_t) + g(\tau_t)] \ln t \\
&= \frac{1}{T/2} \sum_{t \text{ odd}} \varepsilon_t g(\tau_t) \ln t + \frac{1}{T/2} \sum_{t \text{ odd}} \varepsilon_t \cdot [\hat{g}(\tau_t) - g(\tau_t)] \ln t \\
&\quad - \frac{1}{T/2} \sum_{t \text{ odd}} [\hat{g}(\tau_t) - g(\tau_t)] g(\tau_t) \ln t - \frac{1}{T/2} \sum_{t \text{ odd}} [\hat{g}(\tau_t) - g(\tau_t)]^2 \ln t \\
&:= S_{T,1} + S_{T,2} - S_{T,3} - S_{T,4},
\end{aligned} \tag{B.16}$$

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 focus on  $S_{T,1} - S_{T,3}$  below:

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

$$= \frac{1}{T/2} \sum_{t \text{ odd}} \varepsilon_t g(\tau_t) \ln t - \frac{1 + o_P(1)}{T/2} \sum_{t \text{ even}} \varepsilon_t g(\tau_t) \ln t + o_P(1), \quad (\text{B.17})$$

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 4.4 of the main text.

Based on (B.17), if Assumption 1.2\*.1 holds, we immediately obtain that  $\widehat{LM} \rightarrow_D N(0, 1)$ .

Based on (B.17), if Assumption 1.2\*.2 holds, we obtain that

$$\widehat{LM} \rightarrow_D N(0, 1 + \sigma_1^2)$$

by using, for example, Theorem 2.21 of Fan and Yao (2003), where  $\sigma_1^2 = \lim_{T \rightarrow \infty} \frac{2}{\sigma_\varepsilon^2} \sum_{t=2}^T \sum_{s=1}^{t-1} \gamma(t-s) \omega_{Tt} \omega_{Ts}$ , and  $\omega_{Tt}$  has been defined in Assumption 1.2\*.2. Invoking the condition that  $\sum_{t=2}^T \sum_{s=1}^{t-1} \gamma(t-s) \omega_{Tt} \omega_{Ts} = o(1)$  gives  $\sigma_1^2 = 0$ . Thus,  $\widehat{LM} \rightarrow_D N(0, 1)$ .

The proof is now complete.

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

$$\begin{aligned} |\widehat{g}(u)| &= \left| \frac{\sum_{t=1}^T K_h(u - \tau_t) y_t}{\sum_{t=1}^T K_h(u - \tau_t)} \right| = \left| \frac{\sum_{t=1}^T K_h(u - \tau_t) g(\tau_t) t^{\theta_0}}{\sum_{t=1}^T K_h(u - \tau_t)} + \frac{\sum_{t=1}^T K_h(u - \tau_t) \varepsilon_t}{\sum_{t=1}^T K_h(u - \tau_t)} \right| \\ &= \left| T^{\theta_0} \cdot \frac{\sum_{t=1}^T K_h(u - \tau_t) g(\tau_t) \tau_t^{\theta_0}}{\sum_{t=1}^T K_h(u - \tau_t)} \right| + o_P(1) = T^{\theta_0} \cdot (u^{\theta_0} |g(u)| + o_P(1)) + o_P(1) \\ &\rightarrow_P \infty. \end{aligned} \quad (\text{B.18})$$

In connection with (B.16), it is easy to see that  $S_{T,4}$  is the true leading term due to the involvement of a quadratic term. Then by definition,  $\widehat{LM} \rightarrow \infty$  under the alternative hypothesis, as  $T \rightarrow \infty$ . ■

### Proof of Theorem 3.2:

By Theorems 1 and 2 of Robinson (2012), it is easy to show that  $\widehat{\theta} - \theta_0 = O_P(T^{\chi - \theta_0 - \frac{1}{2}})$  and  $\widehat{\beta} - \beta_0 = O_P((\ln T) T^{\chi - \theta_0 - \frac{1}{2}})$  for any given sufficiently small  $\chi > 0$ . Then the proof of Theorem 3.2 follows from the development of Gao and Hawthorne (2006), thus omitted. ■

## B.2.4 Proofs of Section 6

### Proof of Corollary 6.1:

The proofs are a simplified version of the development of Lemma B.2 and Lemma 4.1, so omitted. ■

### Proof of Corollary 6.2:

(1). By the proof of Lemma 4.1, a faster rate of convergence for  $\widehat{g}(u, a)$  under the null can be achieved as follows:

$$\sup_{u \in [c, 1-h]} |\widehat{g}(u, a) - g(u)| = O_P \left( \frac{\sqrt{\ln T}}{T^{\frac{1}{2} + a} h^{\frac{1}{2}}} \right) + O(h^2). \quad (\text{B.19})$$

Then, for  $S_T$  defined in (6.4), write

$$\begin{aligned}
S_T &= -\frac{1}{T^*/2} \sum_{t \text{ odd} \in B_h} [-\varepsilon_t + \widehat{g}(\tau_t)t^a - g(\tau_t)t^a] \cdot [\widehat{g}(\tau_t) - g(\tau_t) + g(\tau_t)] t^a \ln t \\
&= \frac{2}{T^*} \sum_{t \text{ odd} \in B_h} \varepsilon_t g(\tau_t) t^a \ln t + \frac{2}{T^*} \sum_{t \text{ odd} \in B_h} \varepsilon_t \cdot [\widehat{g}(\tau_t) - g(\tau_t)] t^a \ln t \\
&\quad - \frac{2}{T^*} \sum_{t \text{ odd} \in B_h} [\widehat{g}(\tau_t) - g(\tau_t)] g(\tau_t) t^{2a} \ln t - \frac{2}{T^*} \sum_{t \text{ odd} \in B_h} [\widehat{g}(\tau_t) - g(\tau_t)]^2 t^{2a} \ln t \\
&:= S_{T,1} + S_{T,2} - S_{T,3} - S_{T,4}.
\end{aligned} \tag{B.20}$$

Similar to the proof of Theorem 3.1, 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

$$S_{T,1} - S_{T,3} = \left( \frac{2}{T^*} \sum_{t \text{ odd} \in B_h} \varepsilon_t g(\tau_t) t^a \ln t - \frac{2 + o_P(1)}{T^*} \sum_{t \text{ even} \in B_h} \varepsilon_t g(\tau_t) t^a \ln t \right) \cdot (1 + o_P(1)) \tag{B.21}$$

provided  $h^2 T^{2a} \ln T \rightarrow 0$ . Thus, the first result follows immediately.

(2). The proof of the second result follows from a procedure identical to (B.18), thus omitted.  $\blacksquare$

We now provide Assumption 2 before going through the detailed proofs of Corollary 6.3.

**Assumption 2:**

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

1.  $\{x_t \mid t = 1, \dots, T\}$  is a strictly stationary and  $\alpha$ -mixing error process with a density  $p(w)$ . Moreover,  $\sup_{(w,u) \in \mathbb{R}^d \times [0,1]} p(w) \frac{\partial f(w,u)}{\partial u} < \infty$  and  $E[\sup_{u \in [0,1]} |f(x_1, u)|] < \infty$ ; or
2.  $\{x_t \mid t = 1, \dots, T\}$  is a locally stationary process.<sup>9</sup> Let  $f(\cdot, \cdot)$  be uniformly bounded and satisfy that  $|f(x_1, u) - f(x_2, u)| \leq M \|x_1 - x_2\|$  for  $\forall u \in [0, 1]$ , where  $M$  is a positive constant; or
3. (a) Let  $f(\cdot, \cdot)$  be uniformly bounded, and  $x_t = x_{t-1} + w_t$  for  $t \geq 1$  and  $\|x_0\| = O_P(1)$ ;  
(b) Let  $w_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ , where  $\sum_{j=0}^{\infty} j \|\psi_j\| < \infty$  and  $\psi := \sum_{j=0}^{\infty} \psi_j \neq 0$ ;  
(c) Let  $\{\varepsilon_j \mid -\infty < j < \infty\}$  be a sequence of i.i.d. random variables having an absolutely continuous distribution with respect to the Lebesgue measure and satisfying  $E[\varepsilon_1] = 0_{d \times 1}$ ,  $E[\varepsilon_1 \varepsilon_1'] = I_d$ ,  $E\|\varepsilon_1\|^q < \infty$  for some  $q > 4$ . The characteristic function of  $\varepsilon_1$  is integrable.

**Proof of Corollary 6.3:**

---

<sup>9</sup>We adopt the following definition for a locally stationary process (cf., Vogt, 2012; Dong and Linton, 2018):

**Definition B.4.** The process  $\{x_t \mid t = 1, \dots, T\}$  is locally stationary if for each rescaled time point  $u \in [0, 1]$  there exists an associated process  $\{x_t(u) \mid t = 1, \dots, T\}$  with the following two properties:

- (a)  $\{x_t(u) \mid t = 1, \dots, T\}$  is strictly stationary with density  $f_u(w)$ ;
- (b) It holds that  $\|x_t - x_t(u)\|_r \leq (|\tau_t - u| + T^{-1}) U_t(u)$  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  $\rho > 0$  and  $C < \infty$  independent of  $u, t$ , and  $T$ . Moreover,  $\|\cdot\|_r$  denotes an arbitrary norm on  $\mathbb{R}^d$ .

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

$$\sup_{u \in [0, 1]} \int_{-u/h}^{(1-u)/h} K(w)dw = O(1). \quad (\text{B.22})$$

Before proceeding further, we show  $\sup_{(\theta, u) \in \Theta \times [0, 1]} \left| \frac{1}{T} \sum_{t=1}^T \tau_t^\theta f(x_t, \tau_t) K_h(u - \tau_t) \right| = O_P(1)$  under all three conditions of Assumption 2.

*Case 1:* Under Assumption 2.1, we have

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

where the second inequality follows from the fact that  $0 \leq \tau^\theta \leq 1$  uniformly; the first equality follows from the definition of the Riemann integral; the third and fourth equalities follows from Assumption 2.1.; the third inequality follows from (B.22).

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 2.1.

*Case 2:* Let Assumption 2.2 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

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

where the definitions of  $A_1$  and  $A_2$  should be obvious.

For  $A_1$ , we have

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

where the first inequality follows from Assumption 2.2; the second inequality follows from the definition of a 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

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

where the first inequality follows from Assumption 2.2; and the second equality follows from the definition of the Riemann integral; and the fourth equality follows from (B.22).

Thus, we can conclude that  $\sup_{(\theta, u) \in \Theta \times [0, 1]} \left| \frac{1}{T} \sum_{t=1}^T \tau_t^\theta f(x_t, \tau_t) K_h(u - \tau_t) \right| = O_P(1)$ .

*Case 3:* Let Assumption 2.3 hold. Construct a  $\nu_T$  satisfying that  $\nu_T \rightarrow \infty$  and  $\nu_T / (Th) \rightarrow 0$ . 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  $w$ .

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



where the second inequality follows from Assumption 2.3; and the last equality follows from (B.22) and the fact that  $\phi_t(w)$  is a density function.

Thus, we have  $\sup_{(\theta, u) \in \Theta \times [0, 1]} \left| \frac{1}{T} \sum_{t=1}^T \tau_t^\theta f(x_t, \tau_t) K_h(u - \tau_t) \right| = O_P(1)$  under all three conditions of Assumption 2. Then both results of this corollary can be verified by exactly the same procedure as documented in Appendix A of this paper.  $\blacksquare$

### B.3 Potential Issues

In this subsection we consider two potential issues.

#### B.3.1 Issue 1

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

$$y_t = \sum_{j=1}^d g_j(\tau_t) t^{\theta_{0,j}} + \varepsilon_t, \quad (\text{B.23})$$

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

However, using nonparametric methods to estimate model (B.23) suffers from certain identification issues. We consider the kernel method here, and discuss the sieve method in Section B.3.2 below. To make the explanation clearer and simpler, suppose  $\theta_0$  is known. For  $\forall u \in (0, 1)$ , the kernel based OLS estimator of  $G(u) = (g_1(u), \dots, g_d(u))'$  is

$$\hat{G}(u) = \left( \sum_{t=1}^T z_t z_t' K_h(u - \tau_t) \right)^{-1} \sum_{t=1}^T z_t y_t K_h(u - \tau_t), \quad (\text{B.24})$$

where  $z_t = (t^{\theta_{0,1}}, \dots, t^{\theta_{0,d}})'$ . Normalize the matrix in the inverse of (B.24) as follows:

$$D_{\theta_0}^{-1} \sum_{t=1}^T z_t z_t' K_h(u - \tau_t) D_{\theta_0}^{-1}, \quad (\text{B.25})$$

where  $D_{\theta_0} = \text{diag}\{T^{1/2+\theta_{0,1}}, \dots, T^{1/2+\theta_{0,d}}\}$ . The  $(i, j)^{th}$  element of (B.25) with  $1 \leq i, j \leq d$  can be easily calculated:

$$\frac{1}{Th} \sum_{t=1}^T \tau_t^{\theta_{0,i} + \theta_{0,j}} K\left(\frac{u - \tau_t}{h}\right) = u^{\theta_{0,i} + \theta_{0,j}} (1 + o(1)), \quad (\text{B.26})$$

which suggests that (B.25) can be rewritten as

$$D_{\theta_0}^{-1} \sum_{t=1}^T z_t z_t' K_h(u - \tau_t) D_{\theta_0}^{-1} = (u^{\theta_{0,1}}, \dots, u^{\theta_{0,d}})' (u^{\theta_{0,1}}, \dots, u^{\theta_{0,d}}) (1 + o(1)). \quad (\text{B.27})$$

However, the right hand side of (B.27) is obviously not invertible, i.e., (B.24) is not well defined.

The key difference between parametric and nonparametric models lies in the use of the kernel function. For parametric cases, the kernel function is not present in (B.24), so it yields

$$\frac{1}{T} \sum_{t=1}^T \tau_t^{\theta_{0,i} + \theta_{0,j}} = \int_0^1 u^{\theta_{0,i} + \theta_{0,j}} du \cdot (1 + o(1)) = \frac{1}{\theta_{0,i} + \theta_{0,j} + 1} \cdot (1 + o(1)). \quad (\text{B.28})$$

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. One referee suggested that the matrix rotation technique employed by Phillips et al. (2017) may be helpful to solve this problem. We thank the referee for the suggestion, and now point out the key difference between their model and (B.23). While Phillips et al. (2017) rotate their matrix  $\sum_{t=1}^T x_t x_t' K_h(u - \tau_t)$ , there are no parameters  $\theta_{0,j}$ 's existing as the unknown power terms. If  $\theta_{0,j}$ 's were known, we can implement the rotation to solve the singularity problem. However, as  $\theta_{0,j}$ 's are parameters of interest,  $\theta_{0,j}$ 's existing in the rotation matrix will require more involved matrix operations. It is unclear whether one can estimate all  $\theta_{0,j}$ 's and  $G(u)$  after the rotation. The question raised in this extension is in fact more challenging, although (B.23) looks simple and its majority components are deterministic.

For model (B.23), though it is hard to fully recover all the components, we can at least consistently estimate the power and coefficient function of the leading term (i.e.,  $\theta_d$  and  $g_d(\cdot)$ ). Rewrite (B.23) as  $y_t = g_d(\tau_t) t^{\theta_{0,d}} + e_t$ , where  $e_t = \sum_{j=1}^{d-1} g_j(\tau_t) t^{\theta_{0,j}} + \varepsilon_t$ . We can use (4.6) and (4.1) to consistently estimate  $\theta_{0,d}$  and  $g_d(\cdot)$  respectively. The reason is that while deriving the asymptotics, we need a term  $T^{\theta_{0,d}}$  to normalize  $t^{\theta_{0,d}}$ , and it simultaneously gets  $\sum_{j=1}^{d-1} g_j(\tau_t) t^{\theta_{0,j}}$  smoothed out due to the fact that  $\theta_{0,1} < \dots < \theta_{0,d}$ . This is exactly why we can establish Corollary 6.3. Certainly, the rates of convergence depend on  $\max_{j \in \{1, \dots, d-1\}} \{\theta_{0,d} - \theta_{0,j}\} = \theta_{0,d} - \theta_{0,d-1}$  in this case. One may think that it is then possible to recover  $\theta_{0,j}$  and  $g_j(\cdot)$  recursively. For example, estimate  $\theta_{0,d-1}$  and  $g_{d-1}(\cdot)$  after removing  $t^{\hat{\theta}_d} \hat{g}_d(\tau_t)$  from  $y_t$ , and repeat this process until we estimate all the components of model (B.23). However, by doing so, the biases due to the plug-in procedure will be substantial and stop us further establishing consistent estimators for  $\theta_{0,d-1}$  and  $g_{d-1}(\cdot)$ . How to consistently estimate the other components of model (B.23) is still an open question.

Finally, we would like to point out that rather than estimating  $g_j(\cdot)$ 's and  $\theta_{0,j}$ 's, one may follow Cho and Phillips (2018) and Baek et al. (2015) to establish hypothesis tests. It is worth mentioning that Phillips (2007), Cho and Phillips (2018) and Baek et al. (2015) involve estimating a power of a polynomial term, but an extension involving estimating the unknown powers of multiple polynomial terms may not be an easy job as discussed above.

### B.3.2 Issue 2

We now explain the failure of a sieve based OLS method. Still consider  $y_t = g(\tau_t) t^{\theta_0} + \varepsilon_t$ . Further assume  $\theta_0$  is known. Following Newey (1997), we can expand  $g(\cdot)$  by power series on a 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 (B.28), it is easy to obtain

$$\frac{1}{T} \sum_{t=1}^T (\tau_t^{\theta_0}, \tau_t^{\theta_0+1}, \dots, \tau_t^{\theta_0+k-1}) (\tau_t^{\theta_0}, \tau_t^{\theta_0+1}, \dots, \tau_t^{\theta_0+k-1})'$$

$$= \left\{ \frac{1}{2\theta_0 + i + j + 1} \right\}_{k \times k} \cdot (1 + o(1)) \quad (\text{B.29})$$

for  $0 \leq i, j \leq k - 1$  under proper restrictions on  $k$  and  $T$ . As  $k$  diverges, the right hand side of (B.29) is asymptotically singular, which indicates that the sieve based OLS method does not work for model (1.1) in general. Certainly, the choice of basis functions plays an important role; however, it is not clear to us which series can solve the ill-posed problem at this stage.

## B.4 Extra Numerical Studies

### B.4.1 Simulation Results for Section 4.1

The DGP is identical to Section 5.3, and we take  $\theta_0 = 0.4$  as an example.

In order to examine the failure of the two methods proposed in Section 4.1 and compare with the results in Section 5, we recover  $\theta_0$  by minimizing (4.2) and (4.3) respectively, and then estimate  $g(\tau_t)$  for  $t = \lfloor Th \rfloor + 1, \dots, T$  by (4.1). To put all methods on equal footing, we change (4.2) and (4.3) respectively to

$$Q_T(\theta) = \sum_{t=\lfloor Th \rfloor + 1}^T (y_t - t^\theta \hat{g}(\tau_t, \theta))^2, \quad (\text{B.30})$$

$$Q_T(\beta, \theta | u) = \sum_{t=\lfloor Th \rfloor + 1}^T (y_t - \beta t^\theta)^2 K_h(\tau_t - u). \quad (\text{B.31})$$

For (B.31), we obtain  $\{\hat{\theta}(\tau_t) | t = \lfloor Th \rfloor + 1, \dots, T\}$  as explained in Section 4.1, and further calculate the estimate of  $\theta_0$  by  $\hat{\theta} = \frac{1}{T - \lfloor Th \rfloor} \sum_{t=\lfloor Th \rfloor + 1}^T \hat{\theta}(\tau_t)$ . We refer to these two methods as W1 and W2, and calculate their RMSEs in the same way as explained in the main text. As shown in Table B.7, both W1 and W2 perform rather poorly, which supports our argument in Section 4.1.

Table B.7: Simulation Results for Section 4.1

		RMSE $_{\theta}$			RMSE $_g$		
$h \setminus T$		100	200	400	100	200	400
W1	$T^{-1/3}$	0.399	0.399	0.400	6.152	8.644	12.073
	$T^{-1/5}$	0.395	0.397	0.399	5.895	8.341	11.638
	$T^{-1/8}$	0.380	0.387	0.392	5.562	7.824	10.908
W2	$T^{-1/3}$	0.310	0.330	0.343	4.127	6.135	8.733
	$T^{-1/5}$	0.322	0.341	0.361	4.247	6.225	9.354
	$T^{-1/8}$	0.269	0.316	0.338	3.343	5.385	7.937

### B.4.2 Simulation Results for Corollary 6.1

The DGP is  $y_t = g(\tau_t)t^{\theta_0} + \varepsilon_t$ , where  $\theta_0 = -0.35$ ,  $g(\tau) = 3(\tau - 1)^2 + 1$ , and  $\varepsilon_t \sim \text{i.i.d. } N(0, 1)$ . We firstly estimate  $\theta_0$  as explained in the main section, and then estimate  $g(u)$  for  $u = \lfloor Tc_0 \rfloor + 1, \dots, T$ . By Corollary 6.1, the bandwidth selection procedure reduces to the following one.

- **Bandwidth Selection:** Provide an initial bandwidth (say  $h_0 = T^{-1/3}$ ) to start the iteration process. For the  $k^{th}$  ( $k \geq 1$ ) iteration, use  $h_{k-1}$  obtained from the  $(k-1)^{th}$  iteration to calculate  $\hat{\theta}_k$ . Stop iteration, if  $|\hat{\theta}_k - \hat{\theta}_{k-1}| \leq \epsilon$ , where  $\epsilon$  is sufficiently small (e.g.,  $10^{-6}$ ) and serves as a stopping criteria. Otherwise, update the bandwidth by  $h_k = T^{-\frac{1+2\hat{\theta}_k}{3+4\hat{\theta}_k}} \cdot (\ln T)^{\frac{1}{3+4\hat{\theta}_k}}$ . Then proceed to the  $(k+1)^{th}$  iteration.

Without loss of generality, we focus on  $h_{opt}$  only and let  $c_0 = 0.5$ . Since half of the data is thrown away, we choose  $T = 500, 1000$ . As shown in Table B.8, the estimates are fairly accurate, and the RMSEs decrease as  $T$  goes up.

Table B.8: Simulation Results for Corollary 6.1

RMSE $_{\theta}$		RMSE $_g$	
$T = 500$	$T = 1000$	$T = 500$	$T = 1000$
0.107	0.075	0.396	0.365

### B.4.3 Simulation Results for Corollary 6.2

The DGP is  $y_t = \exp(\tau_t)t^{\theta_0} + \varepsilon_t$  and  $\varepsilon_t \sim$  i.i.d.  $N(0,1)$ , and consider  $\theta_0 = 0.2, 0.4, 0.6, 0.8, 1$ . The bandwidth is set to  $h = (\frac{\ln T/2}{T/2})^{7/10}$ , and we let  $c_0 = 0.3$  without loss of generality. As the Epanechnikov kernel having order 2 requires  $h^2 T^{2\theta_0} \ln T \rightarrow 0$ , we would expect that the size of the test will go wrong when  $\theta_0 \geq 0.7$ . For simplicity, we report the size based on 1000 replications in Figure B.7. The power test can be done as in Section 5.1, so we do not pursue it further.

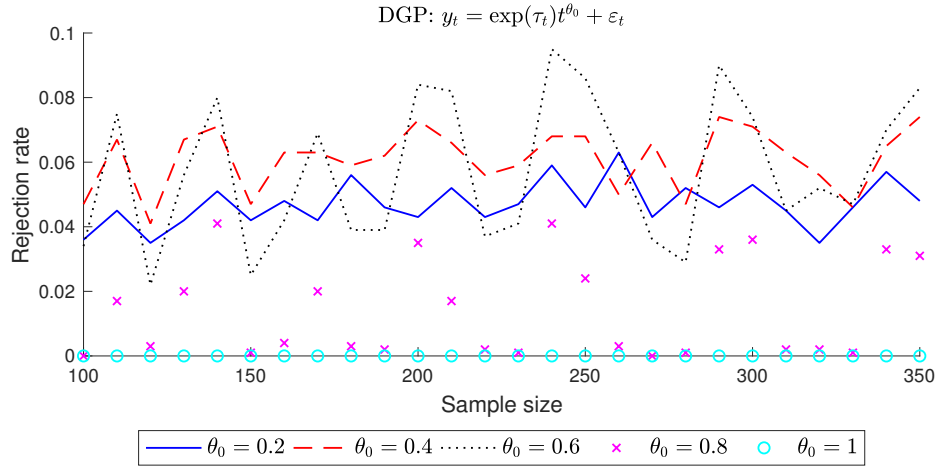


Figure B.7: Size at Nominal Significant Level

As expected, for  $\theta_0 = 0.2, 0.4, 0.6$ , the size is reasonably well controlled. For  $\theta_0 = 0.8, 1$ , the test is clearly undersized. As the value of  $\theta_0$  increases, it can be seen that the consequence of violating  $h^2 T^{2\theta_0} \ln T \rightarrow 0$  becomes more obvious, so it corroborates our arguments on the requirement of  $h^2 T^{2\theta_0} \ln T \rightarrow 0$ .

#### B.4.4 Simulation Results for Corollary 6.3

We now examine Corollary 6.3 and the potential issue discussed in Section B.3.

Specifically, we adopt the following DGPs:

$$\begin{aligned}
 \text{DGP 1: } \quad y_t &= f(x_t, \tau_t) + g(\tau_t)t^{\theta_0} + \varepsilon_t \quad \text{with} \quad g(u) = 3(u-1)^2 + 1, \\
 \text{DGP 2: } \quad y_t &= f(x_t, \tau_t) + g(\tau_t)t^{\theta_0} + \varepsilon_t \quad \text{with} \quad g(u) = 3|u-1|^{0.7} + 1.
 \end{aligned}
 \tag{B.32}$$

The error terms follow  $\varepsilon_t \sim \text{i.i.d. } N(0, 1)$ . Without loss of generality, we set  $d = 1$ , so  $f(\cdot, \cdot)$  and  $\{x_t\}$  are generated as follows:

- *Case 1 (Stationary)*:  $f(x, u) = |x| + 5 \sin(u \cdot \pi)$ , and  $x_t$  follows an AR(1) process  $x_t = 0.5x_{t-1} + v_t$ ;
- *Case 2 (Nonstationary)*:  $f(x, u) = \exp\{-x^2\} + 5 \sin(u \cdot \pi)$ , and  $x_t$  follows an integrated process  $x_t = x_{t-1} + v_t$ .

In both cases,  $x_0 \sim N(0, 1)$  and  $v_t \sim \text{i.i.d. } N(0, 1)$ .

We estimate  $\theta_0$  and  $g(\cdot)$  by our nonparametric method as explained in Section 5 (referred to as NM), and W1 and W2 methods documented above, and report RMSEs in Tables B.9 and B.10 below.

Table B.9: (*DGP1, Case 1*)

		RMSE $_{\theta}$			RMSE $_g$		
	$h \setminus T$	200	500	1000	200	500	1000
NM	$T^{-2/5}$	0.100	0.091	0.085	0.102	0.053	0.033
	$T^{-1/3}$	0.106	0.094	0.086	0.038	0.022	0.016
	$T^{-1/5}$	0.126	0.106	0.095	0.086	0.094	0.092
	$T^{-1/8}$	0.158	0.130	0.114	0.097	0.128	0.139
W1	$T^{-2/5}$	0.300	0.300	0.300	5.451	7.449	9.452
	$T^{-1/3}$	0.300	0.300	0.300	5.338	7.381	9.422
	$T^{-1/5}$	0.300	0.300	0.300	4.948	6.872	8.839
	$T^{-1/8}$	0.300	0.300	0.300	4.802	6.491	8.233
W2	$T^{-2/5}$	0.274	0.268	0.267	4.696	6.057	7.486
	$T^{-1/3}$	0.277	0.280	0.279	4.671	6.492	8.117
	$T^{-1/5}$	0.262	0.279	0.286	3.913	5.940	7.995
	$T^{-1/8}$	0.243	0.272	0.282	3.333	5.349	7.212

Table B.10: (DGP1, Case 2)

		RMSE $_{\theta}$			RMSE $_g$		
	$h \setminus T$	200	500	1000	200	500	1000
NM	$T^{-2/5}$	0.102	0.092	0.085	0.094	0.048	0.028
	$T^{-1/3}$	0.107	0.094	0.086	0.038	0.022	0.017
	$T^{-1/5}$	0.128	0.107	0.095	0.087	0.095	0.093
	$T^{-1/8}$	0.160	0.131	0.114	0.097	0.128	0.139
W1	$T^{-2/5}$	0.300	0.300	0.300	5.357	7.384	9.404
	$T^{-1/3}$	0.300	0.300	0.300	5.252	7.323	9.378
	$T^{-1/5}$	0.300	0.300	0.300	4.871	6.820	8.801
	$T^{-1/8}$	0.300	0.300	0.300	4.729	6.443	8.198
W2	$T^{-2/5}$	0.273	0.267	0.267	4.575	5.926	7.413
	$T^{-1/3}$	0.276	0.280	0.279	4.559	6.437	8.057
	$T^{-1/5}$	0.265	0.278	0.286	3.909	5.860	7.960
	$T^{-1/8}$	0.241	0.272	0.282	3.241	5.284	7.150

As can be seen, the procedure of recovering  $\theta_0$  and  $g(\cdot)$  is not affected by  $f(\cdot, \cdot)$  and  $\{x_t \mid t = 1, \dots, T\}$  too much, which indicates that one can implement our procedure to detrend the data set in a better fashion practically.

#### B.4.5 Simulation Results for Section B.3

Below we focus on DGPs 1 and 2 under Case 1 of Section B.4.4 in order to examine the issue raised in Section B.3. Apart from our proposed method, we also use the sieve based OLS method (referred to as SOLS). In particular, we use power series  $\{1, u, u^2, \dots\}$  to approximate  $g(u)$  in our simulation study (cf., Newey, 1997). Specifically, the new objective function is

$$Q_T(\theta) = \sum_{t=1}^T (y_t - t^\theta \hat{g}_k(\tau_t, \theta))^2, \quad (\text{B.33})$$

where  $\hat{g}_k(\tau_t, \theta) = z_t' \hat{C}(\theta)$ ,  $z_t = (1, \tau_t^1, \dots, \tau_t^{k-1})'$ , and

$$\hat{C}(\theta) = \left( \sum_{t=1}^T [t^\theta z_t] \cdot [t^\theta z_t]' \right)^{-1} \sum_{t=1}^T [t^\theta z_t] y_t.$$

In order to demonstrate our arguments under (B.29), 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.<sup>10</sup> The RMSEs are calculated following the identical procedure of Section 5.3 of the main text.

In Table B.11, it is not surprising to see the best estimate comes from the SOLS method with  $k = 3$ , as this choice of power series perfectly fits the DGP 1. However, when we increase the value of the truncation

<sup>10</sup>The setting of  $h = 1/k$  is indeed reasonable. As for a nonparametric model  $y_t = g(x_t) + e_t$  with  $t = 1, \dots, T$ , 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  $x_t$ .

parameter, the matrix in the inverse is getting closer to singular as explained under (B.29), which is also confirmed by Matlab over the simulation study which warns continuously saying “*Matrix is close to singular or badly scaled*”.

Table B.11: (*DGP 1, Case 1*)

	$h, k \setminus T$	RMSE $_{\theta}$			RMSE $_g$		
		200	500	1000	200	500	1000
NM	$h = 1/2$	0.154	0.137	0.126	0.101	0.118	0.126
	$h = 1/3$	0.124	0.112	0.103	0.082	0.112	0.125
	$h = 1/5$	0.108	0.098	0.091	0.029	0.049	0.065
	$h = 1/10$	0.101	0.092	0.085	0.082	0.028	0.015
	$h = 1/15$	0.100	0.091	0.085	0.104	0.043	0.019
SOLS	$k = 2$	0.300	0.300	0.300	4.749	6.423	8.088
	$k = 3$	0.016	0.005	0.003	0.103	0.036	0.017
	$k = 5$	0.059	0.019	0.009	0.662	0.246	0.131
	$k = 10$	0.240	0.212	0.199	1.088	1.235	1.310
	$k = 15$	0.324	0.316	0.123	1.218	1.476	0.968

Although the power series may work well with a relatively small truncation parameter when  $g(\cdot)$  is a certain polynomial function, it may not 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 B.12, 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.29).

Table B.12: (*DGP 2, Case 1*)

	$h, k \setminus T$	RMSE $_{\theta}$			RMSE $_g$		
		200	500	1000	200	500	1000
NM	$h = 1/2$	0.072	0.065	0.059	0.922	0.942	0.952
	$h = 1/3$	0.043	0.040	0.038	0.864	0.898	0.913
	$h = 1/5$	0.031	0.030	0.028	0.669	0.720	0.743
	$h = 1/10$	0.026	0.026	0.025	0.415	0.480	0.509
	$h = 1/15$	0.026	0.025	0.024	0.294	0.364	0.394
SOLS	$k = 2$	0.187	0.187	0.187	1.246	1.404	1.501
	$k = 3$	0.213	0.219	0.221	4.522	6.073	7.539
	$k = 5$	0.186	0.171	0.165	3.881	4.212	4.690
	$k = 10$	0.273	0.285	0.200	1.663	2.039	1.872
	$k = 15$	0.267	0.257	0.200	1.613	1.982	1.873

## References

- Baek, Y. I., Cho, J. S. and Phillips, P. C. B. (2015), ‘Testing linearity using power transforms of regressors’, *Journal of econometrics* **187**(1), 376–384.
- Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction, second edition*. In: *Lecture Notes in Statistics*, Springer–Verlag.
- Chen, J., Gao, J. and Li, D. (2012), ‘Semiparametric trending panel data models with cross-sectional dependence’, *Journal of Econometrics* **171**(1), 71–85.
- Cho, J. S. and Phillips, P. C. B. (2018), ‘Sequentially testing polynomial model hypotheses using power transforms of regressors’, *Journal of applied econometrics* **33**(1), 141–159.
- 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. (2018), ‘Additive nonparametric models with time variable and both stationary and nonstationary regressors’, *Journal of Econometrics* **207**(1), 212–236.
- Fan, J. and Yao, Q. (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer–Verlag.
- Gao, J. and Hawthorne, K. (2006), ‘Semiparametric estimation and testing of the trend of temperature series’, *Econometrics Journal* **9**(2), 332–355.
- Maria, A. A. and Wen, Y. (2015), ‘Recovery from the great recession has varied around the world’, *The Regional Economist, Federal Reserve Bank of St. Louis* **October**, 10–11.
- Newey, W. K. (1997), ‘Convergence rates and asymptotic normality for series estimators’, *Journal of Econometrics* **97**(1), 147–168.
- Newey, W. K. and Powell, J. L. (2003), ‘Instrumental variable estimation of nonparametric models’, *Econometrica* **71**(5), 1565–1–578.
- Phillips, P. C. B. (2007), ‘Regression slowly varying regressors and nonlinear trends’, *Econometric Theory* **23**(4), 557–614.
- 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. (2012), ‘Inference on power law spatial trends’, *Bernoulli* **18**(2), 644–677.
- 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.