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

JITI GAO* AND OLIVER LINTON⋆ AND BIN PENG†

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

Jan, 2019

Abstract

We consider 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. We first propose two hypothesis tests to detect whether two nested special cases are appropriate. For the case where both null hypotheses are rejected, we propose an estimation method to capture certain 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: Hypothesis testing; Nonparametric Kernel Estimation; Nonstationarity

JEL classification: C14, C22

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

1 Introduction

Trends have been widely studied and used for more than half a century (e.g., Jones, 1943; Anderson, 1971; Hamilton, 2017; Andrews and McDermott, 1995; Phillips, 2001, 2005, 2007, 2009). There is no doubt that time trends exist in many data sets from different fields, so that how to model 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 focuses on the fluctuations about the trend, which is why so many applied works start from detrending the data. 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. Specifically: (1) unit roots and stochastic trends; (2) global deterministic time trends involving a linear term t and/or a quadratic term t^2 (e.g., Feng and Serletis, 2008, Eq. 13 and 19); (3) local deterministic trends under the nonparametric setting, which capture slowly varying long run components (e.g., Engle and Rangel, 2008; Hafner and Linton, 2010); etc. For the third approach, 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).

However, 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. This issue has been raised by Phillips (2007) and Robinson (2012), where power trends have been studied under parametric frameworks. 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. 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 $\theta_0 \geq 0$. The component $g(\cdot)$ 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" behavior that we do not model or estimate. We start from (1.1), and further discuss more generalised settings as well as the associated issues in Section B.3 of the online supplementary file. Our model extends the parametric global trend models considered in[1] Phillips (2007) and Robinson (2012) and the nonparametric local trend model

[1] Phillips (2007) considers multiple regressions with many forms of slowly varying regression functions, which could not be fully covered in this study. Robinson (2012) considers multiple nonlinear power function regressions.

that underpins a lot of statistical trend fitting. In this paper, we are interested in estimating θ_0 and $g(\cdot)$ from a time series dataset $\{y_1, \ldots, y_T\}$. Sornette (2003) proposes deterministic trend and cusp models for modelling stock market crashes with both global trend and bounded trend, but the models are parametric.

We comment briefly on the stochastic trend literature. 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 the two approaches are nonnested, although in practice they achieve similar objectives. 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 components earmarked as trend and the cycle. Our model in contrast owes its flexibility to the nonparametric nature of the deterministic component function. Dahlhaus (1997) introduces a class of locally stationary processes, which combines deterministic local trends with stochastic variation, see also Giraitis, Kapetanios and Yates (2014) who consider a time–varying coefficient model with stochastic variation.

We summarize our contributions: (1) This is the first paper to combine the global and slowly–changing local time trends together; (2) This study provides the practitioner from a variety of fields with a new nonparametric trending method to examine, capture, and remove time effects; (3) We provide the tools to test for the presence of such effects and to estimate its components.

The structure of this paper is as follows. In Section 2 we present the regularity conditions we use in the paper. In Section 3 we propose two hypothesis tests for evaluating the nested parametric and nonparametric models. In Section 4 we propose estimators of both trend components and investigate their asymptotic properties. We provide some simulation studies in Section 5 that examine the finite sample performance of the proposed tests and estimation methods. In Section 6 we discuss some potential extensions and issues. Section 7 concludes. Mathematical proofs of the main results are given in Appendix A. Finally, in the online supplementary file of this paper available at Cambridge Journals Online (journals.cambridge.org/ect), we apply our methodology to study global mean sea level and U.S. GDP data. There can also be found the omitted proofs of the main text and some additional material.

Before proceeding to Section 2, it is convenient to introduce some notation that will be used throughout this paper. The symbol $\to_P$ denotes convergence in probability; $\to_D$ denotes convergence 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;

We refer interested readers to these two papers for more details.

moreover, $K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$.

2 Regularity Conditions

We make the following assumptions we will use to derive our results.

Assumption 1:

1. $0 \leq \theta_0 \in \Theta$, *and* Θ *is a compact set defined on* $\mathbb{R}$. $g(\cdot)$ *is second order differentiable on* $[0,1]$, *and satisfies that* $\sup_{u\in[0,1]} |g(u)| < \infty$, $\inf_{\theta\in[0,1]} \left|\int_0^1 u^{\theta_0+\theta} g(u)du\right| > 0$, *and* $\sup_{(\theta,u)\in\Theta\times[h,1]} \left|\frac{d[u^{\theta+\theta_0}g(u)]}{du}\right| < \infty$ *for the same* h *defined in Assumption 1.4 below.*

2. $\{\varepsilon_t \mid t = 1, \ldots, T\}$ *is an* 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* $\delta > 0$ *satisfying* $\max_{t\geq 1} E|\varepsilon_t|^{2+\delta/2} < \infty$, *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 sigma field generated by* $\{\varepsilon_t \mid j \leq t \leq k\}$. *Moreover, for* $t \geq 1$, $E[\varepsilon_t] = 0$ *and* $E|\varepsilon_t|^2 = \sigma_t^2 \leq c_0 < \infty$.

3. *Let* $K(\cdot)$ *be a function that is 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 sequence* h, *suppose that* $h = O(T^{-\nu})$ *for some* $0 < \nu < \frac{1}{2}$.

Assumption 1.2*:

Suppose that $\{\varepsilon_t\}$ *satisfies either one of the following conditions:*

1. *For* $t \geq 2$, *let* $E[\varepsilon_t \mid \mathcal{F}_t] = 0$, *where* $\mathcal{F}_t \equiv \sigma(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{t-1})$. *In addition,* $E\left[\varepsilon_t^2 \mid \mathcal{F}_t\right] = \sigma_t^2 \leq c_0 < \infty$ *almost surely, and* $\max_{t\geq 1} E[\varepsilon_t^4] < \infty$.

2. *Let Assumption 1.2 hold. Moreover, let* $\sum_{t=2}^T \sum_{s=1}^{t-1} \gamma(t-s)\,\omega_{Tt}\,\omega_{Ts} \to 0$ *as* $T \to \infty$, *where* $\gamma(j) = E[\varepsilon_1\varepsilon_{1+j}]$ *and* $\omega_{Tt} = \frac{g(\tau_t)\ \ln(t)}{\sqrt{\sum_{t=1}^T \sigma_t^2 g^2(\tau_t)\ [\ln t]^2}}$.

Compared to the conditions employed by some of the relevant literature (e.g., Vogt, 2012; Phillips, Li and Gao, 2017), one main difference is that we have to take the power term into consideration when using the kernel method below. This is why we require $\theta_0 \geq 0$ in Assumption 1.1, which is harsher than $\theta_0 > -\frac{1}{2}$ adopted in Robinson (2012) for a parametric model. We will further discuss this issue in detail in Section 4. We also impose some conditions on $g(\cdot)$, which are quite standard. Assumptions 1.1–1.4 are standard in the literature (e.g., Fan and Yao, 2003, Section 2.6).

Assumption 1.2* is a stronger version of Assumption 1.2, and is used only to establish asymptotic properties for the proposed tests in Section 3 below. Assumption 1.2*.1 is a martingale type of condition, and is similar to Assumption A.2 of Su and Chen (2013) and Assumption A.4 of Su, Jin and Zhang (2015). Meanwhile, it allows for the heteroskedasticity, and is analogous to Assumption A1 of Fan and Li (1996). To model more complicated deterministic heteroskedasticity, we refer interested readers to, for example, Section 3.3 of Gao (2007). Assumption 1.2*.2 allows for certain types of weak autocorrelation, and is verifiable in many situations, including the case where $\{\varepsilon_t\}$ follows an ARMA setting.

Either of the two conditions of Assumption 1.2* ensures that the summation of the interaction terms, $\sum_{t=2}^{T}\sum_{s=1}^{t-1}\gamma(t-s)\,\omega_{Tt}\,\omega_{Ts}$, will not create any difficulty while estimating the asymptotic variance in the proof of Theorem 3.1. Although one indeed can consistently estimate the correlation between ε_t and ε_s for any fixed $\ell = t-s \geq 1$ (Fan and Yao, 2003, Chapter 2), one cannot recover, for example, $\sum_{t=2}^{T}\sum_{s=1}^{t-1}\gamma(t-s)\,\omega_{Tt}\,\omega_{Ts}$ as a whole in general without imposing stronger restrictions.

Sections 3 and 4 together provide the main asymptotic results of the paper. In Section 3 we provide two tests of the leading special cases of (1.1). In Section 4 we provide estimation methodology for (1.1). We point out the failure of some intuitive methods in Section 4.1, we discuss how to achieve consistent estimation in general in Section 4.2, and we study the detailed consistent estimators of $g(\cdot)$ and θ_0 based on the least squares method defined in Section 4.3.

3 Two Testing Issues

We first consider two hypothesis tests:

$$\text{(a). Testing } \theta_0\text{:} \begin{cases} H_0:\ \theta_0 = 0 \\ H_1:\ \theta_0 > 0; \end{cases} \tag{3.1}$$

$$\text{(b). Testing } g(\cdot)\text{:} \begin{cases} H_0^*:\ g(\tau) \text{ is a constant function} \\ H_1^*:\ g(\tau) \text{ is a non–constant function.} \end{cases} \tag{3.2}$$

If we fail to reject either of these null hypotheses, everything goes back to some well studied models. (a) Failure to reject H_0 gives the model $y_t = g(\tau_t) + \varepsilon_t$, which, for example, is a special case of Robinson (1997) and Dong and Linton (2018). In addition, $y_t = g(\tau_t)+\varepsilon_t$ nests $y_t = a_0+\varepsilon_t$ as a special case. One can follow Section 3.2 to further test whether $g(\cdot)$ is a constant function, and the procedure can be much simplified. (b) Failure to reject H_0^* leads to $y_t = \beta_0\, t^{\theta_0} + \varepsilon_t$, which has been studied in Phillips (2007) and Robinson (2012).

If both null hypotheses are rejected by the data (at an appropriate significance level), then we may conclude that the general model (1.1) holds or at least we cannot work with either of the (already treated) special cases. In the next subsections we present tests of the two hypotheses (3.1) and (3.2).

3.1 Testing θ_0

If g were known, the Gaussian log–likelihood would be proportional to $Q_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - g(\tau_t)t^\theta\right)^2$, which yields the score function

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

Under the null of (3.1), this reduces to $\frac{\partial Q_T(\theta)}{\partial\theta}\Big|_{\theta=0} = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - g(\tau_t)\right) g(\tau_t)\ln t$. In practice, since $g(\cdot)$ is unknown, we replace $g(\cdot)$ by a kernel based nonparametric estimator $\widehat{g}(\cdot)$. However, we noticed that using the full sample to construct the test will result in two leading terms cancelling with each other, so that further difficulties will arise when deriving the asymptotic distribution. In order to avoid this technical problem, we use sample splitting: we use the even numbered observations to estimate $g(\cdot)$ and we evaluate the score function using the odd numbered observations.[2] Thus, the final version of the score function considered is

$$S_T = \frac{1}{T/2}\sum_{t\text{ odd}}\left(y_t - \widehat{g}(\tau_t)\right)\widehat{g}(\tau_t)\ln t, \tag{3.3}$$

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

Based on the above discussion, a formal hypothesis test is described in the next theorem.

Theorem 3.1. *Let Assumptions 1.1, 1.2*, 1.3 and 1.4 hold.*

1. *In addition,* $\sup_{u\in[0,1]}\left|\frac{\partial g(u)}{\partial u}\right| < \infty$. *Under the null hypothesis of (3.1), as* $T\to\infty$,

$$\widehat{LM} = \frac{\frac{1}{2\sqrt{T}}\sum_{t\text{ odd}}\left(y_t - \widehat{g}(\tau_t)\right)\widehat{g}(\tau_t)\ln t}{\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\widetilde{g}(\tau_t)\ln t\right]^2\widetilde{e}_t^2\right\}^{1/2}} \to_D N(0,1),$$

where $\widetilde{e}_t = y_t - \widetilde{g}(\tau_t)$, *and* $\widetilde{g}(u) = \frac{\sum_{t=1}^{T} K_h(u-\tau_t)y_t}{\sum_{t=1}^{T} K_h(u-\tau_t)}$.

[2] One can also use the even indexed sample to construct S_T of (3.3), and estimate $\widehat{g}(\cdot)$ with the odd indexed sample. Theoretically speaking, both methods of splitting sample lead to the same asymptotic distribution in Theorem 3.1. However, it may cause some difference when using real data, so, in applied works, one may try both methods to see if they reach the same conclusion, which is exactly what we do in the empirical study. We thank one referee for raising this possible confusion due to splitting sample.

2. *Under the alternative hypothesis of (3.1), as* $T \to \infty$, $\widehat{LM} \to \infty$.

We will further provide a generalized version of the test (i.e., $H_0:\ \theta_0 = a$ vs. $H_1:\ \theta_0 > a$) with discussion on establishing inference for θ_0 in Section 6 after providing the consistent estimators of θ_0 and $g(\cdot)$ in Section 4.

3.2 Testing $g(\cdot)$

We now consider the hypothesis (3.2). Notice that, under H_0^*, we have a parametric model of the form $y_t = \beta_0\, t^{\theta_0} + \varepsilon_t$, and the unknown parameters (β_0, θ_0) can be estimated by

$$(\widehat{\beta}, \widehat{\theta}) = \arg\min_{(\beta,\theta)} \sum_{t=1}^{T} \left(y_t - \beta\, t^{\theta}\right)^2, \tag{3.4}$$

which has been fully studied in Phillips (2007) and Robinson (2012).

We now propose a multiscale test of the form proposed by Gao and Hawthorne (2006):

$$\widehat{L} = \max_{h\in\mathcal{H}} L(h) \quad \text{with} \quad L(h) = \frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} K\left(\frac{\tau_t-\tau_s}{h}\right)\widehat{e}_s\,\widehat{e}_t}{\sqrt{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} K^2\left(\frac{\tau_t-\tau_s}{h}\right)\widehat{e}_s^2\,\widehat{e}_t^2}}, \tag{3.5}$$

where $\mathcal{H} = \{h = h_{max}a^k : h \geq h_{min}, k = 0.1, 2, \ldots\}$ with $0 < h_{min} < h_{max}$ and $0 < a < 1$, and $\widehat{e}_t = y_t - \widehat{\beta}\, t^{\widehat{\theta}}$. The associated critical values can be obtained 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 some mean zero unit variance distribution, such as $N(0, 1)$.

2. Use $\{y_t^* \mid t = 1, \ldots, T\}$ to implement (3.4) in order to obtain $(\widetilde{\beta}, \widetilde{\theta})$, and compute the statistic L^* by replacing y_t and $(\widehat{\beta}, \widehat{\theta})$ with y_t^* and $(\widetilde{\beta}, \widetilde{\theta})$, respectively, in (3.5).

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

Theorem 3.2. *Let Assumptions 1.1, 1.2*.1, 1.3, and 1.4 hold. For* $\mathcal{H}$ *of (3.5), suppose that* $c_0[\ln(\ln T)]^{-1} = h_{max} > h_{min} \geq T^{-\vartheta} > 0$ *with some constants* c_0 *and* ϑ *such that* $0 < \vartheta < \frac{1}{3}$.

1. *Under the null of (3.2),* $L(h) \to_D N(0, 1)$, *and* $\lim_{T\to\infty} \Pr(\widehat{L} > l_\alpha) = \alpha$;

2. *Under the alternative of (3.2),* $\lim_{T\to\infty} \Pr(\widehat{L} > l_\alpha) = 1$.

Theorem 3.2 follows from developments similar to the earlier studies by Fan and Li (1996) and Li (1999). The second conclusion of Theorem 3.2 is the same as that of Proposition 1 of Gao and Hawthorne (2006). The same principle of this nonparametric test has also been employed in Su and Chen (2013) and Su et al. (2015) to study panel data models.

We will examine the finite sample performance of Theorems 3.1 and 3.2 in the simulation study of Section 5.

4 Estimation Method and Theory

We now consider estimating (1.1) for the case where $\theta_0 > 0$ and $g(\cdot)$ is a non–constant function. For all (θ, u), the profile least squares estimator of $g(u)$ is defined as

$$\widehat{g}(u,\theta) = \left[\sum_{t=1}^{T} t^{2\theta} K_h\left(u - \tau_t\right)\right]^{-1} \sum_{t=1}^{T} t^{\theta} y_t K_h\left(u - \tau_t\right). \tag{4.1}$$

The key question is how to recover θ_0. Once we have obtained a consistent estimator for θ_0, we need only to plug it in (4.1) to estimate $g(u)$. We first explain why two intuitive least squares methods fail to deliver consistent estimates of θ_0.

4.1 Failure of Some Intuitive Methods

First, we may use the global profile method (e.g., Robinson, 2012; Dong, Gao and Tjøstheim, 2016), with objective function defined as follows:

$$Q_T(\theta) = \sum_{t=1}^{T} \left(y_t - t^{\theta}\widehat{g}(\tau_t,\theta)\right)^2, \tag{4.2}$$

where $\widehat{g}(u,\theta)$ is denoted in (4.1). According to Lemma 4.1 below, we find 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 appears in the residual. Thus, it would be difficult to recover θ_0 from (4.2), as the first order limit of $Q_T(\theta)$ does not depend on θ.

Alternatively, we may use a local profile method, following Section 6 of Phillips (2007). Define the objective function for any given u as

$$Q_T(\beta,\theta\,|u) = \sum_{t=1}^{n} \left(y_t - \beta\, t^{\theta}\right)^2 K_h\left(\tau_t - u\right). \tag{4.3}$$

For all u, the estimators $\left(\widehat{\beta}(u), \widehat{\theta}(u)\right)$ are obtained by minimizing $Q_T(\beta, \theta\,|u)$. Finally, the estimator of θ_0 is obtained by $\widehat{\theta} = \int_0^1 \widehat{\theta}(u)\psi(u)du$, where $\psi(\cdot)$ serves as a weight function. Note that, to minimize $Q_T(\beta, \theta\,|u)$, the first order conditions $\frac{\partial Q_T(\beta,\theta\,|u)}{\partial \beta}\Big|_{(\beta,\theta)=(\widehat{\beta}(u),\widehat{\theta}(u))} = 0$ and $\frac{\partial Q_T(\beta,\theta\,|u)}{\partial \theta}\Big|_{(\beta,\theta)=(\widehat{\beta}(u),\widehat{\theta}(u))} = 0$ must hold, and the first equation yields

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

which has the same form as (4.1), and indicates that the leading term of $Q_T(\widehat{\beta}(u), \widehat{\theta}(u)\,|u)$ is independent of $\widehat{\theta}(u)$ by the same discussion under (4.2). In other words, we can find different θ's belonging to Θ (say, $\widehat{\theta}_1(u)$ and $\widehat{\theta}_2(u)$) to ensure $Q_T(\widehat{\beta}(u), \widehat{\theta}_1(u)\,|u)$ and $Q_T(\widehat{\beta}(u), \widehat{\theta}_2(u)\,|u)$ are asymptotically equivalent. This concludes why the second approach fails.

We leave the numerical examination of these two methods in the online supplementary file of this paper, as they are not our main focus.

4.2 Consistent Estimation

We first provide a result about the performance of the profiled g estimator, which supports our estimation strategy for θ_0.

Lemma 4.1. *Consider $\widehat{g}(u,\theta)$ defined by (4.1), and let Assumption 1 hold. In addition, (1) let $B_T(\theta_0) = [\theta_0 - \frac{M}{\ln T}, \theta_0 + \frac{M}{\ln T}]$, where M is a positive constant; (2) let $B_{\epsilon_1}(h) = [(1+\epsilon_1)h,\ 1]$, where ϵ_1 is a sufficiently small positive constant. As $T \to \infty$,*

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

The constant ϵ_1 controls the minimum value that u is permitted to take, and serves the same purpose as C_1 of Theorem 4.2 of Vogt (2012). Lemma 4.1 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 the boundary of $B_T(\theta_0)$. In Section 4.3, we show that $\widehat{\theta}$ defined by (4.6) indeed falls in $B_T(\theta_0)$ with probability approaching one in Theorem 4.2, and further deal with the unknown constant in Theorem 4.3.

We next explain in general terms our estimation strategy for model (1.1) and some issues that arise. By Lemma 4.1, we write $u^{\theta}\widehat{g}(u,\theta) \simeq u^{\theta_0} T^{\theta_0-\theta} g(u)$, so that

$$\int \left(u^{\theta}\widehat{g}(u,\theta)\right)^2 du \simeq T^{2\theta_0-2\theta} \int u^{2\theta_0} g^2(u) du$$

$$
\begin{aligned}
&\Rightarrow \quad \frac{1}{\ln T^2} \ln \int \left(u^\theta \widehat{g}(u,\theta)\right)^2 du \simeq (\theta_0 - \theta) + e_T \\
&\Rightarrow \quad \left(\frac{1}{\ln T^2} \ln \int \left(u^\theta \widehat{g}(u,\theta)\right)^2 du\right)^2 \simeq (\theta_0 - \theta)^2 + e'_T,
\end{aligned} \tag{4.4}
$$

where e_T, e'_T are $O(1/\ln T)$.[3] Moreover, the expectation of the "true error term" of (4.4) (i.e., e_T) is not 0, but goes to 0 at the rate $\frac{1}{\ln T}$. This reveals why we achieve only a slow rate $\frac{1}{\ln T}$ in Theorem 4.2 below. The verification can easily be done considering the traditional OLS estimator, so it is omitted. Last but not least, although e_T serves as an error term and converges to 0 asymptotically, e_T itself is not random at all and is made of deterministic components. That is why the first result of Theorem 4.4 is a constant instead of a distribution.

4.3 Asymptotic Results for Least Squares Method

We focus on the least squares method due to its popularity and simplicity. It allows for the possibility that $g(\cdot)$ may take negative values. Define the objective function

$$
R_T(\theta) = \left\{ \lambda_T \cdot \ln \left[\frac{1}{T} \sum_{t=\lfloor Th \rfloor + 1}^{T} \tau_t^{2\theta} \widehat{g}(\tau_t, \theta) \right]^2 \right\}^2, \tag{4.5}
$$

where $\lambda_T = \frac{1}{\ln T}$ serves as a normalizer, and $\widehat{g}(\cdot,\cdot)$ is defined in (4.1). The estimator of θ_0 is given by

$$
\widehat{\theta} = \arg\min_{\theta \in \Theta} R_T(\theta). \tag{4.6}
$$

Other methods like least absolute deviations or quantile regression deserve to be considered in separate papers. We leave them to future research.

Remark: *Further to our discussion of Section 4.2, the term $\tau_t^{2\theta}$ in (4.5) serves the purpose of solving a technical issue when recovering the normalizer of Theorem 4.4. A short explanation is that without $\tau_t^{2\theta}$, the term $\frac{1}{T}\sum_{t=\lfloor Th \rfloor+1}^{T} \frac{\partial \widehat{g}(\tau_t,\theta_0)}{\partial \theta}$ will yield a simple average $\frac{1}{T}\sum_{t=\lfloor Th \rfloor+1}^{T} \tau_t^{-2\theta_0}$ in the denominator, when considering the score function generated by (4.5). 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, because $\int_0^1 u^{-2\theta_0} du$ does not exist for $\theta_0 > \frac{1}{2}$.*

We summarize the corresponding asymptotic results in the next theorem.

Theorem 4.2. *Suppose that Assumption 1 holds. As $T \to \infty$,*

[3]Note that we can also take absolute value rather than squared value in the last step of (4.4), which then would lead to a least absolute deviations estimator.

1. $\widehat{\theta} \to_P \theta_0$;

2. $\widehat{\theta} - \theta_0 = O_P\left(\frac{1}{\ln T}\right)$;

3. $\sup_{u \in B_{\epsilon_1}(h)} \left| \widehat{g}(u, \widehat{\theta}) - (uT)^{\theta_0 - \widehat{\theta}} g(u) \right| = O_P\left(\frac{\sqrt{\ln T}}{T^{\frac{1}{2}+\theta_0} h^{\frac{1}{2}+2\theta_0}} \right) + O(h^{\min\{2\theta_0, 1\}})$, *where* $B_{\epsilon_1}(h)$ *is defined in Lemma 4.1.*

Before proceeding further, we explain two issues. Firstly, we consider the difference between our nonparametric model and some parametric models. Having said why we achieve only a slow rate $\frac{1}{\ln T}$ for (4.6) in the end of Section 4.2, we now show why for parametric models one need not take the logarithm, so that fast rates can be achieved. Consider a simple model even without an error term, say $y_t = \tau_t^{\theta_0}$. Simple calculation yields

$$
\begin{aligned}
Q_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T}(y_t - \tau_t^{\theta})^2 = \frac{1}{T}\sum_{t=1}^{T}\tau_t^{2\theta_0} - \frac{2}{T}\sum_{t=1}^{T}\tau_t^{\theta_0+\theta} + \frac{1}{T}\sum_{t=1}^{T}\tau_t^{2\theta} \\
&= \left(\int_0^1 u^{2\theta_0} du - 2\int_0^1 u^{\theta_0+\theta} du + \int_0^1 u^{2\theta} du \right) \cdot (1 + o(1)) \\
&= \left(\frac{1}{2\theta_0 + 1} - \frac{2}{\theta_0 + \theta + 1} + \frac{1}{2\theta + 1} \right) \cdot (1 + o(1)) \\
&= \frac{2(\theta_0 - \theta)^2}{(2\theta_0 + 1)(\theta_0 + \theta + 1)(2\theta + 1)} \cdot (1 + o(1)) \qquad (4.7)
\end{aligned}
$$

under minor restrictions. By the right hand side of (4.7), we can conclude that:

1. Without requiring any transformation, $Q_T(\theta)$ of (4.7) converges to a function having a unique minimum at $\theta = \theta_0$ asymptotically;

2. For $\theta_0 \leq -\frac{1}{2}$, the limit of $Q_T(\theta)$ no longer reaches its minimum value at $\theta = \theta_0$. That is one reason why Robinson (2012) only considers the power term on $(-\frac{1}{2}, \infty)$.

Secondly, we take a careful look at the estimation of $g(\cdot)$, and explain the identification issue of $g(\cdot)$ mentioned under Lemma 4.1. 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 4.2, we let $\theta = \theta^* + \frac{M}{\ln T}$ with M being a constant. Then we can write

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

$$= T^{2\theta^*} \sum_{t=1}^{T} \tau_t^{2\theta^*} \left\{ e^M g(\tau_t) \tau_t^{M/\ln T} - f(\tau_t) \right\}^2,$$

so any sequence $f_T(u) = e^M g(u) u^{M/\ln T}$ will set this objective function exactly zero.

In order to identify the unknown constant, we let $|g(1)| = 1$ in the rest of this paper. For those functions $g(\cdot)$ 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 Dong and Linton (2018) for similar settings on the functional component. To further establish the normality, we define for all $u \in (0,1)$

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

$$\widehat{\Sigma} = \frac{1}{Th} \sum_{t=\lfloor Th \rfloor+1}^{T} \left(y_t - t^{\widehat{\theta}} \widehat{g}(\tau_t, \widehat{\theta}) \right)^2 K^2 \left(\frac{u - \tau_t}{h} \right),$$

$$\kappa_{1T}(\widehat{\theta}, u) = |\widehat{g}(1, \widehat{\theta})|^{-1} \cdot \left(\sum_{t=1}^{T} t^{2\widehat{\theta}} K_h(u - \tau_t) \right)^{-1} \sum_{t=1}^{T} t^{\widehat{\theta}+\theta_0} g(\tau_t) K_h(u - \tau_t) - g(u). \tag{4.8}$$

Theorem 4.3. *Let Assumption 1 hold, and further let $\sigma_t^2 = \sigma^2(\tau_t)$ for $t \geq 1$. For $\forall u \in (0,1)$, as $T \to \infty$,*

1. $\frac{T^{\theta_0+\frac{1}{2}} h^{\frac{1}{2}} u^{\widehat{\theta}}}{\widehat{\eta}_T \sqrt{\widehat{\Sigma}}} \left(\frac{\widehat{g}(u,\widehat{\theta})}{|\widehat{g}(1,\widehat{\theta})|} - g(u) - \kappa_{1T}(\widehat{\theta}, u) \right) \to_D N(0,1)$, *where* $\kappa_{1T}(\widehat{\theta}, u) = O_P(h)$.
2. *Suppose further* $\sup_{\theta\in\Theta} \left| \frac{d^2[w^{\theta+\theta_0} g(w)]}{dw^2} \Big|_{w=u} \right| < \infty$, *and* $h = O(T^{-\nu})$ *with* $0 < \nu \leq 1 - \frac{2+\theta_0}{2.5+2\theta_0}$. *Then* $\kappa_{1T}(\widehat{\theta}, u) = O_P(h^2)$.

The fact that $\lim_{T\to\infty} |\widehat{\eta}_T| = |\int_0^1 u^{2\theta_0} g(u) du| > 0$ has been verified in the proof of Theorem 4.2. The bias term $\kappa_{1T}(\widehat{\theta}, u)$ is due to the use of the smoothing method, and the extra conditions required by the second result of Theorem 4.3 make certain that $\kappa_{1T}(\widehat{\theta}, u)$ will have the usual order $O_P(h^2)$ as in the literature of nonparametric regression (e.g., Vogt, 2012).

We are now ready to consider the asymptotic distribution of $\widehat{\theta}$. By (4.6), Theorem 4.2 and the Mean Value Theorem, we write

$$0 = (\ln T) \frac{\partial R_T(\theta)}{\partial \theta} \Big|_{\theta=\widehat{\theta}} = (\ln T) \frac{\partial R_T(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} + \frac{\partial^2 R_T(\theta)}{\partial \theta^2} \Big|_{\theta=\widetilde{\theta}} \cdot (\ln T)(\widehat{\theta} - \theta_0), \tag{4.9}$$

where $\widetilde{\theta}$ lies between $\widehat{\theta}$ and θ_0. We summarize the asymptotic results in the next theorem.

Theorem 4.4. *Let Assumption 1 hold. As $T \to \infty$,* *1.* $(\ln T)(\widehat{\theta}-\theta_0) \to_P \ln \left|\int_0^1 u^{2\theta_0} g(u)du\right|$;

2. Given $\left|\int_0^1 u^{2\theta_0} g(u)du\right| \neq 1$, $\frac{\ln T}{\ln|\widehat{\eta}_T|}(\widehat{\theta}-\theta_0) \to_P 1$, *where* $\widehat{\eta}_T$ *has been defined in (4.8).*

Theorem 4.4 shows that the limit of $(\ln T)(\widehat{\theta}-\theta)$ is a constant rather than a distribution, which confirms our discussion at the end of Section 4.2. Moreover, without the terms A_1, A_3 and A_5 in the proof of Theorem 4.4, the right hand side of (A.9) would lead to 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 4.4 as it is. In order to conduct inference on θ_0, we further provide Corollary 6.2 in Section 6.2, in which we provide a confidence interval for θ_0 under some strong restrictions.

5 Numerical Studies

We next conduct some simulation studies to examine the asymptotic results established in Sections 3 and 4. Due to space limitations, we report some selected results below and provide extra results in the online supplementary file of this paper. Throughout this paper, we stick to the Epanechnikov kernel only.

5.1 Testing θ_0

To examine the hypothesis test provided in Section 3.1 and account for the heteroskedasticity, the data generating process (DGP) is $y_t = g(\tau_t)t^{\theta_0} + \varepsilon_t$, where ε_t is independently generated from $N(0,\sigma_t^2)$, and σ_t^2 is drawn from a uniform distribution $U(1, 2.25)$. We consider the following cases under different sample sizes in order to evaluate the size and power of the test.

- Case 1 – Size: $\theta_0 = 0$
    1. Case 1.1: $g(w) = \exp(w^2/2)$; Case 1.2: $g(w) = w^2 + 1$
- Case 2 – Power: $\theta_0 = 0.3,\ 0.5,\ 0.7$
    1. Case 2.1: $g(w) = \exp(w^2/2)$; Case 2.2: $g(w) = w^2 + 1$

For each generated data set, we calculate $\widehat{LM}$ of Theorem 3.1, and let $\alpha_{LM} = 1(\widehat{LM} > 1.6449)$ (i.e., rejecting the null at 5% 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. We choose $J = 1000$. In view of (B.15) of the online supplementary file, the estimation error reaches the minimum value when $h = O\left(\left(\frac{\ln T}{T}\right)^{1/3}\right)$. Thus, we let $h = \left(\frac{\ln T}{T}\right)^{1/3}$, which is the "optimal" one under the null subject to an unknown constant. We plot the values of $\bar{\alpha}_{LM}$ (i.e., rejection rate) at different sample sizes in Figures 5.1 and 5.2 instead of reporting them in tables.

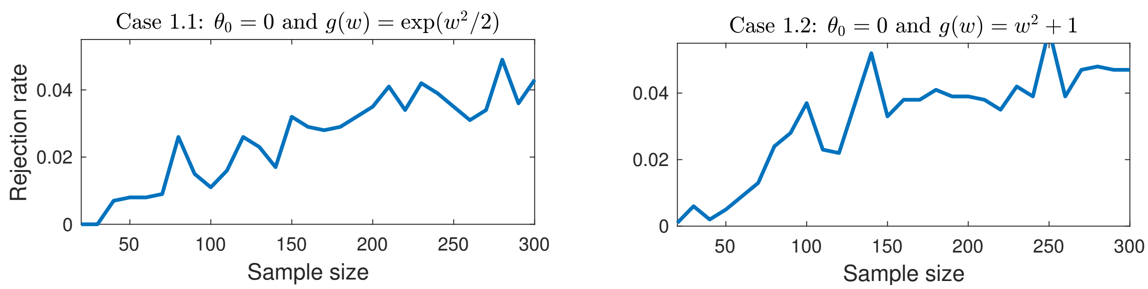

Figure 5.1: Testing θ_0: Case 1 – Size

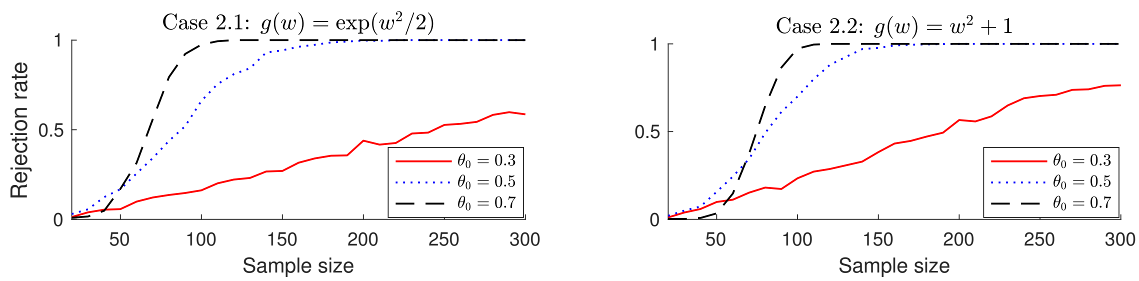

Figure 5.2: Testing θ_0: Case 2 – Power

According to Figures 5.1 and 5.2, the proposed test in general has good finite sample performance. In addition, Figure 5.2 suggests that as θ_0 gets far away from the null, the power 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.

5.2 Testing $g(\cdot)$

In this subsection, we study the test proposed in Section 3.2. It is worthwhile to mention that the principle of this 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; Gao and Hawthorne, 2006; Li, 1999; Su and Chen, 2013; 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 ε_t is independently generated from $N(0, \sigma_t^2)$, and σ_t^2 is drawn from a uniform distribution $U(1, 2.25)$. 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: $\theta_0 = 0.5,\ 1$

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

For each generated data set, we calculate the statistic value by (3.5), and 95% critical values by Theorem 3.2 based on 299 bootstrap replications. Similar to the above subsection, if we reject the null at 5% 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 5.3 and 5.4 below.

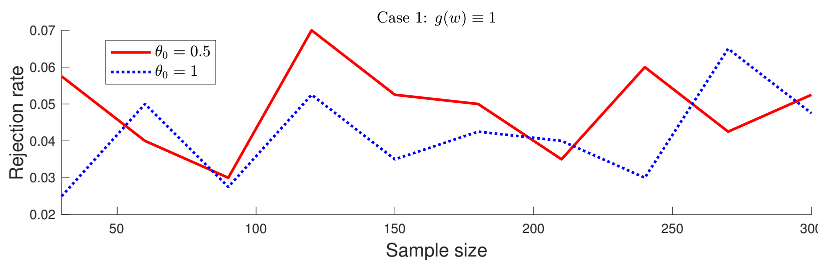

Figure 5.3: Testing $g(\cdot)$: Case 1 – Size

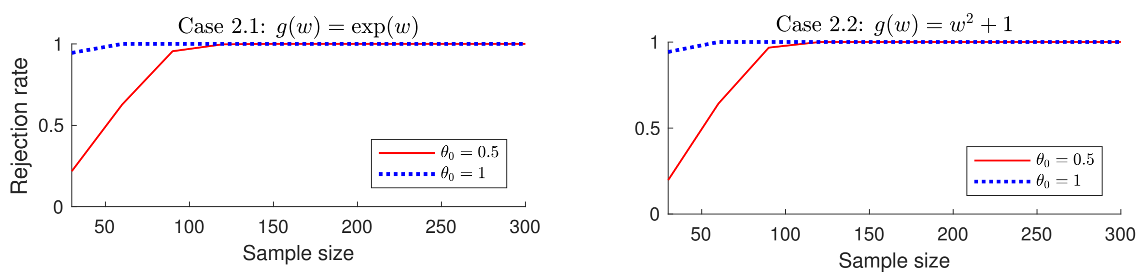

Figure 5.4: Testing $g(\cdot)$: Case 2 – Power

The size is still as good as expected by Figure 5.3, while, according to Figure 5.4, the power of the test is much better than what we see from the previous subsection.

5.3 Evaluation of the Estimates

Before proceeding further, we firstly provide a bandwidth selection procedure based on Theorem 4.2.[4]

[4]While designing the Monte Carlo study, we also tried to use the traditional cross–validation method to select the bandwidth. The criteria function is defined by $CV(h) = \sum_{t=\lfloor Th \rfloor+1}^{T} (y_t - \widehat{y}_{-t})^2$, where $\widehat{y}_{-t} = t^{\widehat{\theta}_{-t}} \widehat{g}_{-t}(\tau_t, \widehat{\theta}_{-t})$,

- **Bandwidth Selection:** It is easy to see that the rate of convergence of Theorem 4.2.2 will reach the minimum value at $h = O\left(T^{-\frac{1+2\theta_0}{3+4\theta_0}} \cdot (\ln T)^{\frac{1}{3+4\theta_0}}\right)$ for $\theta_0 \geq \frac{1}{2}$, and at $h = O\left(T^{-\frac{1+2\theta_0}{1+8\theta_0}} \cdot (\ln T)^{\frac{1}{1+8\theta_0}}\right)$ for $0 < \theta_0 < \frac{1}{2}$. In view of this relationship, we adopt the following iteration procedure, which yields an "optimal" bandwidth up to an unknown constant.

  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 $\widehat{\theta}_k$. Stop iteration, if $|\widehat{\theta}_k - \widehat{\theta}_{k-1}| \leq \epsilon$, where ϵ is sufficiently small (e.g., 10^{-6}) and serves as a stopping criteria. Otherwise, update the bandwidth by $h_k = T^{-\frac{1+2\widehat{\theta}_k}{3+4\widehat{\theta}_k}} \cdot (\ln T)^{\frac{1}{3+4\widehat{\theta}_k}}$ for $\widehat{\theta}_k \geq \frac{1}{2}$, and $h_k = T^{-\frac{1+2\widehat{\theta}_k}{1+8\widehat{\theta}_k}} \cdot (\ln T)^{\frac{1}{1+8\widehat{\theta}_k}}$ for $0 < \widehat{\theta}_k < \frac{1}{2}$. Then proceed to the $(k+1)^{th}$ iteration.

In order to examine the above bandwidth selection procedure as well as the asymptotic results of Section 4.3, the DGP is specified as $y_t = g(\tau_t)t^{\theta_0} + \varepsilon_t$, where we let θ_0 be 0.4 and 0.8 respectively. $\varepsilon_t = 0.5\,\varepsilon_{t-1} + N(0,1)$ and $g(u) = 3(u-1)^2 + 1$. We recover θ_0 by (4.6), and estimate $g(\tau_t)$ for $t = \lfloor Th \rfloor + 1, \ldots, T$ by $\widetilde{g}(u) = (uT)^{-\ln_T |\widehat{g}(1,\widehat{\theta})|}\widehat{g}(u,\widehat{\theta})$ as specified in (4.8). In addition, we calculate $\frac{\ln T}{\ln|\widehat{\eta}_T|}(\widehat{\theta} - \theta_0) - 1$ in order to examine Theorem 4.4. For each generated series $\{y_t\}$, three squared errors are recorded: $\mathrm{se}_\theta = (\widehat{\theta} - \theta_0)^2$, $\mathrm{se}^*_\theta = \left(\frac{\ln T}{\ln|\widehat{\eta}_T|}(\widehat{\theta} - \theta_0) - 1\right)^2$, and $\mathrm{se}_g = \frac{1}{T - \lfloor Th \rfloor}\sum_{t=\lfloor Th \rfloor + 1}^{T}\left(\widetilde{g}(\tau_t) - g(\tau_t)\right)^2$. After repeating the aforementioned procedure J times, we calculate the corresponding root mean squared errors, and label them as RMSE_θ, RMSE^*_θ and RMSE_g, respectively.[5]

Finally, we let $J = 1000$, $T = 100,\ 200,\ 400$ and $h = h_{opt},\ T^{-1/3},\ T^{-1/5},\ T^{-1/8}$, where "$h_{opt}$" is obtained by the procedure mentioned in the beginning of this subsection. The results are reported in Table 5.1. For $h = h_{opt},\ T^{-1/3}$, all RMSEs decrease, when the sample size increases. For $h = T^{-1/5}$ and $\theta_0 = 0.8$, RMSE^*_θ increases when the sample size increases. For $h = T^{-1/8}$, RMSE_g increases when the sample size increases. It suggests that $h = h_{opt},\ T^{-1/3}$ should be preferred practically when using our model and method. As expected, h_{opt} in general provides relatively good estimates in terms of RMSE_g and RMSE_θ. Although h_{opt} does not yield the

and $\widehat{\theta}_{-t}$ and $\widehat{g}_{-t}(\tau_t, \widehat{\theta})$ are obtained by (4.6) and (4.1) respectively but leaving the t^{th} observation out. However, the minimization process always causes our Matlab program to break down, not to mention that the cross–validation method is practically time–consuming. The possible reason is as follows. Suppose we search the optimal h on the set $(0, T^{-\nu_0}]$, where ν_0 is a sufficiently small positive number. It is not hard to see that both $\widehat{\theta}_{-t}$ and $\widehat{g}_{-t}(\tau_t, \widehat{\theta})$ will yield consistent estimates, which then suggests that $\widehat{y}_{-t} = t^{\widehat{\theta}_{-t}}\widehat{g}_{-t}(\tau_t, \widehat{\theta}_{-t})$ converges to $t^{\theta_0}g(\tau_t)$ by Lemma 4.1. In this case, the leading term of the cross–validation criteria function becomes $\sum_{t=\lfloor Th \rfloor+1}^{T}\left(y_t - t^{\theta_0}g(\tau_t)\right)^2$ in which the terms in the bracket are independent of h, so that the minimization process never converges to a possible solution.

As one referee kindly pointed out the popularity of the cross–validation method in applied research, we would like to share our experience and provide possible explanation here.

[5] Take RMSE_θ as an example. It is calculated by $\mathrm{RMSE}_\theta = \left(\frac{1}{J}\sum_{j=1}^{J}\mathrm{se}_{\theta,j}\right)^{1/2}$, where $\mathrm{se}_{\theta,j}$ stands for the value of se_θ obtained from the j^{th} replication.

best estimate in terms of RMSE^*_θ, the difference only happens at the second or third decimal, so negligible.

Table 5.1: Simulation Results

		RMSE_g			RMSE_θ			RMSE^*_θ		
	$h \setminus T$	100	200	400	100	200	400	100	200	400
$\theta_0 = 0.4$	h_{opt}	0.120	0.088	0.059	0.048	0.036	0.028	0.328	0.289	0.232
	$T^{-1/3}$	0.116	0.086	0.059	0.053	0.040	0.031	0.265	0.230	0.183
	$T^{-1/5}$	0.103	0.097	0.089	0.098	0.076	0.058	0.111	0.080	0.055
	$T^{-1/8}$	0.057	0.076	0.090	0.155	0.121	0.097	0.107	0.098	0.093
$\theta_0 = 0.8$	h_{opt}	0.075	0.055	0.038	0.134	0.115	0.101	0.100	0.095	0.090
	$T^{-1/3}$	0.083	0.065	0.049	0.136	0.116	0.102	0.092	0.088	0.085
	$T^{-1/5}$	0.130	0.130	0.124	0.164	0.137	0.117	0.017	0.019	0.024
	$T^{-1/8}$	0.081	0.111	0.133	0.205	0.169	0.142	0.038	0.038	0.035

6 Extensions with Discussion

In this section, we discuss some potential extensions with the corresponding issues. Due to space limitations, the associated proofs and simulation studies of these extensions are provided in the online supplementary file of this paper.

6.1 Extension 1

So far, we have been considering $0 \leq \theta_0 < \infty$ for our nonparametric case, which is stricter than the requirement of the parametric case of Robinson (2012). We now explain how to account for the case where $\theta_0 \in (-\frac{1}{2}, 0)$. In view of the development of Lemma B.2, it is not hard to see that if we sacrifice the range of u that $\widehat{g}(u, \theta)$ (defined by (4.1)) is permitted to take, then we can allow the wider range for θ_0.

Corollary 6.1. *Consider $\widehat{g}(u, \theta)$ defined by (4.1), let Assumption 1 hold, and relax the restriction of θ_0 to $-\frac{1}{2} < \theta_0 < \infty$. In addition, (1) let $B_T(\theta_0) = [\theta_0 - \frac{M}{\ln T}, \theta_0 + \frac{M}{\ln T}]$, where M is a positive constant; (2) let $B_{c_0} = [c_0,\ 1]$, where $0 < c_0 < 1$ is a positive constant. As $T \to \infty$,*

$$\sup_{(\theta,u) \in B_T(\theta_0) \times B_{c_0}} \left| \widehat{g}(u, \theta) - (uT)^{\theta_0 - \theta} g(u) \right| = O_P\left(\frac{\sqrt{\ln T}}{T^{\frac{1}{2}+\theta_0} h^{\frac{1}{2}+2\theta_0}} \right) + O(h).$$

Then, we can rewrite the objective function (4.5) as

$$R_T(\theta) = \left\{ \lambda_T \cdot \ln \left[\frac{1}{T} \sum_{t=\lfloor Tc_0 \rfloor + 1}^{T} \tau_t^{2\theta} \widehat{g}(\tau_t, \theta) \right]^2 \right\}^2. \tag{6.1}$$

The estimator of θ_0 is still $\widehat{\theta} = \arg\min_\theta R_T(\theta)$. All the main theorems still hold after minor modification. However, in this case, $100\, c_0\%$ data are not used at all, and as a consequence, we can no longer estimate $g(u)$ for $0 < u < c_0$.

6.2 Extension 2

We now provide a more generalized version of (3.1), which also indicates how to carry out inference about θ_0. To be precise, the test is specified as follows:

$$H_0 : \ \theta_0 = a \quad \text{vs.} \quad H_1 : \ \theta_0 > a, \tag{6.2}$$

where a is a positive constant. For example $a = 1$ is commonly adopted in some applied settings. For this test, we are able to state the next result.

Corollary 6.2. *Let Assumptions 1.1, 1.2*, 1.3 and 1.4 hold, and suppose $h^2 T^{2a} \ln T \to 0$.*

1. *Under the null of (6.2), as $T \to \infty$,*

$$\widehat{LM} = \frac{\frac{\sqrt{T^*}}{2} S_T}{\left\{ \frac{1}{T^*} \sum_{t \in B_h} \left[\widehat{e}_t \widehat{g}(\tau_t) t^a \ln t \right]^2 \right\}^{1/2}} \to_D N(0,1), \tag{6.3}$$

where $\widehat{e}_t = y_t - \widehat{g}(\tau_t)$, $B_h = \{t \mid \lfloor c_0 T \rfloor \leq t \leq \lfloor (1-h)T \rfloor \}$, T^* is the cardinality of B_h, $c_0 \in (0,1)$ is a fixed constant and

$$S_T = \frac{1}{T^*/2} \sum_{t \ \text{odd} \in B_h} (y_t - \widehat{g}(\tau_t) t^a)\, \widehat{g}(\tau_t) t^a \ln t,$$

$$\widehat{g}(u) = \left[\sum_{t \ \text{even} \in B_h} t^{2a} K_h (u - \tau_t) \right]^{-1} \sum_{t \ \text{even} \in B_h} t^a y_t K_h (u - \tau_t). \tag{6.4}$$

2. *Under the alternative of (6.2), as $T \to \infty$, $\widehat{LM} \to \infty$.*

Suppose that the condition $h^2 T^{2a} \ln T \to 0$ is satisfied, and let θ_α be the largest value of a satisfying $\widehat{LM} \leq z_\alpha$. By Corollary 6.2, we can construct a $(1-2\alpha)/2$ coverage interval for θ_0 of model (1.1) as $[\widehat{\theta}, \theta_\alpha]$, where $\widehat{\theta}$ is obtained by (4.6). If $2\widehat{\theta} - \theta_\alpha \geq 0$, then $[2\widehat{\theta} - \theta_\alpha, \theta_\alpha]$ further provides a $(1-2\alpha)$ coverage interval.

Remark: *In view of the development of Theorem 3.1 and Corollary 6.2, if a higher–order kernel is employed (i.e., $\int u^{\xi}K(u)du > 0$ for a given $\xi > 2$ and $\int u^{j}K(u)du = 0$ for $j < \xi$) and g is smooth enough and satisfies $\sup_{(\theta,u)\in\Theta\times[c_0,1-h]}\left|\frac{\partial^{\xi}[u^{\theta+\theta_0}g(u)]}{\partial u^{\xi}}\right| < \infty$, the condition $h^2T^{2a}\ln T \to 0$ can be further relaxed to $h^{\xi}T^{2a}\ln T \to 0$. In this case, we can establish the inference for θ_0 in a wider range. However, how to fully solve the inference issue for θ_0 remains unknown.*

6.3 Extension 3

In some applications it is of interest to allow for the effect of covariates. Consider a generalized trending model of the form

$$y_t = f(x_t, \tau_t) + g(\tau_t)t^{\theta_0} + \varepsilon_t, \tag{6.5}$$

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).

For model (6.5), the main results of this paper still hold.

Corollary 6.3. *Under Assumptions 1 and 2, consider model (6.5), and obtain $\widehat{\theta}$ and $\widehat{g}(u,\theta)$ by (4.6) and (4.1), respectively. As $T \to \infty$,*

1. $\widehat{\theta} - \theta_0 = O_P\left(\frac{1}{\ln T}\right)$;
2. $\sup_{u\in B_{\epsilon_1}(h)}\left|\widehat{g}(u,\widehat{\theta}) - (uT)^{\theta_0-\widehat{\theta}}g(u)\right| = O_P\left(\frac{1}{T^{\theta_0}h^{2\theta_0}}\right) + O(h^{\min\{2\theta_0,1\}})$, *where $B_{\epsilon_1}(h)$ is defined in Lemma 4.1.*

Assumption 2 is stated in the online supplementary file of this paper right before the detailed proofs of this corollary.

However, there are some issues when recovering $f(\cdot)$. 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 (2018) do in their paper; (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 leave detailed analysis of $f(\cdot,\cdot)$ to future studies.

Apart from the above extensions, we point out that Baek, Cho and Phillips (2015) and Cho and Phillips (2018) develop omnibus specification tests using general power functions and power trends, including specification tests for order estimation in polynomial regressions. An extension following Baek et al. (2015) and Cho and Phillips (2018) may be doable.

7 Conclusion

In summary, this paper provides the practitioner from a variety of fields with a new nonparametric trending method to examine, capture, and remove time effects. We firstly study two hypothesis tests. Then we consider the case where both of these special cases are not supported by the data. We provide consistent estimators and their corresponding asymptotic properties in the general model. Moreover, we examine the proposed hypothesis tests, estimation methods through both simulated and real data examples.

Finally, we acknowledge some limitations in the end of this paper, which may guide our future research. We assume smoothness on $g(\cdot)$, but it may be possible to extend the methodology to consider a finite number of trend breaks or discontinuities in $g(\cdot)$, see Delgado and Hidalgo (2000). Likewise the global trend may be subject to some breaks, Bai and Perron (1998). In addition, the specification does not nest the commonly–used parametric specifications (e.g., Phillips, 2007; Robinson, 2012), and the inference on the key parameter θ_0 is not fully solved.

Acknowledgements

The authors would like to thank Professor Peter C. B. Phillips, Professor Liangjun Su, and the three referees for their constructive suggestions and comments on earlier versions of this paper. The first author would like to acknowledge the Australian Research Council Discovery Grants Program for its support under Grant numbers: DP150101012 & DP170104421.

Appendix A

In this appendix, we provide the proofs for Theorems 4.2–4.4. The rest of the proofs are given in the online supplementary file of this paper. In addition, we provide some empirical studies, extra discussion and simulation studies in the online supplementary file.

Proof of Theorem 4.2:

(1). Firstly, we show $\widehat{\theta} \to_P \theta_0$. By Lemmas 4.1 and B.2, write

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

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

Thus, $\widehat{\theta}\to_P\theta_0$ follows immediately.

(2). After establishing the consistency, we focus on the rate of convergence. Note that $R_T(\theta)=\lambda_T^2R_T^*(\theta)$, 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

$$
\liminf_T \Pr\left\{R_T^*(\theta_0+\lambda_TC)>R_T^*(\theta_0)\right\}\geq 1-\epsilon, \tag{A.1}
$$

$$
\liminf_T \Pr\left\{R_T^*(\theta_0-\lambda_TC)>R_T^*(\theta_0)\right\}\geq 1-\epsilon. \tag{A.2}
$$

Both (A.1) and (A.2) holding true implies with probability at least $1-\epsilon$ that there exists a local minimum in the interval $U_T(\theta_0)=[\theta_0-\lambda_TC,\theta_0+\lambda_TC]$. 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 as the proof of Lemma A.1 of Wang and Xia (2009).

Write

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

where the definitions of $B_{1T}(\theta)$, $B_{2T}(\theta)$, $B_{3T}(\theta)$ and $B_{4T}(\theta_0)$ should be obvious; the second equality follows from Lemma 4.1; and we use $\approx$ in the third step due to dropping the term $(1+o_P(1))$.

Note that, for $\left|\int_h^1 u^{\theta_0+\theta} g(u)du\right|^2$, as $h \to 0$, $\left|\int_h^1 u^{\theta_0+\theta} g(u)du\right| > 0$ by Assumption 1.1, and

$$\begin{aligned}\left|\int_h^1 u^{\theta_0+\theta} g(u)du\right|^2 &\le \int_0^1 u^{2(\theta_0+\theta)}du \int_0^1 g^2(u)du \le O(1)\int_0^1 u^{2(\theta_0+\theta)}du \\ &= O(1)\frac{u^{2(\theta_0+\theta)+1}\big|_0^1}{2(\theta_0+\theta)+1} \le O(1)\frac{1}{2\inf_{\theta\in\Theta}(\theta_0+\theta)+1} < \infty. \end{aligned} \tag{A.3}$$

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

Based on the above development, we 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 indicates that (A.1) and (A.2) hold true with sufficiently large C. The proof of the second result is now complete.

(3). By Lemma 4.1 and the second result of this theorem, the third result follows. ■

Proof of Theorem 4.3:

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

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

where the definitions of A_1 and A_2 should be obvious.

After noting that u is fixed, it is easy to show that $A_1 = O_P(h)$ by proofs similar to (4) and (5) of Lemma B.2 (but much simpler). We then just need to focus on the normalized version of $\sum_{t=1}^T t^{\widehat{\theta}} \varepsilon_t K_h(u-\tau_t)$ and write

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

To investigate B_2, denote $B_T(\theta) = \frac{1}{T}\sum_{t=1}^T \tau_t^{\theta} \varepsilon_t K_h(u-\tau_t)$ and it is easy to see that the first derivative of $B_T(\theta)$ is $B_T^{(1)}(\theta) = \frac{1}{T}\sum_{t=1}^T \tau_t^{\theta}(\ln \tau_t) \varepsilon_t K_h(u-\tau_t)$, which is identical to the term considered in (3) of Lemma B.2. Then we can write

$$B_2 = B_T(\widehat{\theta}) - B_T(\theta_0) = (\widehat{\theta} - \theta_0)\cdot B_T^{(1)}(\theta^*) = (\widehat{\theta} - \theta_0)\cdot O_P\left(\frac{(\ln T)^{\frac{3}{2}}}{\sqrt{Th}}\right),$$

where θ^* lies between θ_0 and $\widehat{\theta}$; the second equality follows from the Mean Value Theorem; and the third equality follows from (3) of Lemma B.2.

By some standard arguments of time series analysis (e.g., Section 2.6.4 of Fan and Yao, 2003), we can prove $\sqrt{Th}B_1 \to_D N(0, \Sigma^*)$, 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

$$\begin{aligned}
&\frac{1}{Th} \sum_{t=1}^{T} \sum_{s=1}^{T} \tau_t^{\theta_0} \tau_s^{\theta_0} K\left(\frac{u-\tau_t}{h}\right) K\left(\frac{u-\tau_s}{h}\right) E[\varepsilon_t \varepsilon_s] \\
&= \frac{1}{Th} \sum_{t=1}^{T} \tau_t^{2\theta_0} K^2\left(\frac{u-\tau_t}{h}\right) E[\varepsilon_t^2] + \frac{2}{Th} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \tau_t^{\theta_0} \tau_s^{\theta_0} K\left(\frac{u-\tau_t}{h}\right) K\left(\frac{u-\tau_s}{h}\right) E[\varepsilon_t \varepsilon_s] \\
&= \frac{1}{Th} \sum_{t=1}^{T} \tau_t^{2\theta_0} K^2\left(\frac{u-\tau_t}{h}\right) \sigma^2(\tau_t) + \frac{2}{Th} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \tau_t^{\theta_0} \tau_s^{\theta_0} K\left(\frac{u-\tau_t}{h}\right) K\left(\frac{u-\tau_s}{h}\right) E[\varepsilon_t \varepsilon_s] \\
&:= V_{1T} + V_{2T}. \qquad\qquad (A.4)
\end{aligned}$$

It is easy to show that as $T \to \infty$, $V_{1T} = (1 + o(1))\sigma^2(u)u^{2\theta_0} \int_{-1}^{1} K^2(x)dx$. Note that V_{2T} is equivalent to the second term on the right hand side of (A.4) of Su, Chen and Ullah (2009). Using the truncation technique employed in (A.4)–(A.7) of Su et al. (2009), we obtain that $|V_{2T}| = o(1)$. Furthermore, by the first result of Theorem 4.4 (the details are temporarily omitted for now, as the order of these proofs does not matter), $|\widehat{g}(1, \widehat{\theta})| = T^{\theta_0 - \widehat{\theta}} \to_P \left|\int_0^1 u^{2\theta_0} g(u)du\right|^{-1}$, and simple calculation yields

$$\begin{aligned}
\widehat{\eta}_T &= \frac{1}{T} \sum_{t=\lfloor Th \rfloor+1}^{T} \tau_t^{2\widehat{\theta}} g(\tau_t) + \frac{1}{T} \sum_{t=\lfloor Th \rfloor+1}^{T} \tau_t^{2\widehat{\theta}} \left(\widetilde{g}(\tau_t) - g(\tau_t)\right) \\
&= \frac{1}{T} \sum_{t=\lfloor Th \rfloor+1}^{T} \tau_t^{2\widehat{\theta}} g(\tau_t) + o_P(1) = \int_0^1 u^{2\theta_0} g(u)du + o_P(1), \qquad (A.5)
\end{aligned}$$

where $\widetilde{g}$ has been defined in the body of this theorem; and the last equality follows from development similar to (B.8).

Based on the above analyses, the first result follows.

(2). Using the extra conditions imposed for the second result of this theorem, it is easy to show the second result follows. ■

Before proving Theorem 4.4, we denote some variables for notational simplicity and provide some discussions.

$$\Omega = \lim_{T\to\infty} \sum_{t=1}^{T} \sum_{s=1}^{T} E[V_t V_s], \quad V_t = V_{1t} + V_{2t}, \quad V_{1t} = -\frac{1}{T^{3/2}} \sum_{u=\lfloor Th \rfloor+1}^{T} \tau_u^{\theta_0} \varepsilon_u K_h(\tau_u - \tau_t),$$

$$V_{2t} = \frac{1}{T^{3/2}\ln T}\sum_{v=\lfloor Th\rfloor+1}^{T}\tau_v^{\theta_0}(\ln\tau_v)\varepsilon_t K_h(\tau_v-\tau_t). \tag{A.6}$$

We now verify the existence of Ω. 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

$$\int_h^1 K_h(w-u)dw = \begin{cases} \int_{-c}^1 K(w)dw, & u = h+ch\in[h,2h) \quad (\text{i.e., } c\in[0,1)) \\ 1, & u\in[2h,1-h] \\ \int_{-1}^c K(w)dw, & u = 1-ch\in(1-h,1] \quad (\text{i.e., } c\in[0,1)) \end{cases}, \tag{A.7}$$

which indicates $0\le \sup_{u\in[0,1]}\int_h^1 K_h(w-u)dw\le 1$. 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 equality follows from the definition of the Riemann integral; and the right hand side converges by (A.7) and standard arguments of time series analysis.

Proof of Theorem 4.4:

(1). By (B.2), it is easy to obtain that

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

$$
\begin{aligned}
&+\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}(\ln\tau_u)\tau_u^{2\theta_0}\frac{\sum_{t=1}^{T}t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)}{\sum_{t=1}^{T}t^{2\theta_0}K_h(\tau_u-\tau_t)}\\
&:=-2A_1-2A_2+A_3+A_4+2A_5+2A_6,
\end{aligned}
\tag{A.8}
$$

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 B.2, we are able to write

$$
\begin{aligned}
&\frac{T^{\theta_0+\frac{1}{2}}}{\ln T}(-2A_2+A_4+2A_6)\\
&=-(1+o(1))\cdot\frac{T^{\frac{1}{2}}}{\ln T}\cdot\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}(\ln\tau_u+\ln T)\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\\
&+(1+o(1))\cdot\frac{T^{\frac{1}{2}}}{\ln T}\cdot\frac{1}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}\frac{1}{T}\sum_{t=1}^{T}\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)(\ln\tau_t+\ln T)\\
&+(1+o(1))\cdot\frac{T^{\frac{1}{2}}}{\ln T}\cdot\frac{2}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}\frac{\ln\tau_u}{T}\sum_{t=1}^{T}\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\\
&=(1+o_P(1))\cdot\frac{1}{T^{3/2}}\sum_{u=\lfloor Th\rfloor+1}^{T}\sum_{t=1}^{T}\left\{-2\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)+\tau_t^{\theta_0}\varepsilon_t K_h(\tau_u-\tau_t)\right\}\\
&+(1+o_P(1))\cdot\frac{1}{T^{3/2}\ln T}\sum_{u=\lfloor Th\rfloor+1}^{T}\sum_{t=1}^{T}\tau_t^{\theta_0}(\ln\tau_t)\varepsilon_t K_h(\tau_u-\tau_t)\\
&=(1+o_P(1))\cdot\sum_{t=1}^{T}V_t,
\end{aligned}
\tag{A.9}
$$

where V_t has been defined in (A.6).

We then can use the large block and small block technique (e.g., Fan and Yao, 2003) to show that $\sum_{t=1}^{T}V_t\to_D N(0,\Omega)$, where Ω has been defined in (A.6). Thus, we know that

$$
-2A_2+A_4+2A_6=O_P\left(\frac{\ln T}{T^{\theta_0+\frac{1}{2}}}\right).
\tag{A.10}
$$

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

$$
\xi_T\to_P\int_0^1 u^{2\theta_0}g(u)du.
\tag{A.11}
$$

Thus, rearranging (4.9) using the decomposition (A.8) gives

$$
\begin{aligned}
&\left[\frac{\partial^2 R_T(\theta)}{\partial\theta^2}\Big|_{\theta=\widetilde{\theta}}\right]^{-1}\left\{\frac{-4\lambda_T^2\cdot\ln\xi_T^2}{\xi_T}\cdot(\ln T)(-2A_2+A_4+2A_6)\right\}\\
&=(\ln T)\left\{(\widehat{\theta}-\theta_0)-\left[\frac{\partial^2 R_T(\theta)}{\partial\theta^2}\Big|_{\theta=\widetilde{\theta}}\right]^{-1}\frac{4\lambda_T^2\cdot\ln\xi_T^2}{\xi_T}(2A_1-A_3-2A_5)\right\}.
\end{aligned}
\tag{A.12}
$$

Note that (A.10) and (7) of Lemma B.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.12) to obtain

$$\begin{aligned}(\ln T)(\widehat{\theta}-\theta_0)&=(\ln T)\left[\frac{\partial^2 R_T(\theta)}{\partial\theta^2}\Big|_{\theta=\widetilde{\theta}}\right]^{-1}\frac{4\lambda_T^2\cdot\ln\xi_T^2}{\xi_T}(2A_1-A_3-2A_5)+O_P\left(\frac{1}{T^{\theta_0+\frac{1}{2}}}\right)\\&=\lambda_T\frac{\ln|\xi_T|}{\xi_T}(2A_1-A_3-2A_5)+O_P\left(\frac{1}{T^{\theta_0+\frac{1}{2}}}\right).\end{aligned}\tag{A.13}$$

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

$$\begin{aligned}A_1&=\frac{1}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}\tau_u^{2\theta_0}\frac{\sum_{t=1}^T\sum_{s=1}^T\tau_t^{2\theta_0}\tau_s^{2\theta_0}g(\tau_s)K_h(\tau_u-\tau_t)K_h(\tau_u-\tau_s)(\ln\tau_t+\ln T)}{\left[\sum_{t=1}^T\tau_t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^2}\\&=(\ln T)\cdot\frac{1}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}\tau_u^{2\theta_0}\frac{\sum_{t=1}^T\sum_{s=1}^T\tau_t^{2\theta_0}\tau_s^{2\theta_0}g(\tau_s)K_h(\tau_u-\tau_t)K_h(\tau_u-\tau_s)}{\left[\sum_{t=1}^T\tau_t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^2}\\&+\frac{1}{T}\sum_{u=\lfloor Th\rfloor+1}^{T}\tau_u^{2\theta_0}\frac{\sum_{t=1}^T\sum_{s=1}^T\tau_t^{2\theta_0}\tau_s^{2\theta_0}g(\tau_s)K_h(\tau_u-\tau_t)K_h(\tau_u-\tau_s)\ln\tau_t}{\left[\sum_{t=1}^T\tau_t^{2\theta_0}K_h(\tau_u-\tau_t)\right]^2}\\&:=A_{11}+A_{12}.\end{aligned}$$

By Lemma B.2 and the definition of the Riemann integral, simple calculation yields

$$A_{11}=(\ln T)\int_0^1 g(u)du+o(1)\quad\text{and}\quad A_{12}=\int_0^1 u^{2\theta_0}g(u)(\ln u)du+o(1).$$

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

$$2A_1-A_3-2A_5=(\ln T)\int_0^1 u^{2\theta_0}g(u)du\cdot(1+O_P(\lambda_T)).\tag{A.14}$$

In connection with (A.13) and (A.11), 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 in the proof of Theorem 4.2. Thus, the proof of the first result of this theorem is now complete.

(2). The second result follows from (A.5) straight away. ■

References

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

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.

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.

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

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.

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. (2018), 'Additive nonparametric models with time variable and both stationary and nonstationary regressors', *Journal of Econometrics* **207**(1), 212–236.

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 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 Hawthorne, K. (2006), 'Semiparametric estimation and testing of the trend of temperature series', *Econometrics Journal* **9**(2), 332–355.

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

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. Working paper available at "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**(224), 453–465.

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

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 Chen, Q. (2013), 'Testing homogeneity in panel data models with interactive fixed effects', *Econometric Theory* **29**(6), 1079–1135.

Su, L., Chen, Y. and Ullah, A. (2009), 'Functional coefficient estimation with both categorical and continuous data', *Advances in Econometrics* **25**, 131–167.

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.