Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 98 (2007) 1539–1557
www.elsevier.com/locate/jmva

Statistical inference of partially linear regression models with heteroscedastic errors

Jinhong You[a,*], Gemai Chen[b], Yong Zhou[c]

[a] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7400, USA
[b] Department of Mathematics & Statistics, University of Calgary, Calgary, AL, Canada T2N 1N4
[c] Institute of Applied Mathematics, Academy of Mathematics and System Science, Chinese Academy of Science, Beijing 100080, PR China

Received 28 September 2005
Available online 5 July 2007

Abstract

The authors study a heteroscedastic partially linear regression model and develop an inferential procedure for it. This includes a test of heteroscedasticity, a two-step estimator of the heteroscedastic variance function, semiparametric generalized least-squares estimators of the parametric and nonparametric components of the model, and a bootstrap goodness of fit test to see whether the nonparametric component can be parametrized.

AMS 1980 subject classification: primary 62H12; secondary 62A10

Keywords: Semiparametric regression model; Heteroscedasticity; Local polynomial; Asymptotic normality; Model selection

1. Introduction

In regression modeling, when the mean response has a known relationship to some variables and an unknown relationship to additional variables, a semiparametric approach can be called in to reduce the high risk of model misspecification relative to a fully parametric model and avoid some serious drawbacks of a fully nonparametric model. One such a semiparametric model is the partially linear regression model introduced by Engle et al. [3] to study the effect of weather on electricity demand. Formally, a partially linear regression model can be defined as

$$y_i = \mathbf{x}_i^\tau \boldsymbol{\beta} + g(t_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

* Corresponding author.
E-mail address: jyou@bios.unc.edu (J. You).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2007.06.011

where y_i are responses, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\tau$ are design points, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\tau$ is a p-vector of unknown parameters, $t_i \in [0, 1]$ are additional design points for an unknown nonlinear relationship $g(\cdot)$ defined on [0, 1], ε_i are unobservable random errors and the prime τ denotes the transpose of a vector or matrix.

Model (1.1) has been studied extensively by many researchers; see, for example, the book by Härdle et al. [6] and the references therein. Usually, the errors are assumed to be i.i.d. to start. However, heteroscedasticity is often found in residuals from both cross-sectional and time series modeling in applications; see Baltagi [1]. It is well known that if the errors are heteroscedastic, the least-squares estimator of $\boldsymbol{\beta}$ is inefficient and the conventional estimator of the covariance matrix is usually inconsistent. When $g(\cdot) \equiv 0$ in model (1.1), namely, the standard linear regression model, many authors have addressed the related problems such as how to detect heteroscedasticity, how to construct efficient estimators of $\boldsymbol{\beta}$, how to construct consistent estimators of the corresponding covariance matrices and so on. However, for model (1.1) little has been discussed on how to detect heteroscedasticity, and, when heteroscedasticity does exist, on how to estimate it and conduct inference subsequently.

In this paper we consider the following heteroscedastic partially linear regression model:

$$y_i = \mathbf{x}_i^\tau \boldsymbol{\beta} + g(t_i) + \sigma(t_i)\varepsilon_i, \quad i = 1, \ldots, n, \tag{1.2}$$

where ε_i are i.i.d. with $E\varepsilon_1 = 0$ and $E\varepsilon_1^2 = 1$, $\sigma(\cdot)$ is an unknown function defined on [0, 1] and the rest are defined similarly to model (1.1).

To fix idea, we assume that both $\mathbf{x}_i$ and t_i are fixed, and $\{t_i\}_{i=1}^n$ forms an asymptotically regular sequence [11] in the sense that

$$\int_0^{t_i} p(t)\, dt = \frac{i-1}{n-1}, \tag{1.3}$$

where $p(\cdot)$ is a positive density function on [0, 1]. This setup is more for preference and ease of presentation than necessary. For example, Robinson [8] allowed both $\mathbf{x}_i$ and t_i to be random, and allowed dependence of $\sigma(\cdot)$ on both $\mathbf{x}_i$ and t_i. In principle, the results we shall establish in this paper can be translated to the setup of Robinson [8].

The plan of this paper is as below. In Section 2 we develop a test of heteroscedasticity under model (1.2), and a two-step estimator of $\sigma(\cdot)$ when it is not constant. In Section 3, we construct semiparametric generalized least-squares estimators (SGLSEs) of $\boldsymbol{\beta}$ and $g(\cdot)$, respectively. In Section 4 we develop a bootstrap goodness of fit test to see whether $g(\cdot)$ can be parametrized. In Section 5 we study the finite sample behaviors of our proposed tests and estimators, and illustrate their use with a real data set. In Section 6, we offer some discussions. The technical details are given in the Appendix.

2. A test of heteroscedasticity

We adopt the test in Dette [2] to test

$$H_0: \quad \sigma(\cdot) = \sigma \quad \text{versus} \quad H_1: \quad \sigma(\cdot) \neq \sigma \tag{2.1}$$

for model (1.2), where $\sigma > 0$ is an unknown constant. Dette's test was constructed for model (1.2) with $\boldsymbol{\beta} = 0$. In order to adopt Dette's test, we proceed as below. From (1.3), the spacing $t_{i+1} - t_i$ is of order $O(1/n)$, so $g(t_{i+1}) - g(t_i) = O(1/n)$ holds for $i = 1, \ldots, n-1$ under

Assumption 2 in the Appendix. Then, we have

$$\begin{aligned} y_{i+1} - y_i &= g(t_{i+1}) - g(t_i) + (x_{1i+1} - x_{1i})\beta_1 + \cdots + (x_{pi+1} - x_{pi})\beta_p \\ &\quad + \sigma(t_{i+1})\varepsilon_{i+1} - \sigma(t_i)\varepsilon_i \\ &\approx (x_{1i+1} - x_{1i})\beta_1 + \cdots + (x_{pi+1} - x_{pi})\beta_p + \varepsilon_i^\star, \end{aligned} \tag{2.2}$$

where $\varepsilon_i^\star = \sigma(t_{i+1})\varepsilon_{i+1} - \sigma(t_i)\varepsilon_i$. Expression (2.2) allows us to estimate $\boldsymbol{\beta}$ with

$$\tilde{\boldsymbol{\beta}}_n = \left(\sum_{i=1}^{n-1}(x_{i+1} - x_i)(x_{i+1} - x_i)^\tau\right)^{-1} \sum_{i=1}^{n-1}(x_{i+1} - x_i)(y_{i+1} - y_i).$$

Now we define $y_i^\star = y_i - \mathbf{x}_i^\tau \tilde{\boldsymbol{\beta}}_n$, $i = 1, \ldots, n$, and construct our test statistic by applying Dette's procedure to $\{y_i^\star\}$. Specifically, for a fixed positive integer $v < n$, using a vth order difference sequence $\{\omega_j^{(1)} : j = 1, \ldots, v\}$ that satisfies $\sum_{j=0}^{v} \omega_j^{(1)} = 0$ and $\sum_{j=0}^{v}(\omega_j^{(1)})^2 = 1$, such as $(1, -1)/\sqrt{2}$ for $v = 1$ and $(1, -2, 1)/\sqrt{6}$ for $v = 2$ (see [2] for more details), we define our test statistic $T_n^\star$ by

$$T_n^\star = \frac{1}{(n-v)(n-v-1)h_1} \sum_{|i-j| \geqslant v+1} K\left(\frac{t_i - t_j}{h_1}\right)(R_{n,v}^{\star 2} - \bar{R}_{i,v}^{\star 2})(R_{j,v}^{\star 2} - \bar{R}_{j,v}^{\star 2}),$$

where $K(\cdot)$ is a kernel defined on $[-1, 1]$, h_1 is the bandwidth,

$$R_{i,v}^{\star 2} = \left(\sum_{j=0}^{v} \omega_j^{(1)} y_{i-j}^\star\right)^2, \quad i = v+1, \ldots, n, \qquad \bar{R}_{n,v}^{\star 2} = \frac{1}{n-v} \sum_{i=v+1}^{n} R_{i,v}^{\star 2}.$$

Let

$$\nabla(\cdot) = \sigma^2(\cdot) - \int_0^1 \sigma^2(t)p(t)\,dt, \qquad \overline{\nabla p} = \int_0^1 \nabla(t)p(t)\,dt.$$

We have the following theorem to support our test.

Theorem 2.1. *If Assumptions* 1–4 *in the Appendix hold*, *then we have*:

(i) *Under the null hypothesis* H_0,

$$n\sqrt{h}T_n^\star \to_{\mathrm{D}} N(0, \zeta_0^2(v)) \quad \text{as } n \to \infty,$$

where

$$\zeta_0^2(v) = 2\sigma^8 \int_{-1}^1 K^2(t)\,dt \int_0^1 (\sigma^4(t)E\varepsilon_1^4 - 1 + 4\delta_v)^2 p^2(t)\,dt,$$

$$\delta_v = \sum_{k=1}^{v}\left(\sum_{j=0}^{v-k} \omega_j^{(1)}\omega_{j+k}^{(1)}\right)^2.$$

(ii) *Under the alternative hypothesis* H_1,

$$\sqrt{n}\left\{T_n^\star - \frac{1}{h}\int_0^1\int_0^1 K\left(\frac{t-t^*}{h}\right)\nabla(t)\nabla(t^*)p(t)p(t^*)\,dt\,dt^*\right\}$$
$$\to_{\mathrm{D}} N(0, \zeta_1^2(v)) \quad \text{as } n \to \infty,$$

where the asymptotic variance is given by

$$\zeta_1^2(v) = 4\int_0^1 (\sigma^4(t)E\varepsilon_1^4 - 1 + 4\delta_v)\sigma^4(t)((\nabla p)(t) - \overline{\nabla p})^2 p(t)\,dt.$$

According to Dette [2], we have a consistent test if we reject the null hypothesis H_0 when $n\sqrt{h}T_n^\star > u_{1-\alpha}\widehat{\zeta}_0(v)$, where $u_{1-\alpha}$ denotes the $(1-\alpha)$ quantile of the standard normal distribution and $\widehat{\zeta}_0^2(v)$ is a consistent estimator of $\zeta_0^2(v)$. As in Dette [2], one such $\widehat{\zeta}_0^2(v)$ is

$$\widehat{\zeta}_0^2(v)=2\left\{\frac{2}{n-3}\sum_{i=2}^{n-2} R_{i,v}^{\star 4}R_{i+2,v}^{\star 4}-\frac{12}{n-5}\sum_{i=2}^{n-4} R_{i,v}^{\star 4}R_{i+2,v}^{\star 2}R_{i+4,v}^{\star 2}+9(\bar{R}_{n,v}^{\star 2})^4\right\}\int_{-1}^1 K^2(t)\,dt.$$

If heteroscedasticity is present, in principle we can construct a consistent estimator of $\sigma^2(\cdot)$ from the residuals after fitting model (1.2) to the data. However, we here choose to apply the two-step method of Müller and Stadtmüller [7] to estimate $\sigma^2(\cdot)$ based on $\{y_i^\star = y_i - \mathbf{x}_i^\tau\tilde{\boldsymbol{\beta}}_n\}_{i=1}^n$ to avoid estimating the nonparametric component $g(\cdot)$. The first step is to construct some initial estimators $\tilde{\sigma}^2(\cdot)$ and the second step is to improve these initial estimators through nonparametric smoothing.

Let t_{v_0} be an interior point of [0, 1]. We follow Müller and Stadtmüller [7] to construct a local variance estimator at t_{v_0} to be

$$\tilde{\sigma}^2(t_{v0}) = \left(\sum_{j=j_1}^{j_2}\omega_j^{(2)}(y_{j+v_0} - \mathbf{x}_{j+v_0}^\tau\tilde{\boldsymbol{\beta}}_n)\right)^2,$$

where $m \geqslant 2$ is a fixed integer, $j_1 = -[m/2]$, $j_2 = [m/2 - \frac{1}{4}]$, $[a]$ denotes the largest integer $\leqslant a$ and the $\omega_j^{(2)}$ satisfy

$$\sum_{j=j_1}^{j_2}\omega_j^{(2)} = 0 \quad \text{and} \quad \sum_{j=j_1}^{j_2}(\omega_j^{(2)})^2 = 1.$$

Two popular choices for $\omega_j^{(2)}$ are again $(1, -1)/\sqrt{2}$ for $m = 2$ and $(1, -2, 1)/\sqrt{6}$ for $m = 3$.

Of the many nonparametric smoothing methods available, we choose to use the local linear smoother of Chiou and Müller [12], which has the form

$$\widehat{\sigma}_n^2(t) = \sum_{i=1}^n W_{h_2 i}(t)\tilde{\sigma}^2(t_i), \tag{2.3}$$

where the weight functions $W_{h_2 i}(\cdot)$ have the following explicit form (Fan [13]; Fan and Gijbels [14]):

$$W_{h_2 i}(\cdot; t_1, \ldots, t_n) = \frac{(nh_2)^{-1}K(h_2^{-1}(t_i - \cdot))\{A_{n,2}(\cdot) - (t_i - \cdot)A_{n,1}(\cdot)\}}{A_{n,0}(\cdot)A_{n,2}(\cdot) - A_{n,1}^2(\cdot)},$$

with

$$A_{n,j}(\cdot) = \frac{1}{nh_2}\sum_{i=1}^n K\left(\frac{t_i - \cdot}{h_2}\right)(t_i - \cdot)^j, \quad j = 0, 1, 2,$$

and h_2 as the bandwidth.

Theorem 2.2. *If Assumptions* 1–3 *and* 5 *in the Appendix hold*, *then we have*

$$\sup_{0\leqslant t\leqslant 1} |\widehat{\sigma}_n^2(t) - \sigma^2(t)| = O_p\left[h_2^2 + (\log n/(nh_2))^{1/2}\right].$$

The bandwidth h_2 can be chosen by cross-validation suitably modified for leave-out $(m-1)$ data points. See Müller and Stadtmüller [7] for details.

3. Semiparametric generalized least-squares estimation

Without considering heteroscedasticity, a *semiparametric least-squares estimator* (SLSE) of $\boldsymbol{\beta}$ can be constructed as below. Assume that $\{\mathbf{x}_i^\tau, t_i, y_i; i = 1, \ldots, n\}$ satisfy model (1.2). If $\boldsymbol{\beta}$ is known to be the true parameter, then by $E\varepsilon_i = 0$ we have $g(t_i) = E(y_i - \mathbf{x}_i^\tau\boldsymbol{\beta})$ for $i = 1, \ldots, n$. Hence, a natural nonparametric estimator of $g(\cdot)$ given $\boldsymbol{\beta}$ is

$$\tilde{g}(t, \boldsymbol{\beta}) = \sum_{i=1}^n W_{h_3 i}(t)(y_i - \mathbf{x}_i^\tau\boldsymbol{\beta}),$$

where the weight functions $W_{h_3 i}(\cdot)$ have the same form as $W_{h_2 i}(\cdot)$ in Section 2 except that the bandwidth h_2 is replaced by h_3. To estimate $\boldsymbol{\beta}$, we seek to minimize

$$SS(\boldsymbol{\beta}) = \sum_{i=1}^n \left[y_i - \mathbf{x}_i^\tau\boldsymbol{\beta} - \tilde{g}(t_i, \boldsymbol{\beta})\right]^2. \tag{3.1}$$

The minimizer to (3.1) is found as

$$\widehat{\boldsymbol{\beta}}_n = (\widehat{X}^\tau\widehat{X})^{-1}\widehat{X}^\tau\widehat{\mathbf{y}},$$

where $\widehat{\mathbf{y}} = (\widehat{y}_1, \ldots, \widehat{y}_n)^\tau$, $\widehat{X} = (\widehat{\mathbf{x}}_1, \ldots, \widehat{\mathbf{x}}_n)^\tau$, $\widehat{y}_i = y_i - \sum_{j=1}^n W_{h_3 j}(t_i)y_j$ and $\widehat{\mathbf{x}}_i = \mathbf{x}_i - \sum_{j=1}^n W_{h_3 j}(t_i)\mathbf{x}_j$ for $i = 1, \ldots, n$.

To improve upon $\widehat{\boldsymbol{\beta}}_n$, we construct an SGLSE of $\boldsymbol{\beta}$ by taking the heteroscedasticity of model (1.2) into consideration. Let $\widehat{\Sigma} = \mathrm{diag}(\widehat{\sigma}^2(t_1), \ldots, \widehat{\sigma}^2(t_n))$. Our SGLSE of $\boldsymbol{\beta}$ is defined as

$$\widehat{\boldsymbol{\beta}}_n^w = (\widehat{X}^\tau\widehat{\Sigma}^{-1}\widehat{X})^{-1}\widehat{X}^\tau\widehat{\Sigma}^{-1}\widehat{\mathbf{y}}.$$

With $\boldsymbol{\beta}$ estimated, our SGLSE estimator of the nonparametric component $g(\cdot)$ is

$$\widehat{g}_n^w(t) = \sum_{i=1}^n W_{h_3 i}(t)(y_i - \mathbf{x}_i^\tau\widehat{\boldsymbol{\beta}}_n^w).$$

Theorem 3.1. *Suppose that Assumptions* 1–3, 5 *and* 6 *in the Appendix hold*. *Then we have*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n^w - \boldsymbol{\beta}) \to_{\mathrm{D}} N(0, V^{-1}) \quad \textit{as } n \to \infty,$$

where $V = \lim_{n\to\infty} \frac{1}{n} U^\tau\Sigma U$ *with* $\Sigma = \mathrm{diag}(\sigma^2(t_1), \ldots, \sigma^2(t_n))$ *and* $U = (\mathbf{u}_1, \ldots, \mathbf{u}_n)^\tau$ *is defined in Assumption* 1.

Similar to Theorem 2.2.1 of You [10] the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ is

$$V_0^{-1} = \left(\lim_{n\to\infty} n^{-1}U^\tau U\right)^{-1}\left(\lim_{n\to\infty} n^{-1}U^\tau\Sigma U\right)\left(\lim_{n\to\infty} n^{-1}U^\tau U\right)^{-1}.$$

Since $\Sigma^{-1/2}U\left(U^\tau\Sigma^{-1}U\right)^{-1}U^\tau\Sigma^{-1/2}$ is an idempotent matrix with rank p, we have

$$V_0^{-1} - V^{-1} = \lim_{n\to\infty} n\left\{(U^\tau U)^{-1}U^\tau\Sigma^{1/2}\left(I - \Sigma^{-1/2}U\left(U^\tau\Sigma^{-1}U\right)^{-1}U^\tau\Sigma^{-1/2}\right)\right.$$
$$\left.\times\,\Sigma^{1/2}U(U^\tau U)^{-1}\right\} \geqslant 0.$$

This implies that $\widehat{\boldsymbol{\beta}}_n^w$ is asymptotically more efficient than $\widehat{\boldsymbol{\beta}}_n$ in terms of asymptotic covariance matrix.

To make statistical inference based on $\widehat{\boldsymbol{\beta}}_n^w$, a consistent estimator of the asymptotic covariance matrix V^{-1} is needed. Define

$$\widehat{V}_n = \frac{1}{n}\widehat{X}^\tau \mathrm{diag}(\widehat{\sigma}_n^2(t_1), \ldots, \widehat{\sigma}_n^2(t_n))^{-1}\widehat{X}.$$

Theorem 3.2. *Suppose that Assumptions* 1–3, 5 *and* 6 *in the Appendix hold. Then* $\widehat{V}_n$ *is a consistent estimator of* V, *namely*, $\widehat{V}_n - V \to_p 0$ *as* $n \to \infty$.

Theorems 3.1 and 3.2 can be used to construct tests and confidence intervals for the parametric component $\boldsymbol{\beta}$. For the SGLSE estimator $\widehat{g}_n^w(\cdot)$ of the nonparametric component $g(\cdot)$, we have the following result.

Theorem 3.3. *Suppose that Assumptions* 1–3, 5 *and* 6 *in the Appendix hold. Then we have*

$$\sqrt{nh_3}\left[\widehat{g}_n^w(t_0) - g(t_0) - \frac{h_3^2}{2}\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2}g''(t_0)\right] \to_{\mathrm{D}} N(0, \zeta(t_0)) \quad \text{as } n \to \infty$$

provided that $p(t_0) \neq 0$, *where*

$$\mu_j = \int_{-1}^{1} t^j K(t)\,dt, \quad v_j = \int_{-1}^{1} t^j K^2(t)\,dt$$

and

$$\zeta(t_0) = \frac{\sigma^2(t_0)(\alpha_0^2 v_0 + 2\alpha_0\alpha_1 v_1 + \alpha_1^2 v_2)}{p(t_0)} \quad \text{with } \alpha_0 = \mu_2/(\mu_2 - \mu_1^2),\ \alpha_1 = -\mu_1/(\mu_2 - \mu_1^2).$$

Bandwidth selection is important to estimate $g(\cdot)$ and much less so to estimate $\boldsymbol{\beta}$. From the difference estimate $\widetilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, we can rewrite model (1.2) approximately as

$$Y_i - \mathbf{x}_i^\tau\widetilde{\boldsymbol{\beta}} = g(t_i) + \varepsilon_i^\star.$$

Therefore, we can get a good starting value for h_3 by applying the usual bandwidth selection methods, such as the pre-asymptotic substitution method, the cross-validation method or the plug in bandwidth selector.

4. Bootstrap goodness of fit test

The nonparametric component $g(\cdot)$ in model (1.2) is meant for capturing any unknown relationship between y_i and t_i. When evidence suggests that such a relationship exists in initial modeling, it is worth the effort to identify this relationship so that model (1.2) can be improved. Here, we extend the generalized likelihood technique in Fan et al. [5] to model (1.2).

Consider the null hypothesis

$$H_0: \quad g(t) = \alpha(t, \boldsymbol{\theta}), \tag{4.1}$$

where $\alpha(\cdot, \boldsymbol{\theta})$ is a given family of functions indexed by an unknown parameter vector $\boldsymbol{\theta}$. Let $\widehat{\boldsymbol{\theta}}_n$ be an estimator of $\boldsymbol{\theta}$. The weighted residual sum of squares under the null hypothesis is

$$\mathrm{RSS}_0 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^{\tau}\widehat{\boldsymbol{\beta}}_n - \alpha(t_i, \widehat{\boldsymbol{\theta}}_n))^2 w(t_i),$$

where $w(t)$ is a weight function. Analogously, the weighted residual sum of squares corresponding to model (1.2) is

$$\mathrm{RSS}_1 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^{\tau}\widehat{\boldsymbol{\beta}}_n - \widehat{g}_n(t_i))^2 w(t_i),$$

where $\widehat{g}_n(t_i) = \sum_{j=1}^{n} W_{h_3 j}(t_i)(y_j - \mathbf{x}_i^{\tau}\widehat{\boldsymbol{\beta}}_n)$. Our test statistic is a quasi-likelihood ratio statistic that is defined as

$$Q_n = \frac{n}{2}\left(\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}\right) = \frac{n}{2}\left(\frac{\mathrm{RSS}_0}{\mathrm{RSS}_1} - 1\right) \approx \frac{n}{2}\log\frac{\mathrm{RSS}_0}{\mathrm{RSS}_1} \tag{4.2}$$

and we reject the null hypothesis (4.1) for large values of Q_n. If the weight function $w(t)$ is continuous with a compact support contained in [0, 1], then it is easy to show by similar argument to that used in Fan et al. [5] that the distribution of $a_K Q_n$ can be approximated by $\chi^2_{b_n}$, where r_K and b_n are defined by

$$a_K = r_K \int_0^1 \sigma^2(t)p(t)\,dt \int_0^1 \sigma^2(t)w(t)\,dt \left[\int_0^1 \sigma^4(t)w(t)\,dt\right]^{-1},$$

$$b_n = r_K c_K h^{-1} \int_0^1 \sigma^2(t)w(t)\,dt \left[\int_0^1 \sigma^4(t)w(t)\,dt\right]^{-1}$$

and

$$r_K = \frac{K(0) - 2^{-1}\int_{-1}^{1} K^2(t)\,dt}{\int_{-1}^{1}[K(t) - 2^{-1}K * K(t)]^2\,dt}, \qquad c_K = K(0) - 2^{-1}\|K\|_2^2,$$

where $K * K$ denotes the convolution of K. This implies that

$$(2b_n)^{-1/2}(a_k Q_n - b_n) \to_{\mathrm{D}} N(0, 1).$$

An important consequence of the above result is that one does not have to derive the theoretical constants b_n and a_K to be able to use the generalized likelihood ratio test. As long as the above Wilk's type of result holds, one can simply simulate the null distribution of the test statistic by bootstrap sampling.

Because of the heteroscedasticity in model (1.2), we apply the wild bootstrap method of Wu [9] to compute the p-value of our test as shown below:

Step 1: Fitting model (1.2) to compute Q_n and estimate the residuals $\{\sigma(t_i)\varepsilon_i\}_{i=1}^n$ by $\{\widehat{e}_i\}_{i=1}^n$ where

$$\widehat{e}_i = y_i - \mathbf{x}_i^\tau \widehat{\boldsymbol{\beta}}_n - \alpha(t_i, \widehat{\theta}), \quad i = 1, \ldots, n.$$

Step 2: Let η be a random variable with distribution function $F(\cdot)$ such that $E\eta = 0$, $E\eta^2 = 1$ and $E|\eta|^3 < \infty$. We stress that $F(\cdot)$ is chosen independently of model (1.2). Generate a random sample $\{\eta_i\}_{i=1}^n$ from $F(\cdot)$.

Step 3: Define $y_i^* = \mathbf{x}_i^\tau \widehat{\boldsymbol{\beta}}_n + \alpha(t_i, \widehat{\theta}) + \widehat{e}_i \eta_i$ for $i = 1, \ldots, n$, and compute the bootstrap test statistic Q_n^* according to (4.2) based on the sample $\{t_i, \mathbf{x}_i, y_i^*\}$.

Step 4: Repeat Steps 2 and 3 a large number M of times and approximate the p-value of our test with the percentage of times the event $\{Q_n^* \geqslant Q_n\}$ occurs. Reject H_0 in (4.1) at level α if this p-value is less than α.

5. Simulation and application

We carry out a Monte Carlo study in this section to investigate the performance of our proposed test and estimators for finite samples.

Test of heteroscedasticity: Observations are generated from

$$y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + g(t_i) + \sigma(t_i)\varepsilon_i, \quad i = 1, \ldots, n, \tag{5.1}$$

where $\beta_1 = 1.5$, $\beta_2 = 2.0$, $t_i = (i - 0.5)/n$, x_{1i} are i.i.d $N(0, 1)$, x_{2i} are i.i.d χ^2 with 1 degree of freedom, $g(t) = \sin(2\pi t)$ or $1.5t^2/(t^2 + 1)$, ε_i are i.i.d $N(0, 1)$ and $\sigma(t)$ follows two models: (I) $\sigma(t) = 1 + c\cos(2\pi t)$ and (II) $\sigma(t) = 1 + c\sin(4\pi t)$, with c taking values from $\{0, 0.5, 0.8\}$. We generate 5000 samples from each setup (the x_{1i} and x_{2i} values are generated once for each n, c and model combination) and calculate the empirical size and power of our test at 0.025, 0.05 and 0.10 levels. The kernel is $K(t) = \frac{3}{4}(1 - t^2)I\{|t| \leqslant 1\}$, both $\omega_j^{(1)}$ and $\omega_j^{(2)}$ use $(1, -1)/\sqrt{2}$ and bandwidth h_1 is the one in equation (4.4) of Dette [2]. The results are in Table 1.

Table 1
Empirical sizes and powers of our test for heteroscedasticity

		c	$n = 50$			$n = 100$			$n = 200$		
			0.025	0.05	0.10	0.025	0.05	0.10	0.025	0.05	0.10
$g(t) = \sin(2\pi t)$	I	0.0	0.031	0.051	0.090	0.035	0.055	0.099	0.038	0.060	0.102
		0.5	0.167	0.217	0.298	0.322	0.389	0.481	0.534	0.609	0.689
		0.8	0.333	0.398	0.486	0.626	0.695	0.765	0.873	0.905	0.937
	II	0.0	0.028	0.047	0.090	0.033	0.053	0.096	0.037	0.060	0.104
		0.5	0.204	0.259	0.350	0.382	0.456	0.549	0.581	0.654	0.739
		0.8	0.417	0.496	0.595	0.672	0.738	0.810	0.890	0.927	0.956
$g(t) = (1.5t^2)/(t^2 + 1)$	I	0.0	0.031	0.054	0.092	0.033	0.053	0.098	0.034	0.058	0.100
		0.5	0.185	0.236	0.312	0.324	0.392	0.487	0.533	0.617	0.701
		1.0	0.350	0.421	0.520	0.625	0.691	0.766	0.870	0.902	0.932
	II	0.0	0.034	0.052	0.089	0.034	0.057	0.097	0.042	0.063	0.110
		0.5	0.221	0.280	0.369	0.382	0.454	0.545	0.588	0.669	0.749
		1.0	0.419	0.501	0.598	0.695	0.760	0.822	0.902	0.929	0.954

Table 2
Sample means, sample standard deviations (SDs) and sample relative efficiencies (REs) of $\widehat{\boldsymbol{\beta}}_n$, $\widehat{\boldsymbol{\beta}}_n^w$ and $\bar{\boldsymbol{\beta}}_n^w$

		$g(t) = \sin(2\pi t)$					$g(t) = (1.5t^2)/(t^2+1)$				
		I			II		I			II	
		$c = 0$	$c = 0.5$	$c = 0.8$	$c = 0.5$	$c = 0.8$	$c = 0$	$c = 0.5$	$c = 0.8$	$c = 0.5$	$c = 0.8$
$\widehat{\beta}_{1n}$	Mean	1.500	1.496	1.496	1.499	1.498	1.498	1.501	1.502	1.498	1.499
	SD	0.074	0.075	0.080	0.077	0.089	0.073	0.077	0.084	0.076	0.081
	RE	1.000	1.731	5.125	1.671	6.539	1.000	1.719	6.071	1.658	5.266
$\widehat{\beta}_{2n}$	Mean	2.000	2.003	2.000	2.003	2.000	2.001	1.999	2.001	1.998	1.998
	SD	0.053	0.059	0.060	0.058	0.061	0.054	0.056	0.061	0.056	0.062
	RE	1.000	1.567	5.435	1.713	5.722	1.000	1.742	5.559	1.683	5.486
$\widehat{\beta}_{1n}^w$	Mean	1.503	1.498	1.497	1.498	1.497	1.498	1.501	1.500	1.500	1.499
	SD	0.075	0.060	0.037	0.062	0.047	0.073	0.062	0.038	0.062	0.041
	RE	1.021	1.086	1.113	1.089	1.810	1.015	1.113	1.224	1.103	1.376
$\widehat{\beta}_{2n}^w$	Mean	2.001	2.002	2.000	2.002	1.999	2.001	2.004	2.000	1.999	2.000
	SD	0.054	0.048	0.027	0.046	0.032	0.055	0.045	0.029	0.047	0.029
	RE	1.021	1.030	1.113	1.095	1.571	1.032	1.118	1.297	1.164	1.246
$\bar{\beta}_{1n}^w$	Mean	1.500	1.498	1.497	1.499	1.498	1.498	1.500	1.499	1.500	1.499
	SD	0.074	0.057	0.035	0.060	0.035	0.073	0.059	0.034	0.059	0.035
$\bar{\beta}_{2n}^w$	Mean	2.000	2.002	2.000	2.002	2.000	2.001	2.000	2.000	2.000	2.000
	SD	0.0537	0.047	0.025	0.044	0.025	0.054	0.042	0.026	0.043	0.026

We see from Table 1 that even for $n = 50$, the normal approximation to the finite sample distribution of our test statistic is very good, and the power of our test is decent.

Estimation of $\boldsymbol{\beta}$: Observations are generated from (5.1) as above and, during model fitting, the bandwidth h_2 for the estimation of $\sigma(t) = 1 + c\cos(2\pi t)$ or $1 + c\sin(4\pi t)$ and the bandwidth h_3 for the estimation of $g(t) = \sin(2\pi t)$ or $1.5t^2/(t^2+1)$ are selected by grid search, aided by the idea discussed in Section 3.

We compare the performance of the SLSE $\widehat{\boldsymbol{\beta}}_n = (\widehat{\beta}_{1n}, \widehat{\beta}_{2n})^\tau$ with that of the SGLSE $\widehat{\boldsymbol{\beta}}_n^w = (\widehat{\beta}_{1n}^w, \widehat{\beta}_{2n}^w)^\tau$ in terms of sample mean, sample standard deviation (SD) and sample relative efficiency (RE), where the sample RE of an estimator is the ratio of the mean square error of the estimator to the mean square error of the benchmark estimator $\bar{\boldsymbol{\beta}}_n^w = (\bar{\beta}_{1n}^w, \bar{\beta}_{2n}^w)^\tau$, where $\bar{\boldsymbol{\beta}}_n^w$ has the same form as that of $\widehat{\boldsymbol{\beta}}_n^w$ except that the true error variances are used in the weighting. The results for $n = 200$ are given in Table 2. We see from Table 2 that both $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\beta}}_n^w$ are asymptotically unbiased, but $\widehat{\boldsymbol{\beta}}_n^w$ generally has smaller SDs than $\widehat{\boldsymbol{\beta}}_n$ as expected. Also the improvement of $\widehat{\boldsymbol{\beta}}_n^w$ over $\widehat{\boldsymbol{\beta}}_n$ increases as the level of heteroscedasticity becomes high.

The power of the bootstrap goodness of fit test: To demonstrate the power of our bootstrap goodness of fit test, we consider the null hypothesis

$$H_0: \quad g(t_i) = \theta t_i \quad \text{for all } i = 1, \ldots, n,$$

namely a linear trend, against the alternative

$$H_1: \quad g(t_i) \neq \theta t_i \quad \text{for at least one } i.$$

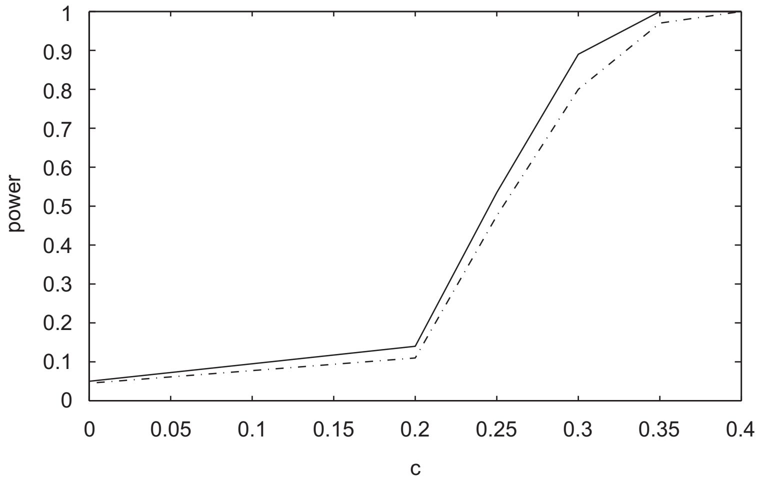

Fig. 1. Power curves of our bootstrap goodness of fit test with sample size $n = 200$. Solid curve is under $\sigma(t) = 1 + 0.5\cos(2\pi t)$; dash-dotted curve is under $\sigma(t) = 1 + 0.8\cos(2\pi t)$.

At 5% significance level, the power function is evaluated under a sequence of alternative models indexed by c, namely, H_1: $g(t_i) = \theta t_i + \sin(c\pi t_i)$, $i = 1, \ldots, n$.

The distribution $F(\cdot)$ we use is a two-point distribution that assigns $\eta_i = -(\sqrt{5}-1)/2$ with probability $(\sqrt{5}+1)/(2\sqrt{5})$ and $\eta_i = (\sqrt{5}+1)/2$ with probability $1 - (\sqrt{5}+1)/(2\sqrt{5})$. We simulate data from (5.1) and for each setup we conduct our bootstrap goodness of fit test with weight function $w(x) = 1$ and $M = 500$ bootstrap samples. Fig. 1 contains some representative results. We see from Fig. 1 that when the null hypothesis holds ($c = 0$) the powers are very close to the significant level 5%. This demonstrates that the bootstrap estimate of the null distribution of the test is approximately correct. When we move away from the null hypothesis ($c > 0$), the power of our test increases to 1 quickly.

An application: We illustrate the use of our proposed tests and estimators by analyzing a data set from the National Survey of Youth (NLSY). NLSY is a widely used panel survey containing a wealth of demographic and labor market information on young males and females in the USA. We here analyze 151 white males from NLSY by the courtesy of Professor Gary Koop of University of Strathclyde, UK. Our dependent variable y_i is the log of hourly wage, our parametric explanatory variables are tenure (weeks on current job) (x_{i1}), ability based on standardized AFQT test score (x_{i2}), father's education (years of schooling) (x_{i3}), indicate variable for urban versus rural residence (x_{i4}) and our nonparametric explanatory variable t_i is education (years of schooling) scaled into [0, 1].

We first check for heteroscedasticity. The p-value of our test is 0.032 and Fig. 2 shows our estimate of $\sigma^2(\cdot)$. It is clear that we need to assume heteroscedasticity. Our SGLSE of $(\beta_1, \beta_2, \beta_3, \beta_4)$ is $(0.069, -0.010, 0.216, -0.072)$ with standard errors (0.028, 0.057, 0.085, 0.227), respectively. Therefore, only tenure and urban versus rural residence are significantly related to log of hourly wage. Fig. 3 plots our SGLSE of $g(\cdot)$, which suggests a quadratic function of t. As a possibility to simplify the model, we conduct our bootstrap goodness of fit test of H_0: $g(t) = \theta_0 + \theta_1 t + \theta_2 t^2$. The p-value from $M = 1000$ bootstrap samples is 0.378, confirming the reasonableness of H_0

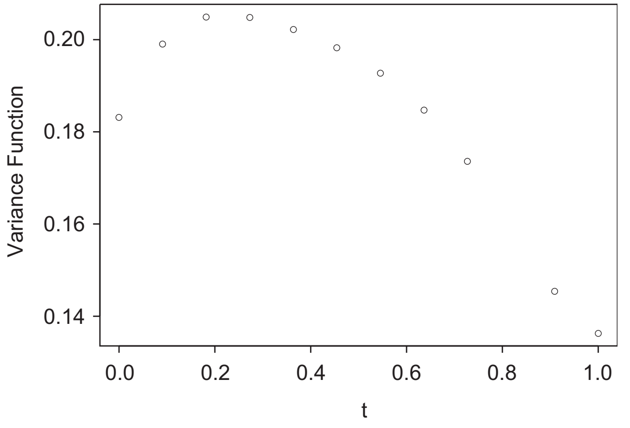

Fig. 2. Estimate $\widehat{\sigma}_{151}(t)$ of the variance function $\sigma(t)$.

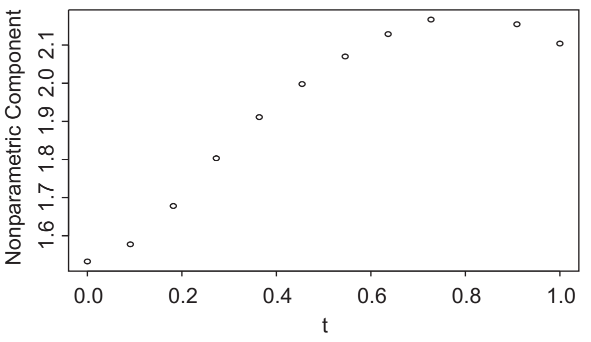

Fig. 3. Estimate $\widehat{g}^{w}_{151}(t)$ of the nonparametric function $g(t)$.

and leading to the simplified linear model (standard errors are in parentheses)

$$\widehat{y_i} = \underset{(0.028)}{0.073x_{i1}} + \underset{(0.085)}{0.221x_{i4}} + \underset{(0.208)}{1.382} + \underset{(0.762)}{1.802t} - \underset{(0.474).}{1.060t^2}$$

6. Discussions

In this paper we have developed a procedure to conduct statistical inference on a partially linear regression model with independent but heteroscedastic errors. The large sample test of heteroscedasticity is shown to perform well, the SGLSE of $\boldsymbol{\beta}$ significantly improves the usual SLSE and the bootstrap goodness of fit test to identify potential relationships captured by the nonparametric component of the model turns out to be powerful.

The procedure we have developed consists of adaptations and/or extensions of the existing methodologies. Our goal is to offer the practitioners a user-friendly procedure. For this, we avoid estimating the nonparametric component of the model whenever we can, and we bootstrap when the normal approximation is complicated.

In times series or cross-sectional studies, heteroscedasticity and autocorrelation may stem from a common cause and thus occur simultaneously. For example, the effects of omitted variables may give rise to both occurrences. When this is the case, the procedure developed in this paper may need modifications to be valid. We shall address this issue in a future study.

Appendix A. Assumptions and proofs of the main results

The following assumptions are needed in this paper.

Assumption 1. There exist some bounded functions $h_j(\cdot)$, $j = 1, \ldots, p$, over [0, 1] such that

$$x_{ij} = h_j(t_i) + u_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, p, \tag{a}$$

where $\mathbf{u}_i = (u_{i1}, \ldots, u_{ip})^\tau$ are real sequences satisfying

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{u}_i \mathbf{u}_i^\tau = B \tag{b}$$

and

$$\max_{1\leqslant j\leqslant n} \left\| \sum_{i=1}^{n} W_{h_3 i}(t_j)\mathbf{u}_i \right\| = o(\delta_n), \tag{c}$$

where B is a $p \times p$ positive definite matrix, $\|\cdot\|$ denotes the Euclidean norm and δ_n satisfies $O(\delta_n) \cdot O[h_3^2 + (\log n/(nh_3))^{1/2}] = o(n^{-1/2})$.

Assumption 2. The functions $g(\cdot)$, $\sigma(\cdot)$ and $h_j(\cdot)$, $j = 1, \ldots, p$, have the continuous second derivatives on [0, 1].

Assumption 3. The kernel $K(\cdot)$ is a density function defined on $[-1, 1]$ with zero mean and finite variance, and is Lipschitz continuous of order $\geqslant 0.5$.

Assumption 4. The bandwidth h_1 satisfies $n^{1/2}h_1^8 \to 0$ and $nh_1^2/(\log n)^2 \to \infty$ as $n \to \infty$.

Assumption 5. The bandwidth h_2 satisfies $n^{1/2}h_2^8 \to 0$ and $nh_2^2/(\log n)^2 \to \infty$ as $n \to \infty$.

Assumption 6. The bandwidth h_3 satisfies $n^{1/2}h_3^8 \to 0$ and $nh_3^2/(\log n)^2 \to \infty$ as $n \to \infty$. Moreover,

$$O[h_2^2 + (\log n/(nh_2))^{1/2}] \cdot O[h_3^2 + (\log n/(nh_3))^{1/2}] = o(n^{-1/2}).$$

In order to prove the main results in the paper, we first present several lemmas. Lemma A.1 is a standard result in nonparametric regression estimation.

Lemma A.1. *Suppose that Assumptions* 2, 3, 5 *and* 6 *hold. Then we have*

$$\sup_{0\leqslant t\leqslant 1} \left| \sigma^2(t) - \sum_{i=1}^{n} W_{h_2 i}(t)\sigma^2(t) \right| = O\left[h_2^2 + (\log n/(nh_2))^{1/2}\right]$$

and

$$\sup_{0\leqslant t\leqslant 1}\left|g(t)-\sum_{i=1}^{n} W_{h_3 i}(t)g(t)\right| = O\left[h_3^2+(\log n/(nh_3))^{1/2}\right].$$

The following lemma is from Shi and Lau [15].

Lemma A.2. *Let* $\{\varepsilon_i\}_{i=1}^n$ *be an independent random variable sequence with mean zero and* $\max_{1\leqslant i\leqslant n} E|\varepsilon_i|^s < c < \infty$ *for some* $s > 1$, *where* c *is a constant. Let* $\{a_{nij}, 1\leqslant i, j\leqslant n\}$ *be a series of real numbers such that* $\max_{1\leqslant j\leqslant n} \sum_{i=1}^n |a_{nij}|\leqslant c_0 < \infty$ *where* c_0 *is a constant. Then we have*

$$\max_{1\leqslant i\leqslant n}\left|\sum_{j=1}^{n} a_{nij}\varepsilon_j\right| = O_p\left(p_n \log n\right),$$

where $p_n = \min(n^{1/s}d_n, d_n^{1/2})$ *and* $d_n = \max_{1\leqslant i,j\leqslant n} |a_{nij}|$.

Lemma A.3. *Suppose that Assumptions* 1–3, 5 *and* 6 *hold. Then we have*

$$\sqrt{n}\left(\bar{\boldsymbol{\beta}}_n^w-\boldsymbol{\beta}\right)\to_{\mathrm{D}} N(0, V^{-1}) \quad as\ \ n\to\infty,$$

where $\bar{\boldsymbol{\beta}}_n^w = (\widehat{X}^\tau\Sigma^{-1}\widehat{X})\widehat{X}^\tau\Sigma^{-1}\widehat{\mathbf{y}}$.

A proof of Lemma A.3 can be found in You [10].

Lemma A.4. *Suppose that Assumptions* 1–3, 5 *and* 6 *hold. Then, with probability tending to* 1, *for any given* $\boldsymbol{\beta}_1^\star$ *satisfying* $\|\boldsymbol{\beta}_1^\star-\boldsymbol{\beta}_{10}\| = O_p(n^{-1/2})$ *and any constant* c,

$$\mathcal{L}\{(\boldsymbol{\beta}_1^{\star\tau}, \mathbf{0})^\tau\} = \min_{\|\boldsymbol{\beta}_2\|\leqslant cn^{-1/2}} \mathcal{L}\{(\boldsymbol{\beta}_1^{\star\tau}, \boldsymbol{\beta}_2^\tau)^\tau\}.$$

A proof of Lemma A.4 follows the arguments in proving Lemma 4.4 of Fan and Li [4]. We here omit the detail.

Proof of Theorem 2.1. For simplicity, we prove the theorem when $v = 1$. The case where $v > 1$ can be proved similarly. Let $R_{i,2} = (y_i - \mathbf{x}_i^\tau\boldsymbol{\beta} - y_{i-1} + \mathbf{x}_{i-1}^\tau\boldsymbol{\beta})$ for $i = 2,\ldots,n$ and $\bar{R}_2^2 = \sum_{i=2}^n R_{i,2}^2/(n-1)$. Then we have

$$R_{i,2}^\star = R_{i,2}^2 + \frac{1}{2}\left[(\mathbf{x}_i-\mathbf{x}_{i-1})^\tau(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n)\right]^2 + (g(t_i)-g(t_{i-1})+\varepsilon_i-\varepsilon_{i-1})\left[(\mathbf{x}_i-\mathbf{x}_{i-1})^\tau(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n)\right]$$

and

$$\bar{R}_2^\star = \bar{R}_2^2 + \frac{1}{2(n-1)}\sum_{i=2}^{n}\left[(\mathbf{x}_i-\mathbf{x}_{i-1})^\tau(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n)\right]^2 + \frac{1}{(n-1)}\sum_{i=2}^{n}(g(t_i)-g(t_{i-1})+\varepsilon_i-\varepsilon_{i-1})\left[(\mathbf{x}_i-\mathbf{x}_{i-1})^\tau(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n)\right].$$

Define

$$S_i = \tfrac{1}{2}\left[(\mathbf{x}_i - \mathbf{x}_{i-1})^\tau(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_n)\right]^2,$$
$$Q_i = (g(t_i) - g(t_{i-1}) + \varepsilon_i - \varepsilon_{i-1})\left[(\mathbf{x}_i - \mathbf{x}_{i-1})^\tau(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_n)\right],$$
$$\bar{S} = \frac{1}{2(n-1)}\sum_{i=2}^{n}\left[(\mathbf{x}_i - \mathbf{x}_{i-1})^\tau(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_n)\right]^2$$

and

$$\bar{Q} = \frac{1}{(n-1)}\sum_{i=2}^{n}(g(t_i) - g(t_{i-1}) + \varepsilon_i - \varepsilon_{i-1})(\mathbf{x}_i - \mathbf{x}_{i-1})^\tau(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_n).$$

Then

$$\begin{aligned}
T_n^\star &= \frac{1}{(n-1)(n-2)h_1}\sum_{|i-j|\geqslant 2} K\left(\frac{t_i - t_j}{h_1}\right)(R_{i,2}^2 - \bar{R}_2^2 + S_i + Q_i - \bar{S}_i - \bar{Q}_i) \\
&\quad \cdot (R_{j,2}^2 - \bar{R}_2^2 + S_j + Q_j - \bar{S}_j - \bar{Q}_j) \\
&= T_n + \frac{1}{(n-1)(n-2)h_1}\sum_{|i-j|\geqslant 2} K\left(\frac{t_i - t_j}{h_1}\right)(S_i + Q_i - \bar{S}_i - \bar{Q}_i)(S_j + Q_j - \bar{S}_j - \bar{Q}_j) \\
&\quad + \frac{1}{(n-1)(n-2)h_1}\sum_{|i-j|\geqslant 2} K\left(\frac{t_i - t_j}{h_1}\right)(R_{i,2}^2 - \bar{R}_2^2)(S_j + Q_j - \bar{S}_j - \bar{Q}_j) \\
&\quad + \frac{1}{(n-1)(n-2)h_1}\sum_{|i-j|\geqslant 2} K\left(\frac{t_i - t_j}{h_1}\right)(S_i + Q_i - \bar{S}_i - \bar{Q}_i)(R_{j,2}^2 - \bar{R}_2^2) \\
&= T_n + J_1 + J_2 + J_3 \quad \text{say.}
\end{aligned}$$

From Theorem 3.4 in Dette [2] we just need to show that $J_i = o(n^{-1}h_1^{-1/2})$. By the root-$n$ consistency of $\tilde{\boldsymbol{\beta}}_n$ it is easy to see that

$$\frac{1}{(n-1)(n-2)h_1}\sum_{|i-j|\geqslant 2} K\left(\frac{t_i - t_j}{h_1}\right) S_i S_j = o(n^{-1}h_1^{-1/2}).$$

Moreover,

$$\begin{aligned}
&\frac{1}{(n-1)(n-2)h_1}\sum_{|i-j|\geqslant 2} K\left(\frac{t_i - t_j}{h_1}\right) Q_i Q_j \\
&= \sum_{s_1=1}^{p}\sum_{s_2=1}^{p}(\beta_{s_1} - \tilde{\beta}_{ns_1})(\beta_{s_2} - \tilde{\beta}_{ns_2})\frac{1}{(n-1)(n-2)h_1} \\
&\quad \cdot \sum_{|i-j|\geqslant 2} K\left(\frac{t_i - t_j}{h_1}\right)(g(t_i) - g(t_{i-1}) + \varepsilon_i - \varepsilon_{i-1})(x_{is_1} - x_{i-1,s-1}) \\
&\quad \cdot (g(t_j) - g(t_{j-1}) + \varepsilon_j - \varepsilon_{j-1}) \cdot (x_{js_2} - x_{j-1,s-2}) \\
&= \sum_{s_1=1}^{p}\sum_{s_2=1}^{p}(\beta_{s_1} - \tilde{\beta}_{ns_1})(\beta_{s_2} - \tilde{\beta}_{ns_2})\frac{1}{(n-1)(n-2)h_1} \times D_{s_1 s_2}.
\end{aligned}$$

By the fact that $g(t_i) - g(t_{i-1}) = O(n^{-1})$ and $x_{js_2} - x_{j-1,s-2} = O(1)$, it holds that

$$\begin{aligned} ED^2_{s_1 s_2} = & \sum_{|i_1-j_1|\geqslant 2} \sum_{|i_2-j_2|\geqslant 2} K\left(\frac{t_{i_1}-t_{j_1}}{h_1}\right) K\left(\frac{t_{i_2}-t_{j_2}}{h_1}\right) \cdot O(1) \\ & \cdot E\left\{\left[\left(O(n^{-1}) + \varepsilon_{i_1} - \varepsilon_{i_1-1}\right)\left(O(n^{-1}) + \varepsilon_{j_1} - \varepsilon_{j_1-1}\right)\right]\right. \\ & \left.\cdot \left[\left(O(n^{-1}) + \varepsilon_{i_2} - \varepsilon_{i_2-1}\right)\left(O(n^{-1}) + \varepsilon_{j_2} - \varepsilon_{j_2-1}\right)\right]\right\} = O(n^2 h_1^2), \end{aligned}$$

So the root-n consistency of $\widehat{\boldsymbol{\beta}}_n$ leads to

$$\frac{1}{(n-1)(n-2)h_1} \sum_{|i-j|\geqslant 2} K\left(\frac{t_i - t_j}{h_1}\right) Q_i Q_j = o(n^{-1} h_1^{-1/2}).$$

This implies that

$$\frac{1}{(n-1)(n-2)h_1} \sum_{|i-j|\geqslant 2} K\left(\frac{t_i - t_j}{h_1}\right) (S_i + Q_i)(S_j + Q_j) = o(n^{-1} h_1^{-1/2}).$$

Similarly, we can show that

$$\frac{1}{(n-1)(n-2)h_1} \sum_{|i-j|\geqslant 2} K\left(\frac{t_i - t_j}{h_1}\right) (S_i + Q_i)(\bar{S}_j + \bar{Q}_j) = o(n^{-1} h_1^{-1/2})$$

and

$$\frac{1}{(n-1)(n-2)h_1} \sum_{|i-j|\geqslant 2} K\left(\frac{t_i - t_j}{h_1}\right) (\bar{S}_i + \bar{Q}_i)(\bar{S}_j + \bar{Q}_j) = o(n^{-1} h_1^{-1/2}).$$

Therefore, $J_1 = o(n^{-1} h_1^{-1/2})$. By the same argument, we can show that

$$J_2 = o(n^{-1} h_1^{-1/2}) \quad \text{and} \quad J_3 = o(n^{-1} h_1^{-1/2}).$$

All together, the proof is complete. □

Proof of Theorem 2.2. From the definition of $\widehat{\sigma}_n^2(\cdot)$ it holds that

$$\begin{aligned} & \widehat{\sigma}_n^2(t_i) \\ & = \sum_{j=1}^n W_{h_2 j}(t_i) \left[\sum_{k=j_1}^{j_2} \omega_k^{(2)} \mathbf{x}_{k+j}^\tau (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_n) + \sum_{k=j_1}^{j_2} \omega_k^{(2)} \sigma(t_{k+j}) \varepsilon_{k+j} + \sum_{k=j_1}^{j_2} \omega_k^{(2)} g(t_{k+j})\right]^2 \\ & = \sum_{j=1}^n W_{h_2 j}(t_i) \left[\sum_{k=j_1}^{j_2} \omega_k^{(2)} \mathbf{x}_{k+j}^\tau (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_n)\right]^2 + \sum_{j=1}^n W_{h_2 j}(t_i) \left(\sum_{k=j_1}^{j_2} \omega_k^{(2)} \sigma(t_{k+j}) \varepsilon_{k+j}\right)^2 \\ & \quad + \sum_{j=1}^n W_{h_2 j}(t_i) \left(\sum_{k=j_1}^{j_2} \omega_k^{(2)} g(t_{k+j})\right)^2 + 2 \sum_{j=1}^n W_{h_2 j}(t_i) \left[\sum_{k=j_1}^{j_2} \omega_k^{(2)} \mathbf{x}_{k+j}^\tau (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_n)\right] \end{aligned}$$

$$\cdot\left(\sum_{k=j_1}^{j_2}\omega_k^{(2)}\varepsilon_{k+j}\right)+2\sum_{j=1}^{n}W_{h_2j}(t_i)\left[\sum_{k=j_1}^{j_2}\omega_k^{(2)}\mathbf{x}_{k+j}^{\tau}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n)\right]\left(\sum_{k=j_1}^{j_2}\omega_k^{(2)}g(t_{k+j})\right)$$
$$+2\sum_{j=1}^{n}W_{h_2j}(t_i)\left(\sum_{k=j_1}^{j_2}\omega_k^{(2)}\sigma(t_{k+j})\varepsilon_{k+j}\right)\left(\sum_{k=j_1}^{j_2}\omega_k^{(2)}g(t_{k+j})\right)=II_1+\cdots+II_6.$$

According to Lemma 8.3 of Chiou and Müller (1999), there is

$$W_{h_2j}(t_i)=\frac{1}{nh_2}K\left(\frac{t_i-t_j}{h_2}\right)\Big/ p(t_i)+O\left(\frac{1}{n}\right).$$

Thus we have $\max_{1\leqslant i,j\leqslant n} W_{h_2j}(t_i)=O(n^{-1}h_2^{-1})$. Therefore, by Assumption 1 it holds that

$$|II_1|\leqslant \max_{1\leqslant i,j\leqslant n} W_{h_2j}(t_i)\cdot O(n)\cdot O\left(n^{-1}\right)=O(h_2^{-1}n^{-1})\quad \text{a.s.}$$

Moreover, II_2 can be decomposed as

$$II_2=\sum_{j=1}^{n}W_{h_2j}(t_i)\sum_{k_1=j_1}^{j_2}\sum_{k_2\neq k-1}\omega_{k_1}^{(2)}\omega_{k_2}^{(2)}\sigma(t_{k_1+j})\sigma(t_{k_2+j})\varepsilon_{k_1+j}\varepsilon_{k_2+j}+\sum_{j=1}^{n}W_{h_2j}(t_i)$$
$$\cdot\sum_{k=j_1}^{j_2}\omega_k^{(2)2}\sigma^2(t_{k+j})(\varepsilon_{k+j}^2-1)+\sum_{j=1}^{n}W_{h_2j}(t_i)\sum_{k=j_1}^{j_2}\omega_k^{(2)2}\sigma^2(t_{k+j})$$
$$=II_{21}+II_{22}+II_{23}\quad \text{say.}$$

Similar to the proof of Lemma A.1 we have $II_{21}=O[(\log n/nh_2)^{1/2}+h_2^2]$ a.s. and $II_{22}=O[(\log n/nh_2)^{1/2}+h_2^2]$ a.s. Further,

$$II_{23}=\sum_{j=1}^{n}W_{h_2j}(t_i)\sigma^2(t_j)+\sum_{j=1}^{n}W_{h_2j}(t_i)\sum_{k=j_1}^{j_2}\omega_k^{(2)2}(\sigma^2(t_{k+j})-\sigma^2(t_j))$$
$$=\sigma^2(t_i)+O(h_2^2)+O(n^{-1}).$$

By the same argument and using the smoothness of $g(\cdot)$ we can show $II_3=O(n^{-1})$. Further, combining Lemma A.1, the root-n consistency of $\tilde{\boldsymbol{\beta}}_n$ and the smoothness of $g(\cdot)$ it holds that $II_4=O_p[(\log n/(n^2h_2))^{1/2}]$ and $II_6=O_p[(\log n/(n^3h_2))^{1/2}]$. By Cauchy–Schwarz inequality we have $II_5=O_p(n^{-3/2})$. Therefore, the proof is complete. □

Proof of Theorem 3.1. By the definition of $\widehat{\boldsymbol{\beta}}_n^w$, we have

$$\widehat{\boldsymbol{\beta}}_n^w-\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}_n^w-\boldsymbol{\beta}+(\widehat{X}^{\tau}\widehat{\Sigma}^{-1}\widehat{X})^{-1}\left[\widehat{X}^{\tau}\widehat{\Sigma}^{-1}\tilde{\mathbf{g}}-\widehat{X}^{\tau}\Sigma^{-1}\tilde{\mathbf{g}}\right]$$
$$+\left[(\widehat{X}^{\tau}\widehat{\Sigma}^{-1}\widehat{X})^{-1}-(\widehat{X}^{\tau}\Sigma^{-1}\widehat{X})^{-1}\right]\widehat{X}^{\tau}\widehat{\Sigma}^{-1}\tilde{\mathbf{g}}$$
$$+(\widehat{X}^{\tau}\widehat{\Sigma}^{-1}\widehat{X})^{-1}\left[\widehat{X}^{\tau}\widehat{\Sigma}^{-1}(\boldsymbol{\varepsilon}-\tilde{\boldsymbol{\varepsilon}})-\widehat{X}^{\tau}\Sigma^{-1}(\boldsymbol{\varepsilon}-\tilde{\boldsymbol{\varepsilon}})\right]$$
$$+\left[(\widehat{X}^{\tau}\widehat{\Sigma}^{-1}\widehat{X})^{-1}-(\widehat{X}^{\tau}\Sigma^{-1}\widehat{X})^{-1}\right]\widehat{X}^{\tau}\widehat{\Sigma}^{-1}(\boldsymbol{\varepsilon}-\tilde{\boldsymbol{\varepsilon}}),$$

where $\tilde{\mathbf{g}}=(\tilde{g}(t_1),\ldots,\tilde{g}(t_n))$ with $\tilde{g}(t_i)=g(t_i)-\sum_{j=1}^{n}W_{h_3j}(t_i)g(t_j)$ and $\tilde{\boldsymbol{\varepsilon}}=(\tilde{\varepsilon}_1,\ldots,\tilde{\varepsilon}_n)$ with $\tilde{\varepsilon}_i=\sum_{j=1}^{n}W_{h_3j}(t_i)\varepsilon_j$. From Lemma A.3 and the following fact

$$(A+aB)^{-1}=A^{-1}-aA^{-1}BA^{-1}+O(a^2)\quad \text{as } a\to 0, \tag{A.1}$$

in order to prove the asymptotic normality for $\widehat{\boldsymbol{\beta}}_n^w$, we only need to show that

$$\frac{1}{n}\left(\widehat{X}^\tau\widehat{\Sigma}^{-1}\widehat{X} - \widehat{X}^\tau\Sigma^{-1}\widehat{X}\right) = O_p\left[(\log n/(nh_2))^{1/2} + h_2^2\right], \tag{A.2}$$

$$\frac{1}{n}\widehat{X}^\tau\widehat{\Sigma}^{-1}\tilde{\mathbf{g}} - \frac{1}{n}\widehat{X}^\tau\Sigma^{-1}\tilde{\mathbf{g}} = o_p\left(n^{-1/2}\right), \tag{A.3}$$

$$\frac{1}{n}\widehat{X}^\tau\widehat{\Sigma}^{-1}(\boldsymbol{\varepsilon} - \tilde{\boldsymbol{\varepsilon}}) - \frac{1}{n}\widehat{X}^\tau\Sigma^{-1}(\boldsymbol{\varepsilon} - \tilde{\boldsymbol{\varepsilon}}) = o_p\left(n^{-1/2}\right), \tag{A.4}$$

$$n(\widehat{X}^\tau\widehat{\Sigma}^{-1}\widehat{X})^{-1} = O_p(1), \quad n(\widehat{X}^\tau\Sigma^{-1}\widehat{X})^{-1} = O_p(1) \tag{A.5}$$

and

$$\frac{1}{n}\widehat{X}^\tau\widehat{\Sigma}^{-1}\tilde{\mathbf{g}} = o_p\left(n^{-1/2}\right), \quad \frac{1}{n}\widehat{X}'\widehat{\Sigma}^{-1}(\boldsymbol{\varepsilon} - \tilde{\boldsymbol{\varepsilon}}) = O_p(n^{-1/2}). \tag{A.6}$$

The absolute value of the (s_1, s_2)th element of $\frac{1}{n}\widehat{X}^\tau\widehat{\Sigma}^{-1}\widehat{X} - \frac{1}{n}\widehat{X}^\tau\Sigma^{-1}\widehat{X}$ is

$$\begin{aligned}
&\frac{1}{n}\left|\sum_{i=1}^n \widehat{x}_{is_1}\widehat{x}_{is_2}(\widehat{\sigma}_n^{-2}(t_i) - \sigma^{-2}(t_i))\right| \\
&\leqslant \frac{1}{n}\sum_{i=1}^n(\widehat{x}_{is_1}^2 + \widehat{x}_{is_2}^2)\widehat{\sigma}_n^{-2}(t_i)\cdot\sigma^{-2}(t_i)\cdot\max_{1\leqslant i\leqslant n}|\widehat{\sigma}_n^2(t_i) - \sigma^2(t_i)| \\
&= O_p\left[(\log n/(nh_2))^{1/2} + h_2^2\right]
\end{aligned}$$

by Assumption 1, Lemmas A.1, A.2 and Theorem 2.2. This implies that (A.2) holds. Moreover, the absolute value of the sth element of $\frac{1}{n}\widehat{X}^\tau\widehat{\Sigma}^{-1}\tilde{\mathbf{g}} - \frac{1}{n}\widehat{X}^\tau\Sigma^{-1}\tilde{\mathbf{g}}$ is

$$\begin{aligned}
&\frac{1}{n}\left|\sum_{i=1}^n \widehat{x}_{si}\tilde{g}(t_i)(\widehat{\sigma}_n^{-2}(t_i) - \sigma^{-2}(t_i))\right| \\
&\leqslant \left\{\max_{1\leqslant i\leqslant n}\max_{1\leqslant s\leqslant p}|\tilde{h}_s(t_i)| + \left[2 + \max_{1\leqslant i\leqslant n}\left(\sum_{j=1}^n W_{h_2j}(t_i) - 1\right)\right]\max_{1\leqslant s\leqslant p}\sqrt{\frac{1}{n}\sum_{i=1}^n u_{si}^2}\right\} \\
&\quad\cdot\max_{1\leqslant i\leqslant n}|\tilde{g}(t_i)|\cdot\max_{1\leqslant i\leqslant n}\left|\widehat{\sigma}_n^{-2}(t_i) - \sigma^{-2}(t_i)\right| \\
&= O_p\left[(\log n/(nh_2))^{1/2} + h_2^2\right]\cdot O_p\left[(\log n/(nh_3))^{1/2} + h_3^2\right],
\end{aligned}$$

where $\tilde{h}_s(t)$ has the same definition as $\tilde{g}(t)$. This implies that (A.3) holds. The sth element of $\frac{1}{n}\widehat{X}^\tau\widehat{\Sigma}^{-1}(\boldsymbol{\varepsilon} - \tilde{\boldsymbol{\varepsilon}}) - \frac{1}{n}\widehat{X}^\tau\Sigma^{-1}(\boldsymbol{\varepsilon} - \tilde{\boldsymbol{\varepsilon}})$ can be decomposed as

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \widehat{x}_{si}\left(\varepsilon_i - \sum_{j=1}^n W_{h_3j}(t_i)\varepsilon_j\right)(\widehat{\sigma}_n^{-2}(t_i) - \sigma^{-2}(t_i)) \\
&= \frac{1}{n}\sum_{i=1}^n \widehat{x}_{si}\left(\varepsilon_i - \sum_{j=1}^n W_{nj}(t_i)\varepsilon_j\right) \\
&\quad\cdot(\widehat{\sigma}_n^2(t_i) - \sigma^2(t_i))\sigma^{-2}(t_i) + \frac{1}{n}\sum_{i=1}^n \widehat{x}_{si}\left(\varepsilon_i - \sum_{j=1}^n W_{h_3j}(t_i)\varepsilon_j\right)(\widehat{\sigma}_n^2(t_i) - \sigma^2(t_i)) \\
&\quad\cdot(\widehat{\sigma}_n^{-2}(t_i) - \sigma^{-2}(t_i))\sigma^{-2}(t_i) = J_1 + J_2 \quad \text{say.}
\end{aligned}$$

Note that

$$
\begin{aligned}
|J_2| &\leqslant O(1)\cdot\left(\max_{1\leqslant i\leqslant n}|\widehat{\sigma}_n^2(t_i)-\sigma^2(t_i)|\right)^2\cdot\left(\varepsilon_i+\sum_{j=1}^n W_{h_3j}(t_i)\varepsilon_j\right)\\
&\quad\cdot\frac{1}{n}\sum_{i=1}^n\left(\tilde{h}_s(t_i)+\sum_{j=1}^n W_{h_3j}(t_i)u_{si}+u_{si}\right)\\
&=O_p\left[(\log n/(nh_2))^{1/2}+h_2^2\right]\cdot O_p(\delta_n).
\end{aligned}
$$

On the other hand, J_1 can be decomposed as

$$
\begin{aligned}
J_1=&\frac{1}{n}\sum_{i=1}^n\widehat{x}_{si}\varepsilon_i(\widehat{\sigma}_n^2(t_i)-\sigma^2(t_i))\sigma^{-2}(t_i)\\
&-\frac{1}{n}\sum_{i=1}^n\widehat{x}_{si}\left(\sum_{j=1}^n W_{nj}(t_i)\varepsilon_j\right)(\widehat{\sigma}_n^2(t_i)-\sigma^2(t_i))\sigma^{-2}(t_i)=J_{11}+J_{12}.
\end{aligned}
$$

By Lemmas A.1, A.2 and Theorem 2.2 it is easy to show that $J_{12}=o_p(n^{-1/2})$. In addition,

$$
\begin{aligned}
J_{11}=&\frac{1}{n}\sum_{i=1}^n\widehat{x}_{is}\varepsilon_i\sigma^{-4}(t_i)\left\{\sum_{j=1}^n W_{h_2j}(t_i)\left[\sum_{k=j_1}^{j_2}\omega_k^{(2)}\mathbf{x}_{k+j}^{\tau}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n)\right]^2\right.\\
&+\left[\sum_{j=1}^n W_{h_2j}(t_i)\left(\sum_{k=j_1}^{j_2}\omega_k^{(2)}\sigma(t_{k+j})\varepsilon_{k+j}\right)^2-\sigma^2(t_i)\right]\\
&+\sum_{j=1}^n W_{h_2j}(t_i)\left(\sum_{s=1}^p\sum_{k=j_1}^{j_2}\omega_k^{(2)}g(t_{k+j})\right)^2\\
&+2\sum_{j=1}^n W_{h_2j}(t_i)\left[\sum_{k=j_1}^{j_2}\omega_k^{(2)}\mathbf{x}_{k+j}^{\tau}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n)\right]\left(\sum_{k=j_1}^{j_2}\omega_k^{(2)}\varepsilon_{k+j}\right)\\
&+2\sum_{j=1}^n W_{h_2j}(t_i)\left[\sum_{k=j_1}^{j_2}\omega_k^{(2)}\mathbf{x}_{k+j}^{\tau}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n)\right]\left(\sum_{k=j_1}^{j_2}\omega_k^{(2)}g(t_{k+j})\right)\\
&\left.+2\sum_{j=1}^n W_{h_2j}(t_i)\left(\sum_{k=j_1}^{j_2}\omega_k^{(2)}\sigma(t_{k+j})\varepsilon_{k+j}\right)\left(\sum_{k=j_1}^{j_2}\omega_k^{(2)}g(t_{k+j})\right)\right\}\\
=&J_{111}+\ldots+J_{116}\quad\text{say.}
\end{aligned}
$$

Using the root-n consistency of $\tilde{\boldsymbol{\beta}}_n$, the smoothness of $g(\cdot)$, Lemmas A.1 and A.2 we can show $J_{11i}=o_p(n^{-1/2})$ for $i=1,3,\ldots,6$. Therefore, in order to complete the proof we just need to show that $J_{112} = o_p(n^{-1/2})$. J_{112} can be decomposed as

$$\begin{aligned} J_{112} &= \frac{1}{n}\sum_{i=1}^{n} \widehat{x}_{is}\varepsilon_i\sigma^{-4}(t_i)\sum_{j=1}^{n} W_{h_2j}(t_i)\sum_{k_1=j_1}^{j_2}\sum_{k_2\neq k_1} \omega_{k_1}^{(2)}\omega_{k_2}^{(2)}\sigma(t_{k_1+j})\sigma(t_{k_2+j})\varepsilon_{k_1+j}\varepsilon_{k_2+j} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} \widehat{x}_{is}\varepsilon_i\sigma^{-4}(t_i)\sum_{j=1}^{n} W_{h_2j}(t_i)\sum_{k=j_1}^{j_2} \omega_k^{(2)2}\sigma^2(t_{k+j})(\varepsilon_k^2 - 1) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} \widehat{x}_{is}\varepsilon_i\sigma^{-4}(t_i)\left(\sum_{j=1}^{n} W_{nj}(t_i)\sum_{k=j_1}^{j_2} \omega_k^{(2)2}\sigma^2(t_{k+j}) - \sigma^2(t_j)\right) \\ &= J_{1121} + J_{1122} + J_{1123}. \end{aligned}$$

Similar to the proof of Lemma A.4 it can be seen that $J_{1121} = o_p(n^{-1/2})$ and $J_{1122} = o_p(n^{-1/2})$. By the smoothness of $\sigma(\cdot)$ it is easy to show $J_{1123} = O_p(n^{-1})$. In summary, the proof of (A.4) is complete. Combining Theorem 2.2 and the proof of Lemma A.3 it is easy to show that (A.5) holds. By the proofs of (A.3) and (A.4) it is easy to show that (A.6) holds. Thus the proof is complete. □

Proof of Theorem 3.2. Applying Lemmas A.1 and A.2, and using the root-n consistency of $\widehat{\boldsymbol{\beta}}_n^w$, it is easy to show that Theorem 3.2 holds. □

References

[1] B.H. Baltagi, Econometric Analysis of Panel Data, Wiley, New York, 1995.
[2] H. Dette, A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. C.R. Rao 80th birthday felicitation volume, part I, J. Statist. Plann. Inference 103 (2002) 311–329.
[3] R.F. Engle, W.J. Granger, J. Rice, A. Weiss, Semi parametric estimates of the relation between weather and electricity sales, J. Amer. Statist. Assoc. 80 (1986) 310–319.
[4] J. Fan, R. Li, New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis, J. Amer. Statist. Assoc. 99 (2004) 710–723.
[5] J. Fan, C. Zhang, J. Zhang, Generalized likelihood ratio statistics and Wilks phenomenon, Ann. Statist. 29 (2001) 153–193.
[6] W. Härdle, H. Liang, J. Gao, Partially Linear Models, Physica-Verlag, Heidelberg, 2000.
[7] H.G. Müller, U. Stadtmüller, Estimation of heteroscedasticity in regression analysis, Ann. Statist. 15 (1987) 610–625.
[8] P. Robinson, Root-N-consistent semiparametric regression, Econometrica 56 (1988) 931–954.
[9] C.F.J. Wu, Jackknife, bootstrap and other resampling methods in regression analysis, Ann. Statist. 14 (1986) 1261–1295.
[10] J. You, Semiparametric regression models with serially correlated and/or heteroscedastic errors, Unpublished Dissertation, University of Regina, Canada, 2002.
[11] J. Sacks, D. Ylvisacker, Design for regression problems with correlated errors III, Ann. Math. Statist. 41 (1970) 2057–2074.
[12] J.M. Chiou, H.G. Müller, Nonparametric quasi-likelihood, Ann. Statist. 27 (1999) 36–64.
[13] J. Fan, Local linear regression smoothers and their minimax efficiency, Ann. Statist. 21 (1993) 196–216.
[14] J. Fan, I. Gijbels, Local Polynomial Modeling and its Applications, Chapman & Hall, New York, 1996.
[15] J. Shi, T.S. Lau, Empirical likelihood for partially linear models, J. Multivariate Anal. 72 (2000) 132–149.