Ann Inst Stat Math (2018) 70:1147–1175
https://doi.org/10.1007/s10463-017-0619-8

CrossMark

A generalized partially linear framework for variance functions

Yixin Fang[1] · Heng Lian[2] · Hua Liang[3]

Received: 10 January 2017 / Revised: 29 May 2017 / Published online: 4 October 2017

Abstract When model the heteroscedasticity in a broad class of partially linear models, we allow the variance function to be a partial linear model as well and the parameters in the variance function to be different from those in the mean function. We develop a two-step estimation procedure, where in the first step some initial estimates of the parameters in both the mean and variance functions are obtained and then in the second step the estimates are updated using the weights calculated based on the initial estimates. The resulting weighted estimators of the linear coefficients in both the mean and variance functions are shown to be asymptotically normal, more efficient than the initial un-weighted estimators, and most efficient in the sense of semiparametric efficiency for some special cases. Simulation experiments are conducted to examine the

Heng Lian's research was supported by City University of Hong Kong Start Up Grant 7200521. Hua Liang's research was partially supported by NSF Grants DMS-1418042 and DMS-1620898, and by Award Number 11529101, made by National Natural Science Foundation of China. The authors thank an associate editor and two referees for their valuable comments and suggestions.

✉ Hua Liang
hliang@gwu.edu

Yixin Fang
yixin.fang@njit.edu

Heng Lian
henglian@cityu.edu.hk

1 Department of Mathematical Sciences, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA

2 Department of Mathematics, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Hong Kong

3 Department of Statistics, The George Washington University, Washington, DC 20052, USA

numerical performance of the proposed procedure, which is also applied to data from an air pollution study in Mexico City.

Keywords Efficiency · Generalized least squares · Generalized partially linear model · Kernel regression · Profiling · Variance function

1 Introduction

The partially linear model (PLM) is an important generalization of the linear model. Since it was proposed by Engle et al. (1986), it has gained a lot of efforts during the past decades (Härdle et al. 2004; Liang et al. 1999; Robinson 1988; Speckman 1988) and become a useful tool in statistical analysis for parsimoniously reflecting nonlinear trend of some continuous covariate. More importantly, it has been applied to explore data in various disciples such as econometrics (Yatchew and No 2001), biomedicine (Hunsberger et al. 2002; Liang et al. 2008; Zeger and Diggle 1994), and environmetrics (Prada-Sánchez et al. 2000).

The model has the flexibility of a nonparametric regression model, while contains a linear combination of coefficients, whose estimators can achieve nice asymptotic normality as it were a pure linear model under mild conditions. Several methods have been proposed to estimate linear coefficients, including profile (Speckman 1988), backfitting (Opsomer and Ruppert 1997), regression spline (Chen 1988), and smoothing spline (Engle et al. 1986; Heckman 1986). The generalized partially linear model (GPLM), a generalization of the PLM, was also well studied (Carroll et al. 1997; Härdle et al. 1998; Wang et al. 2011; Xia and Härdle 2006).

However, most existing work focuses on statistical inferences for the parameters in the mean function and variance function estimation has received much less attention in the literature than it deserves. Although a wealth of work has been done to take heteroscedasticity into account for enhancing the efficiency of estimating the parameters in the mean function, estimating variance function is of independent interest. A simple example is when one derives confidence intervals/bands for the mean function, an appropriate estimator of the variance is needed (Cai and Wang 2008). Alternative examples where the variance function estimation plays an important role include a study of kinetic rate parameters (Box and Hill 1974), quality control (Box and Meyer 1986), and a study of social inequality (Western and Bloome 2009). More recently, Thomas et al. (2012) demonstrated that individual variability in longitudinal measurements for an individual can be predictive of a health outcome, and Teschendorff and Widschwendter (2012) argued that differential variability can be as important as differential means for predicting disease phenotypes in cancer genomics.

In response to these demonstrations of the importance of variance function, flexible and efficient methods for variance function estimation are in demand. Here we give a brief survey on variance function estimation in models related to the GPLM; see Carroll (2003) and Carroll and Ruppert (1988) for comprehensive surveys. Representative work on modeling heteroscedasticity in linear or nonlinear models includes Carroll and Ruppert (1982), Carroll (1982), and Bickel (1978). Along with these, many parametric and nonparametric approaches have been developed (Carroll and Härdle 1989; Fuller and Rao 1978; Hall and Carroll 1989). Recently, Ma et al. (2006) studied the heteroscedastic partially linear model and Ma and Zhu (2012) extended their strategy to the heteroscedastic partially linear single index model. While they emphasized heteroscedasticity, their focus was still on the estimation of the mean function and the variance in their model was assumed to be a function of the mean function.

In this paper, we consider the variance function generalized partially linear model (VFGPLM), a heteroscedastic regression model where the mean function is a partially linear model and the variance function depends upon a generalized partially linear model. Unlike the classical generalized partially linear model, here we do not insist that the variance function depends only upon the mean function. The model under consideration is

$$Y = \mu(\boldsymbol{X}, Z; \boldsymbol{\alpha}_0, m_\mu) + g\{\nu(\boldsymbol{X}, Z; \boldsymbol{\beta}_0, m_\nu)\}\epsilon, \tag{1a}$$

$$\mu(\boldsymbol{X}, Z; \boldsymbol{\alpha}_0, m_\mu) = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\alpha}_0 + m_\mu(Z), \tag{1b}$$

$$\nu(\boldsymbol{X}, Z; \boldsymbol{\beta}_0, m_\nu) = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0 + m_\nu(Z), \tag{1c}$$

where $g(\cdot)$ is a *known* function, while $m_\mu(\cdot)$ and $m_\nu(\cdot)$ are two *unknown* smoothing functions, ϵ is independent of $\boldsymbol{X}$ and Z, $E(\epsilon) = 0$ and $E(|\epsilon|) = 1$. Generally, either $g(\nu) = \nu^2$ or $g(\nu) = \exp(\nu)$.

Lian et al. (2015) studied the variance function partially linear single index model (VFPLSIM), in which the variance function is a function of the sum of linear and single index functions and the parameters in the variance function are allowed to be different from those in the mean function. They developed efficient and practical estimators for the parameters in the mean and variance functions, and weighted the objective function to obtain more efficient estimators for the parameters in the mean function. Although model (1) looks similar to a special version of the model studied in Lian et al. (2015), it is still worth studying in detail for the following reasons. First, model (1) is of its own importance and interest because of the popularity of the PLM. Second, in the mean and variance functions in model (1), the nonparametric and parameter parts involve different predictors, whereas the same set of predictors are involved in both the nonparametric and parametric parts in the model of Lian et al. (2015). Third, through model (1), it is easier to demonstrate the efficiency gain of the second step in our proposed procedure. Fourth, the weighted least squares procedure for the variance function estimation proposed was not discussed in Lian et al. (2015) because of the complexity of their model, while we give a comprehensive discussion in Sect. 3.

We organize the paper as follows. In Sect. 2, we describe the estimation procedures for the variance function generalized partially linear model. In Sect. 3, we present the main theoretical results and their implications. We examine numerical performance of the proposed method in Sect. 4 through simulation studies and analysis of a real dataset. Some discussion is presented in Sect. 5 and all the technical assumptions and proofs of the theoretical results are placed in “Appendix”.

2 Estimation methods

2.1 Outline and notation

The main goal is to develop efficient and practical estimators of the variance function, which is of its own interest and can improve the efficiency of estimating the mean function. Our approach proceeds in two steps.

- *Step 1*: (1) Obtain an initial estimate of the mean function ignoring the heteroscedasticity; (2) Then obtain an initial estimate of the variance function using the absolute residuals from the initial estimate of the mean function.
- *Step 2*: (1) Update the estimate of the mean function using weights based on the initial variance function estimation; (2) Then update the estimate of the variance function using absolute residuals from the updated estimate of the mean function and using weights based on the initial variance function.

For this goal, let $(Y_i, \boldsymbol{X}_i, Z_i)$, $i = 1, \ldots, n$, be independent and identically distributed realizations of $(Y, \boldsymbol{X}, Z)$. Denote the error term as $\varepsilon_i = Y_i - \mu(\boldsymbol{X}_i, Z_i; \boldsymbol{\alpha}_0, m_\mu)$ and its absolute value as $R_i = |Y_i - \mu(\boldsymbol{X}_i, Z_i; \boldsymbol{\alpha}_0, m_\mu)|$. Denoting $G_i = g\{\nu(\boldsymbol{X}_i, Z_i; \boldsymbol{\beta}_0, m_\nu)\}$, because $E(|\epsilon_i|) = 1$, we have $E\{R_i|\boldsymbol{X}_i, Z_i\} = G_i$. Also denote $D_i = I_{(\varepsilon_i>0)} - I_{(\varepsilon_i\le 0)} = \text{sign}(\varepsilon_i)$ and $\delta_i = R_i - G_i$. Let $g^{(1)}$ denote the first derivative of g and define $g^{(1)2} = (g^{(1)})^2$. Consequently, let $G_i^{(1)} = g^{(1)}\{\nu(\boldsymbol{X}_i, Z_i; \boldsymbol{\beta}_0, m_\nu)\}$. Let $\widetilde{\boldsymbol{X}}_i = \boldsymbol{X}_i - E\{\boldsymbol{X}_i|Z_i\}$, $\check{\boldsymbol{X}}_i = \boldsymbol{X}_i - E\{\boldsymbol{X}_i/G_i^2|Z_i\}/E\{1/G_i^2|Z_i\}$, $\breve{\boldsymbol{X}}_i = \boldsymbol{X}_i - E\{G_i^{(1)2}\boldsymbol{X}_i|Z_i\}/E\{G_i^{(1)2}|Z_i\}$, and $\bar{\boldsymbol{X}}_i = \boldsymbol{X}_i - E\{G_i^{(1)2}\boldsymbol{X}_i/G_i^2|Z_i\}/E\{G_i^{(1)2}/G_i^2|Z_i\}$.

The population counterparts are the above terms with subscript i suppressed. For example, $\varepsilon = Y - \mu(\boldsymbol{X}, Z; \boldsymbol{\alpha}_0, m_\mu)$, $D = \text{sign}(\varepsilon)$, $R = |\varepsilon|$, $G = g\{\nu(\boldsymbol{X}, Z; \boldsymbol{\beta}_0, m_\nu)\}$, $\delta = R - G$, $G^{(1)} = g^{(1)}\{\nu(\boldsymbol{X}, Z; \boldsymbol{\beta}_0, m_\nu)\}$, $\widetilde{\boldsymbol{X}} = \boldsymbol{X} - E\{\boldsymbol{X}|Z\}$, $\check{\boldsymbol{X}} = \boldsymbol{X} - E\{\boldsymbol{X}/G^2|Z\}/E\{1/G^2|Z\}$, $\breve{\boldsymbol{X}} = \boldsymbol{X} - E\{G^{(1)2}\boldsymbol{X}|Z\}/E\{G^{(1)2}|Z\}$, and $\bar{\boldsymbol{X}} = \boldsymbol{X} - E\{G^{(1)2}\boldsymbol{X}/G^2|Z\}/E\{G^{(1)2}/G^2|Z\}$. In addition, let $\sigma^2 = Var(\epsilon)$, $p(z)$ be the density function of Z, and $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$ for any matrix $\mathbf{A}$.

2.2 Initial estimator of the mean function

Methods for estimating $m_\mu(\cdot)$ and $\boldsymbol{\alpha}_0$ in the mean function $\mu(\boldsymbol{X}, Z; \boldsymbol{\alpha}_0, m_\mu)$ have already been well-established if the potential heteroscedasticity is ignored (Härdle et al. 2000; Speckman 1988). For given $\boldsymbol{\alpha}$, $m_\mu(z; \boldsymbol{\alpha}) = \arg\min_\xi E\{[Y - (\xi + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\alpha})]^2|Z = z\}$ can be estimated by kernel regression of $Y - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\alpha}$ on Z; that is, by

$$\widehat{m}_\mu(z; \boldsymbol{\alpha}) = \underset{\xi}{\operatorname{argmin}} \sum_{i=1}^{n} \left[Y_i - \left(\xi + \boldsymbol{X}_i{}^{\mathrm{T}}\boldsymbol{\alpha}\right)\right]^2 K_h(Z_i - z), \tag{2}$$

where K is a kernel function, $K_h(\cdot) = K(\cdot/h)/h$ and h is the bandwidth. Then $\boldsymbol{\alpha}_0$ can be estimated by profiling; that is, by

$$\widehat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \sum_{i=1}^{n} \left[Y_i - \left\{ \widehat{m}_{\mu}(Z_i; \boldsymbol{\alpha}) + \boldsymbol{X}_i{}^{\mathrm{T}} \boldsymbol{\alpha} \right\} \right]^2 . \tag{3}$$

The final estimate of $m_\mu(z)$ is $\widehat{m}_\mu(z) = \widehat{m}_\mu(z; \widehat{\boldsymbol{\alpha}})$.

Local constant regression (2) can be replaced by local linear regression, but all the asymptotic properties derived in Sect. 3 remain almost the same. More efficient estimators for $\{m_\mu(\cdot), \boldsymbol{\alpha}_0\}$ can be obtained via generalized least squares, which we discuss in Sect. 2.4.

2.3 Initial estimator of the variance function

Davidian and Carroll (1987) developed some general methodology and theory for variance function estimation in the parametric case. They distinguished between methods based on squared residuals and those based on absolute residuals, the former being more efficient if the regressions errors ϵ_i's are normally distributed, but called this potential efficiency gain "tenuous" because it is less robust to outliers. Here we use absolute residuals and follow a profiling approach analogous to the one in Sect. 2.2.

Define absolute residuals $\widehat{R}_i = |Y_i - \{\boldsymbol{X}_i{}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}} + \widehat{m}_\mu(Z_i)\}|$ and $\widehat{R} = |Y - \{\boldsymbol{X}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}} + \widehat{m}_\mu(Z)\}|$. Recall that $E(|\epsilon|) = 1$. Then, approximately, $E\{\widehat{R}|\boldsymbol{X}, Z\} \approx g\{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0 + m_\nu(Z)\}$. A very quick way to estimate $\{m_\nu(\cdot), \boldsymbol{\beta}_0\}$ is to regress $\widehat{R}$ on $g\{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta} + m_\nu(Z)\}$.

For given $\boldsymbol{\beta}$, $m_\nu(z; \boldsymbol{\beta}) = \arg\min_\zeta E\{(R - g\{\zeta + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}\})^2 | Z = z\}$ can be estimated by

$$\widehat{m}_\nu(z; \boldsymbol{\beta}) = \underset{\zeta}{\operatorname{argmin}} \sum_{i=1}^{n} \left[\widehat{R}_i - g\left\{ \zeta + \boldsymbol{X}_i{}^{\mathrm{T}} \boldsymbol{\beta} \right\} \right]^2 K_h(Z_i - z). \tag{4}$$

Here kernel K and bandwidth h could be different from the ones used in (2) as long as they satisfy the assumptions, but for notational simplicity, we use same K and h throughout. Then $\boldsymbol{\beta}_0$ can be estimated by profiling,

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \left[\widehat{R}_i - g\left\{ \widehat{m}_\nu(Z_i; \boldsymbol{\beta}) + \boldsymbol{X}_i{}^{\mathrm{T}} \boldsymbol{\beta} \right\} \right]^2 . \tag{5}$$

The final estimate of $m_\nu(z)$ is $\widehat{m}_\nu(z) = \widehat{m}_\nu(z; \widehat{\boldsymbol{\beta}})$ and the estimate of variance function G_i is

$$\widehat{G}_i = g\left\{ \widehat{m}_\nu(Z_i) + \boldsymbol{X}_i{}^{\mathrm{T}} \widehat{\boldsymbol{\beta}} \right\}. \tag{6}$$

Again local constant regression (4) can be replaced by local linear regression, but all the asymptotic properties remain almost the same. More efficient estimators for estimating $\{m_\nu(\cdot), \boldsymbol{\beta}_0\}$ can be obtained using absolute residuals from the more efficient estimate of the mean function and/or using weights based on the initial variance function estimate, as we discuss in Sect. 2.4.

2.4 More efficient estimators

The parameters $\{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0\}$ can be estimated more efficiently via generalized least squares. After the initial estimate of variance function is obtained, the estimators (2) and (3) for estimating $\{m_\mu(z; \boldsymbol{\alpha}), \boldsymbol{\alpha}_0\}$ can be replaced by, respectively,

$$\widehat{m}_{\mu,\text{wls}}(z; \boldsymbol{\alpha}) = \underset{\xi}{\text{argmin}} \sum_{i=1}^{n} \left[Y_i - \left(\xi + \boldsymbol{X}_i{}^{\text{T}} \boldsymbol{\alpha} \right) \right]^2 K_h(Z_i - z) \Big/ \widehat{G}_i^2, \tag{7}$$

$$\widehat{\boldsymbol{\alpha}}_{\text{wls}} = \underset{\boldsymbol{\alpha}}{\text{argmin}} \sum_{i=1}^{n} \left[Y_i - \left\{ \widehat{m}_{\mu,\text{wls}}(Z_i; \boldsymbol{\alpha}) + \boldsymbol{X}_i{}^{\text{T}} \boldsymbol{\alpha} \right\} \right]^2 \Big/ \widehat{G}_i^2. \tag{8}$$

Then an updated estimator for the variance function can be obtained via the same method in Sect. 2.3 except that the absolute residuals R_i are the ones from the above updated estimates, $\widehat{R}_{i,\text{wls}} = |Y_i - \{\boldsymbol{X}_i{}^{\text{T}} \widehat{\boldsymbol{\alpha}}_{\text{wls}} + \widehat{m}_{\mu,\text{wls}}(Z_i)\}|$. That is, the estimators (4) and (5) for estimating $\{m_\nu(z; \boldsymbol{\beta}), \boldsymbol{\beta}_0\}$ can be replaced by, respectively,

$$\widehat{m}_{\nu,R}(z; \boldsymbol{\beta}) = \underset{\zeta}{\text{argmin}} \sum_{i=1}^{n} \left[\widehat{R}_{i,\text{wls}} - g \left\{ \zeta + \boldsymbol{X}_i{}^{\text{T}} \boldsymbol{\beta} \right\} \right]^2 K_h(Z_i - z), \tag{9}$$

$$\widehat{\boldsymbol{\beta}}_R = \underset{\boldsymbol{\beta}}{\text{argmin}} \sum_{i=1}^{n} \left[\widehat{R}_{i,\text{wls}} - g \left\{ \widehat{m}_{\nu,R}(Z_i; \boldsymbol{\beta}) + \boldsymbol{X}_i{}^{\text{T}} \boldsymbol{\beta} \right\} \right]^2, \tag{10}$$

where subscript "R" means the improvement comes from the improved estimates of R_i. The estimators (4) and (5) for estimating $\{m_\nu(z; \boldsymbol{\beta}), \boldsymbol{\beta}_0\}$ can be further improved using weights based on the initial variance function. That is, the estimators (4) and (5) can be replaced by

$$\widehat{m}_{\nu,\text{wls}}(z; \boldsymbol{\beta}) = \underset{\zeta}{\text{argmin}} \sum_{i=1}^{n} \left[\widehat{R}_{i,\text{wls}} - g \left\{ \zeta + \boldsymbol{X}_i{}^{\text{T}} \boldsymbol{\beta} \right\} \right]^2 K_h(Z_i - z) / \widehat{G}_i^2, \tag{11}$$

$$\widehat{\boldsymbol{\beta}}_{\text{wls}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \sum_{i=1}^{n} \left[\widehat{R}_{i,\text{wls}} - g \left\{ \widehat{m}_{\nu,\text{wls}}(Z_i; \boldsymbol{\beta}) + \boldsymbol{X}_i{}^{\text{T}} \boldsymbol{\beta} \right\} \right]^2 / \widehat{G}_i^2. \tag{12}$$

The counterparts of the procedure consisting of (7) and (8) and the one of (9) and (10) for the VFPLSIM were discussed in Lian et al. (2015). However, Lian et al. (2015) did not investigate the counterpart of the procedure consisting of (11) and (12), because of the complexity of the VFPLSIM. We believe the theoretical properties of the procedure consisting of (11) and (12) derived in Sect. 3 could help understand the potential properties of its counterpart for the VFPLSIM.

3 Theoretical results

Theorem 1 *Suppose that Assumptions A in "Appendix" hold. Define* $Q_\alpha = E(\widetilde{\boldsymbol{X}}^{\otimes 2})$. *As* $n \to \infty$, $nh^4 \to \infty$, *and* $nh^6 \to 0$, *we have*

$$\widehat{m}_{\mu}(z, \widehat{\boldsymbol{\alpha}}) - m_{\mu}(z) = \frac{1}{n}\sum_{i=1}^{n} K_h(Z_i - z)\frac{\varepsilon_i}{p(z)} - E\{\boldsymbol{X}^{\mathrm{T}}|Z = z\}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$$
$$+ b_{\mu}(z)\kappa_{\mu}h^2 + o_p(1/\sqrt{n}), \tag{13}$$

where $b_{\mu}(z) = m_{\mu}^{(2)}(z)/2 + m_{\mu}^{(1)}(z)p^{(1)}(z)/p(z)$ *if local constant regression is used in* (2), $b_{\mu}(z) = m_{\mu}^{(2)}(z)/2$ *if local linear regression is used, and* $o_p(1/\sqrt{n})$ *is uniform over* z, *and*

$$\sqrt{n}Q_{\alpha}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \varepsilon_i \widetilde{\boldsymbol{X}}_i + o_p(1). \tag{14}$$

Accordingly, $\sqrt{n}Q_{\alpha}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \xrightarrow{D} \mathrm{Normal}(\mathbf{0}, \Sigma_{\alpha})$, *where* $\Sigma_{\alpha} = \sigma^2 E\{(G\widetilde{\boldsymbol{X}})^{\otimes 2}\}$.

Theorem 2 *Suppose that Assumptions A and B in "Appendix" hold. Define* $Q_{\beta} = E\{(G^{(1)}\check{\boldsymbol{X}})^{\otimes 2}\}$. *As* $n \to \infty$, $nh^4 \to \infty$, *and* $nh^6 \to 0$, *we have*

$$\widehat{m}_{v}(z, \widehat{\boldsymbol{\beta}}) - m_{v}(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{K_h(Z_i - z)G_i^{(1)}\delta_i}{p(z)E\{G^{(1)2}|Z = z\}} - \left[\frac{E\{G^{(1)2}\boldsymbol{X}^{\mathrm{T}}|Z = z\}}{E\{G^{(1)2}|Z = z\}}\right](\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$
$$+ b_{v}(z)\kappa_{v}h^2 - \left[\frac{E\{G^{(1)}D\widetilde{\boldsymbol{X}}^{\mathrm{T}}|Z = z\}}{E\{G^{(1)2}|Z = z\}}\right](\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$$
$$+ o_p(1/\sqrt{n}), \tag{15}$$

where $b_{v}(z)$, *which depends on whether local constant regression or local linear regression is used in* (4), *is defined in "Appendix A.3" and* $o_p(1/\sqrt{n})$ *is uniform over* z, *and*

$$\sqrt{n}Q_{\beta}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\delta_i G_i^{(1)}\check{\boldsymbol{X}}_i - \varepsilon_i\left[E\{G^{(1)}D\check{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\}Q_{\alpha}^{-1}\widetilde{\boldsymbol{X}}_i\right.\right.$$
$$\left.\left.+E\left\{G_i^{(1)}D_i\check{\boldsymbol{X}}_i|Z_i\right\}\right]\right) + o_p(1). \tag{16}$$

Accordingly, $\sqrt{n}Q_{\beta}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{D} \mathrm{Normal}(\mathbf{0}, \Sigma_{\beta})$, *where*

$$\Sigma_{\beta} = E\left(\delta G^{(1)}\check{\boldsymbol{X}} - \varepsilon\left[E\{G^{(1)}D\check{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\}Q_{\alpha}^{-1}\widetilde{\boldsymbol{X}} + E\{G^{(1)}D\check{\boldsymbol{X}}|Z\}\right]\right)^{\otimes 2}.$$

Theorem 3 *Suppose that Assumptions A and B "Appendix" hold. Define* $Q_{\alpha,\mathrm{wls}} = E\{(\check{\boldsymbol{X}}/G)^{\otimes 2}\}$. *As* $n \to \infty$, $nh^4 \to \infty$, *and* $nh^6 \to 0$, *we have*

$$\widehat{m}_{\mu,\mathrm{wls}}(z,\widehat{\boldsymbol{\alpha}}_{\mathrm{wls}}) - m_\mu(z) = \frac{1}{n}\sum_{i=1}^{n} K_h(Z_i - z)\frac{\varepsilon_i/G_i^2}{p(z)E\{1/G^2|Z=z\}} + b_\mu(z)\kappa_\mu h^2 - \left[\frac{E\{\boldsymbol{X}^{\mathrm{T}}/G^2|Z=z\}}{E\{1/G^2|Z=z\}}\right](\widehat{\boldsymbol{\alpha}}_{\mathrm{wls}} - \boldsymbol{\alpha}) + o_p(1/\sqrt{n}), \tag{17}$$

where $b_\mu(z)$ is the same as the one in Theorem 1 and $o_p(1/\sqrt{n})$ is uniform over z, and

$$\sqrt{n}Q_{\alpha,\mathrm{wls}}(\widehat{\boldsymbol{\alpha}}_{\mathrm{wls}} - \boldsymbol{\alpha}_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i \check{\boldsymbol{X}}_i/G_i^2 + o_p(1). \tag{18}$$

Accordingly, $\sqrt{n}(\widehat{\boldsymbol{\alpha}}_{\mathrm{wls}} - \boldsymbol{\alpha}_0) \xrightarrow{D} \mathrm{Normal}(\mathbf{0}, \sigma^2 Q_{\alpha,\mathrm{wls}}^{-1})$. *Further, when ϵ is normally distributed, $\widehat{\boldsymbol{\alpha}}_{\mathrm{wls}}$ is the most efficient estimator in the sense of semiparametric efficiency.*

Theorem 4 *Suppose that Assumptions A and B in "Appendix" hold. Define $Q_\beta = E\{(G^{(1)}\check{\boldsymbol{X}})^{\otimes 2}\}$. As $n \to \infty$, $nh^4 \to \infty$, and $nh^6 \to 0$, we have*

$$\widehat{m}_{\nu,R}(z,\widehat{\boldsymbol{\beta}}_R) - m_\nu(z) = \frac{1}{n}\sum_{i=1}^{n}\frac{K_h(Z_i - z)G_i^{(1)}\delta_i}{p(z)E\{G^{(1)2}|Z=z\}} - \left[\frac{E\{G^{(1)2}\boldsymbol{X}^{\mathrm{T}}|Z=z\}}{E\{G^{(1)2}|Z=z\}}\right](\widehat{\boldsymbol{\beta}}_{\mathrm{wls}} - \boldsymbol{\beta}_0) + b_\nu(z)\kappa_\nu h^2 - \left[\frac{E\{G^{(1)}D\check{\boldsymbol{X}}^{\mathrm{T}}|Z=z\}}{E\{G^{(1)2}|Z=z\}}\right](\widehat{\boldsymbol{\alpha}}_{\mathrm{wls}} - \boldsymbol{\alpha}_0) + o_p(1/\sqrt{n}), \tag{19}$$

where $b_\nu(z)$ is the same as the one defined in Theorem 2 and $o_p(1/\sqrt{n})$ is uniform over z, and

$$\sqrt{n}Q_\beta(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\delta_i G_i^{(1)}\check{\boldsymbol{X}}_i - \varepsilon_i\left[E\{G^{(1)}D\check{\boldsymbol{X}}\check{\boldsymbol{X}}^{\mathrm{T}}\}Q_{\alpha,\mathrm{wls}}^{-1}\check{\boldsymbol{X}}_i/G_i^2 + E\left\{G_i^{(1)}D_i\check{\boldsymbol{X}}_i|Z_i\right\}\right]\right) + o_p(1). \tag{20}$$

Accordingly, $\sqrt{n}Q_\beta(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) \xrightarrow{D} \mathrm{Normal}(\mathbf{0}, \Sigma_{\beta,R})$, *where*

$$\Sigma_{\beta,R} = E\left(\delta G^{(1)}\check{\boldsymbol{X}} - \varepsilon\left[E\{G^{(1)}D\check{\boldsymbol{X}}\check{\boldsymbol{X}}^{\mathrm{T}}\}Q_{\alpha,\mathrm{wls}}^{-1}\check{\boldsymbol{X}}/G^2 + E\{G^{(1)}D\check{\boldsymbol{X}}|Z\}\right]\right)^{\otimes 2}.$$

Theorem 5 *Suppose that Assumptions A and B in "Appendix" hold. Define* $Q_{\beta,\mathrm{wls}} = E\{(G^{(1)}\bar{\boldsymbol{X}}/G)^{\otimes 2}\}$. *As* $n \to \infty$, $nh^4 \to \infty$, *and* $nh^6 \to 0$, *we have*

$$\begin{aligned}\widehat{m}_{\nu,\mathrm{wls}}(z, \widehat{\boldsymbol{\beta}}_{\mathrm{wls}}) - m_\nu(z) &= \frac{1}{n}\sum_{i=1}^{n} \frac{K_h(Z_i - z)G_i^{(1)}\delta_i/G_i^2}{p(z)E\{G^{(1)2}/G^2|Z=z\}} \\ &- \left[\frac{E\{G^{(1)2}\boldsymbol{X}^{\mathrm{T}}/G^2|Z=z\}}{E\{G^{(1)2}/G^2|Z=z\}}\right] \\ &\times (\widehat{\boldsymbol{\beta}}_{\mathrm{wls}} - \boldsymbol{\beta}_0) + b_{\nu,\mathrm{wls}}(z)\kappa_\nu h^2 \\ &- \left[\frac{E\{G^{(1)}D\check{\boldsymbol{X}}^{\mathrm{T}}/G^2|Z=z\}}{E\{G^{(1)2}/G^2|Z=z\}}\right](\widehat{\boldsymbol{\alpha}}_{\mathrm{wls}} - \boldsymbol{\alpha}_0) \\ &+ o_p(1/\sqrt{n}), \end{aligned} \quad (21)$$

where $b_{\nu,\mathrm{wls}}(z)$ *is defined in "Appendix A.5" and* $o_p(1/\sqrt{n})$ *is uniform over z, and*

$$\begin{aligned}\sqrt{n}Q_{\beta,\mathrm{wls}}(\widehat{\boldsymbol{\beta}}_{\mathrm{wls}} - \boldsymbol{\beta}_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\delta_i G_i^{(1)}\bar{\boldsymbol{X}}_i/G_i^2 \\ &- \varepsilon_i\Big[E\{G^{(1)}D\bar{\boldsymbol{X}}\check{\boldsymbol{X}}^{\mathrm{T}}/G^2\}Q_{\alpha,\mathrm{wls}}^{-1}\check{\boldsymbol{X}}_i/G_i^2 \\ &+ E\Big\{G_i^{(1)}D_i\bar{\boldsymbol{X}}_i/G_i^2|Z_i\Big\}\Big]\Big) + o_p(1). \end{aligned} \quad (22)$$

Accordingly, $\sqrt{n}Q_{\beta,\mathrm{wls}}(\widehat{\boldsymbol{\beta}}_{\mathrm{wls}} - \boldsymbol{\beta}_0) \xrightarrow{D} \mathrm{Normal}(\boldsymbol{0}, \Sigma_{\beta,\mathrm{wls}})$, *where*

$$\begin{aligned}\Sigma_{\beta,\mathrm{wls}} = E\Big(&\delta G^{(1)}\bar{\boldsymbol{X}}/G^2 - \varepsilon\Big[E\{G^{(1)}D\bar{\boldsymbol{X}}\check{\boldsymbol{X}}^{\mathrm{T}}/G^2\}Q_{\alpha,\mathrm{wls}}^{-1}\check{\boldsymbol{X}}/G^2 \\ &+ E\{G^{(1)}D\bar{\boldsymbol{X}}/G^2|Z\}\Big]\Big)^{\otimes 2}.\end{aligned}$$

Now we compare the asymptotic covariance of the initial estimator for $\boldsymbol{\alpha}_0$, $\widehat{\boldsymbol{\alpha}}$, with that of the weighted estimator, $\widehat{\boldsymbol{\alpha}}_{\mathrm{wls}}$. In this and next comparisons, for simplicity, we ignore the factor n. The asymptotic variance of $\widehat{\boldsymbol{\alpha}}$ is $\sigma^2[E(\widetilde{\boldsymbol{X}}^{\otimes 2})]^{-1}E\{(G\widetilde{\boldsymbol{X}})^{\otimes 2}\}[E(\widetilde{\boldsymbol{X}}^{\otimes 2})]^{-1}$ and the asymptotic variance of $\widehat{\boldsymbol{\alpha}}_{\mathrm{wls}}$ is $\sigma^2[E\{(\check{\boldsymbol{X}}/G)^{\otimes 2}\}]^{-1}$. Noting that

$$\boldsymbol{M}_1 = E\left\{\begin{pmatrix}\check{\boldsymbol{X}}/G \\ G\widetilde{\boldsymbol{X}}\end{pmatrix}^{\otimes 2}\right\} = \begin{pmatrix} E\{(\check{\boldsymbol{X}}/G)^{\otimes 2}\} & E(\widetilde{\boldsymbol{X}}^{\otimes 2}) \\ E(\widetilde{\boldsymbol{X}}^{\otimes 2}) & E\{(G\widetilde{\boldsymbol{X}})^{\otimes 2}\}\end{pmatrix} \geq 0,$$

we see that $E\{(\check{\boldsymbol{X}}/G)^{\otimes 2}\} \geq E(\widetilde{\boldsymbol{X}}^{\otimes 2})[E\{(G\widetilde{\boldsymbol{X}})^{\otimes 2}\}]^{-1}E(\widetilde{\boldsymbol{X}}^{\otimes 2})$, where the strict inequality holds as long as the above covariance matrix $\boldsymbol{M}_1$ is positive definite. This means that weighted estimator $\widehat{\boldsymbol{\alpha}}_{\mathrm{wls}}$ is more efficient than initial estimator $\widehat{\boldsymbol{\alpha}}$.

Next we compare the asymptotic covariances of those estimators for $\boldsymbol{\beta}_0$, $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\beta}}_R$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{wls}}$. For simplicity, we only consider the special case where ϵ is symmetric. In this special case, $E(D) = 0$ and therefore the last two terms in each of Σ_β, Σ_R and $\Sigma_{\rm wls}$ become zero. Noting that $Var(\delta/G) = Var(|\epsilon| - 1) = \sigma^2 - 1$, the asymptotic variances of $\widehat{\boldsymbol\beta}$ and $\widehat{\boldsymbol\beta}_R$ are both equal to $(\sigma^2 - 1)[E\{(G^{(1)}\check{\boldsymbol X})^{\otimes 2}\}]^{-1}E\{(GG^{(1)}\check{\boldsymbol X})^{\otimes 2}\}[E\{(G^{(1)}\check{\boldsymbol X})^{\otimes 2}\}]^{-1}$. This means that in this special case, partially updated estimator $\widehat{\boldsymbol\beta}_R$ is asymptotically as efficient as initial estimator $\widehat{\boldsymbol\beta}$. The asymptotic variance of $\widehat{\boldsymbol\beta}_{\rm wls}$ is $(\sigma^2-1)[E\{(G^{(1)}\bar{\boldsymbol X}/G)^{\otimes 2}\}]^{-1}$. Noting that

$$\boldsymbol M_2 = E\left\{\begin{pmatrix} G^{(1)}\bar{\boldsymbol X}/G \\ GG^{(1)}\check{\boldsymbol X}\end{pmatrix}^{\otimes 2}\right\} = \begin{pmatrix} E\{(G^{(1)}\bar{\boldsymbol X}/G)^{\otimes 2}\} & E\{(G^{(1)}\check{\boldsymbol X})^{\otimes 2}\} \\ E\{(G^{(1)}\check{\boldsymbol X})^{\otimes 2}\} & E\{(GG^{(1)}\check{\boldsymbol X})^{\otimes 2}\}\end{pmatrix} \geq 0,$$

we see that $E\{(G^{(1)}\bar{\boldsymbol X}/G)^{\otimes 2}\} \geq E\{(G^{(1)}\check{\boldsymbol X})^{\otimes 2}\}[E\{(GG^{(1)}\check{\boldsymbol X})^{\otimes 2}\}]^{-1}E\{(G^{(1)}\check{\boldsymbol X})^{\otimes 2}\}$, where the strict inequality holds as long as the above covariance matrix $\boldsymbol M_2$ is positive definite. This means that weighted estimator $\widehat{\boldsymbol\beta}_{\rm wls}$ is more efficient than the other two estimators.

We conclude this section with some discussion on estimating the asymptotic covariances. We only provide estimators for the asymptotic covariance matrices of $\widehat{\boldsymbol\alpha}_{\rm wls}$ and $\widehat{\boldsymbol\beta}_{\rm wls}$. The asymptotic covariance matrices of the other estimators can be estimated similarly. To abuse notation, let $\widehat{G}_i = g\{\boldsymbol X_i{}^{\rm T}\widehat{\boldsymbol\beta}_{\rm wls} + \widehat{m}_{\nu,{\rm wls}}(Z_i)\}$ and $\widehat{G}_i^{(1)} = g^{(1)}\{\boldsymbol X_i{}^{\rm T}\widehat{\boldsymbol\beta}_{\rm wls} + \widehat{m}_{\nu,{\rm wls}}(Z_i)\}$.

The asymptotic covariance of $\sqrt{n}\widehat{\boldsymbol\alpha}_{\rm wls}$ is $\sigma^2 Q^{-1}_{\alpha,{\rm wls}}$. Since $Var(\epsilon) = \sigma^2$, a consistent estimator of σ^2 is $\widehat{\sigma}^2 = \sum\widehat{\epsilon}_i^2/(n-p)$, where $\widehat{\epsilon}_i = [Y_i - \{\boldsymbol X_i{}^{\rm T}\widehat{\boldsymbol\alpha}_{\rm wls} + \widehat{m}_{\mu,{\rm wls}}(Z_i)\}]/\widehat{G}_i$. Recall that $Q_{\alpha,{\rm wls}} = E\{(\check{\boldsymbol X}/G)^{\otimes 2}\}$. A consistent estimator of $E\{(\check{\boldsymbol X}/G)^{\otimes 2}\}$ is $n^{-1}\sum(\check{\boldsymbol X}_i^*/\widehat{G}_i)^{\otimes 2}$, where

$$\check{\boldsymbol X}_i^* = \boldsymbol X_i - \frac{\sum_{j=1}^n \boldsymbol X_j K_h(Z_j - Z_i)/\widehat{G}_j^2}{\sum_{j=1}^n K_h(Z_j - Z_i)/\widehat{G}_j^2}.$$

The asymptotic covariance of $\sqrt{n}\widehat{\boldsymbol\beta}_{\rm wls}$ is $Q^{-1}_{\beta,{\rm wls}}\Sigma_{\beta,{\rm wls}}Q^{-1}_{\beta,{\rm wls}}$. Recall that $Q_{\beta,{\rm wls}} = E\{(G^{(1)}\bar{\boldsymbol X}/G)^{\otimes 2}\}$. A consistent estimator of $E\{(G^{(1)}\bar{\boldsymbol X}/G)^{\otimes 2}\}$ is $n^{-1}\sum(\widehat{G}_i^{(1)}\bar{\boldsymbol X}_i^*/\widehat{G}_i)^{\otimes 2}$, where

$$\bar{\boldsymbol X}_i^* = \boldsymbol X_i - \frac{\sum_{j=1}^n \widehat{G}_j^{(1)2}\boldsymbol X_j K_h(Z_j - Z_i)/\widehat{G}_j^2}{\sum_{j=1}^n \widehat{G}_j^{(1)2} K_h(Z_j - Z_i)/\widehat{G}_j^2}.$$

If ϵ is symmetric, then $\Sigma_{\beta,{\rm wls}} = (\sigma^2 - 1)Q_{\beta,{\rm wls}}$. Otherwise, in general, $\Sigma_{\beta,{\rm wls}}$ can be estimated by

$$\frac{1}{n}\sum_{i=1}^n\left[(|\widehat{\epsilon}_i| - 1)\widehat{G}_i^{(1)}\bar{\boldsymbol X}^*/\widehat{G}_i - \widehat{\epsilon}_i M\widehat{Q}^{-1}_{\alpha,{\rm wls}}\check{\boldsymbol X}^*/\widehat{G}_i - \widehat{\epsilon}_i\pi(Z_i)\widehat{G}_i\right]^{\otimes 2},$$

where $M = n^{-1}\sum {\rm sign}(\widehat{\epsilon}_i)\widehat{G}_i^{(1)}\bar{\boldsymbol X}_i^*\check{\boldsymbol X}_i^{*\rm T}/\widehat{G}_i$ and $\pi(z) = \frac{\sum {\rm sign}(\widehat{\epsilon}_j)\widehat{G}_j^{(1)}\bar{\boldsymbol X}_j^* K_h(Z_j - z)/\widehat{G}_j^2}{\sum K_h(Z_j - z)}$.

4 Numerical results

4.1 Simulations

We generated data from model (1), with $g(\nu) = \exp(\nu)$ and $\epsilon \sim \text{Normal}(0, \sigma^2)$. Note that the parameter identification assumption that $E(|\epsilon|) = 1$ made below model (1) can be satisfied by the following re-parametrization: $\widetilde{\epsilon} = \epsilon/E(|\epsilon|)$, $\widetilde{g} = E(|\epsilon|)g$, and $\widetilde{\sigma}^2 = Var(|\widetilde{\epsilon}|)$. The covariates $(X_0, X_1, \ldots, X_8)^{\mathrm{T}}$ were generated from a multivariate Gaussian distribution with covariance given by $\text{cov}(X_i, X_j) = (0.3)^{|i-j|}$, setting $Z = X_0$ and $\boldsymbol{X} = (X_1, \ldots, X_8)^{\mathrm{T}}$. We considered two simulation examples. In the first example, we set

$$\boldsymbol{\alpha}_0 = (1, -1, 1, -1, 1, -1, 1, -1)^{\mathrm{T}}, m_\mu(z) = 10\sin(z),$$
$$\boldsymbol{\beta}_0 = (-0.1, 0.2, -0.1, 0.2, 0.1, 0.2, 0.2, 0.1)^{\mathrm{T}}, m_\nu(z) = \cos(z) + 1.$$

In the second example, we set

$$\boldsymbol{\alpha}_0 = (1, 1, 1, 1, -0.5, -0.5, -0.5, -0.5)^{\mathrm{T}}, m_\mu(z) = \exp(z),$$
$$\boldsymbol{\beta}_0 = (0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2)^{\mathrm{T}}, m_\nu(z) = \log\{(z-1)^2 + 1\}.$$

We considered sample size $n = 200$ and four noise levels, $\sigma = 0.1, 0.2, 0.5, 1.0$. For each scenario, we randomly generated 100 datasets. We considered the following five groups of estimates described in Sect. 2.

- E1. The initial unweighted estimates for the mean function, (2) and (3),
- E2. The initial unweighted estimates for the variance function, (4) and (5),
- E3. The updated weighted estimates for the mean function, (7) and (8),
- E4. The partially updated estimates for the variance function, (9) and (10),
- E5. The updated weighted estimates for the variance function, (11) and (12).

We compared these estimators with the following three groups of infeasible estimates, "infeasible" indicating they are not real estimators because some unknown components are utilized.

- IE1. The infeasible weighted estimates for the mean function when the variance function is known in calculation of the weights,
- IE2. The infeasible unweighted estimates for the variance function when the mean function is known in calculation of the absolute residuals,
- IE3. The infeasible weighted estimates for the variance function when the mean function is known in calculation of the absolute residuals and the variance function is known in calculation of the weights.

The numerical results reported here were based on locally linear regression. The bandwidths were chosen simply by Silverman's rule of thumb. Although this choice was non-optimal for regression, it still performed well in simulations and was empirically more stable than data-driven methods such as cross-validation or plug-in methods, significantly shortening the computational time. The results are given in

Table 1 RMSE of eight groups of estimates in Example 1

σ	$\boldsymbol{\alpha}_0$	m_μ
E1. Initial mean est		
0.1	0.145 (0.048)	0.185 (0.040)
0.2	0.285 (0.092)	0.266 (0.074)
0.5	0.712 (0.222)	0.521 (0.188)
1	1.421 (0.439)	0.866 (0.363)
E3. Weighted mean est		
0.1	0.122 (0.036)	0.184 (0.037)
0.2	0.220 (0.064)	0.265 (0.066)
0.5	0.525 (0.149)	0.512 (0.175)
1	1.120 (0.326)	0.860 (0.345)
IE1. Infeasible weighted mean est		
0.1	0.118 (0.050)	0.184 (0.037)
0.2	0.219 (0.066)	0.264 (0.065)
0.5	0.494 (0.143)	0.509 (0.169)
1	1.081 (0.325)	0.867 (0.331)
σ	$\boldsymbol{\beta}_0$	m_ν
E2. Initial var est		
0.1	0.288 (0.078)	0.176 (0.074)
0.2	0.283 (0.073)	0.173 (0.074)
0.5	0.287 (0.079)	0.182 (0.079)
1	0.287 (0.080)	0.181 (0.087)
E4. Partially updated var est		
0.1	0.286 (0.076)	0.168 (0.081)
0.2	0.281 (0.070)	0.162 (0.079)
0.5	0.282 (0.074)	0.178 (0.087)
1	0.283 (0.080)	0.179 (0.091)
E5. Weighted var est		
0.1	0.262 (0.071)	0.148 (0.062)
0.2	0.268 (0.072)	0.165 (0.085)
0.5	0.274 (0.077)	0.159 (0.084)
1	0.272 (0.080)	0.165 (0.085)
IE2. Infeasible unweighted var est		
0.1	0.263 (0.073)	0.147 (0.047)
0.2	0.261 (0.059)	0.144 (0.051)
0.5	0.258 (0.057)	0.140 (0.062)
1	0.259 (0.059)	0.138 (0.049)

Tables 1 and 2, which report the root mean squared errors (RMSE) of different quantities, for the two examples respectively. For example, for $\boldsymbol{\alpha}_0$ the RMSE is $\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|$ and for m_μ the RMSE is $\sqrt{\sum_{t=1}^{50}\{\widehat{m}_\mu(z_t) - m_\mu(z_t)\}^2/50}$, where $(z_1, \ldots, z_{50})$ are

Table 1 continued

σ	$\boldsymbol{\beta}_0$	m_ν
IE3. Infeasible weighted var est		
0.1	0.248 (0.051)	0.134 (0.055)
0.2	0.245 (0.048)	0.133 (0.060)
0.5	0.247 (0.050)	0.134 (0.056)
1	0.248 (0.052)	0.134 (0.053)

Numbers in parentheses are standard deviations of RMSE from 100 repetitions

Table 2 RMSE of eight groups of estimates in Example 2

σ	$\boldsymbol{\alpha}_0$	m_μ
E1. Initial mean est		
0.1	0.105 (0.045)	0.102 (0.036)
0.2	0.146 (0.091)	0.169 (0.065)
0.5	0.412 (0.226)	0.340 (0.139)
1	1.037 (0.460)	0.610 (0.263)
E3. Weighted mean est		
0.1	0.071 (0.038)	0.099 (0.033)
0.2	0.123 (0.061)	0.163 (0.063)
0.5	0.321 (0.185)	0.321 (0.119)
1	0.841 (0.489)	0.555 (0.217)
IE1. Infeasible weighted mean est		
0.1	0.064 (0.032)	0.099 (0.032)
0.2	0.094 (0.044)	0.162 (0.060)
0.5	0.264 (0.188)	0.319 (0.112)
1	0.795 (0.510)	0.552 (0.209)
σ	$\boldsymbol{\beta}_0$	m_ν
E2. Initial var est		
0.1	0.355 (0.095)	0.235 (0.094)
0.2	0.357 (0.097)	0.192 (0.077)
0.5	0.357 (0.094)	0.196 (0.076)
1	0.356 (0.090)	0.198 (0.077)
E4. Partially updated var est		
0.1	0.351 (0.102)	0.222 (0.074)
0.2	0.331 (0.097)	0.196 (0.069)
0.5	0.336 (0.093)	0.200 (0.071)
1	0.343 (0.092)	0.191 (0.070)

equally-spaced grid points on an interval with lower bound being the 0.01 quantile of the sampled Z values and upper bound the 0.99 quantile.

First, from Tables 1 and 2, we see that the noise level σ has a large impact on estimating parameters $\{\boldsymbol{\alpha}_0, m_\mu\}$ in the mean function, whereas it has a much smaller impact on estimating parameters $\{\boldsymbol{\beta}_0, m_\nu\}$ in the variance function. Therefore, we ver-

Table 2 continued

σ	$\boldsymbol{\beta}_0$	m_ν
E5. Weighted var est		
0.1	0.346 (0.121)	0.213 (0.129)
0.2	0.330 (0.109)	0.191 (0.116)
0.5	0.329 (0.098)	0.190 (0.100)
1	0.339 (0.101)	0.182 (0.111)
IE2. Infeasible unweighted var est		
0.1	0.339 (0.130)	0.204 (0.100)
0.2	0.321 (0.125)	0.185 (0.084)
0.5	0.311 (0.103)	0.183 (0.082)
1	0.323 (0.116)	0.188 (0.081)
IE3. Infeasible weighted var est		
0.1	0.312 (0.113)	0.178 (0.070)
0.2	0.297 (0.077)	0.175 (0.087)
0.5	0.293 (0.062)	0.167 (0.077)
1	0.293 (0.066)	0.168 (0.081)

Numbers in parentheses are standard deviations of RMSE from 100 repetitions

ify this well-known phenomenon of the variance function estimation firstly discussed in Davidian and Carroll (1987).

By comparing the RMSEs in Tables 1 and 2, we see that the mean function estimation (E3) with weighting is more efficient than that without weighting (E1), as implied by the theoretical results. Similarly, we also see that the variance function estimation with weighting (E5) is more efficient than the partially updated variance function (E4), which is more efficient than the initial variance function estimation (E2).

In addition, although asymptotically the weighted estimators are as efficient as the infeasible counterparts (E3 vs IE1, E4 vs IE2, and E5 vs IE3), from Tables 1 and 2 we can still see some differences between them with finite sample sizes, infeasible estimators exhibiting more efficiency than their counterparts.

To further examine the performance of estimating non-parametric components m_μ and m_ν, for each scenario we present the three estimated curves whose RMSE correspond to the 1st, 2nd and 3rd quartiles of the RMSEs among the 100 replicates. We display these estimated curves in Figs. 1 and 2 (in red, green, blue, respectively), along with the true non-parametric estimands shown in black solid curves. From these two figures, we see that the nonparametric functions are estimated reasonably well. We also see that the noise level σ has a large impact on estimating m_μ, whereas it has a small impact on estimating m_ν.

Theoretically, the asymptotic distributions in Theorems 3 and 5 are the same as the asymptotic distributions that would be obtained when the true weights G_i are known, and thus further updating of the estimates will have no effects on the asymptotic results. We have also tried using the estimated variance function to further update the mean estimates and then the variance estimates, but did not observe improvement in estimation accuracy.

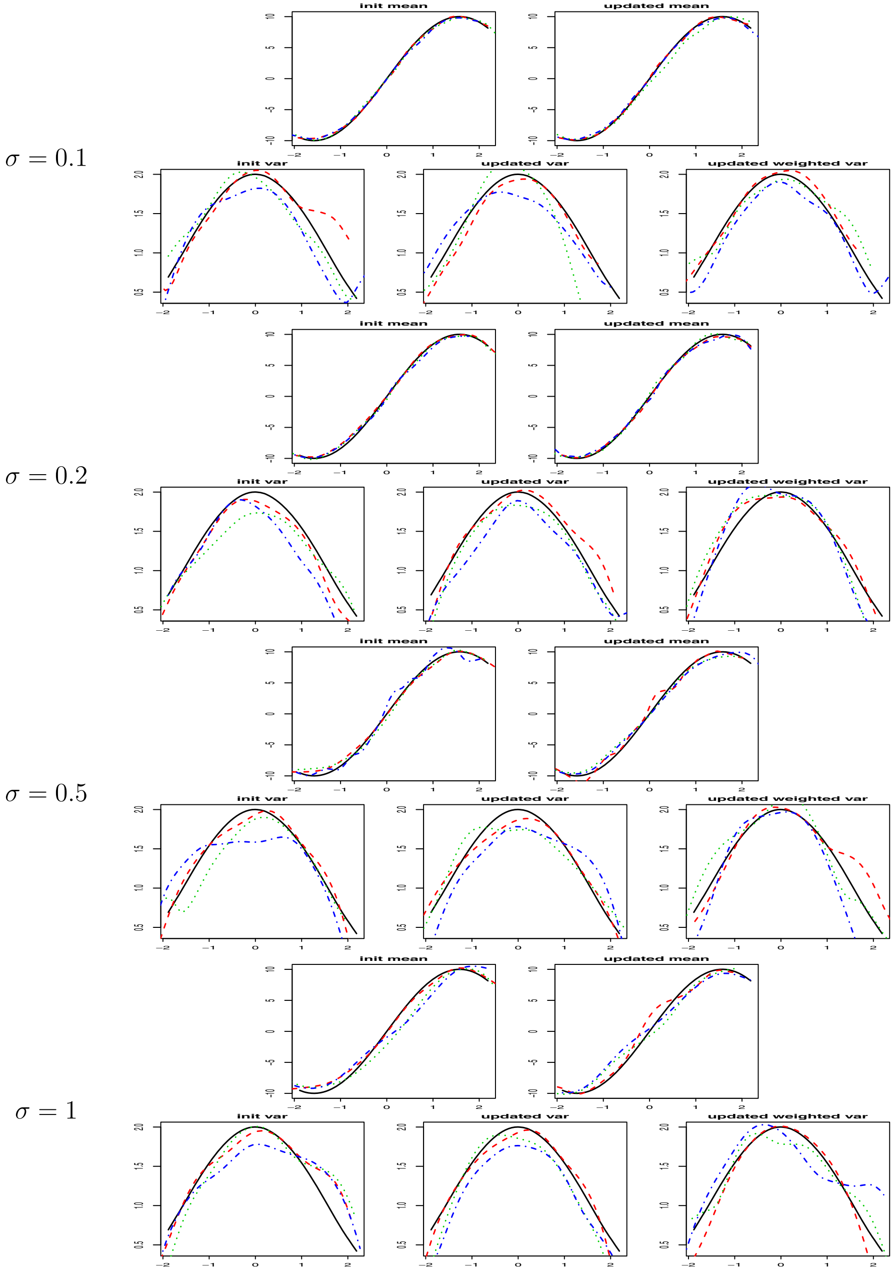

Fig. 1 The simulation results for Example 1. For each value of σ, the five figures correspond clockwisely to five feasible non-parametric estimates E1-E5, with the corresponding estimands (black solid), 1st quartile (red dashed), 2nd quartile (median; green dotted) and 3rd quartile (blue dash-dotted) of RMSE in estimating the functions

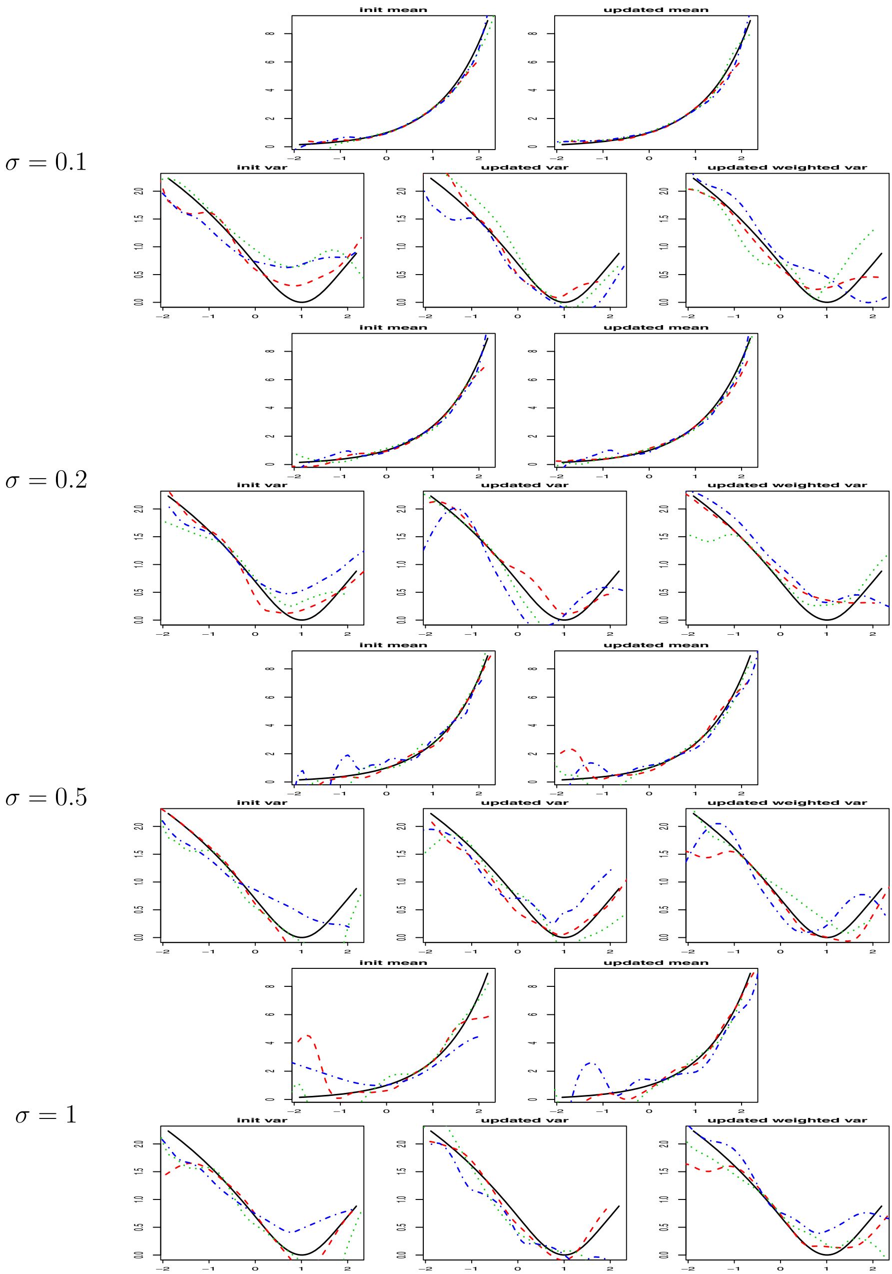

Fig. 2 The simulation results for Example 2. The caption is the same as in Fig. 1

4.2 An empirical example

The proposed methods are applied to an air pollution study with data collected in Mexico City. We use a subset of the data that contains the daily mortality and air

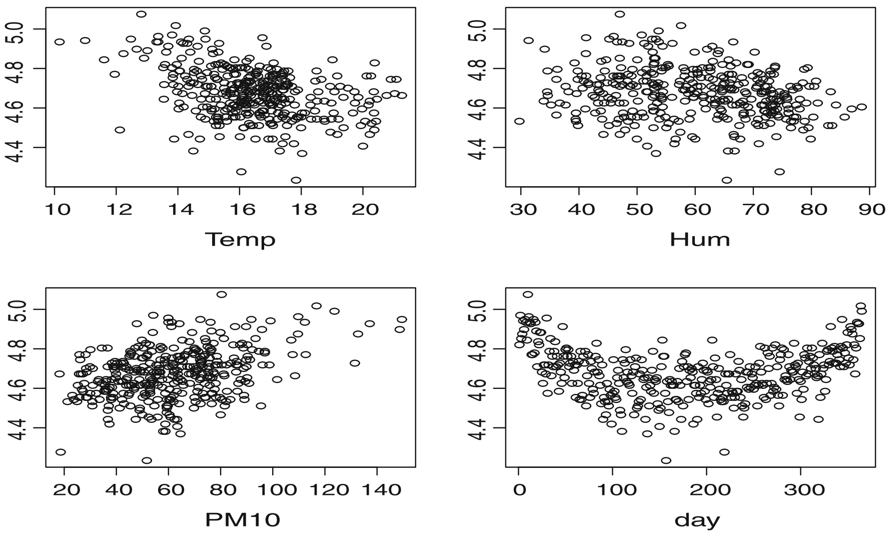

Fig. 3 Scatterplots for the air pollution data

pollution data in the year 1994. The goal of the study was to relate the daily counts of accidental death (D) to three predictors including temperature (Temp), humidity (Hum) and levels of ambient concentration of PM10 (PM10), while accounting for the temporal trend. For 365 days indexed by $i = 1, \ldots, 365$, the scatterplots of $\log(D_i)$ versus the three predictors as well as the day index i are shown in Fig. 3. Nonlinear temporal effect is clearly seen in the lower-right panel. We fit the following model to the data:

$$\log(D_i) = m_\mu(i) + \alpha_1 \text{Temp}_i + \alpha_2 \text{Hum}_i + \alpha_3 \text{PM10}_i + \varepsilon_i,$$
$$\varepsilon_i = \exp\{m_\nu(i) + \beta_1 \text{Temp}_i + \beta_2 \text{Hum}_i + \beta_3 \text{PM10}_i\}\epsilon_i.$$

The estimated non-parametric components in the mean function by different methods are shown on the top row of Fig. 4, with $\widehat{m}_\mu$ in the top-left and $\widehat{m}_{\mu,\text{wls}}$ in the top-right. Both estimated curves show a dip in the middle of the year. The estimated linear coefficients involved in the weighted estimate of the mean function are $\widehat{\boldsymbol{\alpha}}_{\text{wls}} = (-0.18, -0.10, 0.19)^{\mathrm{T}}$. The estimated non-parametric components in the variance function by different methods are shown on the middle row of Fig. 4, with $\widehat{m}_\nu$ in the middle-right, $\widehat{m}_{\nu,R}$ in the middle-middle and $\widehat{m}_{\nu,\text{wls}}$ in the middle-right. In contrast to that in the mean function, all the three estimates of the non-parametric components in the variance function show a bump in the middle of the year. The estimated linear coefficients involved in the weighted estimate of the variance function $\widehat{\boldsymbol{\beta}}_{\text{wls}} = (-1.56, -1.72, -2.34)^{\mathrm{T}}$. Examining the signs of these coefficients for the mean and the variance, we see that higher temperature and higher humidity is associated with lower death rate and lower variance of death rate, while higher level of PM10 is associated with higher death rate but lower variance of death rate.

Finally, on the bottom row of Fig. 4, we show the plots of $\log(\widehat{R}_i)$ versus $\widehat{\nu}_i = \widehat{m}_\nu(i) + \widehat{\beta}_1\text{Temp}_i + \widehat{\beta}_2\text{Hum}_i + \widehat{\beta}_3\text{PM10}_i$. In the bottom-left, the estimates are obtained from the initial estimates of the variance function (E2), while from the partially estimates in the bottom-middle (E4) and from the weighted estimates in the bottom-right (E5). The solid line in each figure on the bottom row is the straight line going through origin with slope one, which serves as a reference since we expect that $\log(\hat{R}_i)$ would be roughly linear against $\widehat{\nu}_i$ if the model fit were reasonable. And we see that our model fit the data reasonably well. In practice, this step can be used for model checking and diagnosis. Numerically, the sum of squared residuals of $\log(\hat{R}_i) - \hat{\nu}_i$ for the bottom plots are 402.4, 372.4 and 372.2, respectively, indicating slightly better fit for the updated variance estimates.

5 Discussions

In this paper we investigate a broad class of models, variance function generalized partially linear models. The flexibility of such models comes from that the variance is not limited to be a known function of the mean. The models are useful for the settings where estimating the variance function is of its own interest. The models are also useful for the settings where estimating the mean function is of main interest, because taking into account the heteroscedasticity would improve the efficiency of estimating the mean function.

The asymptotic properties of weighted estimators for the mean function have been studied in the literature by many authors and it is well known that in general weighted estimators are more efficient than unweighted estimators. However, the asymptotic properties of weighted estimators for the variance function draw much less attention. In this paper, we investigate the asymptotic properties of weighted estimators for the variance function for this class of models and show that weighted estimators are more efficient that unweighted estimators.

If the number of predictors is large and curse of high-dimensionality is a concern, we should consider variable selection. Fortunately, there is a wealthy of literature on the topic of variable selection in the past two decades and many existing variable selection procedures can be easily extended to our setting. For example, we can consider the penalized profiling procedure proposed in Liang et al. (2010) by adding some sparsity penalties onto the profiling objective functions in (3) and (5), respectively. Alternatively, we can apply the method of least squares approximation proposed by Wang and Leng (2007) straightforwardly to our setting. The method of least squares approximation has been applied successfully for variable selection in generalized partially linear models by Leng et al. (2011).

Finally, we emphasize that a great deal of effort has been put on deciding on which predictors should be the linear components of both the mean and variance functions. Scatterplots of the response variable versus those predictors could help us make such decision for the mean function, as demonstrated in Fig. 3. In this paper, we assume that the same subset of predictors are put in the linear components of both the mean and variance function, but using different structures for mean and variance functions would offer more flexible alternatives. The model (1) can be easily extended to allow

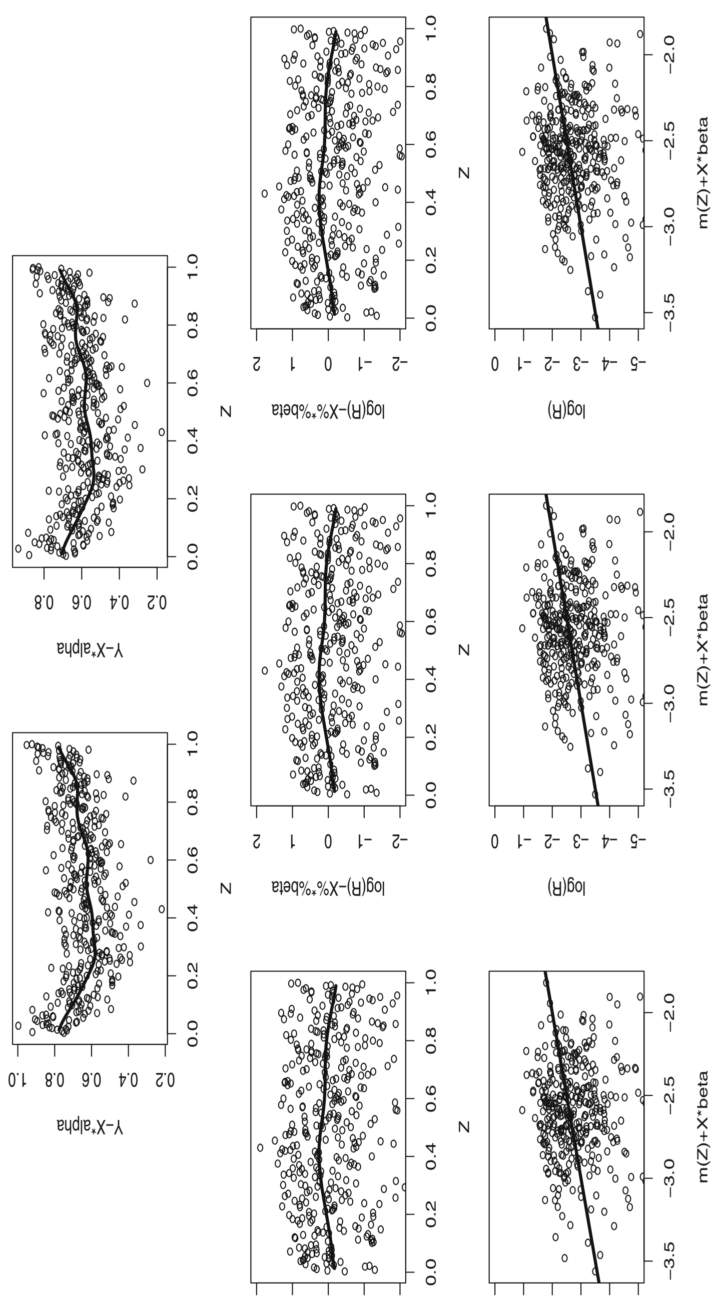

Fig. 4 Estimation results for the air pollution data

different subsets of predictors to be put in the linear components of the mean and variance functions. Scatterplots of the absolute residuals versus those predictors might help us make a decision on which predictors should be in the linear component of the variance function.

Appendix

A.1 Assumptions

Assumption A

A1. The density function $p(z)$ of Z is bounded away from zero and continuously differentiable on [0, 1]. Also, $\boldsymbol{X}$ is in a compact set $\mathfrak{X} \subset \mathbb{R}^p$.
A2. The function $m_\mu(\cdot)$ is twice continuously differentiable.
A3. The kernel $K = K_\mu$ is a bounded and symmetric probability density function, satisfying

$$\kappa_\mu = \int_{-\infty}^{\infty} u^2 K_\mu(u)\mathrm{d}u \neq 0, \quad \int_{-\infty}^{\infty} |u|^i K_\mu(u)\mathrm{d}u < \infty, \quad i = 1, 2, \ldots.$$

A4. The matrix Q_α is positive definite.

Assumption B

B1. The functions $m_\nu(\cdot)$ and $g(\cdot)$ are twice continuously differentiable.
B2. There exist positive constants c and C such that $c \leq g\{\boldsymbol{X\beta}_0 + m_\nu(Z)\} \leq C$.
B3. The kernel $K = K_\nu$ is a bounded and symmetric probability density function, satisfying

$$\kappa_\nu = \int_{-\infty}^{\infty} z^2 K_\nu(z)dz \neq 0, \quad \int_{-\infty}^{\infty} |z|^i K_\nu(z)dz < \infty, \quad i = 1, 2, \ldots.$$

B4. The matrix Q_β is positive definite.

A.2 Proof of Theorem 1

The asymptotic results of $\{\widehat{m}_\mu(z; \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\}$ have been well studied in the literature (Härdle et al. 2000). Here we still give a sketch of the proof, which helps us understand the more complicated proof for Theorem 2. Assume we have shown that $\widehat{\boldsymbol{\alpha}}$ is $\sqrt{n}$-consistent (Speckman 1988).

Consider (13). It suffices to show that, for any $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}_0 + O(1/\sqrt{n})$,

$$\widehat{m}_\mu(z; \boldsymbol{\alpha}^*) - m_\mu(z) = \frac{1}{n}\sum_{i=1}^{n} K_h(Z_i - z)\frac{\varepsilon_i}{p(z)} - E(\boldsymbol{X}^{\mathrm{T}}|Z = z)(\boldsymbol{\alpha}^* - \boldsymbol{\alpha}) + b_\mu(z)\kappa_\mu h^2 + o_p(1/\sqrt{n}), \quad (23)$$

where $o_p(1/\sqrt{n})$ is uniform over z. Noting that $\widehat{m}_\mu(z;\boldsymbol{\alpha}^*)$ is the solution to equation

$$\sum_{i=1}^n \left(Y_i - \widehat{\boldsymbol{\xi}} - \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\alpha}^*\right) K_h(Z_i - z) = 0,$$

we have

$$\widehat{m}_\mu(z;\boldsymbol{\alpha}^*) = \left(\sum_{i=1}^n K_h(Z_i - z)Y_i - \sum_{i=1}^n K_h(Z_i - z)\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\alpha}^*\right) \Big/ \sum_{i=1}^n K_h(Z_i - z).$$

The right-hand-side of the above equation equals

$$\left(\sum_{i=1}^n K_h(Z_i - z)m_\mu(Z_i) - \sum_{i=1}^n K_h(Z_i - z)\boldsymbol{X}_i^{\mathrm{T}}(\boldsymbol{\alpha}^* - \boldsymbol{\alpha}_0) + \sum_{i=1}^n K_h(Z_i - z)\varepsilon_i\right) \Big/ \sum_{i=1}^n K_h(Z_i - z).$$

It is well-known that, as $nh^4 \to \infty$ and $nh^6 \to 0$,

$$\sum_{i=1}^n K_h(Z_i - z)m_\mu(Z_i) \Big/ \sum_{i=1}^n K_h(Z_i - z) = m_\mu(z) + b_\mu(z)\kappa_\mu h^2 + o_p(1/\sqrt{n}),$$
$$\sum_{i=1}^n K_h(Z_i - z)\boldsymbol{X}_i \Big/ \sum_{i=1}^n K_h(Z_i - z) = E(\boldsymbol{X}|Z = z) + o_p(1),$$

where $o_p(\cdot)$ is uniform over $z \in [0, 1]$. Then (23) follows.

Next consider (14). By (23), $\widehat{\boldsymbol{\alpha}}$ of (3) is the solution to

$$\sum_{i=1}^n \widetilde{\boldsymbol{X}}_i \left(Y_i - \widehat{m}_\mu(Z_i;\widehat{\boldsymbol{\alpha}}) - \boldsymbol{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}\right) = \mathbf{0}.$$

By (13), the above equation becomes

$$\sum_{i=1}^n \widetilde{\boldsymbol{X}}_i\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) = \sum_{i=1}^n \widetilde{\boldsymbol{X}}_i\varepsilon_i - \sum_{i=1}^n b_\mu(Z_i)\widetilde{\boldsymbol{X}}_i\kappa_2 h^2 - \sum_{i=1}^n \widetilde{\boldsymbol{X}}_i \left(\frac{1}{n}\sum_{j=1}^n K_h(Z_j - Z_i)\frac{\varepsilon_j}{p(Z_j)}\right).$$

The last term equals $\sum_{i=1}^{n} \varepsilon_i \left(\frac{1}{n} \sum_{j=1}^{n} K_h(Z_j - Z_i) \widetilde{\boldsymbol{X}}_j / p(Z_i) \right)$ and $n^{-1} \sum_{j=1}^{n} K_h(Z_j - z) \widetilde{\boldsymbol{X}}_j \big/ p(z) = o_p(1)$. In the next term to the last, $\sum_{i=1}^{n} b_\mu(Z_i) \widetilde{\boldsymbol{X}}_i = O_p(\sqrt{n})$. Then (14) follows from $\sum_{i=1}^{n} \widetilde{\boldsymbol{X}}_i \widetilde{\boldsymbol{X}}_i^{\mathrm{T}} \big/ n \to Q_\alpha$.

A.3 Proof of Theorem 2

In what follows, we assume that

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + O_p(n^{-1/2}) \quad \text{and} \quad \sup_z |\widehat{m}_\nu(z; \boldsymbol{\beta}^*) - m_\nu(z)| = o_p(n^{-1/4}), \tag{24}$$

for any $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0 + O(n^{-1/2})$. Follow the same arguments as in the proof of Propositions 1 and 2 in Severini and Staniswalis (1994), we can prove such consistency. However, such a proof would be long, detailed, and essentially noninformative. Therefore, as in Davidian and Carroll (1987) and Lian et al. (2015), we skip such proof and assume that we have shown the above consistency.

To show (15), it suffices to show that, for any $\boldsymbol{\beta}^* = \boldsymbol{\beta} + O(1/\sqrt{n})$,

$$\begin{aligned}
\widehat{m}_\nu(z, \boldsymbol{\beta}^*) - m_\nu(z) = &\frac{1}{n} \sum_{i=1}^{n} \frac{K_h(Z_i - z) G_i^{(1)} \delta_i}{p(z) E\{G^{(1)2} | Z = z\}} \\
&- \left[\frac{E\{G^{(1)2} \boldsymbol{X}^{\mathrm{T}} | Z = z\}}{E\{G^{(1)2} | Z = z\}} \right] (\boldsymbol{\beta}^* - \boldsymbol{\beta}_0) \\
&+ \kappa_2 b_\nu(z) h^2 - \left[\frac{E\{G^{(1)} D \widetilde{\boldsymbol{X}}^{\mathrm{T}} | Z = z\}}{E\{G^{(1)2} | Z = z\}} \right] (\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \\
&+ o_p(1/\sqrt{n}),
\end{aligned} \tag{25}$$

where $o_p(1/\sqrt{n})$ is uniform over z. Given $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, the minimizer of (4) is the solution to equation

$$\sum_{i=1}^{n} K_h(Z_i - z) \left(\widehat{R}_i - g\left(\widehat{\zeta} + \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}^* \right) \right) g^{(1)} \left(\widehat{\zeta} + \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}^* \right).$$

By (24) and the assumption of h, the left hand side is

$$\begin{aligned}
&n^{-1} \sum K_h(Z_i - z)(\widehat{R}_i - R_i) G_i^{(1)} + n^{-1} \sum K_h(Z_i - z)(R_i - G_i) G_i^{(1)} \\
&\quad + n^{-1} \sum K_h(Z_i - z) \left(G_i - g\left(\widehat{\zeta} + \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}^* \right) \right) g^{(1)} \left(\widehat{\zeta} + \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}^* \right) + o_p(1/\sqrt{n}),
\end{aligned} \tag{26}$$

where $o_p(\cdot)$ is uniform over z. Consider the first term of (26). We give an expression for $\widehat{R}_i - R_i$. Using an identity in Knight (1998, p. 758),

$$\widehat{R}_i - R_i = -\widehat{S}_i\{I_{(\varepsilon_i>0)} - I_{(\varepsilon_i\le 0)}\} + 2\int_0^{\widehat{S}_i}\{I_{(\varepsilon_i<s)} - I_{(\varepsilon_i\le 0)}\}\mathrm{d}s,$$

where $\widehat{S}_i = \big(\widehat{m}_\mu(Z_i;\widehat{\boldsymbol{\alpha}}) + \boldsymbol{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}\big) - \big(m_\nu(Z_i) + \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\alpha}_0\big)$. By Theorem 1, $\widehat{R}_i - R_i$ equals

$$-D_i\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) - \frac{D_i}{n}\sum_{j=1}^{n}K_h(Z_j - Z_i)\frac{\varepsilon_j}{p(Z_i)}$$
$$+2\int_0^{\widehat{S}_i}\{I_{(\varepsilon_i<s)} - I_{(\varepsilon_i\le 0)}\}\mathrm{d}s + o_p(1/\sqrt{n}). \tag{27}$$

Then, by (27), the first term of (26) is equal to

$$-\frac{1}{n}\sum_{i=1}^{n}D_iG_i^{(1)}\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}K_h(Z_i - z)(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$$
$$-\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}D_iG_i^{(1)}K_h(Z_i - z)K_h(Z_j - Z_i)\frac{\varepsilon_j}{p(Z_i)}$$
$$+\frac{2}{n}\sum_{i=1}^{n}K_h(Z_i - z)G_i^{(1)}\int_0^{\widehat{S}_i}\{I_{(\varepsilon_i<s)} - I_{(\varepsilon_i\le 0)}\}\mathrm{d}s. \tag{28}$$

Following the same arguments as in the proof of Theorem 3.2 in Lian et al. (2015), we have

$$n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}D_iG_i^{(1)}K_h(Z_i - z)K_h(Z_j - Z_i)\frac{\varepsilon_j}{p(Z_i)} = o_p(n^{-1/2}),$$
$$n^{-1}\sum_{i=1}^{n}K_h(Z_i - z)G_i^{(1)}\int_0^{\widehat{S}_i}\{I_{(\varepsilon_i<s)} - I_{(\varepsilon_i\le 0)}\}\mathrm{d}s = o_p(n^{-1/2}).$$

Then the first term of (26) equals $-n^{-1}\sum_{i=1}^{n}K_h(Z_i - z)G_i^{(1)}D_i\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + o_p(n^{-1/2})$, which is further equal to

$$-E\{G^{(1)}D\widetilde{\boldsymbol{X}}^{\mathrm{T}}|Z = z\}p(z)(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + o_p(n^{-1/2}).$$

The second term of (26) equals $n^{-1}\sum_{i=1}^{n}K_h(Z_i - z)G_i^{(1)}\delta_i$. By Taylor expansion, the third term of (26) equals

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} K_h(Z_i - z)\left[g\left(m_\nu(Z_i) + \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right) - g\left(m_\nu(z) + \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right)\right] g^{(1)} \\
&\times\left(m_\nu(z) + \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right) - \frac{1}{n}\sum_{i=1}^{n} K_h(Z_i - z) g^{(1)2}\left(m_\nu(z) + \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right) \\
&\times\left[(\widehat{\zeta} - m_\nu(z)) + \boldsymbol{X}_i^{\mathrm{T}}(\boldsymbol{\beta}^* - \boldsymbol{\beta}_0)\right] + o_p(n^{-1/2}). \qquad (29)
\end{aligned}
$$

Following the similar arguments of deriving $b_\mu(z)$, we can show that the first term of (29) equals $\kappa_\nu b_\nu(z)h^2 + o_p(n^{-1/2})$, where if local constant regression is used in (4),

$$
b_\nu(z) = \frac{m^{(1)}(z)p^{(1)}(z)}{p(z)} + \frac{m_\nu^{(2)}(z)}{2} + \frac{E\{G^{(2)}G^{(1)}|Z=z\}}{2E\{G^{(1)2}|Z=z\}} m_\nu^{(1)}(z),
$$

while if local linear regression is used in (4),

$$
b_\nu(z) = \frac{m_\nu^{(2)}(z)}{2} + \frac{E\{G^{(2)}G^{(1)}|Z=z\}}{2E\{G^{(1)2}|Z=z\}} m_\nu^{(1)}(z).
$$

And the second term of (29) equals

$$
-E\{G^{(1)2}|Z=z\}p(z)\left(\widehat{\boldsymbol{\zeta}} - m_\nu(z)\right) - E\{G^{(1)2}\boldsymbol{X}^{\mathrm{T}}|Z=z\}p(z)(\boldsymbol{\beta}^* - \boldsymbol{\beta}_0) + o_p(n^{-1/2}).
$$

Therefore, combining final expressions of those three terms of (26), (25) follows.

Next, consider (16). By (25), minimizer $\widehat{\boldsymbol{\beta}}$ of (5) is the solution to

$$
\sum_{i=1}^{n} \breve{X}_i\left(\widehat{R}_i - g\left(\widehat{m}_\nu(Z_i;\widehat{\boldsymbol{\beta}}) + \boldsymbol{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\right)\right) g^{(1)}\left(\widehat{m}_\nu(Z_i;\widehat{\boldsymbol{\beta}}) + \boldsymbol{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\right) = \mathbf{0}.
$$

Similarly, using Taylor expansion and the assumptions of h, we know that the left hand side is

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^{n} \breve{\boldsymbol{X}}_i(\widehat{R}_i - R_i)G_i^{(1)} + n^{-1}\sum_{i=1}^{n} \breve{\boldsymbol{X}}_i(R_i - G_i)G_i^{(1)} \\
&\quad - n^{-1}\sum_{i=1}^{n} \breve{\boldsymbol{X}}_i\left(g\left(\widehat{m}_\nu(Z_i;\widehat{\boldsymbol{\beta}}) + \boldsymbol{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\right) - G_i\right) G_i^{(1)} + o_p(n^{-1/2}). \qquad (30)
\end{aligned}
$$

For the first term of (30), by (27), it is equal to

$$
\begin{aligned}
&-n^{-1}\sum_{i=1}^{n} G_i^{(1)} D_i \breve{\boldsymbol{X}}_i \widetilde{\boldsymbol{X}}_i^{\mathrm{T}}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) - n^{-2}\sum_{i=1}^{n} G_i^{(1)} D_i \breve{\boldsymbol{X}}_i \sum_{j=1}^{n} K_h(Z_j - Z_i)\frac{\varepsilon_j}{p(Z_i)} \\
&+ n^{-1}\sum_{i=1}^{n} 2G_i^{(1)}\breve{\boldsymbol{X}}_i \int_0^{\widehat{S}_i} \{I_{(\varepsilon_i < s)} - I_{(\varepsilon_i \le 0)}\} \mathrm{d}s.
\end{aligned}
$$

Following the same arguments as in the proof of Theorem 3.2 in Lian et al. (2015), we have

$$
\begin{aligned}
& n^{-2}\sum_{i=1}^{n} D_i G_i^{(1)} \breve{\boldsymbol{X}}_i \sum_{j=1}^{n} K_h(Z_j - Z_i)\frac{\varepsilon_j}{p(Z_i)} \\
& \quad = n^{-1}\sum_{i=1}^{n} \varepsilon_i E\left\{D_i G_i^{(1)} \breve{\boldsymbol{X}}_i | Z_i\right\} + o_p(n^{-1/2}), \\
& n^{-1}\sum_{i=1}^{n} G_i^{(1)} \breve{\boldsymbol{X}}_i \int_0^{\widehat{S}_i} \{I_{(\varepsilon_i < s)} - I_{(\varepsilon_i \le 0)}\} \mathrm{d}s = o_p(n^{-1/2}).
\end{aligned}
$$

Then the first term of (30) equals

$$
-E\{G^{(1)} D \breve{\boldsymbol{X}} \widetilde{\boldsymbol{X}}^{\mathrm{T}}\}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) - n^{-1}\sum_{i=1}^{n} \varepsilon_i E\left\{D_i G_i^{(1)} \breve{\boldsymbol{X}}_i | Z_i\right\} + o_p(n^{-1/2}).
$$

The second term of (30) equals $n^{-1}\sum_{i=1}^{n} \delta_i G_i^{(1)} \breve{\boldsymbol{X}}_i$. The third term of (30) equals

$$
-n^{-1}\sum_{i=1}^{n} \breve{\boldsymbol{X}}_i G_i^{(1)2}\left(\left(\widehat{m}_v(Z_i, \widehat{\boldsymbol{\beta}}) - m_v(Z_i)\right) + \boldsymbol{X}_i^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right) + o_p(n^{-1/2}),
$$

which, by (15), is further equal to

$$
\begin{aligned}
& -\frac{1}{n}\sum_{i=1}^{n} \frac{G_i^{(1)2} \breve{\boldsymbol{X}}_i}{p(Z_i) E\left\{G_i^{(1)2} | Z_i\right\}} \frac{1}{n}\sum_{j=1}^{n} K_h(Z_j - Z_i) G_j^{(1)} \delta_j \\
& + \frac{1}{n}\sum_{i=1}^{n} G_i^{(1)2} \breve{\boldsymbol{X}}_i \left[\frac{E\left\{G_i^{(1)2} \boldsymbol{X}_i^{\mathrm{T}} | Z_i\right\}}{E\left\{G_i^{(1)2} | Z_i\right\}}\right](\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
& + \frac{1}{n}\sum_{i=1}^{n} G_i^{(1)2} \breve{\boldsymbol{X}}_i \left[\frac{E\left\{G_i^{(1)} D_i \widetilde{\boldsymbol{X}}_i^{\mathrm{T}} | Z_i\right\}}{E\left\{G_i^{(1)2} | Z_i\right\}}\right](\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \\
& - \frac{1}{n}\sum_{i=1}^{n} G_i^{(1)2} \breve{\boldsymbol{X}}_i \kappa_v b_v(z) h^2 - \frac{1}{n}\sum_{i=1}^{n} G_i^{(1)2} \breve{\boldsymbol{X}}_i \boldsymbol{X}_i^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(1/\sqrt{n}).
\end{aligned}
$$

Noting that $E\{G^{(1)2} \breve{\boldsymbol{X}} | Z = z\} = 0$, we see that the second term to the last in the above formula is $o_p(n^{-1/2})$. Then the third term of (30) equals

$$E\left[\frac{G_i^{(1)2}\boldsymbol{X}_i E\left\{G_i^{(1)} D_i \widetilde{\boldsymbol{X}}_i^{\mathrm{T}}|Z_i\right\}}{E\left\{G_i^{(1)2}|Z_i\right\}}\right](\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0) - E\{(G^{(1)}\breve{\boldsymbol{X}})^{\otimes 2}\}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$$
$$-\frac{1}{n}\sum_{i=1}^{n}\delta_i G_i^{(1)}\left[\frac{E\left\{G_i^{(1)2}\boldsymbol{X}_i|Z_i\right\}}{E\left\{G_i^{(1)2}|Z_i\right\}}\right] + o_p(1/\sqrt{n}),$$

which further equals $-E\{(G^{(1)}\breve{\boldsymbol{X}})^{\otimes 2}\}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+o_p(1/\sqrt{n})$, noting that $E\{G^{(1)2}\breve{\boldsymbol{X}}|Z\} = 0$.

Combining the final expressions of those three terms of (30), $\sqrt{n}Q_\beta(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ equals

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i G_i^{(1)}\breve{\boldsymbol{X}}_i - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i E\left\{D_i G_i^{(1)}\breve{\boldsymbol{X}}_i|Z_i\right\}$$
$$-\sqrt{n}E\{G^{(1)}D\breve{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0) + o_p(1).$$

Thus, by the expression of $\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0$ in Theorem 1, (16) follows.

A.4 Proof of Theorem 3

Define the following score function of parameter $\boldsymbol{\alpha}$ with nuisance parameters $\{\mathcal{H}, m_\mu, G\}$,

$$\Psi(\mathcal{H}, m_\mu, G; \boldsymbol{\alpha}, Y, X) = \left(Y - m_\mu(Z) - X^{\mathrm{T}}\boldsymbol{\alpha}\right)(X - \mathcal{H}(Z))/G^2,$$

where $\mathcal{H}(Z) = E\{X/G^2|Z\}/E\{1/G^2|Z\}$. Following the similar arguments as in the proof of Theorem 3.1 in Lian et al. (2015) and Lemma 5.1 in Newey (1994), we can show that $\widehat{\boldsymbol{\alpha}}_{\mathrm{wls}}$ has the same limit distribution as the solution to the equation

$$\sum_{i=1}^{n}\Psi(\mathcal{H}, m_\mu, G; \boldsymbol{\alpha}, Y_i, X_i).$$

And it can bee seen the solution to the above equation has the same limit distribution of (18). The efficiency of $\widehat{\boldsymbol{\alpha}}_{\mathrm{wls}}$ can be approved by the similar arguments as in the proof of Theorem 3.1 in Lian et al. (2015) and the theory in Bickel et al. (1993). Thus second part of the theorem is proved. The first part of the theorem follows consequently.

A.5 Proof of Theorem 4

First we give an expression of $\widehat{R}_{i,\mathrm{wls}} - R_i$. Because

$$\widehat{R}_{i,\mathrm{wls}} - R_i = -\widehat{S}_{i,\mathrm{wls}}\{I_{(\varepsilon_i>0)} - I_{(\varepsilon_i\le 0)}\} + 2\int_0^{\widehat{S}_{i,\mathrm{wls}}}\{I_{(\varepsilon_i<s)} - I_{(\varepsilon_i\le 0)}\}\mathrm{d}s,$$

where $\widehat{S}_{i,\text{wls}} = \big(\widehat{m}_{\mu,\text{wls}}(Z_i;\widehat{\boldsymbol{\alpha}}_{\text{wls}}) + \boldsymbol{X}_i^{\text{T}}\widehat{\boldsymbol{\alpha}}_{\text{wls}}\big) - \big(m_\nu(Z_i) + \boldsymbol{X}_i^{\text{T}}\boldsymbol{\alpha}_0\big)$. By Theorem 3, the dominant term of $\widehat{R}_{i,\text{wls}} - R_i$ is $-D_i\check{\boldsymbol{X}}_i^{\text{T}}(\widehat{\boldsymbol{\alpha}}_{\text{wls}} - \boldsymbol{\alpha}_0)$.

Then the proof of (19) is identical to that of (15), except that $\widehat{\boldsymbol{\alpha}}_{\text{wls}}$ replaces $\widehat{\boldsymbol{\alpha}}$ everywhere, where the asymptotic expression of $\sqrt{n}(\widehat{\boldsymbol{\alpha}}_{\text{wls}} - \boldsymbol{\alpha}_0)$ is obtained in Theorem 3.

A.6 Proof of Theorem 5

As in Sect. A.3, assume that $\widehat{\boldsymbol{\beta}}_{\text{wls}} = \boldsymbol{\beta}_0 + O_p(n^{-1/2})$ and $\sup_z |\widehat{m}_{\nu,\text{wls}}(z,\boldsymbol{\beta}^*) - m_\nu(z)| = o_p(n^{1/4})$ for any $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0 + O_p(n^{-1/2})$.

The proof of (21) is similar to that of (15), except that $1/G_i^2$ is added to each term and $b_{\nu,\text{wls}}(z)$ is defined differently. If local constant regression is used in (11),

$$b_{\nu,\text{wls}}(z) = \frac{m^{(1)}(z)p^{(1)}(z)}{p(z)} + \frac{m_\nu^{(2)}(z)}{2} + \frac{E\{G^{(2)}G^{(1)}/G^2|Z=z\}}{2E\{G^{(1)2}/G^2|Z=z\}}m_\nu^{(1)}(z),$$

while if local linear regression is used in (11),

$$b_{\nu,\text{wls}}(z) = \frac{m_\nu^{(2)}(z)}{2} + \frac{E\{G^{(2)}G^{(1)}/G^2|Z=z\}}{2E\{G^{(1)2}/G^2|Z=z\}}m_\nu^{(1)}(z).$$

The proof of (22) is similar to that of (16), except that $1/G_i^2$ is added to each term and $\check{\boldsymbol{X}}$ is replaced by $\bar{\boldsymbol{X}}$.

References

Bickel, P. (1978). Using residuals robustly: Tests for heteroscedasticity, nonlinearity. *The Annals of Statistics*, *6*, 266–291.

Bickel, P. J., Klaasen, C. A. J., Ritov, Y., Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models*. Baltimore: The Johns Hopkins University Press.

Box, G., Hill, W. (1974). Correcting inhomogeneity of variance with power transformation weighting. *Technometrics*, *16*, 385–389.

Box, G., Meyer, D. (1986). An analysis for unreplicated fractional factorials. *Technometrics*, *28*, 11–18.

Cai, T. T., Wang, L. (2008). Adaptive variance function estimation in heteroscedastic nonparametric regression. *Annals of Statistics*, *36*(5), 2025–2054.

Carroll, R. J. (1982). Adapting for heteroscedasticity in linear models. *The Annals of Statistics*, *10*, 1224–1233.

Carroll, R. J. (2003). Variances are not always nuisance parameters. *Biometrics*, *59*(2), 211–220.

Carroll, R. J., Fan, J., Gijbels, I., Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*(438), 477–489.

Carroll, R. J., Härdle, W. (1989). Second order effects in semiparametric weighted least squares regression. *Statistics*, *2*, 179–186.

Carroll, R. J., Ruppert, D. (1982). Robust estimation in heteroscedasticity linear models. *The Annals of Statistics*, *10*, 429–441.

Carroll, R., Ruppert, D. (1988). *Transformation and weighting in regression*. New York: Chapman & Hall.

Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *The Annals of Statistics*, *16*, 136–146.

Davidian, M., Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association*, *82*, 1079–1091.

Engle, R. F., Granger, C. W. J., Rice, J., Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, *81*(394), 310–320.
Fuller, W., Rao, J. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. *The Annals of Statistics*, *6*, 1149–1158.
Hall, P., Carroll, R. J. (1989). Variance function estimation in regression—The effect of estimating the mean. *Journal of the Royal Statistical Society Series B-Methodological*, *51*(1), 3–14.
Härdle, W., Mammen, E., Müller, M. (1998). Testing parametric versus semiparametric modeling in generalized linear models. *Journal of the American Statistical Association*, *93*, 1461–1474.
Härdle, W., Liang, H., Gao, J. (2000). *Partially linear models*. Heidelberg: Physica-Verlag.
Härdle, W., Müller, M., Sperlich, S., Werwatz, A. (2004). *Nonparametric and semiparametric models*. New York: Springer.
Heckman, N. E. (1986). Spline smoothing in partly linear models. *Journal of the Royal Statistical Society, Series B*, *48*, 244–248.
Hunsberger, S., Albert, P. S., Follmann, D. A., Suh, E. (2002). Parametric and semiparametric approaches to testing for seasonal trend in serial count data. *Biostatistics*, *3*, 289–298.
Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *The Annals of Statistics*, *26*(2), 755–770.
Leng, C., Liang, H., Martinson, N. (2011). Efficient variable selection for semiparametric generalized partially linear models with applications in study of condom use for HIV patients. *Statistics in Medicine*, *30*, 2015–2027.
Lian, H., Liang, H., Carroll, R. J. (2015). Variance function partially linear single-index models. *Journal of the Royal Statistical Society, Series B*, *77*, 171–194.
Liang, F., Paulo, R., Molina, G., Clyde, M. A., Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, *103*(481), 410–423.
Liang, H., Härdle, W., Carroll, R. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics*, *27*, 1519–1535.
Liang, H., Liu, X., Li, R., Tsai, C. L. (2010). Estimation and testing for partially linear single-index models. *The Annals of Statistics*, *38*, 3811–3836.
Ma, Y., Zhu, L. (2012). Doubly robust and efficient estimators for heteroscedastic partially linear single-index models allowing high dimensional covariates. *Journal of the Royal Statistical Society*: *Series B*, *75*, 305–322.
Ma, Y., Chiou, J.-M., Wang, N. (2006). Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika*, *93*, 75–84.
Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica*, *62*, 1349–1382.
Opsomer, J., Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *The Annals of Statistics*, *25*, 186–211.
Prada-Sánchez, J., Febrero-Bande, M., Cotos-Yáñez, T., González-Manteiga, W., Bermúdez-Cela, J., Lucas-Dominguez, T. (2000). Prediction of so_2 pollution incidents near a power station using partially linear models and an historical matrix of predictor-response vectors. *Environmetrics*, *11*, 209–225.
Robinson, P. M. (1988). Root n-consistent semiparametric regression. *Econometrica*, *56*, 931–954.
Severini, T. A., Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *Journal of the American Statistical Association*, *89*, 501–511.
Speckman, P. E. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B*, *50*, 413–436.
Teschendorff, A. E., Widschwendter, M. (2012). Differential variability improves the identification of cancer risk markers in dna methylation studies profiling precursor cancer lesions. *Bioinformatics*, *28*, 1487–1494.
Thomas, L., Stefanski, L. A., Davidian, M. (2012) . *Measurement error model methods for bias reduction and variance estimation in logistic regression with estimated variance predictors*. Technical report, North Carolina State University.
Wang, H., Leng, C. (2007). Unified lasso estimation via least squares approximation. *Journal of the American Statistical Association*, *101*, 1418–1429.
Wang, L., Liu, X., Liang, H., Carroll, R. (2011). Estimation and variable selection for generalized additive partial linear models. *The Annals of Statistics*, *39*, 1827–1851.
Western, B., Bloome, D. (2009). Variance function regressions for studying inequality. *Sociological Methodology*, *39*, 293–326.

Xia, Y. C., Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis*, *97*, 1162–1184.

Yatchew, A., No, J. A. (2001). Household gasoline demand in Canada. *Econometrica*, *69*(6), 1697–1709.

Zeger, S., Diggle, P. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, *50*, 689–699.