EFFICIENT ESTIMATION OF PARAMETERS AND NONPARAMETRIC FUNCTIONS IN GENERALIZED SEMI/NON-PARAMETRIC REGRESSION MODELS

By Ling Zhou, Huazhen Lin,* Kani Chen, and Hua Liang,†

Southwestern University of Finance and Economics, HKUST, and George Washington University

Efficiency of estimation for the parameters in semiparametric models has been widely and well studied in the literature (Bickel *et al.*, 1993). In this paper, we study efficient estimators for both parametric and nonparametric parts in a class of generalized semi/non-parametric regression models, which cover commonly used semiparametric models such as partially linear models, partially linear single index models, and two-sample semiparametric models. We integrate the local linear regression and maximum likelihood principle after we nonparametrically estimate the the density of the model errors for estimating the regression parameters and nonparametric regression functions. The proposed estimators of the parameters are consistent and asymptotically normal, and optimal ??? in the sense that the estimator of the linear functional form, and the estimators of the nonparametric function are also shown to be *semiparametrically efficient* in the sense of Bickel *et al.* (1993). We also suggest an algorithm to achieve the maximization procedure. We conduct simulation experiments to evaluate the numerical performance of proposed methods and analyze a real data example as an illustration.

*Lin's research was partially supported by the National Natural Science Foundation of China (Nos. 11571282 and 11528102).

†Liang's research was partially supported by NSF grant DMS-1418042, and by Award Number 11529101, made by National Natural Science Foundation of China.

AMS 2000 subject classifications: Primary 62G20, 62G05; secondary 62G10, 62J02, 62F12

Keywords and phrases: Local linear method, Maximum likelihood function, Semiparametric efficiency, Semi/non-parametric regression models, Undersmoothing

1. Introduction. Consider a generalized semiparametric regression model

$$Y = m\{g(\boldsymbol{\vartheta}'\mathbf{X}_1), \boldsymbol{\gamma}'\mathbf{X}_2, \mathbf{X}_3; \boldsymbol{\lambda}\} + \varepsilon, \tag{1.1}$$

where $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2', \mathbf{X}_3')'$, $g(\cdot)$ is the vector of unknown regression functions, $\boldsymbol{\vartheta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are the unknown regression parameters, $m(\cdot)$ is an known function and determined by the type of the response Y, ε is the random error with *unknown* distribution f.

Model (1.1) is flexible to cover a variety of commonly used semiparametric models. When $m\{g(\boldsymbol{\vartheta}'\mathbf{X}_1), \boldsymbol{\gamma}'\mathbf{X}_2, \mathbf{X}_3; \boldsymbol{\lambda}\} = m\{g(\boldsymbol{\vartheta}'\mathbf{X}_1) + \boldsymbol{\gamma}'\mathbf{X}_2\}$, (1.1) is simply the generalized partially linear single-index model (Carroll *et al.*, 1997; Xia and Härdle, 2006; Xia and Li, 1999), and further reduces to the partially linear model when $m(x) = x$ and $\mathbf{X}_1$ is one-dimension (Engle *et al.*, 1986; Härdle, Liang and Gao, 2000; Speckman, 1988). When $m\{g(\boldsymbol{\vartheta}'\mathbf{X}_1), \boldsymbol{\gamma}'\mathbf{X}_2, \mathbf{X}_3; \boldsymbol{\lambda}\} = g(\boldsymbol{\vartheta}'\mathbf{X}_1)$, (1.1) reduces to the single index model (Härdle and Stoker, 1989; Horowitz and Härdle, 1996; Ichimura, 1993; Li, 1991; Powell, Stock and Stoker, 1989) and further reduces the usual nonparametric regression with $\mathbf{X}_1$ being one-dimension. When $m\{g(\boldsymbol{\vartheta}'\mathbf{X}_1), \boldsymbol{\gamma}'\mathbf{X}_2, \mathbf{X}_3; \boldsymbol{\lambda}\} = m\{g(X_1)'\mathbf{X}_3 + \boldsymbol{\gamma}'\mathbf{X}_2\}$ with X_1 being one-dimension, (1.1) reduces to the varying coefficient or semi-varying coefficient regression model (Fan, Lin and Zhou, 2006; Fan and Zhang, 1999; Hastie and Tibshirani, 1993; Xia, Zhang and Tong, 2004). When $m\{g(\boldsymbol{\vartheta}'\mathbf{X}_1), \boldsymbol{\gamma}'\mathbf{X}_2, \mathbf{X}_3; \boldsymbol{\lambda}\} = (1 - X_3)g(X_1) + X_3\lambda g(X_1)$ with X_3 being 0 or 1 and X_1 being one-dimension, (1.1) reduces to the models studied by Härdle and Marron (1990), and the problem refers to a two sample problem. Another interesting example includes the model we will use in Section 5 to explore the electricity distribution costs.

In semi-parametric models, of great interest is the development of estimators of the parameters and consideration of efficiency of these estimators. There are a vast literature on this topic. The parameters $\boldsymbol{\vartheta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are generally estimated by the least squares criterion since $f(\cdot)$ is unknown, and the function $g(\cdot)$ is estimated by *spline smoothing methods* (Engle *et al.*, 1986;

Eubank, 1999; Green, Jennison and Seheult, 1985; Green and Silverman, 1993; Wahba, 1984) or *kernel smoothing methods* (Fan and Gijbels, 1996; Härdle, Liang and Gao, 2000; Speckman, 1988) based on the least squares criterion.

The least squares (LS) criterion based estimator is efficient when the error ε is normally distributed but inefficient if the error ε deviates from the normal distribution. Some adjustments have been suggested, for example, the kernel M-smoother (Härdle, 1989), median smoothing (Tukey, 1977), locally weighted regression (Cleveland, 1979; Stone, 1977), and the local least absolute method (Wang and Scott, 1994). These methods are mathematically convenient or robust, and easily implemented. However they also have their own limitation such as these methods are not optimal.

Cuzick *et al.* (1992) derived an efficient estimator of $\boldsymbol{\gamma}$ when the error distribution is unknown for partially linear models. Schick (1993) developed a procedure for constructing efficient estimates for the parameters in the parametric regression model and the partially linear regression model with the unknown error distribution via deriving the efficient influence functions of least dispersed regular estimators. Brown and Newey (1998) focused on efficient semiparametric estimation of a expectation for the model $y = m(\varepsilon, x, \boldsymbol{\gamma})$ with known function $m(\cdot)$, unknown parameter $\boldsymbol{\gamma}$ and unknown distribution of error ε. Ma, Chiou and Wang (2006) derived a semiparametrically efficient estimator of $\boldsymbol{\gamma}$ using a constant weight function for the partially linear model when the error distribution is unknown. There are also some works focus on the efficient estimator for the case with known error distribution. Wong and Severini (1991) and Newey (1994) considered the semiparametric efficient estimates of $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ for the exponential family model. Maity, Ma and Carroll (2007) studied the semiparametric efficient estimator of the population quantities for a semiparametric regression model when the error distribution is known.

All aforementioned work mainly focuses on efficient estimation of the parameters. It is worth-while that $g(\cdot)$ is generally treated as a nuisance parameter and its estimator is regarded as a bridge for estimating the parameters. However, the nonparametric function $g(\cdot)$ itself is sometimes of great interest. For example, when Engle *et al.* (1986) introduced the partially linear model to explore the relationship between electricity usage and temperature, their primary interest was in the nonparametric function. In our real data example, our primary interest is also the nonparametric function. So it is also of great interest in efficiently estimating the nonparametric function $g(\cdot)$.

Recently, Cheng and Shang (2015) considered joint inference of $g(\cdot)$ and $\boldsymbol{\gamma}$ for a semiparametric regression model with partially linear structure $E(Y \mid X_1, \mathbf{X}_2) = m\{g(X_1) + \boldsymbol{\gamma}'\mathbf{X}_2\}$ by using likelihood when the data follows the exponential family or quasi-likelihood where only the relationship between the conditional mean and variance is specified. However, they did not discuss the efficient property of their estimators.

The purpose of this paper is to introduce an estimation method using an optimal criterion function for developing semiparametrically efficient estimators in the sense of Bickel *et al.* (1993) for the parameters $(\boldsymbol{\vartheta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ and the nonparametric regression function $g(\cdot)$ when the error distribution is unknown.

This paper is organized as follows. In Section 2, we give the criterion function and the estimators of $\boldsymbol{\vartheta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\lambda}$ and $g(\cdot)$. We derive the asymptotic properties including consistency, asymptotic normality and semiparametric efficiency of the proposed estimator in Section 3. We present simulation results of the robustness and efficiency of the estimator in Section 4. Finally, we illustrate our methods using two real examples in Section 5.

2. Estimation. It is worth pointing out that Model (1.1) presents a general structure, and the identifiability condition depends on specific model formula. Basically, let the mean of the error equal to 0, and the intercept term be excluded from linear parts if there exists linear structure and $\|\boldsymbol{\vartheta}\| = 1$, $\vartheta_1 > 0$, where ϑ_1 is the first component of $\boldsymbol{\vartheta}$.

If the unknown function $g(\cdot)$ is parameterized, the parameters can be estimated by minimizing

$$l\{\boldsymbol{\vartheta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, g(\cdot)\} = \sum_{i=1}^{n} \rho[Y_i - m\{g(\boldsymbol{\vartheta}'\mathbf{X}_{i1}), \boldsymbol{\gamma}'\mathbf{X}_{i2}, \mathbf{X}_{i3}; \boldsymbol{\lambda}\}], \tag{2.2}$$

where ρ is a given loss function, for example, the least squares and least absolute criteria; i.e., $\rho(t) = t^2$ and $\rho(t) = |t|$. When the form of the unknown function $g(\cdot)$ is not available, we can only rely on its qualitative trait. Assume that $g(\cdot)$ is smooth so that Taylor's expansion is applicable: for each given x and w around x,

$$g(w) \approx g(x) + \dot{g}(x)(w - x) \equiv \alpha + \beta(w - x), \tag{2.3}$$

where $\dot{g} = dg(t)/dt$. Substituting (2.3) into (2.2), we obtain the following local ρ-objective function:

$$\sum_{i=1}^{n} \rho[Y_i - m\{\alpha + \beta(\boldsymbol{\vartheta}'\mathbf{X}_{i1} - x), \boldsymbol{\gamma}'\mathbf{X}_{i2}, \mathbf{X}_{i3}; \boldsymbol{\lambda}\}]K_i(x; \boldsymbol{\vartheta}), \tag{2.4}$$

where $K_i(x; \boldsymbol{\vartheta}) = K_h(\boldsymbol{\vartheta}'\mathbf{X}_{i1} - x)$, $K_h(u) = K(u/h)/h$, K is a nonnegative symmetric kernel function with support on $[-1, 1]$ and h is the bandwidth. The minimized values of α and β are the estimates of $g(x)$ and its derivative. The localization makes the estimates of (α, β) free of those observations that satisfy $|X_{i1} - x| > h$ because they are trimmed out by the kernel function $K_i(x; \boldsymbol{\vartheta})$ in the above local objective function. The reasonability of the local ρ-objective function is that the non-local observations do not carry sufficient information about $g(x)$ under the given criterion ρ.

The existing kernel-based method uses the objective functions (2.2) and (2.4) with various functions ρ to estimate $\boldsymbol{\vartheta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\lambda}$ and $g(\cdot)$, respectively. If the density function of the error, f, is known, an ideal choice of the loss function $\rho(\cdot)$ is $-\log f(\cdot)$. It is appealing to consider objective functions

$$\sum_{i=1}^{n} \log f[Y_i - m\{g(\boldsymbol{\vartheta}'\mathbf{X}_{i1}), \boldsymbol{\gamma}'\mathbf{X}_{i2}, \mathbf{X}_{i3}; \boldsymbol{\lambda}\}], \text{ and} \tag{2.5}$$

$$\sum_{i=1}^{n} \log f[Y_i - m\{\alpha + \beta(\boldsymbol{\vartheta}'\mathbf{X}_{i1} - x), \boldsymbol{\gamma}'\mathbf{X}_{i2}, \mathbf{X}_{i3}; \boldsymbol{\lambda}\}]K_i(x; \boldsymbol{\vartheta}), \tag{2.6}$$

for $(\boldsymbol{\vartheta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ and (α, β), respectively. The maximizers of objective functions are then the estimators of the parameters and the function $g(x)$, and the maximum likelihood estimator (MLE) and the local MLE then have the desired optimal properties. A contribution of the paper is that we estimate $\boldsymbol{\vartheta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\lambda}$ and $g(\cdot)$ using the optimal objective functions (2.5) and (2.6) even the error distribution is unknown.

This paper proposes an iterative algorithm to achieve maximization of (2.5) and (2.6) with respect to $\boldsymbol{\vartheta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\lambda}$, α and β when $f(\cdot)$ is unknown. Specifically, we replace the unknown density function $f(\cdot)$ in (2.5) and (2.6) by an estimated one derived in the previous step and we estimate the density function $f(\cdot)$ by the Nadaraya-Watson kernel, given by

$$f_n(w) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_{\hbar} \left[Y_i - m\{g(\boldsymbol{\vartheta}'\mathbf{X}_{i1}), \boldsymbol{\gamma}'\mathbf{X}_{i2}, \mathbf{X}_{i3}; \boldsymbol{\lambda}\} - w \right],$$

where $\mathcal{K}_{\hbar}(\cdot) = \mathcal{K}(\cdot/\hbar)/\hbar$, $\mathcal{K}(\cdot)$ is a nonnegative symmetric kernel function with support on $[-1, 1]$ and $\hbar$ is the bandwidth. It is known that under regularity conditions, we have $\sup_w \|f_n(w) - f(w)\| \to 0$. Hence, we expect that the estimators of $(\boldsymbol{\vartheta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ and $g(x)$ based on the following maximum likelihood and local maximum likelihood, respectively,

$$\sum_{i=1}^{n} \log f_n[Y_i - m\{g(\boldsymbol{\vartheta}'\mathbf{X}_{i1}), \boldsymbol{\gamma}'\mathbf{X}_{i2}, \mathbf{X}_{i3}; \boldsymbol{\lambda}\}], \text{ and} \tag{2.7}$$

$$\sum_{i=1}^{n} \log f_n[Y_i - m\{\alpha + \beta(\boldsymbol{\vartheta}'\mathbf{X}_{i1} - x), \boldsymbol{\gamma}'\mathbf{X}_{i2}, \mathbf{X}_{i3}; \boldsymbol{\lambda}\}]K_i(x; \boldsymbol{\vartheta}), \tag{2.8}$$

close to the one based on (2.5) and (2.6). The iterative algorithm is formally presented as follows.

STEP 0. Choose initial estimators of $\boldsymbol{\vartheta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\lambda}$ and $g(\cdot)$, denoted as $\boldsymbol{\vartheta}^{(0)}$, $\boldsymbol{\gamma}^{(0)}$, $\boldsymbol{\lambda}^{(0)}$ and $g^{(0)}(\cdot)$. For example, the usual least squares estimators of $\boldsymbol{\vartheta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$, and the local least squares estimator of $g(\cdot)$.

SETP 1. **Iteration from the $(k-1)^{\text{th}}$ to the k^{th}.**

(a) For each $w_j = Y_j - m\{g^{(k-1)}(\boldsymbol{\vartheta}^{(k-1)\prime}\mathbf{X}_{j1}), \boldsymbol{\gamma}^{(k-1)\prime}\mathbf{X}_{j2}, \mathbf{X}_{j3}; \boldsymbol{\lambda}^{(k-1)}\}, j = 1, \cdots, n$, we estimate $f(w_j)$ by

$$f^{(k)}(w_j) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\hbar}\left[Y_i - m\{g^{(k-1)}(\boldsymbol{\vartheta}^{(k-1)\prime}\mathbf{X}_{i1}), \boldsymbol{\gamma}^{(k-1)\prime}\mathbf{X}_{i2}, \mathbf{X}_{i3}; \boldsymbol{\lambda}^{(k-1)}\} - w_j\right] \tag{2.9}$$

(b) For each $x = \boldsymbol{\vartheta}^{(k-1)\prime}\mathbf{X}_{11}, \cdots, \boldsymbol{\vartheta}^{(k-1)\prime}\mathbf{X}_{n1}$, we estimate $\boldsymbol{\theta}(x) \doteq (\alpha, \beta)' = (g(x), \dot{g}(x))'$ by estimating the local maximum likelihood,

$$\sum_{i=1}^{n} \log f^{(k)}[Y_i - m\{\alpha + \beta(\boldsymbol{\vartheta}^{(k-1)\prime}\mathbf{X}_{i1} - x), \boldsymbol{\gamma}^{(k-1)\prime}\mathbf{X}_{i2}, \mathbf{X}_{i3}; \boldsymbol{\lambda}^{(k-1)}\}]K_i(x; \boldsymbol{\vartheta}) \tag{2.10}$$

with respect to α and β. Let $\widehat{\alpha}(x)$ be the solutions of α. Then $g^{(k)}(\boldsymbol{\vartheta}^{(k-1)\prime}\mathbf{X}_{i1}) = \widehat{\alpha}(\boldsymbol{\vartheta}^{(k-1)\prime}\mathbf{X}_{i1})$ for $i = 1, \cdots, n$.

(c) Estimate $\boldsymbol{\vartheta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ by maximizing

$$\sum_{i=1}^{n} \log f^{(k)}[Y_i - m\{g^{(k)}(\boldsymbol{\vartheta}'\mathbf{X}_{i1}), \boldsymbol{\gamma}'\mathbf{X}_{i2}, \mathbf{X}_{i3}; \boldsymbol{\lambda}\}] \tag{2.11}$$

with respect to $\boldsymbol{\vartheta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ by Newton-Raphson iteration.

(d) Repeat this iteration procedure until convergence.

For each x in the range of $\widehat{\boldsymbol{\vartheta}}'\mathbf{X}_{i1}, i = 1, \cdots, n$, the estimates of $g(x)$ and $\dot{g}(x)$, denoted by $\widehat{g}(x)$ and $\widehat{\dot{g}}(x)$, respectively, are obtained by maximizing (2.10) for α and β by replacing $f^{(k)}(\cdot)$ with their estimators defined as the limits of the above iteration.

Remark 1. In the actual implementation of the algorithm, at each step, only $g^{(k)}(x)$ for x values of $\boldsymbol{\vartheta}^{(k-1)\prime}\mathbf{X}_{11}, \ldots, \boldsymbol{\vartheta}^{(k-1)\prime}\mathbf{X}_{n1}$ need to be computed, since only these quantities are used in the next step.

Remark 2. In step 1 of the above iterations, we use all observations to estimate the density function and then we use the estimated density to obtain an estimate of $g(x)$. As a result, the estimate of $g(\cdot)$ at a single point x also uses the information of all observations inside and outside the neighborhood of x. In this sense, it is a genuine global estimation, different from the standard local estimation in (2.4) using a known loss function, where only the local data inside the neighborhood of x are used. Based on our numerical experience or observations, the impact of the globalization on improving efficiency seems more essential than that of the use of the likelihood principle. In Section 3, the proposed estimator is shown to be semiparametrically efficient in the sense of Bickel *et al.* (1993).

Remark 3. When f is known, our estimates reduce to the maximum likelihood estimator and the local maximum likelihood estimator. The procedure based on (2.10) and (2.11) can be regarded as a nonparametric version of the maximum profile likelihood estimator (Murphy and Van der Vaart, 2000). Similar idea can be traced back to Stone (1975), where the author developed a routine for constructing adaptive maximum likelihood estimator of a location parameter. However, the extension to our model is by no means of straightforward given the complexity of our model and the property of infinity dimension of interested function: the techniques used to establish the asymptotic theory of the traditional maximum profile likelihood estimator is no more feasible for infinity dimension function. Wong and Severini (1991) also has considered nonparametric profile idea and has built the corresponding theory. However, they require either error distribution or the regression function is known and they also focus on the finite parameters.

Remark 4. Our method is similar to the backfitting technique (Hastie and Tibshirani, 1990) in that both methods are defined implicitly as the limit of a complicated iterative algorithm for estimating more than one nonparametric functions. The difference between the two methods is that our method uses optimal criterion function to obtain semiparametrically efficient estimator, while the backfitting method utilizes the criterion functions based on mathematical convenience or robust considerations and is proposed to avoid the curse of dimensionality.

Remark 5. The estimate of the derivative of $f(\cdot)$ is required for estimating $\boldsymbol{\vartheta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\lambda}$ and $g(\cdot)$ when Newton-Raphson iteration algorithm is used. Non-parametric estimators for derivative functions may converge relatively slowly, so a higher-order kernel $\mathcal{K}(\cdot)$ (Müller, 1984) is needed to ensure sufficient convergence rates. In the simulation studies and the real example, we use the fourth-order kernel for $\mathcal{K}(\cdot)$ to satisfy the imposed assumptions for consistency and asymptotic normality of the estimators.

3. Large sample properties.

We now establish the uniform consistency and asymptotic normality of the proposed likelihood estimator. Without loss of generality, we assume the support of $\mathbf{X}_1$ is $[0,1]^{d_1}$. Let $\boldsymbol{\vartheta} = (\vartheta_1, \boldsymbol{\vartheta}'_{-1})'$. By the identifiability condition, $\|\boldsymbol{\vartheta}\| = 1$, we have $\vartheta_1 = \sqrt{1 - \|\boldsymbol{\vartheta}_{-1}\|_2^2}$. Other regularity conditions are stated in the Appendix *A.2*. The uniform consistency of $\widehat{g}(\cdot)$ and consistency of $\widehat{\boldsymbol{\Theta}}_2 = (\widehat{\boldsymbol{\vartheta}}'_{-1}, \widehat{\boldsymbol{\gamma}}', \widehat{\boldsymbol{\lambda}}')'$ are presented in Theorem 1. Denote $\boldsymbol{\Theta}_{20}$ and $\boldsymbol{\vartheta}_0$ to be the true values of $\boldsymbol{\Theta}_2$ and $\boldsymbol{\vartheta}$, respectively.

THEOREM 1. *Under Conditions 1-6 stated in Appendix A.2, we have*

$$\begin{aligned} \|\widehat{\boldsymbol{\Theta}}_2 - \boldsymbol{\Theta}_{20}\| &\xrightarrow{p} 0, \\ \sup_{0<x<1} |\widehat{g}(x) - g(x)| &\xrightarrow{p} 0. \end{aligned}$$

To express explicitly the asymptotic expression of the estimator $\widehat{g}(x) - g(x)$, we denote the density function of $\boldsymbol{\vartheta}'_0\mathbf{X}_1$ by $p_1(\cdot)$. Let $\dot{m}_r(\cdot,\cdot,\cdot;\boldsymbol{\lambda})$ be the first derivative of $m(\cdot,\cdot,\cdot;\boldsymbol{\lambda})$ subject to the rth element for $r = 1, 2$. For example, $\dot{m}_2(u, v, w; \boldsymbol{\lambda}) = \partial m(u, v, w; \boldsymbol{\lambda})/\partial v$. Denote $\mu_i = \int x^i K(x)dx$, $\nu_i = \int x^i K^2(x)dx$, $\tau = E\left\{\dot{f}^2(\varepsilon)/f^2(\varepsilon)\right\}$, $\sigma(x) = \nu_0/\tau p_1(x)\left\{\dot{M}_{1,1}^2(x)\right\}^{-1}$, where $\dot{M}_{1,1}^r(x) = E\left\{\dot{m}_1^r(\mathbf{X}_i) \mid \boldsymbol{\vartheta}'_0\mathbf{X}_{i1} = x\right\}$, $r = 1, 2$, $\dot{m}_r(\mathbf{X}_i) = \dot{m}_r(\mathbf{X}_i; g, \boldsymbol{\Theta}_{20})$ and $\dot{m}_r(\mathbf{X}_i; \alpha, \boldsymbol{\Theta}_2) = \dot{m}_r\{\alpha(\boldsymbol{\vartheta}'\mathbf{X}_{i1}), \boldsymbol{\gamma}'\mathbf{X}_{i2}, \mathbf{X}_{i3}; \boldsymbol{\lambda}\}$.

THEOREM 2. *Under Conditions 1-6 stated in Appendix A.2, for* $0 < x < 1$,

$$(nh)^{1/2}\{\widehat{g}(x) - g(x) - \frac{1}{2}h^2\mu_2(I - \mathcal{P})^{-1}(\ddot{g})(x)\} \xrightarrow{\mathcal{D}} N(0, \Sigma(x)), \tag{3.12}$$

where $\mathcal{P}$ *is the linear operator satisfying* $\mathcal{P}(\phi)(x) = \frac{\dot{M}_{1,1}(x)}{\dot{M}^2_{1,1}(x)} \int_0^1 \dot{M}_{1,1}(u)p_1(u)\phi(u)du$ *for any function* ϕ, I *is the identity operator,* $\Sigma(x) = \{(I - \mathcal{P})^{-1}(\sigma)(x)\}^2$, *and* $\ddot{g}(x) = d^2g(x)/dx^2$.

Hence, the orders of the asymptotic bias and variance of $\widehat{g}(x) - g(x)$ are h^2 and $(nh)^{-1}$, respectively. Consequently, the theoretical optimal bandwidth is of order $n^{-1/5}$.

Denote $\boldsymbol{\theta}(x) = (\alpha(x), \beta(x))'$ and $\boldsymbol{\theta}_0(x) = (g(x), \dot{g}(x))'$ is the true value of $\boldsymbol{\theta}(x)$, $\mathbf{J} = \binom{-(1-\|\boldsymbol{\vartheta}_{-1}\|^2)^{-1/2}\boldsymbol{\vartheta}_{-1}}{I_{d_1-1}}$, I_{d_1-1} is the identity matrix with dimension $d_1 - 1$.

$$\dot{\mathbf{m}}_2\left(\mathbf{X}_i; \boldsymbol{\theta}, \boldsymbol{\Theta}_2\right) = \left\{\dot{m}_1(\mathbf{X}_i; \alpha, \boldsymbol{\Theta}_2)\beta(\boldsymbol{\vartheta}'\mathbf{X}_{i1})\mathbf{X}'_{i1}\mathbf{J}, \dot{m}_2(\mathbf{X}_i; \alpha, \boldsymbol{\Theta}_2)\mathbf{X}'_{i2}, \dot{\mathbf{m}}_3(\mathbf{X}_i; \alpha, \boldsymbol{\Theta}_2)\right\}',$$

$$\dot{\mathbf{m}}_3(u, v, w; \boldsymbol{\lambda}) = \partial m(u, v, w; \boldsymbol{\lambda})/\partial\boldsymbol{\lambda}, \;\; \dot{\mathbf{m}}_2(\mathbf{X}_i) = \dot{\mathbf{m}}_2\left(\mathbf{X}_i; \boldsymbol{\theta}_0, \boldsymbol{\Theta}_{20}\right),$$

$$\dot{\mathbf{M}}_{1,2}(x) = E\left[\dot{m}_1(\mathbf{X}_i)\dot{\mathbf{m}}_2(\mathbf{X}_i) \mid \boldsymbol{\vartheta}_0'\mathbf{X}_{i1} = x\right],$$

$$\dot{\mathbf{M}}_{2,2}(x) = E\left[\dot{\mathbf{m}}_2(\mathbf{X}_i) \mid \boldsymbol{\vartheta}_0'\mathbf{X}_{i1} = x\right], \;\; D(x) = \dot{\mathbf{M}}_{1,2}(x) - \dot{M}_{1,1}(x)E\left(\dot{\mathbf{M}}_{2,2}(\boldsymbol{\vartheta}_0'\mathbf{X}_{i1})\right),$$

$$C(\mathbf{X}_i) = \dot{m}_1(\mathbf{X}_i)\dot{M}_{1,1}(\boldsymbol{\vartheta}_0'\mathbf{X}_{i1})/\dot{M}^2_{1,1}(\boldsymbol{\vartheta}_0'\mathbf{X}_{i1}),$$

$$B(\mathbf{X}_i) = \dot{\mathbf{m}}_2(\mathbf{X}_i) - \dot{m}_1(\mathbf{X}_i)\dot{M}_{1,2}(\boldsymbol{\vartheta}_0'\mathbf{X}_{i1})/\dot{M}^2_{1,1}(\boldsymbol{\vartheta}_0'\mathbf{X}_{i1}),$$

$$\mathcal{B}(x) = \left\{\dot{M}_{1,2}(x) - I\left(E\left\{C(\mathbf{X}_i)\right\} \neq 1\right)\frac{\dot{M}_{1,1}(x)}{1 - E\left\{C(\mathbf{X}_i)\right\}}E\left(B(\mathbf{X}_i)\right)\right\}\frac{D(x)}{\dot{M}^2_{1,1}(x)},$$

$$A = \mathrm{Var}\left\{\dot{\mathbf{m}}_2(\mathbf{X}_i)\right\} - \int \mathcal{B}(u)p_1(u)du, \;\; \text{and}$$

$$B = \mathrm{Var}\left[B(\mathbf{X}_i) + I\left(E\left\{C(\mathbf{X}_i)\right\} \neq 1\right)\frac{E\left\{B(\mathbf{X}_i)\right\}C(\mathbf{X}_i)}{1 - E\left\{C(\mathbf{X}_i)\right\}}\right].$$

THEOREM 3. *Under Conditions 1-6 stated in Appendix A.2, if* $nh^4 \to 0$, $nh^2/(\log n)^2 \to \infty$ *and* $n\hbar^5 h/(\log n)^2 \to \infty$ *as* $n \to \infty$, *we have*

$$\widehat{\boldsymbol{\Theta}}_2 - \boldsymbol{\Theta}_{20} \xrightarrow{\mathcal{D}} N(0, \tau^{-1}A^{-1}BA^{-1}). \tag{3.13}$$

It is worth noting that, $E\{C(\mathbf{X}_i)\} \equiv E\left\{\frac{E^2\left[\dot{m}_1(\mathbf{X}_i)|\boldsymbol{\vartheta}_0'\mathbf{X}_{i1}\right]}{E\left[\dot{m}_1^2(\mathbf{X}_i)|\boldsymbol{\vartheta}_0'\mathbf{X}_{i1}\right]}\right\} \leq 1$, and $E\{C(\mathbf{X}_i)\} = 1$ if and only if $\dot{m}_1(\mathbf{X}_i) = E\left[\dot{m}_1(\mathbf{X}_i) \mid \boldsymbol{\vartheta}_0'\mathbf{X}_{i1}\right]$, a special case of which is the partial linear model where $\dot{m}_1(\mathbf{X}_i) \equiv 1$, so the additive structure for the parameter and non-parameter parts holds,

The following statement may not be true. It is not easy to prove that two estimators are independent——which leads to that the estimators for the parameter and non-parameter are independent –HL

, providing X_{i1} and $(\mathbf{X}_{i2}, \mathbf{X}_{i3})$ are independent. In the other word, under the partial linear model, we can efficiently estimate the parameter even the estimator for the nonparameter part is not efficient. Indeed, the existing most works on efficient estimator in semiparametric model mainly focus on the partial linear model.

To evaluate the optimality of the proposed approach, we provide a justification via semiparametric efficiency in the following theorem. Let $\int_0^1 \psi_1(x)\widehat{g}(x)dx + \boldsymbol{\psi}_2'\widehat{\boldsymbol{\Theta}}_2$ be the proposed estimator of $\int_0^1 \psi_1(x)g(x)dx + \boldsymbol{\psi}_2'\boldsymbol{\Theta}_2$ for any given function $\psi_1(x)$. Then,

THEOREM 4. *Assume Conditions 1-6 in Appendix A.2 hold. If $nh^4 \to 0$, $nh^2/(\log n)^2 \to \infty$ and $n\hbar^5 h/(\log n)^2 \to \infty$, then for any $\psi_1(\cdot)$ and $\boldsymbol{\psi}_2$,*

$$\int_0^1 \psi_1(x)\widehat{g}(x)dx + \boldsymbol{\psi}_2'\widehat{\boldsymbol{\Theta}}_2 \text{ is an efficient estimator of } \int_0^1 \psi_1(x)g(x)dx + \boldsymbol{\psi}_2'\boldsymbol{\Theta}_2.$$

Hence, by taking $\psi_1(x) = 0$, we know that $\widehat{\boldsymbol{\Theta}}_2$ is an efficient estimator of $\boldsymbol{\Theta}_2$. By taking $\boldsymbol{\psi}_2 = 0$, then $\widehat{g}(x)$ is a semiparametrically efficient estimator of $g(x)$ in the sense of Bickel *et al.* (1993).

From Theorems 2 and 3, the estimators of the parameters $\boldsymbol{\Theta}_2$ and $\int_0^1 \psi_1(x)g(x)dx + \boldsymbol{\psi}_2'\boldsymbol{\Theta}_2$ can reach the rate $n^{-1/2}$ subject to undersmoothing with $h = o(n^{-1/4})$. The requirement of undersmoothing to gain $\sqrt{n}$-consistent estimators is common in semi-parametric regression (Carroll *et al.*, 1997; Hastie and Tibshirani, 1990).

4. Simulation Studies. In this section we investigate the performance of the proposed maximum likelihood curve estimator by comparing our method with the local least squares (LLS) estimator. The performance of the estimator $\widehat{g}(\cdot)$ is assessed via

$$bias = \left[\frac{1}{n_{grid}}\sum_{i=1}^{n_{grid}}\{E\widehat{g}(x_i)-g(x_i)\}^2\right]^{1/2}, \quad sd = \left[\frac{1}{n_{grid}}\sum_{i=1}^{n_{grid}}E\{\widehat{g}(x_i)-E\widehat{g}(x_i)\}^2\right]^{1/2}$$

and RMSE $= \left[bias^2+sd^2\right]^{1/2}$, where x_i ($i=1,\ldots,n_{grid}$) are the grid points in which the function $g(\cdot)$ is estimated and $E\widehat{g}(x_i)$ is approximated by its sample mean based on N simulated data sets. In the following examples, the Epanechnikov kernel will be used, and $n_{grid}=200$. The results presented below are based on $N=300$ replications and $n=300$ sample size.

Example 1. Consider the partially linear model $Y=g(X)+\mathbf{Z}'\boldsymbol{\beta}+\varepsilon$, where $X\sim 0.5N(-0.5,0.5^2)+0.5N(0.5;0.5^2)$, $g(x)=\sin(x)$, $\mathbf{Z}=(Z_1,Z_2)'$, $Z_1\sim N(0,0.5^2)$, $Z_2\sim U(0,1)$, $\boldsymbol{\beta}=(\beta_1,\beta_2)'=(0.5,1)'$, and $\varepsilon\sim U[-0.4,0.4]$ (Model 1) or $\varepsilon\sim 0.5N(-0.4,0.1^2)+0.5N(0.4,0.1^2)$ (Model 2).

Table 1 provides the bias, empirical standard deviations (sd) and root mean square errors (RMSE) of the nonparametric estimators of the function $g(\cdot)$ using the proposed method and the LLS method for Models 1 and 2, respectively.

From Table 1, we see that for each model, given a bandwidth h, the biases of the proposed estimator and the LLS estimator are comparable, but the proposed estimator has a much smaller standard deviation than the LLS estimator. The empirical efficiency of the LLS estimator relative to our method is 85.56% on average for Model 1 and 54.29% on average for Model 2. This implies that to achieve the same efficiency our method only requires about 85.56% of the sample size that is required by the LLS estimator for Model 1, and about 54.29% of the sample size that is required by the LLS estimator for Model 2. As a result of the smaller standard deviation, the proposed estimator has better performance than the LLS estimator in each case according to the RMSE. The improvement of the proposed method, compared to the LLS estimator, is much more significant for Model 2 than that for Model 1. This is likely because of the fact that the error term in Model 2 deviates further from the normal random variable distribution than that in Model 1.

In addition, from Table 1, we can see that the optimal bandwidth of the proposed estimator is smaller than that for the LLS estimator, this may be attributed to that the amount of data used by the proposed estimator is more than that used by the LLS estimator.

Table 2 provides the bias, sd and RMSEs of the regression coefficients estimators $\widehat{\boldsymbol{\beta}}$ using the proposed method and the LLS method for Models 1 and 2, respectively. With a given bandwidth, both methods are unbiased but the proposed estimator has a much smaller standard deviation than the LLS estimator. Therefore, the proposed estimator has better performance than the local estimator in terms of the RMSE for the parameters. Finally, we also note that both methods for the parameters are insensitive to the bandwidth, which confirm the theoretical result presented in Theorem 3 that the asymptotic bias and variance are independent of bandwidth as long the bandwidth is located in a certain range.

EXAMPLE 2. In this example, we generate data using the model, which is stimulated from the model for the Electricity Costs analysis in Section 5.2, with the following form:

$$Y = \sin(X_1) + \frac{1}{\lambda_2}\log\left\{\lambda_1 X_2^{\lambda_2} + (1-\lambda_1)X_3^{\lambda_2}\right\} + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3 + \varepsilon, \text{(Model 3)}$$

where $X_1 \sim U(-2,2)$, $X_2 \sim U(1,2)$, $X_3 \sim N(2,0.5^2)$, $Z_1 \sim Bernoulli(0.5)$, $Z_2 \sim U(-1,1)$, $Z_3 \sim N(0,1)$, $\varepsilon \sim 0.5N(-0.2,0.1^2) + 0.5N(0.2,0.1^2)$, and $\lambda_1 = 0.85$, $\lambda_2 = 0.5$, $\boldsymbol{\gamma} = (\gamma_1,\gamma_2,\gamma_3)' = (0.5,0.5,1)'$.

The resulting estimators are presented in Tables 3 and 4. Similar conclusions to those from Tables 1 and 2 can be drawn.

Example 3. Finally, we consider the following partial linear single index model,

$$Y = g(\boldsymbol{\vartheta}'\mathbf{X}) + \boldsymbol{\gamma}'\mathbf{Z} + \varepsilon,$$

where $g(X) = X^2$, $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2)' = (3, -4)'/5$, $\mathbf{X}_i \sim N_2(0, I)$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)' = (1, -1)'$, $Z_1 \sim U(-1, 1)$, $Z_2 \sim N(0, 1)$, $\varepsilon \sim N(0, 0.5^2)$ (Model 4) and $\varepsilon \sim 0.5N(-0.5, 0.5^2) + 0.5t(0.5, 4)$ (Model 5), where $U \sim t(0.5, 4)$ represents $U - 0.5 \sim t(4)$.

The resulting estimators are presented in Tables 5–7. It is worth noting that when the errors follow normal distribution, the traditional least square criterion is optimal. In this case, the proposed method is comparable to the least square criterion. This means the loss caused by the estimation of density function is ignorable. For non-normal situations, our estimator again outperforms least square criterion.

5. Two Real Data Examples.

5.1. *The Prestige data.* To study how prestige depends on education levels and income, Fox (1997) collected a data set from the 1971 Canada Census that contains prestige (Pineo-Porter prestige score for occupation from a social survey in the mid-1960s), income (average income of incumbents, in dollars) and education (average education time of occupation incumbents, in years) of 102 individuals. Fox (1997) pointed out that linear regression analysis of prestige on income and education is improper. He suggested the use of nonparametric methods to obtain prior knowledge, regarding the relationship between prestige and education and income.

Firstly, we investigate the dependence of prestige on income and education separately by using classical nonparametric regression model. Both the LLS estimator and the proposed method suggest that the function g(income) is highly non-constant and non-linear, while prestige basically increases linearly with education. Based on the above preliminary work, partially linear regression model $Y = g(X) + \gamma Z + \varepsilon$ is considered, where Y is prestige, X is the income and

Z is the education. We estimate the nonparametric regression function $g(\cdot)$ and parameter γ using both the LLS and the proposed methods. The bandwidths $h = 16800$ and $h = 11100$ are optimal chosen by 10-fold cross validation for the LLS and the proposed methods, respectively. The variances are estimated by using 500 bootstrap samples.

The estimate values and associated standard deviations for γ based on the proposed method and LLS methods are 4.917 (0.324) and 4.870 (0.427), respectively, and the estimates for $g(\cdot)$ and the pointwise 95% bootstrap confidence interval are shown in Figure 1. The estimated values indicates that high education can significantly increase prestige. Figure 1 suggests that prestige increases linearly with income increasing from $0 to $12000, and is stable when income varies from $12000 to $18000, then increases linearly when income is larger than $18000. Once again, the estimated values and Figure 1 show that the LLS and the proposed methods give similar estimators for the regression function and coefficient, but the proposed method has less standard deviation and hence may be more efficient than the LLS estimator for the parameters and non-parametric functions.

5.2. *Electricity Costs data.* We continue with our example on electricity distribution costs for Ontario municipal distributors (Yatchew, 2000). The response variable is log(*total cost per customer*)(termed TC), and the covariates include log(*number of customers*)(cust), log(*wage of lineman*)(WAGE), log(*accumulated gross investment/kilometers of distribution wire*)(PCAP), log(*kilowatt hour sales per custome*r)(KWH), log(*load factor*)(LF) and log(*kilometers of distribution wire per customer*)(KNWIRE).

Considering a conventional constant elasticity of substitution (CES) cost function (see, e.g., Varian (1992), p.56), we are interested in assessing whether cost per customer is affected by the scale of operation, that is , the number of customers. The researchers suggested the following nonparametric scale effect (Varian, 1992)

$$\begin{aligned}\mathrm{TC} = g(\mathrm{cust}) + \frac{1}{\lambda_2}\log\left\{\lambda_1 \mathrm{WAGE}^{\lambda_2} + (1-\lambda_1)\mathrm{PCAP}^{\lambda_2}\right\} \\ + \gamma_1 KWH + \gamma_2 LF + \gamma_3 KNWIRE + \varepsilon. \end{aligned} \tag{5.14}$$

Both the LLS and the proposed method are used to estimate the parameters and the function. The bandwidths, $h_{LLS} = 1$ and $h_{Prop.} = 1$, are chosen by 16-fold cross-validation. The proposed estimator is not sensitive to the selection of the bandwidth $\hbar$, which is chosen to be 1 for simplicity. The LLS and the proposed method give similar estimators for the regression function and parameters, but the proposed method has less variance and is more efficient than the LLS estimator for the parameters and non-parametric functions. For simplicity, Table 8 and Figure 2 only give the results of the proposed method. The explanation??

6. Discussion. We have proposed a new class of models, generalized semi/non-parametric regression model, that covers widely studied partially linear models, single-index models and two-sample semiparametric models. We have developed an iterative procedure to integrate maximum likelihood principle and the localization procedure, the latter of which is usually applied in semiparametric models, in order that global observations can be used. The corresponding estimators of the parameters and nonparametric functions therefore gain efficiency. The efficiency of the nonparametric function, which has been overlooked in the literature, is of independent interest. Furthermore, the proposed methods have the following features: (i) the proposed methods show promising performance in finite sample situations; and (ii) the implemented algorithm is computationally efficient by the standard Newton-Raphson iteration with the LLS method as the initial value.

In this article, we have focused on modeling with continuous response variable. It is possible to extend the methods to discrete response variable. The theory and implementation of such an extension may be much more complicated, at least technically, and warrants a further study.

In this article, we didn't consider the interactions between $\mathbf{X}_2$ and $\mathbf{X}_3$ or their partial interaction terms. In principle, it is straightforward to extend our theory to these more general cases with complications of notation, and tedious derivations after we address identifiability concerns, whose theory may need more efforts to develop and also warrant a further study.

References.

Bickel, P. J., Klaassen, C. A., Wellner, J. A. and Ritov, Y. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press Baltimore.

Brown, B. W. and Newey, W. K. (1998). Efficient semiparametric estimation of expectations. *Econometrica* **66** 453–464.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *J. Am. Statist. Asso.* **92** 477–489.

Cheng, G. and Shang, Z. (2015). Joint asymptotics for semi-nonparametric regression models with partially linear structure. *Ann. Statist.* **43** 1351–1390.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Am. Statist. Asso.* **74** 829–836.

Cuzick, J. et al. (1992). Efficient estimates in semiparametric additive regression models with unknown error distribution. *Ann. Statist.* **20** 1129–1136.

Engle, R. F., Granger, C. W. J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Am. Statist. Asso.* **81** 310-320.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. CRC press.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications* **66**. CRC Press.

Fan, J., Lin, H. and Zhou, Y. (2006). Local partial-likelihood estimation for lifetime data. *Ann. Statist.* **34** 290–325.

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.* **27** 1491–1518.

Fox, J. (1997). Applied Regression, Linear Models, nad Related Methods. Sage.

Green, P., Jennison, C. and Seheult, A. (1985). Analysis of field experiments by least squares smoothing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **47** 299–315.

Green, P. J. and Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: a roughness penalty approach*. CRC Press.

Härdle, W. (1989). Asymptotic maximal deviation of M-smoothers. *J. Multi. Ana.* **29** 163–179.

Härdle, W., Liang, H. and Gao, J. (2000). *Partially Linear Models*. Physica-Verlag, Heidelberg.

Härdle, W. and Marron, J. S. (1990). Semiparametric comparison of regression curves. *Ann. Statist.* **18** 63–89.

Härdle, W. and Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Am. Statist. Asso.* **84** 986–995.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models* **43**. CRC Press.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **55** 757–796.

Horowitz, J. L. and Härdle, W. (1996). Direct semiparametric estimation of single-index models with discrete covariates. *J. Am. Statist. Asso.* **91** 1632–1640.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* **58** 71-120.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association* **86** 316–342. With discussion and a rejoinder by the author.

Ma, Y., Chiou, J.-M. and Wang, N. (2006). Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika* **93** 75–84.

Maity, A., Ma, Y. and Carroll, R. J. (2007). Efficient estimation of population level summaries in general semiparametric regression models. *J. Am. Statist. Asso.* **102** 123–139.

Müller, H.-G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes. *Ann. Statist.* **12** 766–774.

Murphy, S. A. and Van der Vaart, A. W. (2000). On profile likelihood. *J. Am. Statist. Asso.* **95** 449–465.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica* **62** 1349–1382.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematica Statistics.

Powell, J. L., Stock, J. H. and Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica* **57** 1403–1430.

Schick, A. (1993). On efficient estimation in regression models. *Ann. Statist.* **21** 1486–1521.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **50** 413–436.

Stone, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. *The Annals of Statistics* **3** 267–284.

Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–620.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, Mass.: Addison-Wesley.

Varian, H. R. (1992). Microeconomic analysis. 3rd edn, New York: W. W. Norton.

Wahba, G. (1984). *Cross validated spline methods for the estimation of multivariate functions from data on functionals*. Ames: Iowa State University Press In Statistics: An Appraisal, Proc. 50th Anniversary Conf. Iowa State Statistical Laboratory (Edited by David, H. A. and David, H. T.).

Wang, F. T. and Scott, D. W. (1994). The L_1 method for robust nonparametric regression. *J. Am. Statist. Asso.* **89** 65–76.

Wong, W. H. and Severini, T. A. (1991). On maximum likelihood estimation in infinite dimensional parameter spaces. *Ann. Statist.* **19** 603–632.

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

Xia, Y. and Li, W. K. (1999). On single-index coefficient regression models. *Journal of the American Statistical Association* **94** 1275-1285.

Xia, Y., Zhang, W. and Tong, H. (2004). Efficient estimation for semivarying-coefficient models. *Biometrika* **91** 661–681.

Yatchew, A. (2000). Scale economies in electricity distribution: A semiparametric analysis. *J. Appl. Econ.* **15** 187–210.

Appendices.

A.1. Notation. For given $\boldsymbol{\psi}(x) = \left\{\psi_1(x), \boldsymbol{\psi}_2'\right\}'$, let $\left\{\phi_1(x), \boldsymbol{\phi}_2'\right\}'$ satisfy the following integral equations:

$$\psi_1(x) = \phi_1(x)\dot{M}_{1,1}^2(x)p_1(x) - \int \dot{M}_{1,1}(x)\phi_1(x)p_1(x)dx\dot{M}_{1,1}(x)p_1(x) + \boldsymbol{\phi}_2'D(x)p_1(x),$$
$$\boldsymbol{\psi}_2 = \int D(x)\phi_1(x)p_1(x)dx + \boldsymbol{\phi}_2'Var\left\{\dot{\mathbf{m}}_2\left(\mathbf{X}_i\right)\right\},$$

where $D(x), p_1(x)$, $\dot{M}_{1,1}(x)$ and $\dot{\mathbf{m}}_2\left(\mathbf{X}_i\right)$ are defined in Section 3. Denote

$$\mathcal{C}_0 = \{\alpha(x) : x \in [0,1], \alpha(x) \text{ is continuous on } [0,1], \alpha(0) = 0 \text{ and } \sup_x |\alpha(x)| \le C\},$$

where C is a constant, and $\mathcal{R}$ is the set of continuous functions on $[0,1]$, which are bounded uniformly.

A.2. Conditions. K and $\mathcal{K}$ are orders of 2 and r, respectively. Denote the joint density function of $\mathbf{X}$ by $p(\mathbf{x})$.

1. The kernel functions $K(\cdot)$ and $\mathcal{K}(\cdot)$ are symmetric function with a compact support $[-1,1]$. $K(\cdot)$ has bounded derivative and $\mathcal{K}(\cdot)$ has second derivative. $\int_{-1}^1 u^j\mathcal{K}(u)du$ equals 1 when $j=0$, 0 when $1 \le j \le r-1$, and nonzero when $j=r$.

2. $\mathbf{X}$ are bounded with compact support $[0,1]^q$, where q is the dimension of $\mathbf{X}$. The covariate $\mathbf{X}$ and ε are independent.
3. The density functions $f(\cdot)$ and $p_1(\cdot)$ are positive, and have continuous $(r+2)$th and second derivatives on corresponding supports, respectively.
4. The function g has continuous second derivative on $[0,1]$.
5. $E\left\{\frac{\int\cdots\int_{[0,1]^q} \dot{f}(\varepsilon_i - h(\mathbf{X}) + h(\mathbf{y}))p(\mathbf{y})d\mathbf{y}}{\int\cdots\int_{[0,1]^q} f(\varepsilon_i - h(\mathbf{X}) + h(\mathbf{y}))p(\mathbf{y})d\mathbf{y}}\dot{\mathbf{m}}(\mathbf{X}_i)\right\} = 0$ if and only if $h(\mathbf{x}) \equiv 0$ over $\mathbf{x} \in [0,1]^q$.
6. $n\hbar^5/\log(n) \to \infty$, $h^2 = o(\hbar)$, $\hbar^r = o(h^2)$, $h^2\log(n) \to 0$ and $nh/(\log(n))^2 \to \infty$, as $n \to \infty$.

The second-order and fourth-order kernel functions can be taken from Müller (1984). Since the derivative of f is required to estimate g, and derivative functional converge relatively slowly, the higher-order kernel for $\mathcal{K}$ is needed to insure sufficiently rapid convergence. Condition 5 is used to ensure the identification of the model.

A.3. Preliminary Lemmas. Denote $\boldsymbol{\theta} = (\alpha, \beta)'$ and $\boldsymbol{\Omega} = (\alpha, \boldsymbol{\Theta}_2)$, $\boldsymbol{\theta}_0 = (g, \dot{g})'$ and $\boldsymbol{\Omega}_0 = (g, \boldsymbol{\Theta}_{20})$ to be the true values of $\boldsymbol{\theta}$ and $\boldsymbol{\Omega}$, respectively, $\dot{f}(x) = \partial f(x)/\partial x$, $\ddot{f}(x) = \partial^2 f(x)/\partial x^2$, $f^{(r)}(x) \doteq \partial^r f(x)/\partial x^r$,

$$m(\mathbf{X}_i;\boldsymbol{\Omega}) = m\{\alpha(\boldsymbol{\vartheta}'\mathbf{X}_{i1}), \boldsymbol{\gamma}'\mathbf{X}_{i2}, \mathbf{X}_{i3}; \boldsymbol{\lambda}\},\ m(\mathbf{X}_i) = m(\mathbf{X}_i;\boldsymbol{\Omega}_0),\ \mathcal{Y}_i(\boldsymbol{\Omega}) = Y_i - m(\mathbf{X}_i;\boldsymbol{\Omega}),$$

$$f_i^{(r)}(\boldsymbol{\Omega}) = f^{(r)}(\mathcal{Y}_i(\boldsymbol{\Omega});\boldsymbol{\Omega}),\ f^{(r)}(w;\boldsymbol{\Omega}) = Ef^{(r)}\left[w - m(\mathbf{X}_i) + m(\mathbf{X}_i;\boldsymbol{\Omega})\right],$$

$$\rho_1(\boldsymbol{\Omega};x) = E\left[\frac{\dot{f}_i(\boldsymbol{\Omega})}{f_i(\boldsymbol{\Omega})}\dot{m}_1(\mathbf{X}_i;\boldsymbol{\Omega}) \mid \boldsymbol{\vartheta}'\mathbf{X}_{i1} = x\right],\ \dot{f}_i(\boldsymbol{\Omega}) = f_i^{(1)}(\boldsymbol{\Omega}),\ f_i(\boldsymbol{\Omega}) = f_i^{(0)}(\boldsymbol{\Omega}),$$

$$u_1(\boldsymbol{\Omega};x) = \rho_1(\boldsymbol{\Omega};x)p_1(x),\ \mathbf{u}_2(\boldsymbol{\theta},\boldsymbol{\Theta}_2) = E\left[\frac{\dot{f}_i(\boldsymbol{\Omega})}{f_i(\boldsymbol{\Omega})}\dot{\mathbf{m}}_2\{\mathbf{X}_i;\boldsymbol{\theta},\boldsymbol{\Theta}_2\}\right],$$

$$\widehat{\dot{f}}(w;\boldsymbol{\Omega}) = -\frac{1}{n\hbar^2}\sum_{i=1}^n \dot{\mathcal{K}}\left[\frac{Y_i - m(\mathbf{X}_i;\boldsymbol{\Omega}) - w}{\hbar}\right],$$

$$\widehat{f}^{(r)}(w;\boldsymbol{\Omega}) = \frac{1}{n\hbar^{r+1}}\sum_{i=1}^n \mathcal{K}^{(r)}\left[\frac{Y_i - m(\mathbf{X}_i;\boldsymbol{\Omega}) - w}{\hbar}\right],\ \text{for}\ \ r = 0,1,2.$$

LEMMA A.1. *Under Conditions 1-6 in Appendix A.2, we have*

$$\sup_{\binom{w\in\{-\infty,\infty\}}{\mathcal{B}_{\mathbf{\Omega}}}} |\widehat{f}^{(r)}(w;\mathbf{\Omega}) - f^{(r)}(w;\mathbf{\Omega})| \;=\; O_p\left(\hbar^r + \log(n)^{1/2}(n\hbar^{2r+1})^{-1/2}\right), \qquad \text{(A.1)}$$

for $r = 0, 1, 2$, *where* $\mathcal{B}_{\mathbf{\Omega}} = \{\alpha \in \mathcal{C}_0, \mathbf{\Theta}_2 \in R^{p-1}\}$.

PROOF. Let $F(\cdot)$ be the distribution function corresponding to $f(\cdot)$. By Condition 3, for a given n, there exists an $M > 0$ such that

$$\sup_{|w|>M} |f(w)| < \hbar^r + \log(n)^{1/2}(n\hbar)^{-1/2},$$

$$F(M - C - \hbar) \geq 1 - \hbar^{r+1} - \hbar\log(n)^{1/2}(n\hbar)^{-1/2} \text{ and}$$

$$F(-M + C + \hbar) \leq \hbar^{r+1} + \hbar\log(n)^{1/2}(n\hbar)^{-1/2}.$$

By the condition that g is bounded and $\mathbf{X}$ has bounded support, we get

$$\sup_{\binom{|w|>M}{\mathcal{B}_{\mathbf{\Omega}}}} |f(w;\mathbf{\Omega})| < \hbar^r + \log(n)^{1/2}(n\hbar)^{-1/2}. \qquad \text{(A.2)}$$

Furthermore, by Condition 1, we have

$$\widehat{f}(w;\mathbf{\Omega}) = \frac{1}{n}\sum_{i=1}^{n} I\left(|\mathcal{Y}_i(\mathbf{\Omega}) - w| \leq \hbar\right)\mathcal{K}_\hbar\left[\mathcal{Y}_i(\mathbf{\Omega}) - w\right].$$

Then

$$\sup_{\binom{|w|>M}{\mathcal{B}_{\mathbf{\Omega}}}} |\widehat{f}(w;\mathbf{\Omega})| \leq \frac{C}{\hbar}\sup_{\binom{|w|>M}{\mathcal{B}_{\mathbf{\Omega}}}}\Big|\frac{1}{n}\sum_{i=1}^{n} I(\big|\mathcal{Y}_i(\mathbf{\Omega}) - w\big| \leq \hbar)\Big|.$$

Again by the properties of $g(\cdot)$ and $\alpha(\cdot)$, and $\mathbf{X}$ has bounded support, we obtain

$$\begin{aligned}
\sup_{\binom{|w|>M}{\mathcal{B}_{\mathbf{\Omega}}}} |\widehat{f}(w;\mathbf{\Omega})| &\leq \frac{C}{\hbar}\left\{\sup_{w>M} E\left[I(\big|\mathcal{Y}_i(\mathbf{\Omega}) - w\big| \leq \hbar)\right] + \sup_{w<-M} E\left[I(\big|\mathcal{Y}_i(\mathbf{\Omega}) - w\big| \leq \hbar)\right]\right\} \\
&\leq \frac{C}{\hbar}\left[\{1 - F(-\hbar - C + M)\} + F(\hbar + C - M)\right] \\
&\leq C(\hbar^r + \log(n)^{1/2}(n\hbar)^{-1/2}). \qquad \text{(A.3)}
\end{aligned}$$

It follows from (A.2) and (A.3) that

$$\begin{aligned}\sup_{\binom{|w|>M}{\mathcal{B}_{\mathbf{\Omega}}}} \left|\widehat{f}(w;\mathbf{\Omega}) - f(w;\mathbf{\Omega})\right| &\leq \sup_{\binom{|w|>M}{\mathcal{B}_{\mathbf{\Omega}}}} \left|\widehat{f}(w;\mathbf{\Omega})\right| + \sup_{\binom{|w|>M}{\mathcal{B}_{\mathbf{\Omega}}}} |f(w;\mathbf{\Omega})| \\ &\leq C\left\{\hbar^r + \log(n)^{1/2}(n\hbar)^{-1/2}\right\}. \qquad (A.4)\end{aligned}$$

On the other hand, it can be shown by using Theorem 2.37 and Example 38 in Chapter 2 of Pollard (1984) that

$$\sup_{\binom{|w|\leq M}{\mathcal{B}_{\mathbf{\Omega}}}} |\widehat{f}(w;\mathbf{\Omega}) - f(w;\mathbf{\Omega})| = O_p(\hbar^r + \log(n)^{1/2}(n\hbar)^{-1/2}). \qquad (A.5)$$

(A.1) for $r=0$ follows from (A.4) and (A.5). Similarly, we can prove (A.1) for $r=1,2$. □

Denote $\boldsymbol{\theta}(x) = (\alpha(x), \beta(x))'$, $m_i(x;\boldsymbol{\theta},\mathbf{\Theta}_2) = m\left\{\alpha(x) + \beta(x)(\boldsymbol{\vartheta}'\mathbf{X}_{i1} - x), \boldsymbol{\gamma}'\mathbf{X}_{i2}, \mathbf{X}_{i3}; \boldsymbol{\lambda}\right\}$, $m_i(x) = m_i(x;\boldsymbol{\theta}_0,\mathbf{\Theta}_{20})$, $\dot{m}_{ri}(x;\boldsymbol{\theta},\mathbf{\Theta}_2) = \dot{m}_r\left\{\alpha(x) + \beta(x)(\boldsymbol{\vartheta}'\mathbf{X}_{i1} - x), \boldsymbol{\gamma}'\mathbf{X}_{i2}, \mathbf{X}_{i3}; \boldsymbol{\lambda}\right\}$, $\dot{m}_{ri}(x) = \dot{m}_{ri}(x;\boldsymbol{\theta}_0,\mathbf{\Theta}_{20})$, $r=1,2,3$. Let $\mathcal{Y}_i(x;\boldsymbol{\theta},\mathbf{\Theta}_2) = Y_i - m_i(x;\boldsymbol{\theta},\mathbf{\Theta}_2)$, $m_d(\mathbf{X}_i;x) = m(\mathbf{X}_i) - m_i(x)$, $f_i^{(r)}(x;\boldsymbol{\theta},\mathbf{\Theta}_2) = f^{(r)}(\mathcal{Y}_i(x;\boldsymbol{\theta},\mathbf{\Theta}_2);\mathbf{\Omega})$, $\widehat{f}_i^{(r)}(\mathbf{\Omega}) = \widehat{f}^{(r)}(\mathcal{Y}_i(\mathbf{\Omega});\mathbf{\Omega})$. For any vector function $\boldsymbol{\theta}$ and parameter $\mathbf{\Theta}_2$, set

$$\begin{aligned}\mathbf{U}_1(\boldsymbol{\theta},\mathbf{\Theta}_2;x) &= \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{\dot{f}}_i(x;\boldsymbol{\theta},\mathbf{\Theta}_2)}{\widehat{f}_i(x;\boldsymbol{\theta},\mathbf{\Theta}_2)} \dot{m}_{1i}(x;\boldsymbol{\theta},\mathbf{\Theta}_2) K_i(x;\boldsymbol{\vartheta}) W_i(x;\boldsymbol{\vartheta}), \qquad (A.6)\\ \mathbf{U}_2(\boldsymbol{\theta},\mathbf{\Theta}_2) &= \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{\dot{f}}_i(\mathbf{\Omega})}{\widehat{f}_i(\mathbf{\Omega})} \dot{\mathbf{m}}_2(\mathbf{X}_i;\boldsymbol{\theta},\mathbf{\Theta}_2). \qquad (A.7)\end{aligned}$$

where $W_i(x;\boldsymbol{\vartheta}) = (1, \boldsymbol{\vartheta}'\mathbf{X}_{i1} - x)'$, $\dot{\mathbf{m}}_2(\mathbf{X}_i;\boldsymbol{\theta},\mathbf{\Theta}_2)$ is defined in Section 3.

LEMMA A.2. *Under Conditions 1-6 stated in Appendix A.2, we have*

$$\mathbf{U}_1(\boldsymbol{\theta}_0,\mathbf{\Theta}_{20};x) = \frac{1}{n}\sum_{i=1}^{n} \frac{\dot{f}_i(x)}{f_i(x)} \dot{m}_{1i}(x) K_i(x) W_i(x) - \frac{1}{n^2}\sum_{i\neq j}^{n} \mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x) + O_p(\hbar^r) + o_p(h^2 + n^{-1/2}),$$

$$\mathbf{U}_2(\boldsymbol{\theta}_0,\mathbf{\Theta}_{20}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\dot{f}(\varepsilon_i)}{f(\varepsilon_i)} \dot{\mathbf{m}}_2(\mathbf{X}_i) - \frac{1}{n^2}\sum_{i\neq j}^{n} \mathbf{V}_2(\varepsilon_j,\varepsilon_i,\mathbf{X}_i) + O_p(\hbar^r) + o_p\left(n^{-1/2}\right),$$

where $\dot{f}_i(x) = f_i^{(1)}(x;\boldsymbol{\theta}_0,\boldsymbol{\Theta}_{20}),\ f_i(x) = f_i(x;\boldsymbol{\theta}_0,\boldsymbol{\Theta}_{20}),\ K_i(x) = K_i(x;\boldsymbol{\vartheta}_0),\ W_i(x) = W_i(x;\boldsymbol{\vartheta}_0).$

$$
\begin{aligned}
\mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x) &= \left[\mathcal{K}_\hbar(\varepsilon_j-\varepsilon_i-m_d(\mathbf{X}_i;x)) - E\left\{\mathcal{K}_\hbar(\varepsilon_k-\varepsilon_i-m_d(\mathbf{X}_i;x)) \mid \varepsilon_i,\mathbf{X}_i\right\}\right] \\
&\qquad \times \dot{m}_{1i}\,(x)\,\frac{K_i(x)W_i(x)\dot{f}(\varepsilon_i+m_d(\mathbf{X}_i;x))}{f^2(\varepsilon_i+m_d(\mathbf{X}_i;x))}, \\
\mathbf{V}_2(\varepsilon_j,\varepsilon_i,\mathbf{X}_i) &= \left[\mathcal{K}_\hbar(\varepsilon_j-\varepsilon_i) - E\left\{\mathcal{K}_\hbar(\varepsilon_k-\varepsilon_i) \mid \varepsilon_i\right\}\right]\dot{\mathbf{m}}_2(\mathbf{X}_i)\dot{f}(\varepsilon_i)/f^2(\varepsilon_i).
\end{aligned}
$$

Proof. Denote

$$
\begin{aligned}
\nabla\mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x) &= \left[\dot{\mathcal{K}}_\hbar(\varepsilon_j-\varepsilon_i-m_d(\mathbf{X}_i;x)) - E\left\{\dot{\mathcal{K}}_\hbar(\varepsilon_k-\varepsilon_i-m_d(\mathbf{X}_i;x)) \mid \varepsilon_i,\mathbf{X}_i\right\}\right] \\
&\qquad \times\frac{\dot{m}_{1i}\,(x)\,K_i(x)W_i(x)}{f(\varepsilon_i+m_d(\mathbf{X}_i;x))}, \\
\nabla\mathbf{V}_2(\varepsilon_j,\varepsilon_i,\mathbf{X}_i) &= \left[\dot{\mathcal{K}}_\hbar(\varepsilon_j-\varepsilon_i) - E\left\{\dot{\mathcal{K}}_\hbar(\varepsilon_k-\varepsilon_i) \mid \varepsilon_i\right\}\right]\dot{\mathbf{m}}_2(\mathbf{X}_i)/f(\varepsilon_i), \\
\nabla\boldsymbol{\psi}_x(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j) &= \frac{1}{2}\nabla\mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x)+\frac{1}{2}\nabla\mathbf{V}_1(\varepsilon_i,\varepsilon_j,\mathbf{X}_j;x), \\
\boldsymbol{\psi}_x(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j) &= \frac{1}{2}\mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x)+\frac{1}{2}\mathbf{V}_1(\varepsilon_i,\varepsilon_j,\mathbf{X}_j;x), \\
\nabla\boldsymbol{\psi}(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j) &= \frac{1}{2}\nabla\mathbf{V}_2(\varepsilon_j,\varepsilon_i,\mathbf{X}_i)+\frac{1}{2}\nabla\mathbf{V}_2(\varepsilon_i,\varepsilon_j,\mathbf{X}_j), \\
\boldsymbol{\psi}(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j) &= \frac{1}{2}\mathbf{V}_2(\varepsilon_j,\varepsilon_i,\mathbf{X}_i)+\frac{1}{2}\mathbf{V}_2(\varepsilon_i,\varepsilon_j,\mathbf{X}_j).
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{i,j}^n\nabla\mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x) = \frac{1}{n^2}\sum_{i\neq j}^n\nabla\boldsymbol{\psi}_x(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j)+\frac{1}{n^2}\sum_{i=1}^n\nabla\mathbf{V}_1(\varepsilon_i,\varepsilon_i,\mathbf{X}_i;x) \equiv \nabla I_1+\frac{\nabla I_2}{n}, \\
&\frac{1}{n^2}\sum_{i,j}^n\mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x) = \frac{1}{n^2}\sum_{i\neq j}^n\boldsymbol{\psi}_x(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j)+\frac{1}{n^2}\sum_{i=1}^n\mathbf{V}_1(\varepsilon_i,\varepsilon_i,\mathbf{X}_i;x) \equiv I_1+\frac{I_2}{n}, \\
&\frac{1}{n^2}\sum_{i,j}^n\nabla\mathbf{V}_2(\varepsilon_j,\varepsilon_i,\mathbf{X}_i) = \frac{1}{n^2}\sum_{i\neq j}^n\nabla\boldsymbol{\psi}(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j)+\frac{1}{n^2}\sum_{i=1}^n\nabla\mathbf{V}_2(\varepsilon_i,\varepsilon_i,\mathbf{X}_i) \equiv \nabla I_{2,1}+\frac{\nabla I_{2,2}}{n}, \\
&\frac{1}{n^2}\sum_{i,j}^n\mathbf{V}_2(\varepsilon_j,\varepsilon_i,\mathbf{X}_i) = \frac{1}{n^2}\sum_{i\neq j}^n\boldsymbol{\psi}(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j)+\frac{1}{n^2}\sum_{i=1}^n\mathbf{V}_2(\varepsilon_i,\varepsilon_i,\mathbf{X}_i) \equiv I_{2,1}+\frac{I_{2,2}}{n}, \qquad k=1,2.
\end{aligned}
$$

It is clear that ∇I_1, I_1, $\nabla I_{2,1}$ and $I_{2,1}$ are U-statistics with kernel functions $\nabla\boldsymbol{\psi}_x(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j)$, $\boldsymbol{\psi}_x(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j)$, $\nabla\boldsymbol{\psi}(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j)$ and $\boldsymbol{\psi}(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j)$, respectively. Using Conditions 1 and

2, we obtain

$$
\begin{aligned}
&E\left\{\nabla \mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x)\right\} = \begin{cases} 0 & \text{if } j \neq i \\ O(h^2\hbar^{-3}+h^2)(1,h^2)' & \text{otherwise} \end{cases}, \\
&E\left\{\mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x)\right\} = \begin{cases} 0 & \text{if } j \neq i \\ O(h^2\hbar^{-1})(1,h^2)' & \text{otherwise} \end{cases}, \qquad (A.8)\\
&E\left\{\mathbf{V}_2(\varepsilon_j,\varepsilon_i,\mathbf{X}_i)\right\} = \begin{cases} 0 & \text{if } j \neq i \\ O(\hbar^r) & \text{otherwise} \end{cases}, \;\; E\left\{\nabla \mathbf{V}_2(\varepsilon_j,\varepsilon_i,\mathbf{X}_i)\right\} = 0, \;\; k=1,2, \\
&var\left\{\nabla \mathbf{V}_1(\varepsilon_i,\varepsilon_i,\mathbf{X}_i;x)\right\} = O((h^3+h^4)\hbar^{-6}+h^4\hbar^{-3}+h^{-1}), \\
&var\left\{\mathbf{V}_1(\varepsilon_i,\varepsilon_i,\mathbf{X}_i;x)\right\} = O(h^{-1}\hbar^{-2}), \\
&var\left\{\nabla \mathbf{V}_2(\varepsilon_i,\varepsilon_i,\mathbf{X}_i)\right\} = O(1), \;\; var\left\{\mathbf{V}_2(\varepsilon_i,\varepsilon_i,\mathbf{X}_i)\right\} = O(\hbar^{-2}), \;\; k=1,2, \\
&E\left\{\nabla \boldsymbol{\psi}_x(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j)\right\} = E\left\{\boldsymbol{\psi}_x(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j)\right\} = 0, \\
&E\left\{\nabla \boldsymbol{\psi}(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j)\right\} = E\left\{\boldsymbol{\psi}(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j)\right\} = 0, \\
&\nabla \boldsymbol{\psi}_x(\varepsilon_i,\mathbf{X}_i) \equiv E\left\{\nabla \boldsymbol{\psi}_x(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j) \mid \mathbf{X}_i,\varepsilon_i\right\} = O_p(h^2)(1,h^2)', \\
&\boldsymbol{\psi}_x(\varepsilon_i,\mathbf{X}_i) \equiv E\left\{\boldsymbol{\psi}_x(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j) \mid \mathbf{X}_i,\varepsilon_i\right\} = O_p(1)(1,h^2)', \\
&\nabla \boldsymbol{\psi}(\varepsilon_i,\mathbf{X}_i) \equiv E\left\{\nabla \boldsymbol{\psi}(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j) \mid \mathbf{X}_i,\varepsilon_i\right\} = 0, \\
&\boldsymbol{\psi}(\varepsilon_i,\mathbf{X}_i) \equiv E\left\{\boldsymbol{\psi}(\varepsilon_i,\mathbf{X}_i;\varepsilon_j,\mathbf{X}_j) \mid \mathbf{X}_i,\varepsilon_i\right\} = O_p(1), \;\; cov\left\{\boldsymbol{\psi}(\varepsilon,\mathbf{X}_i)\right\} = O(1), \\
&cov\left\{\nabla \boldsymbol{\psi}_x(\varepsilon_i,\mathbf{X}_i)\right\} = O(h^4), \quad \text{and} \;\; cov\left\{\boldsymbol{\psi}_x(\varepsilon_i,\mathbf{X}_i)\right\} = O(1).
\end{aligned}
$$

Hence, by the central limit theorem of U-statistic, we know

$$
\begin{aligned}
\frac{1}{n^2}\sum_{i,j}^{n} \nabla \mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x) &= E(\nabla I_1) + O_p\left\{\sqrt{var(\nabla I_1)}\right\} + n^{-1}\left[E(\nabla I_2) + O_p\left\{\sqrt{var(\nabla I_2)}\right\}\right] \\
&= O_p\left\{h^2\left(n^{-1/2}+(n\hbar^3)^{-1}\right) + n^{-1}(nh)^{-1/2}\right\}, \\
\frac{1}{n^2}\sum_{i,j}^{n} \mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x) &= E(I_1) + O_p\left\{\sqrt{var(I_1)}\right\} + n^{-1}\left[E(I_2) + O_p\left\{\sqrt{var(I_2)}\right\}\right] \\
&= O_p\left\{n^{-1/2} + n^{-1/2}(nh)^{-1/2}(n\hbar^2)^{-1/2} + h^2(n\hbar)^{-1}\right\},
\end{aligned}
$$

$$\frac{1}{n^2}\sum_{i,j}^{n}\nabla\mathbf{V}_2(\varepsilon_j,\varepsilon_i,\mathbf{X}_i) = E(\nabla I_{2,1}) + O_p\left\{\sqrt{var(\nabla I_{2,1})}\right\} + n^{-1}\left[E(\nabla I_{2,2}) + O_p\left\{\sqrt{var(\nabla I_{2,2})}\right\}\right]$$
$$= O_p\left\{n^{-3/2}\right\},$$
$$\frac{1}{n^2}\sum_{i,j}^{n}\mathbf{V}_2(\varepsilon_j,\varepsilon_i,\mathbf{X}_i) = E(I_{2,1}) + O_p\left\{\sqrt{var(I_{2,1})}\right\} + n^{-1}\left[E(I_{2,2}) + O_p\left\{\sqrt{var(I_{2,2})}\right\}\right]$$
$$= O_p\left(n^{-1/2}\right). \tag{A.9}$$

By (A.9) and Condition 6, we have

$$\begin{aligned}
\mathbf{U}_1(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_{20};x) = &\frac{1}{n}\sum_{i=1}^{n}\frac{f^{(r+1)}(\mathcal{Y}_i(x))f_i(x) - f^{(r)}(\mathcal{Y}_i(x))\dot{f}_i(x)}{f_i^2(x)} \times \dot{m}_{1i}(x)\,\mu_r\hbar^r/r!K_i(x)W_i(x) \\
&-\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\nabla\mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x) \\
&+\frac{1}{n}\sum_{i=1}^{n}\frac{\dot{f}_i(x)}{f_i(x)}\dot{m}_{1i}(x)\,K_i(x)W_i(x) + o_p\left(\hbar^r + n^{-1/2} + h^2\right) \\
= &\frac{1}{n}\sum_{i=1}^{n}\frac{\dot{f}_i(x)}{f_i(x)}\dot{m}_{1i}(x)\,K_i(x)W_i(x) - \frac{1}{n^2}\sum_{i\neq j}^{n}\mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x) + O_p(\hbar^r) + o_p(h^2 + n^{-1/2}),
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{U}_2(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_{20}) = &\frac{1}{n}\sum_{i=1}^{n}\frac{f^{(r+1)}(\varepsilon_i)f(\varepsilon_i) - f^{(r)}(\varepsilon_i)\dot{f}(\varepsilon_i)}{f^2(\varepsilon_i)}\dot{\mathbf{m}}_2(\mathbf{X}_i)\mu_r\hbar^r/r! - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\nabla\mathbf{V}_2(\varepsilon_j,\varepsilon_i,\mathbf{X}_i) \\
&-\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{V}_2(\varepsilon_j,\varepsilon_i,\mathbf{X}_i) + \frac{1}{n}\sum_{i=1}^{n}\frac{\dot{f}(\varepsilon_i)}{f(\varepsilon_i)}\dot{\mathbf{m}}_2(\mathbf{X}_i) + o_p\left(\hbar^r + n^{-1/2}\right) \\
= &\frac{1}{n}\sum_{i=1}^{n}\frac{\dot{f}(\varepsilon_i)}{f(\varepsilon_i)}\dot{\mathbf{m}}_2(\mathbf{X}_i) - \frac{1}{n^2}\sum_{i\neq j}^{n}\mathbf{V}_2(\varepsilon_j,\varepsilon_i,\mathbf{X}_i) + O_p(\hbar^r) + o_p\left(n^{-1/2}\right),
\end{aligned}$$

where $f^{(r)}(x) = \partial^r f(x)/\partial x^r$. Then, the proof of Lemma A.2 is completed. □

A.4. Proof of Theorem 1. Under Model (1.1), we have

$$\mathbf{U}(\boldsymbol{\theta},\boldsymbol{\Theta}_2;x) = \mathbf{u}(\boldsymbol{\theta},\boldsymbol{\Theta}_2;x) + o_p(1),$$

where $\mathbf{U}(\boldsymbol{\theta},\boldsymbol{\Theta}_2;x) = (\mathbf{U}_1(\boldsymbol{\theta},\boldsymbol{\Theta}_2;x)',\mathbf{U}_2(\boldsymbol{\theta},\boldsymbol{\Theta}_2)')'$ and $\mathbf{u}(\boldsymbol{\theta},\boldsymbol{\Theta}_2;x) = (u_1(\boldsymbol{\Omega};x),0,\mathbf{u}_2(\boldsymbol{\theta},\boldsymbol{\Theta}_2)')'$. Suppose there exist two combinations $(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_{20})$ and $(\boldsymbol{\theta}_0+\mathbf{h},\boldsymbol{\Theta}_{20}+\mathbf{h}_{\boldsymbol{\Theta}_2})$ in $\mathcal{C}_0\times\mathcal{R}\times R^{p-1}$ such that $\mathbf{u}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_{20}; x) \equiv 0$ and $\mathbf{u}(\boldsymbol{\theta}_0 + \mathbf{h}, \boldsymbol{\Theta}_{20} + \mathbf{h}_{\boldsymbol{\Theta}_2}; x) \equiv 0$, where $\mathbf{h}(\cdot) = (h_1(\cdot), h_2(\cdot))'$. Denote $\mathbf{h}_{\boldsymbol{\Omega}} = (h_1(\cdot), \mathbf{h}_{\boldsymbol{\Theta}_2})'$, then,

$$
\begin{aligned}
0 &= \left[u_1(\boldsymbol{\Omega}_0 + \mathbf{h}_{\boldsymbol{\Omega}}; x) - u_1(\boldsymbol{\Omega}_0; x)\right]/p_1(x) \\
&= E\left[\frac{\underset{[0,1]^q}{\int\cdots\int} \dot{f}(\varepsilon_i - \Delta_{\boldsymbol{\Omega}_0}(\mathbf{h}_{\boldsymbol{\Omega}}; \mathbf{X}_i) + \Delta_{\boldsymbol{\Omega}_0}(\mathbf{h}_{\boldsymbol{\Omega}}; \mathbf{y}))p(\mathbf{y})d\mathbf{y}}{\underset{[0,1]^q}{\int\cdots\int} f(\varepsilon_i - \Delta_{\boldsymbol{\Omega}_0}(\mathbf{h}_{\boldsymbol{\Omega}}; \mathbf{X}_i) + \Delta_{\boldsymbol{\Omega}_0}(\mathbf{h}_{\boldsymbol{\Omega}}; \mathbf{y}))p(\mathbf{y})d\mathbf{y}} \times \dot{m}_1\left\{\mathbf{X}_i; \boldsymbol{\Omega}_0 + \mathbf{h}_{\boldsymbol{\Omega}}\right\} \mid \boldsymbol{\vartheta}_0'\mathbf{X}_{i1} = x\right], \\
0 &= u_2(\boldsymbol{\theta}_0 + \mathbf{h}, \boldsymbol{\Theta}_{20} + \mathbf{h}_{\boldsymbol{\Theta}_2}) - u_2(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_{20}) \\
&= E\left[\frac{\underset{[0,1]^q}{\int\cdots\int} \dot{f}(\varepsilon_i - \Delta_{\boldsymbol{\Omega}_0}(\mathbf{h}_{\boldsymbol{\Omega}}; \mathbf{X}_i) + \Delta_{\boldsymbol{\Omega}_0}(\mathbf{h}_{\boldsymbol{\Omega}}; \mathbf{y}))p(\mathbf{y})d\mathbf{y}}{\underset{[0,1]^q}{\int\cdots\int} f(\varepsilon_i - \Delta_{\boldsymbol{\Omega}_0}(\mathbf{h}_{\boldsymbol{\Omega}}; \mathbf{X}_i) + \Delta_{\boldsymbol{\Omega}_0}(\mathbf{h}_{\boldsymbol{\Omega}}; \mathbf{y}))p(\mathbf{y})d\mathbf{y}} \times \dot{\mathbf{m}}_2\left\{\mathbf{X}_i; \boldsymbol{\Omega}_0 + \mathbf{h}_{\boldsymbol{\Omega}}, \dot{g} + h\right\}\right],
\end{aligned}
$$

where $\Delta_{\boldsymbol{\Omega}}(\mathbf{h}_{\boldsymbol{\Omega}}; \mathbf{x}) = m\left\{\mathbf{x}; \boldsymbol{\Omega} + \mathbf{h}_{\boldsymbol{\Omega}}\right\} - m\left\{\mathbf{x}; \boldsymbol{\Omega}\right\}$ and $\mathbf{x} = (x_1, \mathbf{x}_2', \mathbf{x}_3')'$. By Condition 5, we have $\Delta_{\boldsymbol{\Omega}_0}(\mathbf{h}_{\boldsymbol{\Omega}}; \mathbf{x}) \equiv 0$ over $\mathbf{x} \in [0,1]^q$. Using identifiability conditions on Model (1.1), we have $h_1(x_1) \equiv 0$ over $x_1 \in [0,1]$ and $\mathbf{h}_{\boldsymbol{\Theta}_2} = 0$. By $\mathbf{U}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2; x) = 0$ and $\mathbf{u}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_{20}; x) = 0$, $(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_{20})$ is the unique root of $\mathbf{u}(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x) = 0$ in $\mathcal{C}_0 \times R^{p-1} \times \mathcal{R}$.

Write $\mathcal{B}_n = \{\alpha : \|\alpha\| \le D, \|\alpha(x_1) - \alpha(x_2)\| \le d[|x_1 - x_2| + b_n], x_1, x_2 \in [0,1]\}$ for some constants $D > 0$ and $d > 0$, where $b_n = h + \log(n)^{1/2}(nh^3)^{-1/2} + \hbar^r + \log(n)^{1/2}(n\hbar^5)^{-1/2}$. Write $\mathcal{A}_{x,\boldsymbol{\theta},\boldsymbol{\Theta}_2} = \{x \in [0,1], \alpha \in \mathcal{B}_n, \beta \in \mathcal{B}_n, \boldsymbol{\Theta}_2 \in R^{p-1}\}$. To show the uniform consistency of $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2)$, it suffices to prove the following (i)-(iii):

(i) For each continuous function vector $\alpha(\cdot), \beta(\cdot)$, and any parameter vector $\boldsymbol{\Theta}_2$,

$$\sup_{0 \le x \le 1} \|\mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x) - \mathbf{u}(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x)\| = o_p(1).$$

(ii) $\sup_{\mathcal{A}_{x,\boldsymbol{\theta},\boldsymbol{\Theta}_2}} \|\mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x) - \mathbf{u}(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x)\| = o_p(1)$.

(iii) $P\{\widehat{g} \in \mathcal{B}_n \cap \mathcal{C}_0, \widehat{\dot{g}} \in \mathcal{B}_n \cap \mathcal{R}\} \to 1$.

Once (i)-(iii) are established, using the similar idea to the Arzela-Ascoli theorem and (i-ii), we can show that for any subsequence of $\widehat{\boldsymbol{\Upsilon}}(\cdot) = \{\widehat{\boldsymbol{\theta}}(\cdot), \widehat{\boldsymbol{\Theta}}_2\}$, there exists a further convergent subsequence $\widehat{\boldsymbol{\Upsilon}}_n(\cdot) = \{\widehat{\boldsymbol{\theta}}_n(\cdot), \widehat{\boldsymbol{\Theta}}_{2n}\}$ such that uniformly in $x \in [0,1]$, $\widehat{\boldsymbol{\Upsilon}}_n(x) \to \boldsymbol{\Upsilon}^*(x) \equiv \{\boldsymbol{\theta}^*(x), \boldsymbol{\Theta}_2^*\} \equiv \{g^*(x), \dot{g}^*(x), \boldsymbol{\Theta}_2^*\}$ in probability. Note that $g^* \in \mathcal{C}_0$, $\dot{g}^* \in \mathcal{R}$, and

$\mathbf{u}(\boldsymbol{\theta}^*, \boldsymbol{\Theta}_2^*; x) = \mathbf{U}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2; x) - \left[\mathbf{U}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2; x) - \mathbf{u}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2; x)\right] - \left[\mathbf{u}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2; x) - \mathbf{u}(\boldsymbol{\theta}^*, \boldsymbol{\Theta}_2^*; x)\right]$, and $\mathbf{U}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2; x) = 0$. It follows from (ii) and (iii) that $\mathbf{u}(\boldsymbol{\theta}^*, \boldsymbol{\Theta}_2^*; x) = 0$. Since $\mathbf{u}(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x) = 0$ has a unique root at $(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_{20})$, we have $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}_2^* \equiv \boldsymbol{\Theta}_{20}$, which yields the uniform consistency of $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2)$.

Proof of (i). Using Lemma A.2 and noting that $\frac{1}{n^2}\sum_{i,j}^n \mathbf{V}_1(\varepsilon_j, \varepsilon_i, X_i; x) = O_p(n^{-1/2})$, by the central limit theorem of U-statistics, for any given $\boldsymbol{\theta}(x) = (\alpha(x), \beta(x))'$ and $\boldsymbol{\Omega} = (\alpha, \boldsymbol{\Theta}_2)$, we have

$$\sup_{x\in[0,1]} \left\|\mathbf{U}_1(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x) - u_1(\boldsymbol{\Omega}; x)(1,0)'\right\| \tag{A.10}$$
$$= \sup_{x\in[0,1]} \left\|\frac{1}{n}\sum_{i=1}^n \frac{\dot{f}_i(x;\boldsymbol{\theta},\boldsymbol{\Theta}_2)}{f_i(x;\boldsymbol{\theta},\boldsymbol{\Theta}_2)} \dot{m}_{1i}\left\{x;\boldsymbol{\theta},\boldsymbol{\Theta}_2\right\} K_i(x;\boldsymbol{\vartheta}) W_i(x;\boldsymbol{\vartheta}) - u_1(\boldsymbol{\Omega}; x)(1,0)'\right\| + o_p(1),$$
$$\sup_{x\in[0,1]} \left\|\mathbf{U}_2(\boldsymbol{\theta}, \boldsymbol{\Theta}_2) - \mathbf{u}_2(\boldsymbol{\theta}, \boldsymbol{\Theta}_2)\right\|,$$
$$= \sup_{x\in[0,1]} \left\|\frac{1}{n}\sum_{i=1}^n \frac{\dot{f}_i(\boldsymbol{\Omega})}{f_i(\boldsymbol{\Omega})} \dot{\mathbf{m}}_2\left(\mathbf{X}_i;\boldsymbol{\theta},\boldsymbol{\Theta}_2\right) - \mathbf{u}_2(\boldsymbol{\theta}, \boldsymbol{\Theta}_2)\right\| + o_p(1). \tag{A.11}$$

It can be shown in the same way as for (A.5) that the first terms of the right sides of ($A.10$) and ($A.11$) are $O_p\left\{h + \log(n)^{1/2}(nh^3)^{-1/2}\right\}$ and $O_p\left\{\log(n)^{1/2}n^{-1/2}\right\}$, respectively. (i) follows.

Proof of (ii). For any given continuous function vector $\boldsymbol{\theta}(x) = (\alpha(x), \beta(x))'$, $\boldsymbol{\eta}(x) = (\eta_1(x), \eta_2(x))'$, we have

$$\begin{aligned}
&\mathbf{U}_1(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x) - \mathbf{U}_1(\boldsymbol{\eta}, \boldsymbol{\Theta}_2; x)\\
&= \frac{1}{n}\sum_{i=1}^n \frac{\{\widehat{\dot{f}}_i(x;\boldsymbol{\theta},\boldsymbol{\Theta}_2) - \widehat{\dot{f}}_i(x;\boldsymbol{\eta},\boldsymbol{\Theta}_2)\}}{\widehat{f}_i(x;\boldsymbol{\theta},\boldsymbol{\Theta}_2)} \dot{m}_{1i}\left\{x;\boldsymbol{\theta},\boldsymbol{\Theta}_2\right\} K_i(x;\boldsymbol{\vartheta}) W_i(x;\boldsymbol{\vartheta})\\
&+ \frac{1}{n}\sum_{i=1}^n \frac{\{\widehat{f}_i(x;\boldsymbol{\eta},\boldsymbol{\Theta}_2) - \widehat{f}_i(x;\boldsymbol{\theta},\boldsymbol{\Theta}_2)\}\widehat{\dot{f}}_i(x;\boldsymbol{\eta},\boldsymbol{\Theta}_2)}{\widehat{f}_i(x;\boldsymbol{\theta},\boldsymbol{\Theta}_2)\widehat{f}_i(x;\boldsymbol{\eta},\boldsymbol{\Theta}_2)} \dot{m}_{1i}\left\{x;\boldsymbol{\theta},\boldsymbol{\Theta}_2\right\} K_i(x;\boldsymbol{\vartheta}) W_i(x;\boldsymbol{\vartheta})\\
&+ \frac{1}{n}\sum_{i=1}^n \frac{\widehat{\dot{f}}_i(x;\boldsymbol{\eta},\boldsymbol{\Theta}_2)}{\widehat{f}_i(x;\boldsymbol{\eta},\boldsymbol{\Theta}_2)} \left[\dot{m}_{1i}\left\{x;\boldsymbol{\theta},\boldsymbol{\Theta}_2\right\} - \dot{m}_{1i}\left\{x;\boldsymbol{\eta},\boldsymbol{\Theta}_2\right\}\right] K_i(x;\boldsymbol{\vartheta}) W_i(x;\boldsymbol{\vartheta}).
\end{aligned} \tag{A.12}$$

It follows from Condition 1 that

$$\sup_{\mathcal{A}_{x,\alpha,\beta,\delta}} |\widehat{\dot{f}}_i(x;\boldsymbol{\theta},\boldsymbol{\Theta}_2) - \widehat{\dot{f}}_i(x;\boldsymbol{\eta},\boldsymbol{\Theta}_2)|$$

$$
\begin{aligned}
&= \sup_{\mathcal{A}_{x,\alpha,\beta,\delta}} \Bigg| \frac{1}{n} \sum_{j=1}^{n} \dot{\mathcal{K}}_{\hbar} \left(Y_j - Y_i - m\left\{ \mathbf{X}_j; \boldsymbol{\Omega} \right\} + m_i \left\{ x; \boldsymbol{\theta}, \boldsymbol{\Theta}_2 \right\} \right) \\
&\qquad\qquad - \frac{1}{n} \sum_{j=1}^{n} \dot{\mathcal{K}}_{\hbar} \left(Y_j - Y_i - m\left\{ \mathbf{X}_j; \eta_1, \boldsymbol{\Theta}_2 \right\} + m_i \left\{ x; \boldsymbol{\eta}, \boldsymbol{\Theta}_2 \right\} \right) \Bigg| \\
&\leq M \sup_{x \in [0,1], \alpha \in \mathcal{B}_n} \|\boldsymbol{\theta}(x) - \boldsymbol{\eta}(x)\|. \qquad \text{(A.13)}
\end{aligned}
$$

where M is a constant and may be different at different places. Similarly,

$$
\sup_{\mathcal{A}_{x,\alpha,\beta,\delta}} |\widehat{f}_i(x; \boldsymbol{\eta}, \boldsymbol{\Theta}_2) - \widehat{f}_i(x; \boldsymbol{\theta}, \boldsymbol{\Theta}_2)| \leq M \sup_{x \in [0,1], \alpha \in \mathcal{B}_n} \|\boldsymbol{\theta}(x) - \boldsymbol{\eta}(x)\|.
$$

Note that, $\sup_{\mathcal{A}_{x,\alpha,\beta,\delta}} |\dot{m}_{1i} \left\{ x; \boldsymbol{\theta}, \boldsymbol{\Theta}_2 \right\} - \dot{m}_{1i} \left\{ x; \boldsymbol{\eta}, \boldsymbol{\Theta}_2 \right\}| \leq M \sup_{x \in [0,1], \alpha \in \mathcal{B}_n} \|\boldsymbol{\eta}(x) - \boldsymbol{\theta}(x)\|$. Then, it follows that

$$
\sup_{\mathcal{A}_{x,\alpha,\beta,\delta}} |\mathbf{U}_1(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x) - \mathbf{U}_1(\boldsymbol{\eta}, \boldsymbol{\Theta}_2; x)| \leq M \sup_{x \in [0,1], \alpha \in \mathcal{B}_n} \|\boldsymbol{\theta}(x) - \boldsymbol{\eta}(x)\|.
$$

Similarly, we can prove

$$
\sup_{\mathcal{A}_{x,\alpha,\beta,\delta}} |\mathbf{U}_2(\boldsymbol{\theta}, \boldsymbol{\Theta}_2) - \mathbf{U}_2(\boldsymbol{\eta}, \boldsymbol{\Theta}_2)| \leq M \sup_{x \in [0,1], \alpha \in \mathcal{B}_n} \|\boldsymbol{\eta}(x) - \boldsymbol{\theta}(x)\|.
$$

Hence, we have

$$
\sup_{\mathcal{A}_{x,\alpha,\beta,\delta}} |\mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x) - \mathbf{U}(\boldsymbol{\eta}, \boldsymbol{\Theta}_2; x)| \leq M \sup_{x \in [0,1], \alpha \in \mathcal{B}_n} \|\boldsymbol{\eta}(x) - \boldsymbol{\theta}(x)\|. \qquad \text{(A.14)}
$$

For any $\varepsilon > 0$, let $\mathcal{H}$ denote the finite set of points $mD/M, m = 0, \pm 1, \cdots, M$, where M is an integer such that $D/M < \varepsilon/3$ ($\mathcal{H}$ is an ε-net for the linear interval $[-D, D]$). Choose k and n large enough so that $d(1/k + b_n) < \varepsilon/3$. Let $\mathcal{Q}$ consist of those elements of the space of continuous functions on $[0, 1]$ that are linear on each subinterval $I_i = [(i-1)/k, i/k], i = 1, \cdots, k$, and the set of the function values at points $i/k, i = 0, 1, \cdots, k$ belong to $\mathcal{H}$. The size of the set $\mathcal{Q}$ is $(2M+1)^{k+1} = O((1/\varepsilon)^{1/\varepsilon})$. For any $\alpha(\cdot) \in \mathcal{B}_n$, there exists an element $\eta(\cdot)$ of $\mathcal{Q}$ such that $|\eta(i/k) - \alpha(i/k)| < \varepsilon/3, i = 0, 1, \cdots, k$. Since $\eta(\cdot)$ is linear on each subinterval I_i, it follows that $sup_{x \in [0,1], \alpha \in \mathcal{B}_n} |\eta(x) - \alpha(x)| < \varepsilon$. That is, for all n sufficiently large, $\mathcal{Q}$ is an ε-net for $\mathcal{B}_n$. As a result, $\log N(\varepsilon, \mathcal{B}_n, \|\cdot\|_\infty) \leq O(1/\varepsilon \log(1/\varepsilon)) = o(n)$, where $N(\varepsilon, \mathcal{B}_n, \|\cdot\|_\infty)$ is the covering number with respect to the norm $\|\cdot\|_\infty$ of the class $\mathcal{B}_n$.

Hence, with (i) and (A.14), as well as

$\sup_{\mathcal{A}_{x,\alpha,\beta,\delta}} |\mathbf{u}(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x) - \mathbf{u}(\boldsymbol{\eta}, \boldsymbol{\Theta}_2; x)| \leq M \sup_{x\in[0,1],\alpha\in\mathcal{B}_n} \|\boldsymbol{\eta}(x) - \boldsymbol{\theta}(x)\|$, by the uniform law of large numbers (Pollard (1990), p.39), (ii) is proved.

Proof of (iii). Note that $\mathbf{U}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2; x) = 0$ for all $x \in R$. Hence,

$$\begin{aligned}
0 &= \mathbf{U}_1(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2; x_1) - \mathbf{U}_1(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2; x_2) \\
&= \frac{1}{n}\sum_{i=1}^n \frac{\widehat{\dot{f}}_i(x_1; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2) - \widehat{\dot{f}}_i(x_2; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2)}{\widehat{f}_i(x_1; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2)} \dot{m}_{1i}\left\{x_1; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2\right\} K_i(x_1; \widehat{\boldsymbol{\vartheta}}) W_i(x_1; \widehat{\boldsymbol{\vartheta}}) \\
&\quad -\frac{1}{n}\sum_{i=1}^n \frac{\left\{\widehat{f}_i(x_1; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2) - \widehat{f}_i(x_2; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2)\right\} \widehat{\dot{f}}_i(x_2; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2)}{\widehat{f}_i(x_1; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2)\widehat{f}_i(x_2; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2)} \dot{m}_{1i}\left\{x_1; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2\right\} K_i(x_1; \widehat{\boldsymbol{\vartheta}}) W_i(x_1; \widehat{\boldsymbol{\vartheta}}) \\
&\quad +\frac{1}{n}\sum_{i=1}^n \frac{\widehat{\dot{f}}_i(x_2; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2)}{\widehat{f}_i(x_2; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2)} K_i(x_1; \widehat{\boldsymbol{\vartheta}}) \left[\dot{m}_{1i}\left\{x_1; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2\right\} - \dot{m}_{1i}\left\{x_2; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2\right\}\right] W_i(x_1; \widehat{\boldsymbol{\vartheta}}) \\
&\quad +\frac{1}{n}\sum_{i=1}^n \frac{\widehat{\dot{f}}_i(x_2; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2)}{\widehat{f}_i(x_2; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2)} \left\{K_i(x_1; \widehat{\boldsymbol{\vartheta}}) W_i(x_1; \widehat{\boldsymbol{\vartheta}}) - K_i(x_2; \widehat{\boldsymbol{\vartheta}}) W_i(x_2; \widehat{\boldsymbol{\vartheta}})\right\} \dot{m}_{1i}\left\{x_2; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2\right\} \\
&= \mathbf{G}_1(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2; x_1; x_0, x_1, x_2) \begin{pmatrix} \widehat{g}(x_1) - \widehat{g}(x_2) \\ \widehat{\dot{g}}(x_1) - \widehat{\dot{g}}(x_2) \end{pmatrix} + \mathbf{G}_2(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2; x_0, x_1, x_2)(x_1 - x_2) + O_p(b_n),
\end{aligned}$$

where x_0 is between x_1 and x_2, and may be different in different functions,

$$\mathbf{G}_1(\cdot) = \begin{pmatrix} G_{11}(\cdot) & G_{12}(\cdot) \\ G_{12}(\cdot) & G_{13}(\cdot) \end{pmatrix}, \quad \mathbf{G}_2(\cdot) = \begin{pmatrix} G_{21}(\cdot) \\ G_{22}(\cdot) \end{pmatrix},$$

$$\begin{aligned}
G_i(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x_0, x_1, x_2) = &\left\{\frac{\dot{f}_i(x_0; \boldsymbol{\theta}, \boldsymbol{\Theta}_2)\dot{f}_i(x_2; \boldsymbol{\theta}, \boldsymbol{\Theta}_2)}{f_i(x_1; \boldsymbol{\theta}, \boldsymbol{\Theta}_2) f_i(x_2; \boldsymbol{\theta}, \boldsymbol{\Theta}_2)} - \frac{\ddot{f}_i(x_0; \boldsymbol{\theta}, \boldsymbol{\Theta}_2)}{f_i(x_1; \boldsymbol{\theta}, \boldsymbol{\Theta}_2)}\right\} \\
&\times \dot{m}_{1i}\left\{x_0; \boldsymbol{\theta}, \boldsymbol{\Theta}_2\right\} \dot{m}_{1i}\left\{x_1; \boldsymbol{\theta}, \boldsymbol{\Theta}_2\right\} + \frac{\dot{f}_i(x_2; \boldsymbol{\theta}, \boldsymbol{\Theta}_2)}{f_i(x_2; \boldsymbol{\theta}, \boldsymbol{\Theta}_2)} \ddot{m}_1\left\{x_0; \boldsymbol{\theta}, \boldsymbol{\Theta}_2\right\},
\end{aligned}$$

$$G_{11}(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x_1; x_0, x_1, x_2) = E\left\{G_i(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x_0, x_1, x_2) \mid \boldsymbol{\vartheta}'\mathbf{X}_{i1} = x_1\right\} p_1(x_1),$$

$$\begin{aligned}
G_{12}(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x_1; x_0, x_1, x_2) = &\ h^2\mu_2 E\left\{\frac{\partial G_i(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x_0, x_1, x_2)}{\partial \boldsymbol{\vartheta}'\mathbf{X}_{i1}} \mid \boldsymbol{\vartheta}'\mathbf{X}_{i1} = x_1\right\} p(x_1) \\
&+ h^2\mu_2 E\left\{G_i(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x_0, x_1, x_2) \mid \boldsymbol{\vartheta}'\mathbf{X}_{i1} = x_1\right\} \dot{p}_1(x_1),
\end{aligned}$$

$$G_{13}(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x_1; x_0, x_1, x_2) = h^2\mu_2 G_{11}(\boldsymbol{\theta}, \boldsymbol{\Theta}_2; x_1; x_0, x_1, x_2),$$

$$G_{21}(\boldsymbol{\theta},\boldsymbol{\Theta}_2;x_0,x_1,x_2)=\beta(x_2)\left[G_{11}(\boldsymbol{\theta},\boldsymbol{\Theta}_2;x_0;x_2,x_2,x_2)-G_{11}(\boldsymbol{\theta},\boldsymbol{\Theta}_2;x_1;x_0,x_1,x_2)\right]$$
$$+E\left[\frac{\dot f_i(x_2;\boldsymbol{\theta},\boldsymbol{\Theta}_2)}{f_i(x_2;\boldsymbol{\theta},\boldsymbol{\Theta}_2)}\dot m_{1i}\left\{x_2;\boldsymbol{\theta},\boldsymbol{\Theta}_2\right\}\mid \boldsymbol{\vartheta}'\mathbf{X}_{i1}=x_0\right]\dot p_1(x_0),$$
$$G_{22}(\boldsymbol{\theta},\boldsymbol{\Theta}_2;x_0,x_1,x_2)=\beta(x_2)\left[G_{12}(\boldsymbol{\theta},\boldsymbol{\Theta}_2;x_0;x_2,x_2,x_2)-G_{12}(\boldsymbol{\theta},\boldsymbol{\Theta}_2;x_1;x_0,x_1,x_2)\right]$$
$$+h^2\mu_2E\left[\beta(x_2)G_i(\boldsymbol{\theta},\boldsymbol{\Theta}_2;x_2,x_2,x_2)\dot p_1(x_0)+\frac{\dot f_i(x_2;\boldsymbol{\theta},\boldsymbol{\Theta}_2)}{f_i(x_2;\boldsymbol{\theta},\boldsymbol{\Theta}_2)}\dot m_{1i}\left\{x_2;\boldsymbol{\theta},\boldsymbol{\Theta}_2\right\}\ddot p_1(x_0)\mid \boldsymbol{\vartheta}'\mathbf{X}_{i1}=x_0\right].$$

Hence, *(iii)* follows. ∎

A.5. Proofs of Theorems 2 and 3. Denote $d_{n,1}=\sup_x|\widehat g(x)-g(x)|$, $d_{n,2}=h\sup_x|\widehat{\dot g}(x)-\dot g(x)|$, $d_{n,3}=\|\widehat{\boldsymbol{\Theta}}_2-\boldsymbol{\Theta}_2\|$ and $\boldsymbol{\theta}_0=(g,\dot g)'$ to be the true value of $\boldsymbol{\theta}$. By simple calculation, we have

$$\begin{aligned}
&\mathbf{U}_1(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Theta}}_2;x)-\mathbf{U}_1(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_{20};x)\\
&=\{\widehat g(x)-g(x)\}\dot M_{1,1}^2(x)\tau p_1(x)\,(1,0)'+h^2\{\widehat{\dot g}(x)-\dot g(x)\}\dot M_{1,1}^2(x)\tau p_1(x)\mu_2(0,1)'\\
&\quad-\int \dot M_{1,1}(u)\{\widehat g(u)-g(u)\}p_1(u)du\dot M_{1,1}(x)\tau p_1(x)(1,0)'\\
&\quad+\left\{\dot{\mathbf{M}}_{1,2}(x)-\dot M_{1,1}(x)E\left\{\dot{\mathbf{M}}_{2,2}(\boldsymbol{\vartheta}_0'\mathbf{X}_{i1})\right\}\right\}'(\widehat{\boldsymbol{\Theta}}_2-\boldsymbol{\Theta}_{20})\tau p_1(x)(1,0)'\\
&\quad+O_p\left\{(d_{n,1}+d_{n,2}+d_{n,3}+h+\hbar^r+(n\hbar^5)^{-1/2}(\log n)^{1/2})(d_{n,1}+d_{n,2}+d_{n,3})\right\},\\
&\mathbf{U}_2(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Theta}}_2)-\mathbf{U}_2(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_{20})\\
&=\tau\int_0^1\dot{\mathbf{M}}_{1,2}(u)\left\{\widehat g(u)-g(u)\right\}p_1(u)du-\tau\int_0^1\dot M_{1,1}(u)\left\{\widehat g(u)-g(u)\right\}p_1(u)du\int_0^1\dot{\mathbf{M}}_{2,2}(u)p_1(u)du\\
&\quad+\tau E\left\{\dot{\mathbf{M}}_{2,2}^2(\boldsymbol{\vartheta}_0'\mathbf{X}_{i1})-\dot{\mathbf{M}}_{2,2}(\boldsymbol{\vartheta}_0'\mathbf{X}_{i1})E\left(\dot{\mathbf{M}}_{2,2}(\boldsymbol{\vartheta}_0'\mathbf{X}_{j1})\right)'\right\}(\widehat{\boldsymbol{\Theta}}_2-\boldsymbol{\Theta}_{20})\\
&\quad+O_p\left\{(d_{n,1}+d_{n,3}+\hbar^r+(n\hbar^5)^{-1/2}(\log n)^{1/2})(d_{n,1}+d_{n,3})\right\}.
\end{aligned}\tag{A.15}$$

On the other hand, by Lemma A.2,

$$\begin{aligned}
\mathbf{U}_1(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_{20};x)&=-\frac{h^2\ddot g(x)\mu_2}{2}\tau p_1(x)\dot M_{1,1}^2(x)(1,0)'-\frac{1}{n^2}\sum_{i\neq j}^n\mathbf{V}_1(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x)\\
&\quad+\frac1n\sum_{i=1}^n\frac{\dot f(\varepsilon_i)}{f(\varepsilon_i)}\dot m_{1i}\left\{x;\boldsymbol{\theta},\boldsymbol{\Theta}_{20}\right\}K_i(x)W_i(x)+O_p\left(\hbar^r\right)+o_p(h^2+n^{-1/2}),
\end{aligned}\tag{A.16}$$

$$\mathbf{U}_2(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_{20}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\dot{f}(\varepsilon_i)}{f(\varepsilon_i)}\dot{\mathbf{m}}_2(\mathbf{X}_i) - \frac{1}{n^2}\sum_{i\neq j}^{n} \mathbf{V}_2(\varepsilon_j, \varepsilon_i, \mathbf{X}_i) + O_p\left(\hbar^r\right) + o_p\left(n^{-1/2}\right),$$

where $\dot{\mathbf{M}}_{j,k}^r(x) = \dot{\mathbf{M}}_{j,k}^r(x)$, $\dot{\mathbf{M}}_{2,2}^r(x;\boldsymbol{\theta},\boldsymbol{\Theta}_2) = E\left[\dot{\mathbf{m}}_2^{\otimes r}\left\{\mathbf{X}_j;\boldsymbol{\theta},\boldsymbol{\Theta}_2\right\} \mid \boldsymbol{\vartheta}'\mathbf{X}_{j1} = x\right], r = 1, 2, q^1 = q,$ $q^{\otimes 2} = q'q$, $\dot{\mathbf{M}}_{1,2}(x;\boldsymbol{\theta},\boldsymbol{\Theta}_2)$, and $\dot{M}_{1,1}^r(x;\boldsymbol{\theta},\boldsymbol{\Theta}_2)$ are defined in Section 3, $\mathbf{V}_1(\varepsilon_j, \varepsilon_i, \mathbf{X}_i; x)$ and $\mathbf{V}_2(\varepsilon_j, \varepsilon_i, \mathbf{X}_i)$ are defined in Lemma A.2.

Let $V_{11}(\varepsilon_j, \varepsilon_i, \mathbf{X}_i; x)$ to be the first component of $\mathbf{V}_1(\varepsilon_j, \varepsilon_i, \mathbf{X}_i; x)$. Substituting ($A.16$) into ($A.15$), noting that $\mathbf{U}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}_2; x) = 0$, we obtain

$$\begin{aligned}
&\dot{M}_{1,1}^2(x)\left(\widehat{g}(x) - g(x)\right) - \int \dot{M}_{1,1}(u)\left(\widehat{g}(u) - g(u)\right)p_1(u)du\dot{M}_{1,1}(x) \\
&\qquad + \left\{\dot{\mathbf{M}}_{1,2}(x) - \dot{M}_{1,1}(x)E\left(\dot{\mathbf{M}}_{2,2}(\boldsymbol{\vartheta}_0'\mathbf{X}_{i1})\right)\right\}'\left(\widehat{\boldsymbol{\Theta}}_2 - \boldsymbol{\Theta}_{20}\right) \\
&\quad = \frac{h^2\ddot{g}(x)\mu_2}{2}\dot{M}_{1,1}^2(x) + \frac{1}{n^2\tau p_1(x)}\sum_{i\neq j}^{n} V_{11}(\varepsilon_j, \varepsilon_i, \mathbf{X}_i; x) - \frac{1}{n\tau p_1(x)}\sum_{i=1}^{n}\frac{\dot{f}(\varepsilon_i)}{f(\varepsilon_i)}\dot{m}_{1i}(x)K_i(x) \\
&\quad\quad + O_p\left(\hbar^r + d_{n,1}^2 + d_{n,2}^2 + (n\hbar^5)^{-1/2}(\log n)^{1/2}(d_{n,1} + d_{n,2})\right) + o_p(h^2 + n^{-1/2}), \\
&d_{n,2} = O_p\left((\log n)^{1/2}(nh)^{-1/2} + \hbar^r\right) + o_p(h^2 + n^{-1/2}), \qquad\qquad \text{(A.17)} \\
&\mathrm{Var}\left[\dot{\mathbf{m}}_2(\mathbf{X}_i)\right]\left(\widehat{\boldsymbol{\Theta}}_2 - \boldsymbol{\Theta}_{20}\right) + \mathcal{A}_m\left(\widehat{g} - g\right) \\
&\quad = -\frac{1}{n^2\tau}\sum_{i\neq j}^{n}\left\{\frac{n^2}{n(n-1)} - \frac{\mathcal{K}_\hbar(\varepsilon_j - \varepsilon_i) - E\left(\mathcal{K}_\hbar(\varepsilon_j - \varepsilon_i) \mid \varepsilon_i\right)}{f(\varepsilon_i)}\right\}\frac{\dot{f}(\varepsilon_i)}{f(\varepsilon_i)}\dot{\mathbf{m}}_2(\mathbf{X}_i) \\
&\quad\quad + O_p\left(\hbar^r + d_{n1}^2 + (n\hbar^5)^{-1/2}(\log n)^{1/2}d_{n,1}\right) + o_p(n^{-1/2}),
\end{aligned}$$

where $\dot{m}_{1i}(x) = \dot{m}_{1i}\{x\}$, $\mathcal{A}_m\left(\widehat{g} - g\right) = \int_0^1 D(u)\left\{\widehat{g}(u) - g(u)\right\}p_1(u)du$ and $D(x)$ is defined in Section 3.

Note that

$$E\left\{C(\mathbf{X}_i)\right\} = E\left\{\frac{E^2\left[\dot{m}_1(\mathbf{X}_i) \mid \boldsymbol{\vartheta}'\mathbf{X}_{i1} = U\right]}{E\left[\dot{m}_1^2(\mathbf{X}_i) \mid \boldsymbol{\vartheta}'\mathbf{X}_{i1} = U\right]}\right\} \leq 1$$

The equality holds if and only if $\dot{m}_1(\mathbf{X}_i) = E\left[\dot{m}_1(\mathbf{X}_i) \mid \boldsymbol{\vartheta}'\mathbf{X}_{i1} = x\right]$.

When $E\left\{C(\mathbf{X}_i)\right\} \neq 1$, we have

$$\left\{\mathrm{Var}\left[\dot{\mathbf{m}}_2(\mathbf{X}_i)\right] - \int \mathcal{B}(u)p_1(u)du\right\}\left(\widehat{\boldsymbol{\Theta}}_2 - \boldsymbol{\Theta}_{20}\right)$$

$$= -\frac{1}{n^2\tau}\sum_{i\neq j}^{n}\left\{\frac{n^2}{n(n-1)} - \frac{\mathcal{K}_\hbar(\varepsilon_j-\varepsilon_i) - E\left(\mathcal{K}_\hbar(\varepsilon_j-\varepsilon_i)\mid\varepsilon_i\right)}{f(\varepsilon_i)}\right\}\frac{\dot f(\varepsilon_i)}{f(\varepsilon_i)}\left[\frac{E\left\{B(\mathbf{X}_i)\right\}C(\mathbf{X}_i)}{1-E\left\{C(\mathbf{X}_i)\right\}}\right.$$
$$\left.+B(\mathbf{X}_i)\right] + O_p\left(h^2+\hbar^r+d_{n,1}^2+(n\hbar^5)^{-1/2}(\log n)^{1/2}d_{n,1}\right)+o_p(n^{-1/2}), \tag{A.18}$$

where $\mathcal{B}(x)$, $B(\mathbf{X}_i)$, $C(\mathbf{X}_i)$ are defined in Section 3.

When $E\left\{C(\mathbf{X}_i)\right\}\equiv 1$, we have $\dot{\mathbf{M}}_{1,2}(x) = \dot M_{1,1}(x)\dot{\mathbf{M}}_{2,2}(x)$ and $\int \mathcal{B}(u)p_1(u)du = \int \dot{\mathbf{M}}_{2,2}(u)\left\{\dot{\mathbf{M}}_{2,2}(u) - E\dot{\mathbf{M}}_{2,2}(\boldsymbol{\vartheta}_0'\mathbf{X}_{i1})\right\}' p_1(u)du$.

Then, we obtain

$$\operatorname{Var}\left[\dot{\mathbf{m}}_2\left\{\mathbf{X}_i\right\} - \dot{\mathbf{M}}_{2,2}(\boldsymbol{\vartheta}_0'\mathbf{X}_{i1})\right](\widehat{\boldsymbol{\Theta}}_2 - \boldsymbol{\Theta}_{20}) \tag{A.19}$$
$$= -\frac{1}{n^2\tau}\sum_{i\neq j}^{n}\left\{\frac{n^2}{n(n-1)} - \frac{\mathcal{K}_\hbar(\varepsilon_j-\varepsilon_i) - E\left(\mathcal{K}_\hbar(\varepsilon_j-\varepsilon_i)\mid\varepsilon_i\right)}{f(\varepsilon_i)}\right\}\frac{\dot f(\varepsilon_i)}{f(\varepsilon_i)}$$
$$\times\left[\dot{\mathbf{m}}_2(\mathbf{X}_i) - \dot{\mathbf{M}}_{2,2}(\boldsymbol{\vartheta}_0'\mathbf{X}_{i1})\right] + O_p\left(\hbar^r + d_{n,1}^2 + (n\hbar^5)^{-1/2}(\log n)^{1/2}d_{n,1}\right) + o_p(n^{-1/2}).$$

On the other hand, we have

$$\widehat g(x) - g(x) - \frac{\dot M_{1,1}(x)}{\dot M_{1,1}^2(x)}\int \dot M_{1,1}(u)\left(\widehat g(u) - g(u)\right)p_1(u)du$$
$$= \frac{h^2\ddot g(x)\mu_2}{2} + \frac{1}{n^2\tau p_1(x)\dot M_{1,1}^2(x)}\sum_{i\neq j}^{n} V_{11}(\varepsilon_j,\varepsilon_i,\mathbf{X}_i;x) - \frac{1}{n\tau p_1(x)}\sum_{i=1}^{n}\frac{\dot f(\varepsilon_i)}{f(\varepsilon_i)}\frac{\dot m_{1i}(x)}{\dot M_{1,1}^2(x)}K_i(x)$$
$$+O_p\left(\hbar^r + d_{n,1}^2 + d_{n,2}^2 + (n\hbar^5)^{-1/2}(\log n)^{1/2}(d_{n,1}+d_{n,2})\right) + o_p(h^2+n^{-1/2}), \tag{A.20}$$

and

$$d_{n,2} = O_p\left(\log(n)^{1/2}(nh)^{-1/2} + \hbar^r\right) + o_p(h^2). \tag{A.21}$$

Using the central limit theorem of U-statistics, we have

$$\frac{1}{n^2\tau p_1(x)\dot M_{1,1}(x)}\sum_{i\neq j}^{n}\mathbf{V}_1(\varepsilon_j,\varepsilon_i,X_i;x) = O_p(n^{-1/2}).$$

Taking supremum norm on both side of (A.20), and using Condition 6, we get $d_{n,1} = O_p\{h^2 + (nh)^{-1/2}\}$. Then Theorems 2 and 3 follow from (A.19) - (A.21). ∎

A.6. Proof of Theorem 4. First we consider the asymptotic variance of $\int_0^1 \widehat{g}(x)\psi_1(x)dx + \widehat{\boldsymbol{\Theta}}_2'\boldsymbol{\psi}_2$. From (A.17), using $nh^4 \to 0$, $nh^2/(\log n)^2 \to \infty$, and $h\hbar^5 h/(\log n)^2 \to \infty$, we have

$$\begin{aligned}
&\int \{\widehat{g}(x) - g(x)\}\,\psi_1(x)dx + (\widehat{\boldsymbol{\Theta}}_2 - \boldsymbol{\Theta}_{20})'\boldsymbol{\psi}_2 \qquad (A.22)\\
&= -\frac{1}{n^2\tau}\sum_{i\neq j}^{n}\left\{\frac{n^2}{n(n-1)} - \frac{\mathcal{K}_\hbar(\varepsilon_j-\varepsilon_i) - E\left(\mathcal{K}_\hbar(\varepsilon_j-\varepsilon_i)\mid\varepsilon_i\right)}{f(\varepsilon_i)}\right\}\frac{\dot{f}(\varepsilon_i)}{f(\varepsilon_i)}\\
&\quad\times\left[\dot{m}_1(\mathbf{X}_i)\phi_1(\boldsymbol{\vartheta}_0'\mathbf{X}_{i1}) + \boldsymbol{\phi}_2'\dot{\mathbf{m}}_2(\mathbf{X}_i)\right] + o_p\left(n^{-1/2}\right).
\end{aligned}$$

Coupling with the central limit theorem of U-statistic, we hence obtain,

$$\sqrt{n}\int\{\widehat{g}(x) - g(x)\}\psi_1(x)dx + (\widehat{\boldsymbol{\Theta}}_2 - \boldsymbol{\Theta}_{20})'\boldsymbol{\psi}_2 \to N(0,\tilde{\sigma}^2), \qquad (A.23)$$

where

$$\tilde{\sigma}^2 \equiv \frac{1}{\tau}\mathrm{Var}\left[\dot{m}_1(\mathbf{X}_i)\phi_1(\boldsymbol{\vartheta}_0'\mathbf{X}_{i1}) + \boldsymbol{\phi}_2'\dot{\mathbf{m}}_2(\mathbf{X}_i)\right].$$

To show the asymptotic efficiency of $\int_0^1 \widehat{g}(x)\psi_1(x)dx + \widehat{\boldsymbol{\Theta}}_2'\boldsymbol{\psi}_2$, we consider a parametric submodel with unknown parameter δ,

$$\boldsymbol{\Theta}_2(\delta) = \boldsymbol{\Theta}_{20} + \delta\tau^{-1}\boldsymbol{\phi}_2,\;\; g(\boldsymbol{\vartheta}_0'\mathbf{x};\delta) = g(\boldsymbol{\vartheta}(\delta)'\mathbf{x}) + \delta\tau^{-1}\phi_1(\boldsymbol{\vartheta}(\delta)'\mathbf{x}),$$

and the only unknown piece of the distribution of ε is a location parameter, i.e., $f(\varepsilon;\alpha) = f(\varepsilon-\alpha)$. Obviously, $\delta_0 = 0$ is the true value of δ and $\alpha_0 = 0$ is the true value of α. The score function for this parametric submodel is

$$\begin{aligned}
S_\delta &= -\frac{1}{\sqrt{n}\tau}\sum_{i=1}^{n}\left\{\frac{\dot{f}\left(Y_i - m\left\{\mathbf{X}_i; g(\cdot;\delta),\boldsymbol{\Theta}_2(\delta)\right\} - \alpha\right)}{f\left(Y_i - m\left\{\mathbf{X}_i; g(\cdot;\delta),\boldsymbol{\Theta}_2(\delta)\right\} - \alpha\right)}\right\}\left[\dot{m}_1\left\{\mathbf{X}_i; g(\cdot;\delta),\boldsymbol{\Theta}_2(\delta)\right\}\right.\\
&\quad\times\left\{\phi_1\left(\boldsymbol{\vartheta}(\delta)'\mathbf{X}_{i1}\right) + \dot{g}(\boldsymbol{\vartheta}(\delta)'\mathbf{X}_{i1})\boldsymbol{\phi}_{21}'\mathbf{J}'\mathbf{X}_{i1} + \delta\tau^{-1}\dot{\phi}_1(\boldsymbol{\vartheta}(\delta)'\mathbf{X}_{i1})\boldsymbol{\phi}_{21}'\mathbf{J}'\mathbf{X}_{i1}\right\}\\
&\qquad\left.+\boldsymbol{\phi}_{22}'\dot{m}_2\left\{\mathbf{X}_i;\boldsymbol{\theta}(\cdot;\delta),\boldsymbol{\Theta}_2(\delta)\right\}\mathbf{X}_{i2} + \boldsymbol{\phi}_{23}'\dot{\mathbf{m}}_3\left\{\mathbf{X}_i;\boldsymbol{\theta}(\cdot;\delta),\boldsymbol{\Theta}_2(\delta)\right\}\right],\\
S_\alpha &= -\frac{1}{\sqrt{n}\tau}\sum_{i=1}^{n}\left\{\frac{\dot{f}\left(Y_i - m\left\{\mathbf{X}_i; g(\cdot;\delta),\boldsymbol{\Theta}_2(\delta)\right\} - \alpha\right)}{f\left(Y_i - m\left\{\mathbf{X}_i; g(\cdot;\delta),\boldsymbol{\Theta}_2(\delta)\right\} - \alpha\right)}\right\},
\end{aligned}$$

where $\boldsymbol{\vartheta}(\delta) = \boldsymbol{\vartheta} + \delta\tau^{-1}\boldsymbol{\phi}_{21}$ and $\boldsymbol{\phi}_2 = (\boldsymbol{\phi}_{21}', \boldsymbol{\phi}_{22}', \boldsymbol{\phi}_{23}')'$. The partial score for δ at (δ_0, α_0) is

$$\begin{aligned} & S_\delta - E\left[S_\delta S_\alpha\right]\left(E\left[S_\alpha^2\right]\right)^{-1} S_\alpha \\ &= \frac{1}{\sqrt{n}\tau}\sum_{i=1}^{n}\frac{\dot{f}(\varepsilon_i)}{f(\varepsilon_i)}\Big[\dot{m}_1\left\{\mathbf{X}_i; g, \boldsymbol{\Theta}_2\right\}\phi_1(\boldsymbol{\vartheta}'\mathbf{X}_{i1}) + \boldsymbol{\phi}_2'\dot{\mathbf{m}}_2\left(\mathbf{X}_i; \boldsymbol{\theta}, \boldsymbol{\Theta}_2\right) \\ &\qquad - \int \dot{M}_{1,1}(x)\phi_1(x)p_1(x)dx - \boldsymbol{\phi}_2'\int \dot{\mathbf{M}}_{2,2}(x)p_1(x)dx\Big]. \end{aligned}$$

Hence the variance of the partial score at (δ_0, α_0) is $\tilde{\sigma}^2$. Thus, the maximum likelihood estimator of δ, denoted as $\tilde{\delta}$, satisfies

$$\sqrt{n}(\tilde{\delta} - \delta_0) \to N(0, \tilde{\sigma}^{-2}). \tag{A.24}$$

Observe that

$$\begin{aligned} & \int \{g(x;\tilde{\delta}) - g(x;\delta_0)\}\psi_1(x)dx + \{\boldsymbol{\Theta}_2(\tilde{\delta}) - \boldsymbol{\Theta}_2(\delta_0)\}'\boldsymbol{\psi}_2 \\ &\qquad = (\tilde{\delta} - \delta_0)\left\{\int \psi_1(x)\phi_1(x)dx + \boldsymbol{\phi}_2'\boldsymbol{\psi}_2\right\}/\tau. \end{aligned} \tag{A.25}$$

Thus, it follows from $(A.24)$ and $(A.25)$ that

$$\sqrt{n}\int_0^1 \{g(x;\tilde{\delta}) - g(x;\delta_0)\}\psi_1(x)dx + \{\boldsymbol{\Theta}_2(\tilde{\delta}) - \boldsymbol{\Theta}_2(\delta_0)\}'\boldsymbol{\psi}_2 \to N(0, \tilde{\sigma}^2).$$

This, together with (A.22) and (A.23), shows that the asymptotic variance of $\int_0^1 \widehat{g}(x)\psi_1(x)dx + \widehat{\boldsymbol{\Theta}}_2'\boldsymbol{\psi}_2$ is the same as that of $\int_0^1 g(x;\tilde{\delta})\psi_1(x)dx + \boldsymbol{\Theta}_2(\tilde{\delta})'\boldsymbol{\psi}_2$. As explained in Bickel *et al.* (1993, pp46), $\int_0^1 \widehat{g}(x)\psi_1(x)dx + \widehat{\boldsymbol{\Theta}}_2'\boldsymbol{\psi}_2$ is asymptotically efficient estimator of $\int_0^1 g(x)\psi_1(x)dx + \boldsymbol{\Theta}_2'\boldsymbol{\psi}_2$. The proof of Theorem 4 is completed. ∎

Table 1

The simulation results for Example 1: nonparametric part

h	Proposed			LLS		
	bias	sd	RMSE	bias	sd	RMSE
	Model 1					
0.2	0.0030	0.0332	0.0333	0.0022	0.0444	0.0444
0.3	0.0055	0.0306	0.0311	0.0037	0.0387	0.0389
0.4	0.0094	0.0291	0.0306*	0.0067	0.0357	0.0363
0.5	0.0132	0.0303	0.0331	0.0105	0.0340	0.0356*
0.6	0.0174	0.0298	0.0345	0.0146	0.0330	0.0361
	Model 2					
0.2	0.0023	0.0339	0.0340	0.0033	0.0779	0.0779
0.3	0.0047	0.0323	0.0327*	0.0037	0.0678	0.0678
0.4	0.0079	0.0317	0.0327	0.0066	0.0622	0.0625
0.5	0.0118	0.0316	0.0337	0.0106	0.0589	0.0598
0.6	0.0159	0.0318	0.0356	0.0148	0.0569	0.0588*
0.7	0.0202	0.0325	0.0382	0.0191	0.0559	0.0591

Note: * stands for the minimum RMSE

Table 2

The simulation results for Example 1: parametric part.

		Proposed			LLS		
	h	bias	sd	RMSE	bias	sd	RMSE
		Model 1					
β_1	0.2	0.0010	0.0214	0.0215	-0.0004	0.0286	0.0286
	0.3	0.0012	0.0209	0.0210	-0.0003	0.0282	0.0282
	0.4	-0.0003	0.0207	0.0207	-0.0002	0.0280	0.0280
	0.5	0.0013	0.0210	0.0211	-0.0002	0.0279	0.0279
	0.6	0.0016	0.0213	0.0213	-0.0001	0.0278	0.0278
β_2	0.2	0.0023	0.0453	0.0454	0.0021	0.0462	0.0462
	0.3	0.0025	0.0457	0.0457	0.0026	0.0462	0.0462
	0.4	0.0043	0.0458	0.0460	0.0028	0.0464	0.0464
	0.5	0.0027	0.0500	0.0500	0.0029	0.0467	0.0468
	0.6	0.0047	0.0495	0.0497	0.0029	0.0471	0.0472
		Model 2					
β_1	0.2	-0.0000	0.0131	0.0131	-0.0004	0.0487	0.0487
	0.3	-0.0001	0.0130	0.0130	-0.0002	0.0482	0.0482
	0.4	0.0001	0.0130	0.0130	-0.0000	0.0478	0.0478
	0.5	-0.0000	0.0131	0.0131	0.0001	0.0476	0.0476
	0.6	-0.0001	0.0132	0.0132	0.0001	0.0475	0.0475
	0.7	-0.0002	0.0133	0.0133	0.0001	0.0473	0.0473
β_2	0.2	0.0002	0.0304	0.0304	-0.0023	0.0780	0.0781
	0.3	-0.0001	0.0291	0.0291	-0.0031	0.0779	0.0779
	0.4	-0.0003	0.0287	0.0287	-0.0030	0.0779	0.0779
	0.5	-0.0004	0.0296	0.0296	-0.0029	0.0781	0.0782
	0.6	-0.0004	0.0309	0.0309	-0.0026	0.0784	0.0785
	0.7	-0.0004	0.0328	0.0328	-0.0022	0.0791	0.0792

Table 3
The simulation results for Example 2: nonparametric part

	Proposed			LLS		
h	bias	sd	RMSE	bias	sd	RMSE
h=0.3	0.0081	0.0553	0.0559	0.0109	0.0760	0.0768
h=0.4	0.0133	0.0496	0.0513	0.0151	0.0677	0.0693
h=0.5	0.0187	0.0436	0.0474*	0.0202	0.0627	0.0659
h=0.6	0.0271	0.0423	0.0503	0.0267	0.0587	0.0645*
h=0.7	0.0348	0.0407	0.0535	0.0339	0.0559	0.0654
h=0.8	0.0428	0.0412	0.0594	0.0416	0.0540	0.0682

Note: * stands for the minimum RMSE

TABLE 4

The simulation results for Example 2: parametric part

	h	Proposed bias	Proposed sd	Proposed RMSE	LLS bias	LLS sd	LLS RMSE
λ_1	0.3	0.0030	0.0607	0.0608	0.0043	0.0832	0.0833
	0.4	0.0023	0.0595	0.0595	0.0030	0.0833	0.0834
	0.5	0.0006	0.0618	0.0618	0.0036	0.0836	0.0837
	0.6	0.0028	0.0616	0.0617	0.0032	0.0825	0.0826
	0.7	-0.0010	0.0565	0.0565	0.0035	0.0828	0.0829
	0.8	-0.0066	0.0652	0.0655	0.0029	0.0826	0.0827
λ_2	0.3	0.0138	1.4930	1.4930	-0.1579	1.9990	2.0052
	0.4	0.0375	1.4731	1.4736	-0.2061	1.9359	1.9468
	0.5	-0.0206	1.5042	1.5043	-0.1790	1.9922	2.0002
	0.6	0.0529	1.5151	1.5160	-0.1507	1.9667	1.9724
	0.7	-0.0562	1.4849	1.4859	-0.1759	2.0116	2.0193
	0.8	-0.1421	1.4557	1.4627	-0.1669	2.0226	2.0294
γ_1	0.3	0.0009	0.0148	0.0148	-0.0004	0.0255	0.0255
	0.4	0.0009	0.0151	0.0151	0.0000	0.0248	0.0248
	0.5	-0.0002	0.0143	0.0143	-0.0002	0.0246	0.0246
	0.6	0.0001	0.0150	0.0150	-0.0000	0.0244	0.0244
	0.7	0.0008	0.0157	0.0157	-0.0003	0.0246	0.0246
	0.8	0.0012	0.0165	0.0166	-0.0003	0.0246	0.0246
γ_2	0.3	-0.0001	0.0254	0.0254	-0.0001	0.0389	0.0389
	0.4	-0.0009	0.0248	0.0248	0.0010	0.0383	0.0383
	0.5	-0.0010	0.0241	0.0241	0.0008	0.0382	0.0382
	0.6	-0.0012	0.0257	0.0257	0.0014	0.0375	0.0375
	0.7	0.0002	0.0263	0.0263	0.0016	0.0379	0.0379
	0.8	0.0026	0.0277	0.0278	0.0012	0.0380	0.0380
γ_3	0.3	-0.0001	0.0142	0.0142	-0.0001	0.0242	0.0242
	0.4	-0.0004	0.0141	0.0141	0.0001	0.0242	0.0242
	0.5	0.0005	0.0138	0.0138	-0.0001	0.0242	0.0242
	0.6	0.0004	0.0141	0.0141	-0.0004	0.0240	0.0240
	0.7	0.0008	0.0148	0.0148	-0.0004	0.0239	0.0239
	0.8	-0.0001	0.0152	0.0152	-0.0004	0.0240	0.0240

Table 5

The simulation results for Example 3: nonparametric part

h	Proposed			LLS		
	bias	sd	RMSE	bias	sd	RMSE
	Model 4					
0.1	0.0163	0.1830	0.1838	0.0145	0.1843	0.1848
0.2	0.0406	0.1427	0.1484*	0.0353	0.1450	0.1492
0.3	0.0736	0.1303	0.1497	0.0690	0.1305	0.1476*
0.4	0.1110	0.1290	0.1702	0.1052	0.1267	0.1647
0.5	0.1471	0.1298	0.1962	0.1423	0.1289	0.1920
	Model 5					
0.2	0.1069	0.2529	0.2746	0.0987	0.3167	0.3317
0.3	0.1206	0.2238	0.2543*	0.1311	0.2695	0.2997
0.4	0.1382	0.2219	0.2615	0.1658	0.2485	0.2987*
0.5	0.2059	0.2025	0.2889	0.1978	0.2344	0.3067
0.6	0.2289	0.2012	0.3047	0.2311	0.2256	0.3229

Note: * stands for the minimum RMSE

Table 6

The simulation results for Example 3: index estimation.

	h	Proposed bias	Proposed sd	Proposed RMSE	LLS bias	LLS sd	LLS RMSE
				Model 4			
ϑ_1	0.1	0.0016	0.0126	0.0127	0.0002	0.0119	0.0119
	0.2	0.0008	0.0122	0.0122	0.0004	0.0118	0.0118
	0.3	0.0007	0.0119	0.0119	0.0002	0.0116	0.0116
	0.4	0.0007	0.0118	0.0118	0.0003	0.0114	0.0114
	0.5	0.0006	0.0116	0.0117	0.0005	0.0112	0.0112
ϑ_2	0.1	0.0014	0.0095	0.0096	0.0003	0.0089	0.0090
	0.2	0.0007	0.0092	0.0092	0.0004	0.0089	0.0089
	0.3	0.0006	0.0090	0.0090	0.0003	0.0087	0.0087
	0.4	0.0007	0.0089	0.0089	0.0003	0.0085	0.0085
	0.5	0.0006	0.0088	0.0088	0.0005	0.0084	0.0084
				Model 5			
ϑ_1	0.2	-0.0007	0.0218	0.0218	-0.0032	0.0279	0.0281
	0.3	-0.0014	0.0229	0.0230	-0.0020	0.0278	0.0278
	0.4	-0.0007	0.0230	0.0230	-0.0018	0.0286	0.0287
	0.5	-0.0008	0.0198	0.0198	-0.0020	0.0286	0.0286
	0.6	-0.0003	0.0201	0.0201	-0.0019	0.0280	0.0281
ϑ_2	0.2	-0.0000	0.0163	0.0163	-0.0016	0.0206	0.0207
	0.3	-0.0005	0.0171	0.0171	-0.0007	0.0205	0.0205
	0.4	0.0000	0.0172	0.0172	-0.0006	0.0211	0.0211
	0.5	-0.0002	0.0148	0.0148	-0.0007	0.0210	0.0211
	0.6	0.0002	0.0151	0.0151	-0.0006	0.0207	0.0207

TABLE 7

The simulation results for Example 3: linear part.

	h	Proposed bias	Proposed sd	Proposed RMSE	LLS bias	LLS sd	LLS RMSE
				Model 4			
γ_1	0.1	-0.0035	0.0630	0.0631	-0.0000	0.0521	0.0521
	0.2	0.0000	0.0540	0.0540	-0.0003	0.0520	0.0520
	0.3	0.0001	0.0531	0.0531	-0.0014	0.0521	0.0521
	0.4	-0.0012	0.0539	0.0539	-0.0014	0.0524	0.0524
	0.5	0.0001	0.0545	0.0545	-0.0014	0.0540	0.0540
γ_2	0.1	0.0005	0.0364	0.0364	0.0003	0.0306	0.0306
	0.2	0.0015	0.0317	0.0318	0.0006	0.0304	0.0305
	0.3	0.0001	0.0316	0.0316	-0.0006	0.0303	0.0303
	0.4	-0.0001	0.0320	0.0320	-0.0008	0.0307	0.0307
	0.5	-0.0005	0.0314	0.0314	-0.0009	0.0314	0.0314
				Model 5			
γ_1	0.2	0.0030	0.0962	0.0963	0.0097	0.1121	0.1125
	0.3	0.0012	0.1007	0.1007	0.0070	0.1120	0.1122
	0.4	0.0014	0.0987	0.0987	0.0060	0.1113	0.1115
	0.5	-0.0019	0.0892	0.0892	0.0026	0.1094	0.1095
	0.6	0.0010	0.0874	0.0874	0.0025	0.1097	0.1097
γ_2	0.2	0.0038	0.0565	0.0566	-0.0005	0.0703	0.0703
	0.3	-0.0065	0.0596	0.0599	-0.0011	0.0700	0.0700
	0.4	-0.0061	0.0611	0.0614	-0.0035	0.0702	0.0703
	0.5	-0.0030	0.0538	0.0539	-0.0025	0.0710	0.0710
	0.6	-0.0023	0.0543	0.0544	-0.0014	0.0709	0.0709

Table 8

The results for the electricity costs data: parametric part

	LLS			Proposed		
	Est.	sd	P-vale	Est.	sd	P-value
λ_1	0.806	0.161	0.000	0.800	0.167	0.000
λ_2	0.796	0.875	0.363	0.747	0.860	0.385
γ_1	-0.081	0.114	0.477	-0.095	0.112	0.396
γ_2	-0.276	0.276	0.317	-0.250	0.268	0.351
γ_3	0.394	0.092	0.000	0.394	0.092	0.000

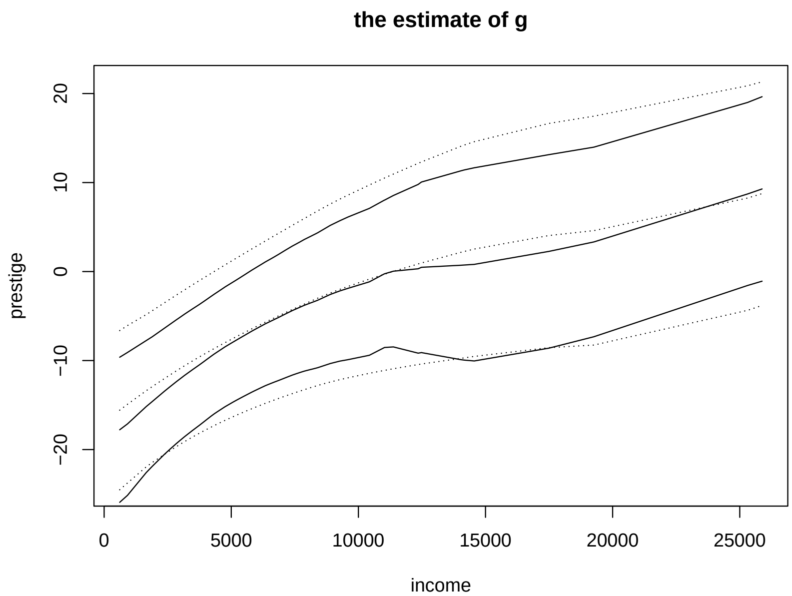

FIG 1. *The estimated regression functions and their* 95% *confidence limits by the LLS estimator(dotted lines) and the proposed estimator(solid lines) when analyzing the relationship between prestige, income and education.*

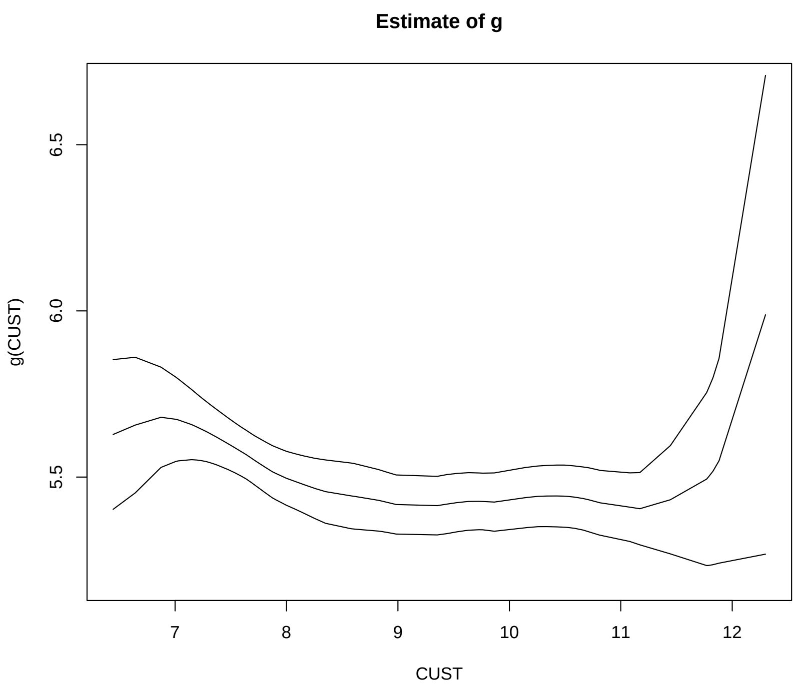

FIG 2. *The results for the electricity costs data: nonparametric part. The estimated curves of $g(\cdot)$ and associated pointwise confidence intervals (solid line: the proposed method; dotted line: the LLS estimators.*

LING ZHOU AND HUAZHEN LIN
CENTER OF STATISTICAL RESEARCH
SCHOOL OF STATISTICS
SOUTHWESTERN UNIVERSITY OF FINANCE AND ECONOMICS
CHENGDU, SICHUAN, CHINA
E-MAIL: zhouling1003@126.com
E-MAIL: linhz@swufe.edu.cn

KANI CHEN
DEPARTMENT OF MATHEMATICS
HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
KOWLOON, HONG KONG
E-MAIL: makchen@ust.hk

HUA LIANG
DEPARTMENT OF STATISTICS
GEORGE WASHINGTON UNIVERSITY
WASHINGTON, D.C. 20052, USA
E-MAIL: hliang@gwu.edu