

Local Rank Inference for Varying Coefficient Models

Lan WANG, Bo KAI, and Runze LI

By allowing the regression coefficients to change with certain covariates, the class of varying coefficient models offers a flexible approach to modeling nonlinearity and interactions between covariates. This article proposes a novel estimation procedure for the varying coefficient models based on local ranks. The new procedure provides a highly efficient and robust alternative to the local linear least squares method, and can be conveniently implemented using existing R software package. Theoretical analysis and numerical simulations both reveal that the gain of the local rank estimator over the local linear least squares estimator, measured by the asymptotic mean squared error or the asymptotic mean integrated squared error, can be substantial. In the normal error case, the asymptotic relative efficiency for estimating both the coefficient functions and the derivative of the coefficient functions is above 96%; even in the worst case scenarios, the asymptotic relative efficiency has a lower bound 88.96% for estimating the coefficient functions, and a lower bound 89.91% for estimating their derivatives. The new estimator may achieve the nonparametric convergence rate even when the local linear least squares method fails due to infinite random error variance. We establish the large sample theory of the proposed procedure by utilizing results from generalized U -statistics, whose kernel function may depend on the sample size. We also extend a resampling approach, which perturbs the objective function repeatedly, to the generalized U -statistics setting, and demonstrate that it can accurately estimate the asymptotic covariance matrix.

KEY WORDS: Asymptotic relative efficiency; Local linear regression; Local rank; Varying coefficient model.

1. INTRODUCTION

As introduced in Cleveland, Crosse, and Shyu (1992) and Hastie and Tibshirani (1993), the varying coefficient model provides a natural and useful extension of the classical linear regression model by allowing the regression coefficients to depend on certain covariates. Due to its flexibility to explore the dynamic features which may exist in the data and its easy interpretation, the varying coefficient model has been widely applied in many scientific areas. It has also experienced rapid developments in both theory and methodology; see Fan and Zhang (2008) for a comprehensive survey. Fan and Zhang (1999) proposed a two-step estimation procedure for the varying coefficient model when the coefficient functions have possibly different degrees of smoothness. Kauermann and Tutz (1999) investigated the use of varying coefficient models for diagnosing the lack-of-fit of regression, regarding the varying coefficient model as an alternative to a parametric null model. Cai, Fan, and Li (2000) developed a more efficient estimation procedure for varying coefficient models in the framework of generalized linear models. As special cases of varying coefficient models, time-varying coefficient models are particularly appealing in longitudinal studies, survival analysis, and time series data since they allow one to explore the time-varying effect of covariates over the response. Pioneering works on novel applications of time-varying coefficient models to longitudinal data include Brumback and Rice (1998), Hoover et al. (1998), Wu, Chiang, and Hoover (1998), and Fan and Zhang (2000), among others. For more details, readers are referred to Fan and Li (2006) and the references therein. Time-varying coefficient

models are also popular in modeling and predicting nonlinear time series data and survival data; see Fan and Zhang (2008) for related literature.

Estimation procedures in the aforementioned papers are built on either local least squares type or local likelihood type methods. Although these estimators remain asymptotically normal for a large class of random error distributions, their efficiency can deteriorate dramatically when the true error distribution deviates from normality. Furthermore, these estimators are very sensitive to outliers. Even a few outlying data points may introduce undesirable artificial features in the estimated functions. These considerations motivate us to develop a novel local rank estimation procedure that is highly efficient, robust, and computationally simple. In particular, the proposed local rank regression estimator may achieve the nonparametric convergence rate even when the local linear least squares method fails to consistently estimate the regression coefficient functions due to infinite random error variance, which occurs for instance when the random error has a Cauchy distribution.

The new approach can substantially improve upon the commonly used local linear least squares procedure for a wide class of error distributions. Theoretical analysis reveals that the asymptotic relative efficiency (ARE), measured by the asymptotic mean squared error (or the asymptotic mean integrated squared error), of the local rank regression estimator in comparison with the local linear least squares estimator has an expression that is closely related to that of the Wilcoxon–Mann–Whitney rank test in comparison with the two-sample t -test. However, different from the two-sample test scenario where the efficiency is completely determined by the asymptotic variance, in the current setting of estimating an infinite-dimensional parameter both bias and variance contribute to the asymptotic efficiency. The value of ARE is often significantly greater than one. For example, the ARE is 167% for estimating the regression coefficient functions when the random error has a t_3 distribution, is 240% for exponential random error distribution, and is 493% for lognormal random error distribution.

Lan Wang is Associate Professor, School of Statistics, University of Minnesota, Minneapolis, MN 55455 (E-mail: lan@stat.umn.edu). Bo Kai is Assistant Professor, Department of Mathematics, College of Charleston, Charleston, SC 29424 (E-mail: kaib@cofc.edu). Runze Li is Professor, Department of Statistics and The Methodology Center, The Pennsylvania State University, University Park, PA 16802 (E-mail: rli@stat.psu.edu). Wang's research is supported by National Science Foundation grant DMS-0706842. Kai's research is supported by National Science Foundation grants DMS 0348869 as a research assistant. Li's research is supported by NIDA, NIH grants R21 DA024260 and P50 DA10075. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA or the NIH. The authors thank an associate editor and two anonymous referees for their insightful and constructive comments.

A striking feature of the local rank procedure is that its pronounced efficiency gain comes with only a little loss when the random error actually has a normal distribution, for which the ARE of the local rank regression estimator relative to the local linear least squares estimator is above 96% for estimating both the coefficient functions and their derivatives. For estimating the regression coefficient functions, the ARE has a sharp lower bound 88.96%, which implies that the efficiency loss is at most 11.04% in the worst case scenario. For estimating the first derivative of the regression coefficient functions, the ARE possesses a lower bound 89.91%. Kim (2007) developed a quantile regression procedure for varying coefficient models when the random errors are assumed to have a certain quantile equal to zero. She used the regression splines method and derived the convergence rate, but the lack of asymptotic normality result does not allow the comparison of the relative efficiency. On the other hand, one may extend the local quantile regression approach (Yu and Jones 1998) to the varying coefficient models. However, this is expected to yield an estimator which still suffers from loss of efficiency and may have near zero ARE relative to the local linear least squares estimator in the worst case scenario.

The new estimator proposed in this article minimizes a convex objective function based on local ranks. The implementation of the minimization can be conveniently carried out using existing functions in the R statistical software package via a simple algorithm (Section 4.1). The objective function has the form of a generalized U -statistic whose kernel varies with the sample size. Under some mild conditions, we establish the asymptotic representation of the proposed estimator and further prove its asymptotic normality. We derive the formula of the asymptotic relative efficiency of the local rank estimator relative to the local linear least squares estimator, which confirms the efficiency advantage of the local rank approach. We also extend a resampling approach, which perturbs the objective function repeatedly, to the generalized U -statistics setting, and demonstrate that it can accurately estimate the asymptotic covariance matrix.

This article is organized as follows. Section 2 presents the local rank procedure for estimating the varying coefficient models. Section 3 discusses its large sample properties and proposes a resampling method for estimating the asymptotic covariance matrix. In Section 4, we address issues related to practical implementation and present Monte Carlo simulation results. We further illustrate the proposed procedure via analyzing an environment dataset. Regularity conditions and technical proofs are presented in the Appendix.

2. LOCAL RANK ESTIMATION PROCEDURE

Let Y be a response variable, and U and \mathbf{X} be the covariates. The varying coefficient model is defined by

$$Y = a_0(U) + \mathbf{X}^T \mathbf{a}(U) + \epsilon, \tag{1}$$

where $a_0(\cdot)$ and $\mathbf{a}(\cdot)$ are both unknown smooth functions. The random error ϵ has probability density function $g(\cdot)$ which has finite Fisher information, that is, $\int \{g(x)\}^{-1} g'(x)^2 dx < \infty$. In this article, it is assumed that U is a scalar and \mathbf{X} is a p -dimensional vector. The proposed procedures can be extended

to the case of multivariate U with more complicated notations by following the same idea in this article.

Suppose that $\{U_i, \mathbf{X}_i, Y_i\}$, $i = 1, \dots, n$, is a random sample from model (1). Write $X_i = (X_{i1}, \dots, X_{ip})^T$ and $\mathbf{a}(\cdot) = (a_1(\cdot), \dots, a_p(\cdot))^T$. For u in a neighborhood of any given u_0 , we locally approximate the coefficient function by a Taylor expansion

$$a_m(u) \approx a_m(u_0) + a'_m(u_0)(u - u_0), \quad m = 0, 1, \dots, p. \tag{2}$$

Denote $\alpha_1 = a_0(u_0)$, $\alpha_2 = a'_0(u_0)$, $\beta_m = a_m(u_0)$, and $\beta_{p+m} = a'_m(u_0)$, for $m = 1, \dots, p$. Based on the above approximation, we obtain the residual for estimating Y_i at $U_i = u_0$

$$e_i = Y_i - \alpha_1 - \alpha_2(U_i - u_0) - \sum_{m=1}^p [\beta_m + \beta_{p+m}(U_i - u_0)]X_{im}. \tag{3}$$

We define the local rank objective function to be

$$Q_n(\boldsymbol{\beta}, \alpha_2) = \frac{1}{n(n-1)} \times \sum_{1 \leq i, j \leq n} |e_i - e_j| K_h(U_i - u_0) K_h(U_j - u_0), \tag{4}$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p, \beta_{p+1}, \dots, \beta_{2p})^T$, and for a given kernel function $K(\cdot)$ and a bandwidth h , $K_h(t) = h^{-1}K(t/h)$. Note that $Q_n(\boldsymbol{\beta}, \alpha_2)$ does not depend on α_1 because α_1 is canceled out in $e_i - e_j$. The objective function $Q_n(\boldsymbol{\beta}, \alpha_2)$ is a local version of Gini's mean difference, which is a classical measure of concentration or dispersion (David 1998). Without the kernel functions, $[n(n-1)]^{-1} \sum_{1 \leq i, j \leq n} |e_j - e_i|$ is the global rank objective function that leads to the classical rank estimator in linear models based on Wilcoxon scores. Rank-based statistical procedures have played a fundamental role in nonparametric analysis of linear models due to its high efficiency and robustness. We refer to the review paper of McKean (2004) for many useful references.

For any given u_0 , minimizing $Q_n(\boldsymbol{\beta}, \alpha_2)$ yields the local Wilcoxon rank estimator for $(\boldsymbol{\beta}_0^T, \alpha_2^T)^T$, where $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(u_0) = (a_1(u_0), \dots, a_p(u_0), a'_1(u_0), \dots, a'_p(u_0))^T$. Denote the minimizer of $Q_n(\boldsymbol{\beta}, \alpha_2)$ by $(\widehat{\boldsymbol{\beta}}^T, \widehat{\alpha}_2^T)^T$. Then for $m = 1, \dots, p$,

$$\widehat{a}_m(u_0) = \widehat{\beta}_m, \quad \widehat{a}'_m(u_0) = \widehat{\beta}_{p+m}, \quad \text{and} \quad \widehat{a}'_0(u_0) = \widehat{\alpha}_2.$$

In the sequel, we also use the vector notation $\widehat{\mathbf{a}}(u_0) = (\widehat{a}_1(u_0), \dots, \widehat{a}_p(u_0))^T$ and $\widehat{\mathbf{a}}'(u_0) = (\widehat{a}'_1(u_0), \dots, \widehat{a}'_p(u_0))^T$ when convenient.

The location parameter $a_0(u_0)$ needs to be estimated separately. This is analogous to the scenario of global rank estimation of intercept in the linear regression model. In order to make the intercept identifiable, it is essential to have additional location constraint on the random errors. We adopt the commonly used constraint that ϵ_i has median zero. Given $(\widehat{\boldsymbol{\beta}}^T, \widehat{\alpha}_2^T)^T$, we estimate $a_0(u_0)$ by $\widehat{\alpha}_1$, the value of α_1 that minimizes

$$n^{-1} \sum_{i=1}^n \left| Y_i - \alpha_1 - \widehat{\alpha}_2(U_i - u_0) - \sum_{m=1}^p [\widehat{\beta}_m + \widehat{\beta}_{p+m}(U_i - u_0)] X_{im} \right| K_h(U_i - u_0), \tag{5}$$

which is a local version of a weighted L_1 -norm objective function.

3. THEORETICAL PROPERTIES

3.1 Large Sample Distributions

In this subsection, we investigate the asymptotic properties of $\hat{\beta}$ and $\hat{\alpha}_2$. The main challenge comes from the nonsmoothness of the objective function $Q_n(\beta, \alpha_2)$. To overcome this difficulty, we first derive an asymptotic representation of $\hat{\beta}$ and $\hat{\alpha}_2$ via a quadratic approximation of $Q_n(\beta, \alpha_2)$, which holds uniformly in a local neighborhood of the true parameter values. Aided with this asymptotic representation, we further establish the asymptotic normality of the local rank estimator.

Let us begin with some new notation. Let $\gamma_n = (nh)^{-1/2}$, and define

$$\begin{aligned}\beta^* &= \gamma_n^{-1}(\beta_1 - a_1(u_0), \dots, \beta_p - a_p(u_0), \\ &\quad h(\beta_{p+1} - a'_1(u_0)), \dots, h(\beta_{2p} - a'_p(u_0)))^T, \\ \alpha^* &= (\alpha_1^*, \alpha_2^*)^T = \gamma_n^{-1}(\alpha_1 - a_0(u_0), h(\alpha_2 - a'_0(u_0)))^T,\end{aligned}$$

$$\begin{aligned}\Delta_i(u_0) &= \sum_{m=1}^p [a_m(U_i) - a_m(u_0) - a'_m(u_0)(U_i - u_0)]X_{im} \\ &\quad + [a_0(U_i) - a_0(u_0) - a'_0(u_0)(U_i - u_0)].\end{aligned}$$

Let $(\hat{\beta}_n^{*T}, \hat{\alpha}_{2n}^{*T})^T$ be the value of $(\beta^{*T}, \alpha_2^*)^T$ that minimizes the following reparametrized objective function

$$\begin{aligned}Q_n^*(\beta^*, \alpha_2^*) &= \frac{1}{n(n-1)} \\ &\quad \times \sum_{1 \leq i < j \leq n} |(\epsilon_i - \gamma_n \alpha_2^*(U_i - u_0)/h - \gamma_n \beta^{*T} \mathbf{Z}_i + \Delta_i(u_0)) \\ &\quad - (\epsilon_j - \gamma_n \alpha_2^*(U_j - u_0)/h - \gamma_n \beta^{*T} \mathbf{Z}_j + \Delta_j(u_0))| \\ &\quad \times K_h(U_i - u_0)K_h(U_j - u_0),\end{aligned}\quad (6)$$

where $\mathbf{Z}_i = (\mathbf{X}_i^T, ((U_i - u_0)/h)\mathbf{X}_i^T)^T$. Let $\mathbf{H} = \text{diag}(1, h) \otimes \mathbf{I}_p$, where \otimes denotes the operation of Kronecker product and \mathbf{I}_p denotes a $p \times p$ identity matrix. Then it can be easily seen that

$$\hat{\beta}_n^* = \sqrt{nh}\mathbf{H}(\hat{\beta} - \beta_0) \quad \text{and} \quad \hat{\alpha}_{2n}^* = \sqrt{nh^3}(\hat{\alpha}_2 - a'_0(u_0)).$$

We next show that the nonsmooth function $Q_n^*(\beta^*, \alpha_2^*)$ can be locally approximated by a quadratic function of $(\beta^{*T}, \alpha_2^*)^T$. Let $\mu_i = \int t^i K(t) dt$, $i = 1, 2$, and $v_i = \int t^i K^2(t) dt$, $i = 0, 1, 2$. In this article, we assume that the kernel function $K(\cdot)$ is symmetric. This is not restrictive considering that most of the commonly used kernel functions, such as the Epanechnikov kernel $K(t) = 0.75(1 - t^2)I(|t| < 1)$, are symmetric. We use $S_n(\beta^*, \alpha_2^*) = (S_{n1}^T(\beta^*, \alpha_2^*), S_{n2}(\beta^*, \alpha_2^*))^T$ to denote the gradient function of $Q_n^*(\beta^*, \alpha_2^*)$, that is, $S_{n1}(\beta^*, \alpha_2^*) = \nabla_{\beta^*} Q_n^*(\beta^*, \alpha_2^*)$ and $S_{n2}(\beta^*, \alpha_2^*) = \nabla_{\alpha_2^*} Q_n^*(\beta^*, \alpha_2^*)$. More specifically,

$$\begin{aligned}S_{n1}(\beta^*, \alpha_2^*) &= 2\gamma_n[n(n-1)]^{-1} \\ &\quad \times \sum_{i \neq j} [I(\epsilon_i - \gamma_n \alpha_2^*(U_i - u_0)/h - \gamma_n \beta^{*T} \mathbf{Z}_i + \Delta_i(u_0)) \\ &\quad \leq \epsilon_j - \gamma_n \alpha_2^*(U_j - u_0)/h - \gamma_n \beta^{*T} \mathbf{Z}_j + \Delta_j(u_0)) - 1/2] \\ &\quad \times (\mathbf{Z}_i - \mathbf{Z}_j)K_h(U_i - u_0)K_h(U_j - u_0)\end{aligned}$$

and

$$\begin{aligned}S_{n2}(\beta^*, \alpha_2^*) &= 2\gamma_n[n(n-1)]^{-1} \\ &\quad \times \sum_{i \neq j} [I(\epsilon_i - \gamma_n \alpha_2^*(U_i - u_0)/h - \gamma_n \beta^{*T} \mathbf{Z}_i + \Delta_i(u_0)) \\ &\quad \leq \epsilon_j - \gamma_n \alpha_2^*(U_j - u_0)/h - \gamma_n \beta^{*T} \mathbf{Z}_j + \Delta_j(u_0)) - 1/2] \\ &\quad \times ((U_i - U_j)/h)K_h(U_i - u_0)K_h(U_j - u_0).\end{aligned}$$

Furthermore, we consider the following quadratic function of $(\beta^{*T}, \alpha_2^*)^T$:

$$\begin{aligned}B_n(\beta^*, \alpha_2^*) &= \gamma_n^{-1}(\beta^{*T}, \alpha_2^*) \begin{pmatrix} S_{n1}(\mathbf{0}, 0) \\ S_{n2}(\mathbf{0}, 0) \end{pmatrix} \\ &\quad + \frac{1}{2}\gamma_n(\beta^{*T}, \alpha_2^*)\mathbf{A} \begin{pmatrix} \beta^* \\ \alpha_2^* \end{pmatrix} + \gamma_n^{-1}Q_n^*(\mathbf{0}, 0),\end{aligned}\quad (7)$$

where

$$\mathbf{A} = 4\tau f^2(u_0) \begin{pmatrix} \Sigma(u_0) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mu_2 \Sigma(u_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mu_2 \end{pmatrix},\quad (8)$$

$\Sigma(u_0) = E[\mathbf{X}_i \mathbf{X}_i^T | U_i = u_0]$, $\mathbf{0}$ denotes a matrix (or vector) of zeroes whose dimension is determined by the context, $\tau = \int g^2(t) dt$ is the Wilcoxon constant, and $g(\cdot)$ is the density function of the random error ϵ .

Lemma 3.1. Suppose that Conditions (C1)–(C4) in the Appendix hold. Then $\forall \epsilon > 0, \forall c > 0$,

$$P\left[\sup_{\|(\beta^{*T}, \alpha_2^*)\| \leq c} |\gamma_n^{-1} Q_n^*(\beta^*, \alpha_2^*) - B_n(\beta^*, \alpha_2^*)| \geq \epsilon\right] \rightarrow 0,$$

where $\|\cdot\|$ denotes the Euclidean norm.

Lemma 3.1 implies that the nonsmooth objective function $Q_n^*(\beta^*, \alpha_2^*)$ can be uniformly approximated by a quadratic function $B_n(\beta^*, \alpha_2^*)$ in a neighborhood around $\mathbf{0}$. In the Appendix, it is also shown that the minimizer of $B_n(\beta^*, \alpha_2^*)$ is asymptotically within a $o(1)$ neighborhood of $(\hat{\beta}_n^{*T}, \hat{\alpha}_{2n}^{*T})^T$. This further allows us to derive the asymptotic distribution.

The local linear Wilcoxon estimator of $\mathbf{a}(u_0) = (a_1(u_0), \dots, a_p(u_0))^T$ is $\hat{\mathbf{a}}(u_0)$. The theorem below provides an asymptotic representation of $\hat{\mathbf{a}}(u_0)$ and the asymptotic normal distribution. Let $S_{n1}(\mathbf{0}, 0) = (S_{n11}^T(\mathbf{0}, 0), S_{n12}^T(\mathbf{0}, 0))^T$, where $S_{n11}(\mathbf{0}, 0)$ and $S_{n12}(\mathbf{0}, 0)$ are both $p \times 1$ vectors.

Theorem 3.2. Suppose that Conditions (C1)–(C4) in the Appendix hold. Then we have the following asymptotic representation

$$\begin{aligned}\sqrt{nh}[\hat{\mathbf{a}}(u_0) - \mathbf{a}(u_0)] &= -\gamma_n^{-2}[4\tau f^2(u_0)\Sigma(u_0)]^{-1}S_{n11}(\mathbf{0}, 0) + o_P(1),\end{aligned}\quad (9)$$

where $f(u)$ is the density function of U . Furthermore,

$$\begin{aligned}\sqrt{nh}\left[\hat{\mathbf{a}}(u_0) - \mathbf{a}(u_0) - \frac{\mu_2}{2}\mathbf{a}''(u_0)h^2 + o(h^2)\right] &\rightarrow \mathbf{N}\left(0, \frac{v_0}{12\tau^2 f(u_0)}\Sigma^{-1}(u_0)\right)\end{aligned}\quad (10)$$

in distribution, where $\mathbf{a}''(u_0) = (a''_1(u_0), \dots, a''_p(u_0))^T$.

Remark. For the estimators of the derivatives of the coefficient functions, we have the following asymptotic representations:

$$\begin{aligned} &\sqrt{nh^3}[\hat{\alpha}_2 - a'_0(u_0)] \\ &= -\gamma_n^{-2}[4\tau f^2(u_0)\mu_2]^{-1}S_{n2}(\mathbf{0}, 0) + o_P(1), \end{aligned} \tag{11}$$

$$\begin{aligned} &\sqrt{nh^3}[\hat{\mathbf{a}}'(u_0) - \mathbf{a}'(u_0)] \\ &= -\gamma_n^{-2}[4\tau f^2(u_0)\mu_2 \Sigma(u_0)]^{-1}S_{n12}(\mathbf{0}, \mathbf{0}) + o_P(1). \end{aligned} \tag{12}$$

Following similar proof as that for Theorem 3.2 in the Appendix, it can be shown that $\sqrt{nh^3}[\hat{\alpha}_2 - a'_0(u_0)]$ and $\sqrt{nh^3}[\hat{\mathbf{a}}'(u_0) - \mathbf{a}'(u_0)]$ are both asymptotically normal. The proof of the asymptotic normality of $\hat{\alpha}_2$ and $\hat{\mathbf{a}}'(u_0)$ is given in the technical report version of this article (Wang, Kai, and Li 2009).

3.2 Asymptotic Relative Efficiency

We now compare the estimation efficiency of the local rank estimator [denoted by $\hat{\mathbf{a}}_R(u_0)$] with that of the local linear least squares estimator [denoted by $\hat{\mathbf{a}}_{LS}(u_0)$] for estimating $\mathbf{a}(u_0)$ in the varying coefficient model. To measure efficiency, we consider both the asymptotic mean squared error (MSE) at a given u_0 and the asymptotic mean integrated squared error (MISE). When evaluating both criteria, we plug in the theoretical optimal bandwidth.

Zhang and Lee (2000) gives the asymptotic MSE of $\hat{\mathbf{a}}_{LS}(u_0)$ for estimating $\mathbf{a}(u_0)$:

$$\begin{aligned} \text{MSE}_{LS}(h; u_0) &= E\|\hat{\mathbf{a}}_{LS}(u_0) - \mathbf{a}(u_0)\|^2 \\ &= \frac{\mu_2^2 \|\mathbf{a}''(u_0)\|^2}{4} h^4 + \frac{v_0 \sigma^2}{f(u_0)} \text{tr}\{\Sigma^{-1}(u_0)\} \frac{1}{nh}, \end{aligned}$$

where $\sigma^2 = \text{Var}(\epsilon)$ is assumed to be finite and positive. Thus, the theoretical optimal bandwidth, which minimizes the asymptotic MSE of $\hat{\mathbf{a}}_{LS}(u_0)$, is

$$h_{LS}^{opt}(u_0) = \left[\frac{v_0 \sigma^2 \text{tr}\{\Sigma^{-1}(u_0)\}}{\mu_2^2 \|\mathbf{a}''(u_0)\|^2 f(u_0)} \right]^{1/5} n^{-1/5}. \tag{13}$$

From (10), the asymptotic MSE of the local rank estimator $\hat{\mathbf{a}}_R(u_0)$ is

$$\begin{aligned} \text{MSE}_R(h; u_0) &= E\|\hat{\mathbf{a}}_R(u_0) - \mathbf{a}(u_0)\|^2 \\ &= \frac{\mu_2^2 \|\mathbf{a}''(u_0)\|^2}{4} h^4 + \frac{v_0}{12\tau^2 f(u_0)} \text{tr}\{\Sigma^{-1}(u_0)\} \frac{1}{nh}. \end{aligned}$$

The theoretical optimal bandwidth for the local rank estimator thus is

$$h_R^{opt}(u_0) = \left[\frac{v_0 \text{tr}\{\Sigma^{-1}(u_0)\}}{12\tau^2 \mu_2^2 \|\mathbf{a}''(u_0)\|^2 f(u_0)} \right]^{1/5} n^{-1/5}. \tag{14}$$

This allows us to calculate the local asymptotic relative efficiency.

Theorem 3.3. The asymptotic relative efficiency of the local rank estimator to the local linear least squares estimator for $\mathbf{a}(u_0)$ is

$$\text{ARE}(u_0) = \frac{\text{MSE}_{LS}\{h_{LS}^{opt}(u_0), u_0\}}{\text{MSE}_R\{h_R^{opt}(u_0), u_0\}} = (12\sigma^2\tau^2)^{4/5}.$$

This asymptotic relative efficiency has a lower bound 0.8896, which is attained at the random error density $f(t) = \frac{3}{20\sqrt{5}}(5 - x^2)I(|x| \leq 5)$.

Remark 1. Alternatively, we may consider the asymptotic relative efficiency obtained by comparing the MISE, which is defined as $\text{MISE}(h) = \int E\|\hat{\mathbf{a}}(u) - \mathbf{a}(u)\|^2 w(u) du$ with a weight function $w(\cdot)$. This provides a global measurement. Interestingly, it leads to the same relative efficiency. This follows by observing that the theoretical optimal global bandwidths for the local linear least squares estimator and the local rank estimator are

$$h_{LS}^{opt} = \left[\frac{v_0 \sigma^2 \int w(u) \text{tr}\{\Sigma^{-1}(u)\} / f(u) du}{\mu_2^2 \int \|\mathbf{a}''(u)\|^2 w(u) du} \right]^{1/5} n^{-1/5} \tag{15}$$

and

$$h_R^{opt} = \left[\frac{v_0 \int w(u) \text{tr}\{\Sigma^{-1}(u)\} / f(u) du}{12\tau^2 \mu_2^2 \int \|\mathbf{a}''(u)\|^2 w(u) du} \right]^{1/5} n^{-1/5}, \tag{16}$$

respectively. Thus, with the theoretical optimal bandwidths,

$$\text{ARE} = \frac{\text{MISE}_{LS}(h_{LS}^{opt})}{\text{MISE}_R(h_R^{opt})} = (12\sigma^2\tau^2)^{4/5}.$$

Define $\phi = (12\sigma^2\tau^2)^{4/5}$. Then $\text{ARE}(u_0) = \text{ARE} = \phi$.

Note that the above ARE is closely related to the asymptotic relative efficiency of the Wilcoxon–Mann–Whitney rank test in comparison with the two-sample t -test. Table 1 depicts the value of ϕ for some commonly used error distributions. It can be seen that the desirable high efficiency of traditional rank methods for estimating a finite-dimensional parameter completely carries over to the local rank method for estimating an infinite dimensional parameter.

By a similar calculation, we can show that the asymptotic relative efficiencies of the local rank estimator to the local linear estimator for $\mathbf{a}'(u_0)$ and $\mathbf{a}'(\cdot)$ both equal $\psi = (12\sigma^2\tau^2)^{8/11}$, which has a lower bound 0.8991. This value is also reported in Table 1 for some common error distributions.

Remark 2. We may also apply the local median approach (Yu and Jones 1998) to estimate the coefficient functions and their first derivatives. Similarly, we can prove that such estimators are asymptotically normal. The ARE of the local median estimator versus the local linear least squares estimator is closely related to that of the sign test versus the t -test. It is known that the ARE of the sign test versus the t -test for the normal distribution is 0.63. Thus we expect the efficiency loss of the local median procedure to be substantial for normal random error.

Table 1. Asymptotic relative efficiency

Error	Normal	Laplace	t_3	Exponential	Log N	Cauchy
ϕ	0.9638	1.3832	1.6711	2.4082	4.9321	∞
ψ	0.9671	1.3430	1.5949	2.2233	4.2661	∞

3.3 Asymptotic Normality of $\widehat{\alpha}_1$

Following (5), $\widehat{\alpha}_1^* = \sqrt{nh}[\widehat{\alpha}_1 - a_0(u_0)]$ is the value of α_1^* that minimizes

$$\begin{aligned} Q_{n0}^*(\alpha_1^*, \widehat{\alpha}_2, \widehat{\beta}) &= n^{-1} \sum_{i=1}^n \left| \epsilon_i - \gamma_n \alpha_1^* - (\widehat{\alpha}_2 - a'_0(u_0))(U_i - u_0) \right. \\ &\quad \left. - \sum_{m=1}^p [(\widehat{\beta}_m - a_m(u_0)) + (\widehat{\beta}_{p+m} - a'_m(u_0)) \right. \\ &\quad \left. \times (U_i - u_0)] X_{im} + \Delta_i(u_0) \right| K_h(U_i - u_0). \end{aligned}$$

Similarly as Lemma 3.1, we can establish the following local quadratic approximation which holds uniformly in a neighborhood around 0:

$$\begin{aligned} \gamma_n^{-1} Q_{n0}^*(\alpha_1^*, \widehat{\alpha}_2, \widehat{\beta}) &= \gamma_n^{-1} \alpha_1^* S_{n0} + \gamma_n g(0) f(u_0) \alpha_1^{*2} + \gamma_n^{-1} Q_{n0}^*(0, a'_0(u_0), \beta_0) \\ &\quad + o_p(1), \end{aligned} \quad (17)$$

where

$$S_{n0} = 2\gamma_n n^{-1} \sum_{i=1}^n [I(\epsilon_i \leq -\Delta_i(u_0)) - 1/2] K_h(U_i - u_0). \quad (18)$$

This further leads to an asymptotic representation of $\widehat{\alpha}_1$:

$$\sqrt{nh}(\widehat{\alpha}_1 - a_0(u_0)) = -\gamma_n^{-2} [2g(0)f(u_0)]^{-1} S_{n0} + o_p(1). \quad (19)$$

The theorem below gives the asymptotic distribution of $\widehat{\alpha}_1$.

Theorem 3.4. Under the conditions of Theorem 3.2, we have

$$\begin{aligned} \sqrt{nh} \left[\widehat{\alpha}_1 - a_0(u_0) - \frac{\mu_2 a''_0(u_0)}{2} h^2 + o(h^2) \right] \\ \rightarrow \mathbf{N}(0, [12g^2(0)f(u_0)]^{-1} v_0). \end{aligned}$$

3.4 Estimation of the Standard Errors

To make statistical inference based on the local rank methodology, one needs to estimate the standard error of the resulting estimator. As indicated by Theorem 3.2, the asymptotic covariance matrix of the local rank estimator is rather complex and involves unknown functions. Here we propose a standard error estimator using a simple resampling method proposed by Jin, Ying, and Wei (2001).

Let V_1, \dots, V_n be independent and identically distributed nonnegative random variables with mean 1/2 and variance 1. We consider a stochastic perturbation of (4):

$$\begin{aligned} \overline{Q}_n(\beta, \alpha_2) &= \frac{1}{n(n-1)} \sum_{1 \leq i, j \leq n} (V_i + V_j) |e_i - e_j| \\ &\quad K_h(U_i - u_0) K_h(U_j - u_0), \end{aligned} \quad (20)$$

where e_i is defined in (3). In $\overline{Q}_n(\beta, \alpha_2)$, the data $\{Y_i, U_i, \mathbf{X}_i\}$ are considered to be fixed, and the randomness comes from the V_i 's. Let $(\overline{\beta}^T, \overline{\alpha}_2)^T$ be the value of $(\beta^T, \alpha_2)^T$ that minimizes $\overline{Q}_n(\beta, \alpha_2)$. It is easy to obtain $(\overline{\beta}^T, \overline{\alpha}_2)^T$ by applying a simple algorithm described in Section 3.1.

Jin, Ying, and Wei (2001) established the validity of the resampling method when the objective function has a U -statistics

structure. Although their theory covers many important applications, they require that the U -statistic has a fixed kernel. We extend their result to our setting, where the U -statistic involves variable kernel due to nonparametric smoothing. Let $\overline{\mathbf{a}}(u_0)$ be the local rank estimator of $\mathbf{a}(u_0)$ based on the perturbed objective function (20), that is, it is the subvector that consists of the first p components of $\overline{\beta}$. Its asymptotic normality is given in the theorem below.

Theorem 3.5. Under the conditions of Lemma 3.1, conditional on almost surely every sequence of data $\{Y_i, U_i, \mathbf{X}_i\}$,

$$\sqrt{nh}[\overline{\mathbf{a}}(u_0) - \widehat{\mathbf{a}}(u_0)] \rightarrow \mathbf{N}\left(0, \frac{v_0}{12\tau^2 f(u_0)} \Sigma^{-1}(u_0)\right)$$

in distribution.

This theorem suggests that to estimate the asymptotic covariance matrix of $\widehat{\mathbf{a}}(u_0)$, one can repeatedly perturb (4) by generating a large number of independent random samples $\{V_i\}_{i=1}^n$. For each perturbed objective function, one solves for $\overline{\mathbf{a}}(u_0)$. The sample covariance matrix of $\overline{\mathbf{a}}(u_0)$ based on a large number of independent perturbations provides a good approximation. The accuracy of the resulting standard error estimate will be tested in the next section.

The perturbed estimator has conditional bias equal to zero. It has been found that standard bootstrap method, which resamples from the empirical distribution of the data, also estimates bias as zero when estimating nonparametric curves (Hall and Kang 2001). It is possible to use more delicate bootstrap technique to estimate the bias of a nonparametric curve estimator. Although some of the ideas may be adapted to the method of disturbing the objective function, this is beyond the scope of our article and is not pursued further here.

4. NUMERICAL STUDIES

4.1 A Pseudo-Observation Algorithm

The local rank estimator can be obtained by applying an efficient and reliable algorithm. Note that the local rank estimator of $(\beta_0^T, a'_0(u_0))^T$ can be solved by fitting a weighted L_1 regression on $\frac{n(n-1)}{2}$ pseudo observations $(\mathbf{x}_i^* - \mathbf{x}_j^*, Y_i - Y_j)$ with weights $w_{ij} = K((U_i - u_0)/h)K((U_j - u_0)/h)$, where $\mathbf{x}_i^* = (U_i - u_0, X_i^T, (U_i - u_0)X_i^T)^T$, $1 \leq i < j \leq n$. Given $(\widehat{\beta}^T, \widehat{\alpha}_2)^T$, the estimator of $a_0(u_0)$ can be obtained by another weighted L_1 regression on $(1, Y_i - \widehat{\alpha}_2(U_i - u_0) - \sum_{m=1}^p [\widehat{\beta}_m + \widehat{\beta}_{p+m}(U_i - u_0)] X_{im})$ with weights $w_i = K((U_i - u_0)/h)$, $1 \leq i \leq n$. Many statistical software packages can implement weighted L_1 regression. In our numerical studies, we use the function ‘‘rq’’ in the R package *quantreg*.

4.2 Bandwidth Selection

Bandwidth selection is an important issue for all statistical models that involve nonparametric smoothing. Although we have derived the theoretical optimal bandwidth for the local rank estimator in (14) and (16), it is difficult to use the ‘‘plug-in’’ method to estimate it due to many unknown quantities.

We propose below an alternative bandwidth selection method that is practically feasible. This approach is based on the relationship between h_R^{opt} and h_{LS}^{opt} . From Section 3.2, we see that

$$\begin{aligned} h_R^{opt}(u_0) &= \left(\frac{1}{12\tau^2\sigma^2} \right)^{1/5} h_{LS}^{opt}(u_0) \quad \text{and} \\ h_R^{opt} &= \left(\frac{1}{12\tau^2\sigma^2} \right)^{1/5} h_{LS}^{opt}. \end{aligned} \quad (21)$$

Thus, we can first use existing bandwidth selectors (e.g., Zhang and Lee 2000) to estimate $h_{LS}^{opt}(u_0)$ or h_{LS}^{opt} . The error variance σ^2 can be estimated based on the residuals; in particular when robustness is of concern it can be estimated using the MAD of the residuals. Hettmansperger and McKean (1998, p. 181) discussed in details how to estimate τ , which can be obtained by the function “wilcoxontau” in the R software developed by Terpstra and McKean (2005). In the end, we plug in these estimators into (21) to get the bandwidth for local rank estimator.

Alternatively, instead of the above two-step procedure, we may directly use the computationally intensive cross-validation to estimate the bandwidth for the local rank procedure. Note that under outlier contamination, standard cross-validation can lead to extremely biased bandwidth estimates because it can be adversely influenced by extreme prediction errors. Robust cross-validation method, such as that developed by Leung (2005), is therefore preferred.

4.3 Examples

We conduct Monte Carlo simulations to assess the finite sample performance, and illustrate the proposed methodology on a real environmental dataset. In the analysis, we use the Epanechnikov kernel $K(u) = 0.75(1 - u^2)I(|u| < 1)$.

Example 1. We generate random data from

$$Y = a_0(U) + a_1(U)X_1 + a_2(U)X_2 + \epsilon,$$

where $a_0(u) = \exp(2u - 1)$, $a_1(u) = 8u(1 - u)$, and $a_2(u) = 2\sin^2(2\pi u)$. The covariate U follows a uniform distribution on $[0, 1]$, and is independent from (X_1, X_2) , where the covariates X_1 and X_2 are standard normal random variables with correlation coefficient $2^{-1/2}$. The coefficient functions and the mechanism to generate U and (X_1, X_2) are the same as those in Cai, Fan, and Li (2000). We consider six different error distributions: $N(0, 1)$, Laplace, standard Cauchy, t -distribution with 3 degrees of freedom, mixture of normals $0.9N(0, 1) + 0.1N(0, 10^2)$ and log-normal distribution. Except for Cauchy error, all the other generated random errors are standardized to have median 0 and variance 1. We consider sample sizes $n = 400$ and 800 , and conduct 400 simulations for each case.

We compare the performance of the local rank estimate with the local least squares estimate using the square root of average squared errors (RASE), defined by

$$\text{RASE} = \left\{ \frac{1}{n_{\text{grid}}} \sum_{m=1}^p \sum_{k=1}^{n_{\text{grid}}} \{\hat{a}_m(u_k) - a_m(u_k)\}^2 \right\}^{1/2},$$

where $\{u_k : k = 1, \dots, n_{\text{grid}}\}$ is a set of grid points uniformly placed on $[0, 1]$ with $n_{\text{grid}} = 200$. The sample mean and standard deviation of the RASEs over 400 simulations are presented

in Figures 1 and 2, for sample size $n = 400$ and 800 , respectively. The two figures clearly demonstrate that the local rank estimator performs almost as well as the local least squares estimator for normal random error, and has smaller RASE for other heavier-tailed error distributions. The efficiency gain can be substantial; see, for example, the mixture normal case where the observed relative efficiency for the local rank estimator versus the local least squares estimator is above 2 for most choices of bandwidth. For Cauchy random error, the local rank estimator yields a \sqrt{n} -consistent estimator but the local least squares estimator is inconsistent, which is reflected by the extremely large value of RASE for the local least squares estimator.

Figure 3 depicts the estimated coefficient functions for the normal random error and the mixture normal random error for a typical sample, which is selected in such a way that its RASE value is the median of the 400 RASE values. For this typical sample, we observe that the local rank estimator is almost identical to the local least squares estimator for normal random error; but falls much closer to the truth than the local least squares estimator does for mixture normal random error. Figure 4 plots the estimated coefficient functions for all 400 simulations when the random error has a mixture of normal distribution. It is clear that the local rank estimator has smaller variance. In these two figures, we set the bandwidth to be the theoretic optimal one h^{opt} , calculated using (15) and (16), for both the local rank estimator and the local least squares estimator.

At the end, we evaluate the resampling method (Section 3.4) for estimating the standard errors. We randomly perturb the objective function 1000 times; each time the random variables V_i in (20) are generated from the Gamma(0.25, 2) distribution. Table 2 summarizes the simulation results at three points $u_0 = 0.25, 0.50$, and 0.75 . In the table, “SD” denotes the standard deviation of 400 estimated $\hat{a}_m(u_0)$ and can be regarded as the true standard error; “SE (std(SE))” denotes the mean (standard deviation) of 400 estimated standard errors from the resampling method. Bandwidths are set to be the optimal ones. We observe that the resampling method estimates the standard error very accurately.

Example 2. As an illustration, we now apply the local rank procedure to the environmental dataset in Fan and Zhang (1999). Of interest is to study the relationship between levels of pollutants and the number of total hospital admissions for circulatory and respiratory problems on every Friday from January 1, 1994 to December 31, 1995. The response variable is the logarithm of the number of total hospital admissions, and the covariates include the level of sulfur dioxide (X_1), the level of nitrogen dioxide (X_2), and the level of dust (X_3). A scatter plot of the response variable over time is given in Figure 5(a). We analyze this dataset using the following varying coefficient model:

$$Y = a_0(u) + a_1(u)X_1 + a_2(u)X_2 + a_3(u)X_3 + \epsilon,$$

where u denotes time and is scaled to the interval $[0, 1]$.

We select the bandwidth via the relation (21). More specifically, we first use 20-fold cross-validation to select a bandwidth \hat{h}_{LS} for the local least squares estimator. We then use the function “wilcoxontau” in the R package for rank regression by Terpstra and McKean to estimate $(12\tau^2)^{-1/2}$ and use the MAD

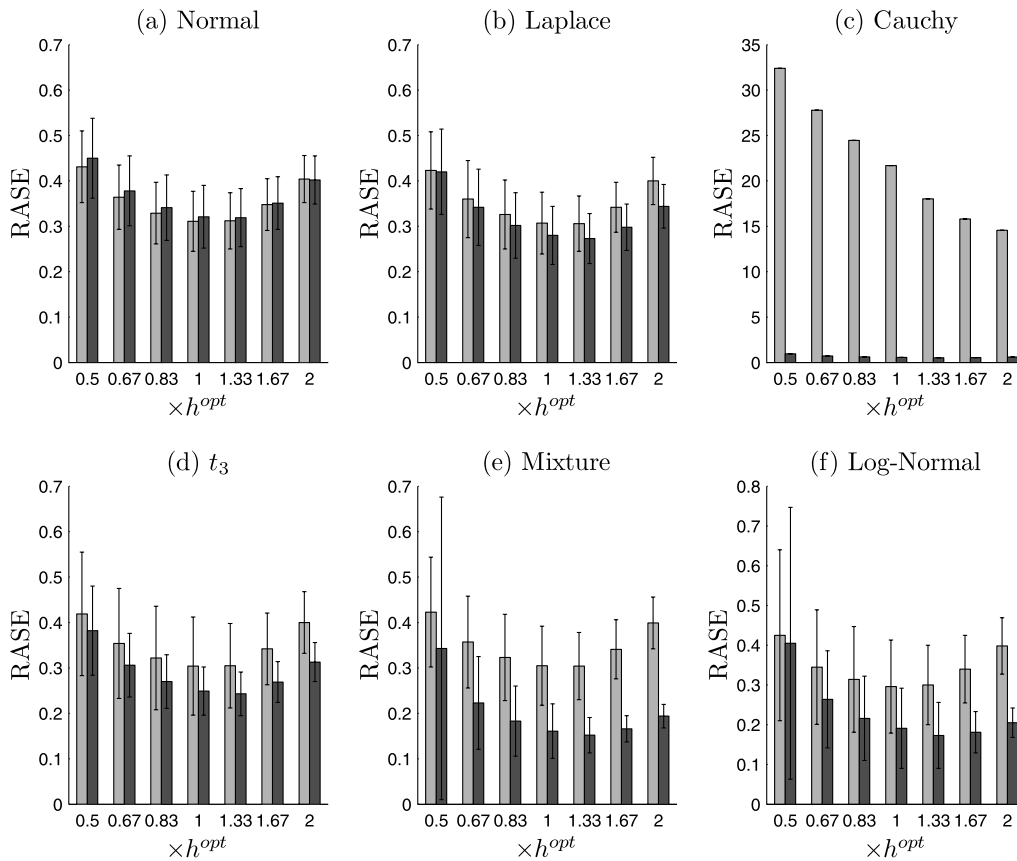


Figure 1. Bar graphs of the RASE with standard error for sample size $n = 400$ over 400 simulations. The light gray bar denotes the local least squares method and the dark gray bar denotes the local rank method. The horizontal axis is in the unit of h^{opt} , which is calculated separately for each method and error specification using either (13) or (14).

of the residuals to robustly estimate σ . All these lead to the selected bandwidth for the local rank estimator: $\hat{h}_R = 0.26$.

The estimated coefficient functions are depicted in Figures 5(b), (c), and (d), where the two dashed curves around the solid line are the estimated function plus/minus twice the standard errors estimated by the resampling method. These two dashed lines can be regarded as a pointwise confidence interval with bias ignored. The figures suggest clearly that the coefficient functions vary with time. The fitted curve is shown in Figure 5(a).

Now we demonstrate the robustness of the local rank procedure. To this end, we artificially perturb the dataset by moving the response value of the 68th observation from 5.89 to 6.89, and the response value of the 34th observation from 5.07 to 3.07. We refit the data with both the local least squares procedure and the local rank procedure; see Figure 6. We observe that the local least squares estimator changes dramatically due to the presence of these two artificial outliers. In contrast, the local rank estimator is nearly unaffected.

APPENDIX: PROOFS

Regularity Conditions.

(C1) Assume that $\{U_i, \mathbf{X}_i, Y_i\}$ are independent and identically distributed, and that the random error ϵ and the covariate $\{U, \mathbf{X}\}$ are independent. Furthermore, assume that ϵ has probability density function $g(\cdot)$ which has finite Fisher information, that

is, $\int \{g(x)\}^{-1} g'(x)^2 dx < \infty$; and that U has probability density function $f(\cdot)$.

- (C2) The function $a_m(\cdot)$, $m = 0, 1, \dots, p$, has continuous second-order derivative in a neighborhood of u_0 .
- (C3) Assume that $E(X_i|U_i = u_0) = 0$ and that $\Sigma(u) = E(X_i X_i^T | U_i = u)$ is continuous at $u = u_0$. The matrix $\Sigma(u_0)$ is positive definite.
- (C4) The kernel function $K(\cdot)$ is symmetric about the origin, and has a bounded support. Assume that $h \rightarrow 0$ and $nh^2 \rightarrow \infty$, as $n \rightarrow \infty$.

These conditions are used to facilitate the proofs, but may not be the weakest ones. The assumptions on the random errors in (C1) are the same as those for multiple linear rank regression (Hettmansperger and McKean 1998). (C2) imposes smoothness requirement on the coefficient functions. In (C3), the assumption $E(X_i|U_i = u_0) = 0$ [also adopted by Kim (2007)] makes the presentation simpler but can be relaxed. It can be shown that the asymptotic normality still holds without this assumption. The conditions on the kernel function and bandwidth in (C4) are common for nonparametric kernel smoothing.

In our proofs, we will use some results on generalized U -statistic, where the kernel function is allowed to depend on the sample size n . The generalized U -statistic has the form $U_n = [n(n-1)]^{-1} \times \sum \sum_{i \neq j} H_n(D_i, D_j)$, where $\{D_i\}_{i=1}^n$ is a random sample and H_n is symmetric in its arguments, that is, $H_n(D_i, D_j) = H_n(D_j, D_i)$. In this article, $D_i = (\mathbf{X}_i^T, U_i, \epsilon_i)^T$. Define $r_n(D_i) = E[H_n(D_i, D_j)|D_i]$, $\bar{r}_n = E[r_n(D_i)]$, and $\hat{U}_n = \bar{r}_n + 2n^{-1} \sum_{i=1}^n [r_n(D_i) - \bar{r}_n]$. We will repeatedly use the following lemma taken from Powell, Stock, and Stoker (1989).

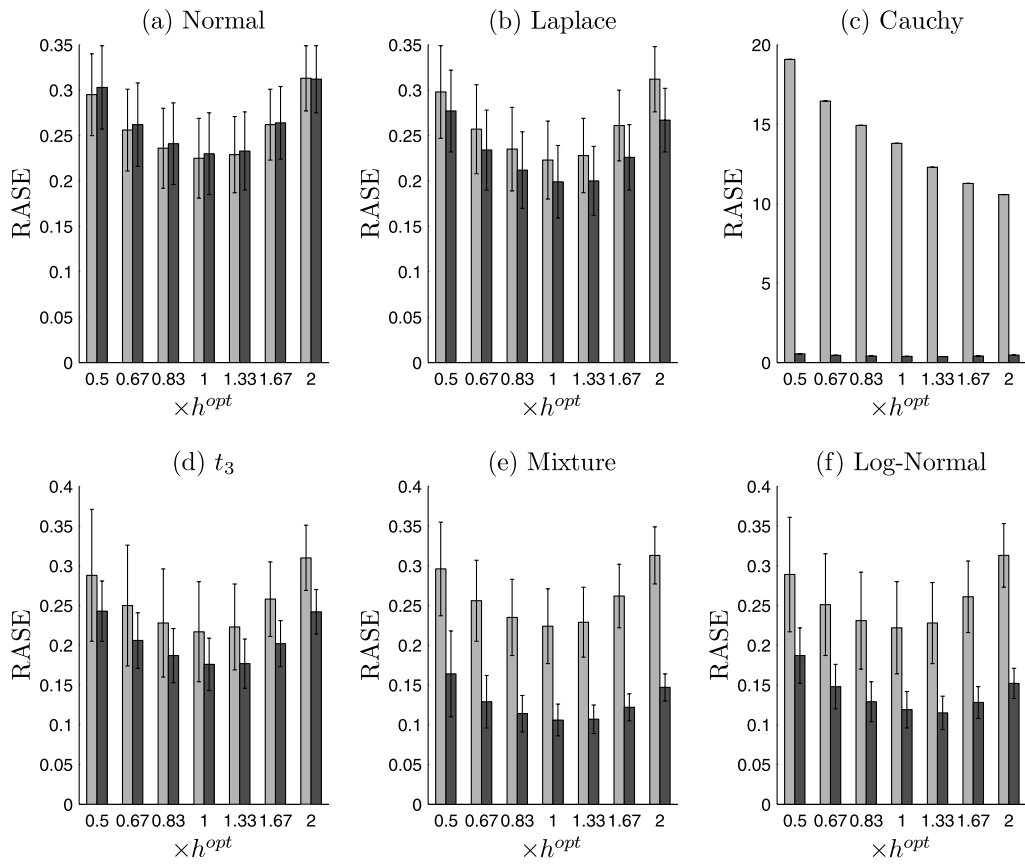


Figure 2. Bar graphs of the RASE with standard error for sample size $n = 800$ over 400 simulations. The light gray bar denotes the local least squares method and the dark gray bar denotes the local rank method. The horizontal axis is in the unit of h^{opt} , which is calculated separately for each method and error specification using either (13) or (14).

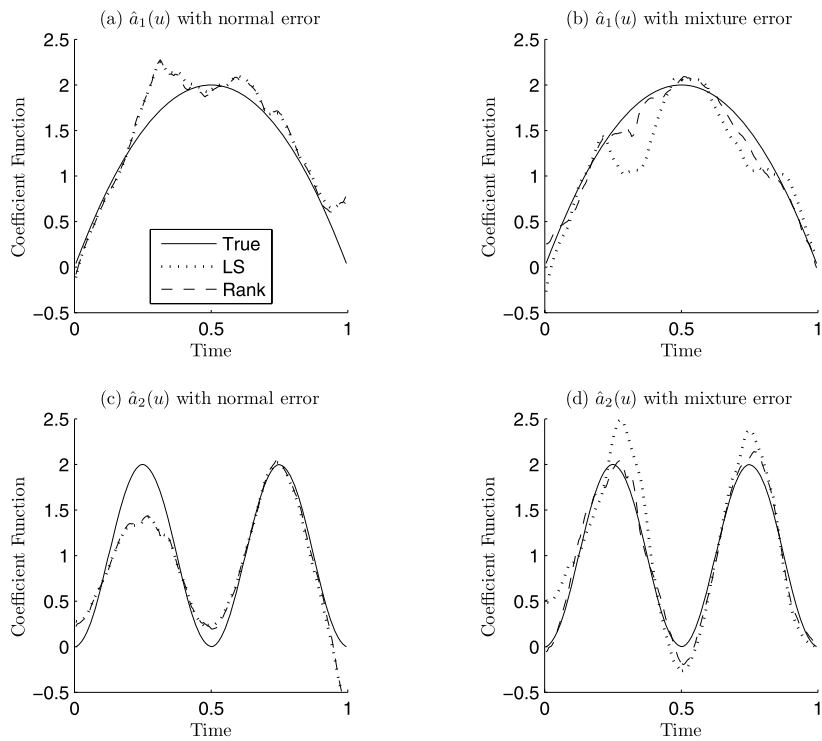


Figure 3. Plot of estimated coefficient functions for a typical dataset.

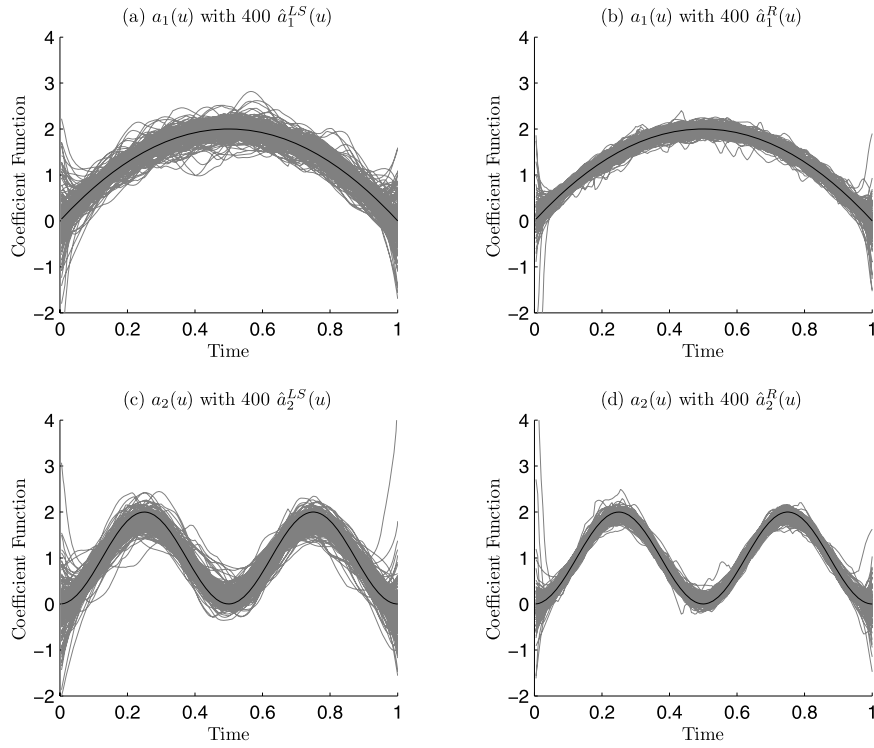


Figure 4. (a) and (c) are plots of 400 local least squares estimators of $a_1(\cdot)$ and $a_2(\cdot)$ over 400 simulation, respectively. (b) and (d) are plots of 400 local rank estimators of $a_1(\cdot)$ and $a_2(\cdot)$, respectively.

Lemma A.1. If $E[\|H_n(D_i, D_j)\|^2] = o(n)$, then $\sqrt{n}(U_n - \widehat{U}_n) = o_p(1)$ and $U_n = \bar{\tau}_n + o_p(1)$.

We need the following two lemmas to prove Lemma 3.1. Denote

$$\mathbf{A}_{n11} = 2h^{-2}E\left\{(\mathbf{Z}_i - \mathbf{Z}_j)(\mathbf{Z}_i - \mathbf{Z}_j)^T K\left(\frac{U_i - u_0}{h}\right)K\left(\frac{U_j - u_0}{h}\right)\right\},$$

$$\mathbf{A}_{n12} = 2h^{-2}E\left\{(\mathbf{Z}_i - \mathbf{Z}_j)[(U_i - U_j)/h]K\left(\frac{U_i - u_0}{h}\right)K\left(\frac{U_j - u_0}{h}\right)\right\},$$

Table 2. Standard deviations of the local rank estimators with $n = 400$

Error	u_0	$\widehat{a}_1(u)$		$\widehat{a}_2(u)$	
		SD	SE (std(SE))	SD	SE (std(SE))
Normal	0.25	0.189	0.159 (0.032)	0.197	0.160 (0.032)
	0.5	0.183	0.159 (0.030)	0.180	0.162 (0.031)
	0.75	0.191	0.162 (0.033)	0.195	0.163 (0.032)
Laplace	0.25	0.175	0.151 (0.037)	0.174	0.151 (0.037)
	0.5	0.168	0.153 (0.039)	0.173	0.154 (0.039)
	0.75	0.168	0.150 (0.037)	0.177	0.150 (0.037)
Mixture	0.25	0.095	0.107 (0.051)	0.092	0.107 (0.049)
	0.5	0.095	0.109 (0.057)	0.091	0.109 (0.055)
	0.75	0.095	0.108 (0.061)	0.093	0.109 (0.055)
t_3	0.25	0.144	0.137 (0.039)	0.145	0.138 (0.036)
	0.5	0.148	0.133 (0.035)	0.152	0.136 (0.037)
	0.75	0.158	0.137 (0.039)	0.155	0.139 (0.042)
Log N	0.25	0.111	0.112 (0.047)	0.112	0.114 (0.049)
	0.5	0.106	0.114 (0.047)	0.107	0.119 (0.050)
	0.75	0.118	0.117 (0.058)	0.118	0.120 (0.060)

$$\mathbf{A}_{n21} = \mathbf{A}_{n12}^T,$$

$$\mathbf{A}_{n22} = 2h^{-2}E\left\{[(U_i - U_j)^2/h^2]K\left(\frac{U_i - u_0}{h}\right)K\left(\frac{U_j - u_0}{h}\right)\right\},$$

and define

$$\mathbf{A}_n = \tau \begin{pmatrix} \mathbf{A}_{n11} & \mathbf{A}_{n12} \\ \mathbf{A}_{n21} & \mathbf{A}_{n22} \end{pmatrix}.$$

Lemma A.2. Suppose that Conditions (C1)–(C4) hold, then $\mathbf{A}_n \rightarrow \mathbf{A}$, where \mathbf{A} is defined in (8).

Proof. We can write

$$\mathbf{A}_{n11} = \begin{pmatrix} \mathbf{A}_{n11}^1 & \mathbf{A}_{n11}^2 \\ \mathbf{A}_{n11}^3 & \mathbf{A}_{n11}^4 \end{pmatrix}.$$

Let

$$\mathbf{A}_{n11}^1 = 2h^{-2}E\left[(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T K\left(\frac{U_i - u_0}{h}\right)K\left(\frac{U_j - u_0}{h}\right)\right].$$

Calculating the expectation by conditional on U_i and U_j first, \mathbf{A}_{n11} becomes

$$2h^{-2} \int E[(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T | U_i = u, U_j = v] \times K\left(\frac{u - u_0}{h}\right)K\left(\frac{v - u_0}{h}\right)f(u)f(v) du dv.$$

Using Condition (C3), straightforward calculation gives $\mathbf{A}_{n11}^1 \rightarrow 4f^2(u_0)\Sigma(u_0)$. Let

$$\mathbf{A}_{n11}^2 = 2h^{-2}E\left\{(\mathbf{X}_i - \mathbf{X}_j)[\mathbf{X}_i(U_i - u_0)/h - \mathbf{X}_j(U_j - u_0)/h]^T \times K\left(\frac{U_i - u_0}{h}\right)K\left(\frac{U_j - u_0}{h}\right)\right\}.$$

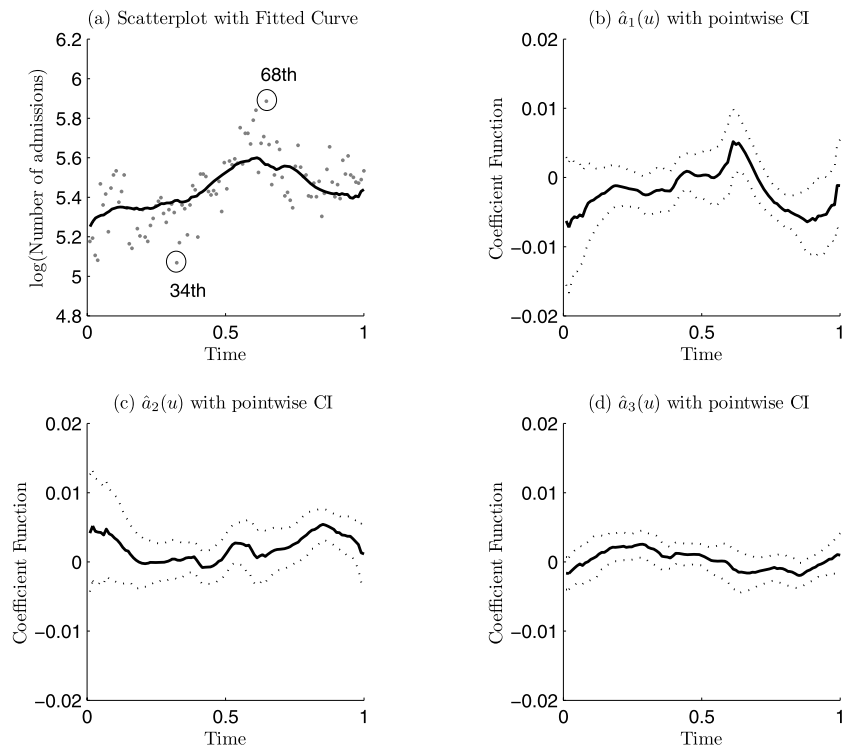


Figure 5. (a) Scatterplot of the log of number of total hospital admissions over time, and the solid curve is an estimator of the expected log of number of hospital admissions over time at the average pollutant levels, that is, $\hat{a}_0(u) + \hat{a}_1(u)\bar{X}_1 + \hat{a}_2(u)\bar{X}_2 + \hat{a}_3(u)\bar{X}_3$, (b), (c), and (d) are the estimated coefficient functions via the local rank estimator for $a_k(\cdot)$, $k = 1, 2$, and 3 , respectively.

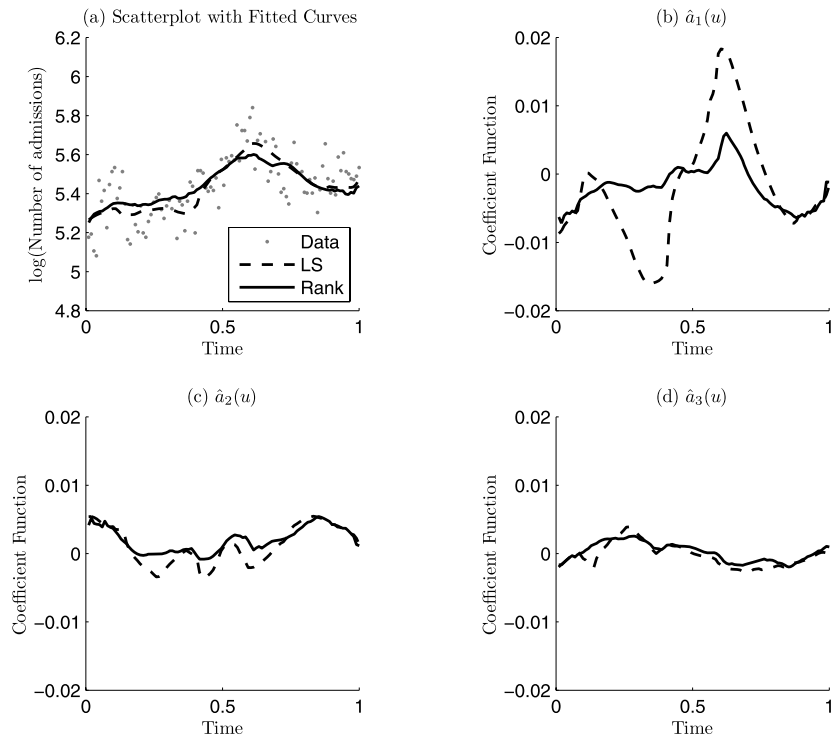


Figure 6. (a) Scatterplot of the perturbed data (without the two outliers shown as they are outside of the range), and the local LS and the local rank estimators of the expected log of number of hospital admissions. (b), (c), and (d) are the local LS and the rank estimators of the coefficient functions for $a_k(\cdot)$, $k = 1, 2$, and 3 , respectively.

Using Condition (C3) and notice that $K(\cdot)$ is symmetric, it can be shown that $\mathbf{A}_{n11}^2 \rightarrow \mathbf{0}$. By symmetry, $\mathbf{A}_{n11}^3 \rightarrow \mathbf{0}$. Similarly, we have

$$\begin{aligned} \mathbf{A}_{n11}^4 &= 2h^{-2}E\left\{\left[\mathbf{X}_i(U_i - u_0)/h - \mathbf{X}_j(U_j - u_0)/h\right] \right. \\ &\quad \times \left. [\mathbf{X}_i(U_i - u_0)/h - \mathbf{X}_j(U_j - u_0)/h]^T \right. \\ &\quad \times \left. K\left(\frac{U_i - u_0}{h}\right)K\left(\frac{U_j - u_0}{h}\right)\right\} \\ &\rightarrow 4f^2(u_0)\Sigma(u_0)\mu_2. \end{aligned}$$

Thus

$$\mathbf{A}_{n11} \rightarrow 4f^2(u_0)\Sigma(u_0)\begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mu_2\mathbf{I}_p \end{pmatrix}.$$

Similarly, we can show that $\mathbf{A}_{n12} = \mathbf{A}_{n21}^T \rightarrow \mathbf{0}$, and

$$\begin{aligned} A_{n22} &= 2 \int (t_1 - t_2)^2 K(t_1)K(t_2)f(u_0 + t_1h)f(u_0 + t_2h) dt_1 dt_2 \\ &\rightarrow 4f^2(u_0)\mu_2. \end{aligned}$$

Lemma A.3. Under Conditions (C1)–(C4), we have

$$\gamma_n^{-1}[S_n(\boldsymbol{\beta}^*, \alpha_2^*) - S_n(\mathbf{0}, 0)] = \gamma_n \mathbf{A} \begin{pmatrix} \boldsymbol{\beta}^* \\ \alpha_2^* \end{pmatrix} + o_p(1).$$

Proof. Let $U_n = \gamma_n^{-1}[S_n(\boldsymbol{\beta}^*, \alpha_2^*) - S_n(\mathbf{0}, 0)] = [n(n-1)]^{-1} \times \sum \sum_{i \neq j} W_n(D_i, D_j)$, where

$$\begin{aligned} W_n(D_i, D_j) &= 2[I(\epsilon_i - \gamma_n \alpha_2^*(U_i - u_0)/h - \gamma_n \boldsymbol{\beta}^{*T} \mathbf{Z}_i + \Delta_i(u_0)) \\ &\quad \leq \epsilon_j - \gamma_n \alpha_2^*(U_j - u_0)/h - \gamma_n \boldsymbol{\beta}^{*T} \mathbf{Z}_j + \Delta_j(u_0)) - 1/2] \\ &\quad \times \begin{pmatrix} \mathbf{Z}_i - \mathbf{Z}_j \\ (U_i - U_j)/h \end{pmatrix} K_h(U_i - u_0)K_h(U_j - u_0). \end{aligned}$$

Let $H_n(D_i, D_j) = [W_n(D_i, D_j) + W_n(D_j, D_i)]/2$, then $U_n = [n(n-1)]^{-1} \sum \sum_{i \neq j} H_n(D_i, D_j)$ has the form of a generalized U -statistic. Note that $E[\|H_n(D_i, D_j)\|^2] \leq \frac{1}{2}E[\|W_n(D_i, D_j)\|^2] + \frac{1}{2}E[\|W_n(D_i, D_j)\|^2] = E[\|W_n(D_i, D_j)\|^2]$. Furthermore,

$$\begin{aligned} E[\|W_n(D_i, D_j)\|^2] &\leq 4h^{-4}E\left\{[(\mathbf{Z}_i - \mathbf{Z}_j)^T(\mathbf{Z}_i - \mathbf{Z}_j) + [(U_i - U_j)/h]^2] \right. \\ &\quad \times \left. K^2\left(\frac{U_i - u_0}{h}\right)K^2\left(\frac{U_j - u_0}{h}\right)\right\} \\ &= O(h^{-2}) = o(n) \quad \text{since } nh^2 \rightarrow \infty \text{ by assumption.} \end{aligned}$$

Thus $U_n = E[H_n(D_i, D_j)] + o_p(1)$ by Lemma A.1. Furthermore,

$$\begin{aligned} E[H_n(D_i, D_j)] &= 2h^{-2}E\left\{\int [G(\epsilon + \Delta_j(u_0) - \Delta_i(u_0)) \right. \\ &\quad - \gamma_n \alpha_2^*(U_j - U_i)/h - \gamma_n \boldsymbol{\beta}^{*T}(\mathbf{Z}_j - \mathbf{Z}_i)] - G(\epsilon)]g(\epsilon) d\epsilon \\ &\quad \times \begin{pmatrix} \mathbf{Z}_i - \mathbf{Z}_j \\ (U_i - U_j)/h \end{pmatrix} K\left(\frac{U_i - u_0}{h}\right)K\left(\frac{U_j - u_0}{h}\right)\right\} \\ &= 2h^{-2}\gamma_n E\left\{\int g[\epsilon + \Delta_j(u_0) - \Delta_i(u_0)]g(\epsilon) d\epsilon \right. \\ &\quad \times \begin{pmatrix} \mathbf{Z}_i - \mathbf{Z}_j \\ (U_i - U_j)/h \end{pmatrix} (\mathbf{Z}_i^T - \mathbf{Z}_j^T, (U_i - U_j)/h) \end{aligned}$$

$$\begin{aligned} &\times K\left(\frac{U_i - u_0}{h}\right)K\left(\frac{U_j - u_0}{h}\right)\} \begin{pmatrix} \boldsymbol{\beta}^* \\ \alpha_2^* \end{pmatrix} (1 + o(1)) \\ &= \gamma_n \mathbf{A}_n \begin{pmatrix} \boldsymbol{\beta}^* \\ \alpha_2^* \end{pmatrix} \{1 + o(1)\}. \end{aligned}$$

The proof is completed by using Lemma A.2.

Proof of Lemma 3.1

In view of Lemma A.3, it follows that

$$\begin{aligned} &\nabla[\gamma_n^{-1}Q_n^*(\boldsymbol{\beta}^*, \alpha_2^*) - B_n(\boldsymbol{\beta}^*, \alpha_2^*)] \\ &= \gamma_n^{-1}[S_n(\boldsymbol{\beta}^*, \alpha_2^*) - S_n(\mathbf{0}, 0)] - \gamma_n \mathbf{A} \begin{pmatrix} \boldsymbol{\beta}^* \\ \alpha_2^* \end{pmatrix} \\ &= o_p(1). \end{aligned}$$

The proof follows along the same lines of the proof of theorem A.3.7. of Hettmansperger and McKean (1998), using a ‘‘diagonal subsequencing’’ argument and convexity.

Proof of Theorem 3.2

By Lemma 3.1, $\gamma_n^{-1}Q_n^*(\mathbf{s}_1, s_2) = B_n(\mathbf{s}_1, s_2) + r_n(\mathbf{s}_1, s_2)$, where $r_n(\mathbf{s}_1, s_2) \xrightarrow{p} 0$ uniformly over any bounded set. Note that $\gamma_n^{-1}Q_n^*(\mathbf{s}_1, s_2)$ is minimized by $(\tilde{\boldsymbol{\beta}}_n^{*T}, \tilde{\alpha}_{2n}^*)^T$, and $B_n(\mathbf{s}_1, s_2)$ is minimized by $(\tilde{\boldsymbol{\beta}}_n^{*T}, \tilde{\alpha}_{2n}^*)^T = -\gamma_n^{-2}\mathbf{A}^{-1}(S_{n1}^T(\mathbf{0}, 0), S_{n2}(\mathbf{0}, 0))^T$. We first establish the asymptotic representation by following similar argument as in Hjort and Pollard (1993). For any constant $c > 0$, define

$$\begin{aligned} T_n &= \inf_{\|(\mathbf{s}_1^T, s_2) - (\tilde{\boldsymbol{\beta}}_n^{*T}, \tilde{\alpha}_{2n}^*)\| \leq c} B_n(\mathbf{s}_1, s_2) - B_n(\tilde{\boldsymbol{\beta}}_n^*, \tilde{\alpha}_{2n}^*), \\ R_n &= \sup_{\|(\mathbf{s}_1^T, s_2) - (\tilde{\boldsymbol{\beta}}_n^{*T}, \tilde{\alpha}_{2n}^*)\| \leq c} |\gamma_n^{-1}Q_n^*(\mathbf{s}_1, s_2) - B_n(\mathbf{s}_1, s_2)|, \end{aligned}$$

then $R_n \xrightarrow{p} 0$ as $n \rightarrow \infty$. Let $(\mathbf{s}_1^T, s_2)^T$ be an arbitrary point outside the ball $\{(\mathbf{s}_1^T, s_2)^T : \|(\mathbf{s}_1^T, s_2) - (\tilde{\boldsymbol{\beta}}_n^{*T}, \tilde{\alpha}_{2n}^*)\| \leq c\}$, then we can write $(\mathbf{s}_1^T, s_2)^T = (\tilde{\boldsymbol{\beta}}_n^{*T}, \tilde{\alpha}_{2n}^*)^T + l\mathbf{1}_{2p+1}$, where $l > c$ is a positive constant and $\mathbf{1}_d$ denotes a unit vector of length d . Write

$$\begin{aligned} &\frac{c}{l}[\gamma_n^{-1}Q_n^*(\mathbf{s}_1, s_2) - \gamma_n^{-1}Q_n^*(\tilde{\boldsymbol{\beta}}_n^*, \tilde{\alpha}_{2n}^*)] \\ &= \frac{c}{l}\gamma_n^{-1}Q_n^*(\mathbf{s}_1, s_2) + \left(1 - \frac{c}{l}\right)\gamma_n^{-1}Q_n^*(\tilde{\boldsymbol{\beta}}_n^*, \tilde{\alpha}_{2n}^*) \\ &\quad - \gamma_n^{-1}Q_n^*(\tilde{\boldsymbol{\beta}}_n^*, \tilde{\alpha}_{2n}^*). \end{aligned}$$

By the convexity of $\gamma_n^{-1}Q_n^*(\mathbf{s}_1, s_2)$, we have

$$\begin{aligned} &\frac{c}{l}[\gamma_n^{-1}Q_n^*(\mathbf{s}_1, s_2) - \gamma_n^{-1}Q_n^*(\tilde{\boldsymbol{\beta}}_n^*, \tilde{\alpha}_{2n}^*)] \\ &\geq \gamma_n^{-1}Q_n^*\left(\frac{c}{l}(\mathbf{s}_1, s_2) + \left(1 - \frac{c}{l}\right)(\tilde{\boldsymbol{\beta}}_n^*, \tilde{\alpha}_{2n}^*)\right) - \gamma_n^{-1}Q_n^*(\tilde{\boldsymbol{\beta}}_n^*, \tilde{\alpha}_{2n}^*). \end{aligned}$$

Thus,

$$\begin{aligned} &\frac{c}{l}[\gamma_n^{-1}Q_n^*(\mathbf{s}_1, s_2) - \gamma_n^{-1}Q_n^*(\tilde{\boldsymbol{\beta}}_n^*, \tilde{\alpha}_{2n}^*)] \\ &\geq \gamma_n^{-1}Q_n^*(\tilde{\boldsymbol{\beta}}_n^* + c\mathbf{1}_{2p}, \tilde{\alpha}_{2n}^* + c) - \gamma_n^{-1}Q_n^*(\tilde{\boldsymbol{\beta}}_n^*, \tilde{\alpha}_{2n}^*) \\ &= B_n(\tilde{\boldsymbol{\beta}}_n^* + c\mathbf{1}_{2p}, \tilde{\alpha}_{2n}^* + c) + r_n(\tilde{\boldsymbol{\beta}}_n^* + c\mathbf{1}_{2p}, \tilde{\alpha}_{2n}^* + c) \\ &\quad - B_n(\tilde{\boldsymbol{\beta}}_n^*, \tilde{\alpha}_{2n}^*) - r_n(\tilde{\boldsymbol{\beta}}_n^*, \tilde{\alpha}_{2n}^*) \\ &\geq T_n - 2R_n. \end{aligned}$$

If $R_n \leq \frac{1}{2}T_n$, then $\gamma_n^{-1}Q_n^*(\mathbf{s}_1, s_2) > \gamma_n^{-1}Q_n^*(\tilde{\boldsymbol{\beta}}_n^*, \tilde{\alpha}_{2n}^*)$ for all $(\mathbf{s}_1^T, s_2)^T$ outside the ball. This implies if $R_n \leq \frac{1}{2}T_n$ then the minimizer of

$\gamma_n^{-1}Q_n^*$ must be inside the ball. Thus

$$\begin{aligned} &P(\|(\hat{\beta}_n^{*T}, \hat{\alpha}_{2n}^{*T})^T - (\tilde{\beta}_n^{*T}, \tilde{\alpha}_{2n}^{*T})^T\| \geq c) \\ &\leq P(R_n \geq \frac{1}{2}T_n) = P(R_n \geq \frac{1}{2}\lambda c^2) \\ &\rightarrow 0, \end{aligned}$$

where λ is the smallest eigenvalue of \mathbf{A} . Therefore, $(\hat{\beta}_n^{*T}, \hat{\alpha}_{2n}^{*T})^T = (\tilde{\beta}_n^{*T}, \tilde{\alpha}_{2n}^{*T})^T + o_p(1)$. This in particular implies the asymptotic representations (9), (11), and (12).

We next show the asymptotic normality of $\hat{\mathbf{a}}(u_0)$. From (9), we have $\sqrt{nh}(\hat{\mathbf{a}}(u_0) - \mathbf{a}(u_0)) = -\gamma_n^{-2}(4\tau f^2(u_0)\Sigma(\mu_0))^{-1}S_{n11}(\mathbf{0}, 0) + o_p(1)$, (A.1)

where

$$\begin{aligned} S_{n11}(\mathbf{0}, 0) &= 2\gamma_n[n(n-1)]^{-1} \\ &\times \sum_{i \neq j} [I(\epsilon_i + \Delta_i(u_0) \leq \epsilon_j + \Delta_j(u_0)) - 1/2](\mathbf{X}_i - \mathbf{X}_j) \\ &\times K_h(U_i - u_0)K_h(U_j - u_0). \end{aligned} \tag{A.2}$$

By (A.2), let us rewrite $-\gamma_n^2 S_{n11}(\mathbf{0}, 0) = S_{na1}(\mathbf{0}, 0) + S_{nb2}(\mathbf{0}, 0)$, where

$$\begin{aligned} S_{na}(\mathbf{0}, 0) &= 2\gamma_n^{-1}[n(n-1)]^{-1} \sum_{i \neq j} [I(\epsilon_i \leq \epsilon_j) - 1/2](\mathbf{X}_j - \mathbf{X}_i) \\ &\times K_h(U_i - u_0)K_h(U_j - u_0), \\ S_{nb}(\mathbf{0}, 0) &= 2\gamma_n^{-1}[n(n-1)]^{-1} \\ &\times \sum_{i \neq j} [I(\epsilon_i + \Delta_i(u_0) \leq \epsilon_j + \Delta_j(u_0)) - I(\epsilon_i \leq \epsilon_j)] \\ &\times (\mathbf{X}_j - \mathbf{X}_i)K_h(U_i - u_0)K_h(U_j - u_0). \end{aligned}$$

We next prove that

$$S_{na}(\mathbf{0}, 0) \rightarrow \mathbf{N}(\mathbf{0}, \frac{4}{3}f^3(u_0)v_0\Sigma(u_0)) \text{ in distribution.} \tag{A.3}$$

Note that we can write $S_{na}(\mathbf{0}, 0) = \sqrt{n}[n(n-1)]^{-1} \sum_{i \neq j} H_n(D_i, D_j)$, where $H_n(D_i, D_j) = W_n(D_i, D_j) + W_n(D_j, D_i)$ with

$$\begin{aligned} W_n(D_i, D_j) &= h^{-3/2}[I(\epsilon_i \leq \epsilon_j) - 1/2](\mathbf{X}_j - \mathbf{X}_i) \\ &\times K\left(\frac{U_i - u_0}{h}\right)K\left(\frac{U_j - u_0}{h}\right). \end{aligned}$$

Similarly to the arguments in the proof of Lemma A.3, it can be shown that $E[\|H_n(D_i, D_j)\|^2] = o(n)$. By Lemma A.1, this implies that $S_{na}(\mathbf{0}, 0) = 2n^{-1} \sum_{i=1}^n r_n(D_i) + o_p(1)$ since it is easy to check that $\bar{r}_n = 0$. We have

$$\begin{aligned} r_n(D_i) &= E[H_n(D_i, D_j)|D_i] \\ &= 2h^{-3/2}[G(\epsilon_i) - 1/2]K\left(\frac{U_i - u_0}{h}\right) \\ &\times E\left\{(\mathbf{X}_i - \mathbf{X}_j)K\left(\frac{U_j - u_0}{h}\right) \mid \mathbf{X}_i, U_i, \epsilon_i\right\} \\ &= 2h^{-1/2}[G(\epsilon_i) - 1/2]K\left(\frac{U_i - u_0}{h}\right) \\ &\times \left[\left(\int K(t)f(u_0 + th) dt \right) \mathbf{X}_i \right. \\ &\left. - \int E(X_j|U_j = u_0 + th)K(t)f(u_0 + th) dt \right]. \end{aligned}$$

Furthermore,

$$\begin{aligned} &E[r_n(D_i)r_n(D_i)^T] \\ &= \frac{1}{3}h^{-1}E\left\{K^2\left(\frac{U_i - u_0}{h}\right)\left[\left(\int K(t)f(u_0 + th) dt\right) \mathbf{X}_i \right. \right. \\ &\quad \left. \left. - \int E(X_j|U_j = u_0 + th)K(t)f(u_0 + th) dt\right] \right. \\ &\quad \times \left[\left(\int K(t)f(u_0 + th) dt\right) \mathbf{X}_i^T \right. \\ &\quad \left. \left. - \int E(X_j^T|U_j = u_0 + th)K(t)f(u_0 + th) dt\right] \right\} \\ &\rightarrow \frac{1}{3}f^3(u_0)v_0\Sigma(u_0). \end{aligned}$$

To prove the asymptotic normality of $S_{na}(\mathbf{0}, 0)$, it is sufficient to check the Lindeberg–Feller condition: $\forall \epsilon > 0, n^{-1} \sum_{i=1}^n E\{r_n(D_i)r_n(D_i)^T \times I(\|r_n(D_i)\| > \epsilon\sqrt{n})\} \rightarrow 0$. This can be easily verified by applying the dominated convergence theorem. Next we show that

$$S_{nb}(\mathbf{0}, 0) = \frac{2h^2}{\gamma_n}[\tau f^2(u_0)\mu_2\Sigma(u_0)\mathbf{a}''(u_0) + o(1)] + o_p(1). \tag{A.4}$$

We may write $S_{nb}(\mathbf{0}, 0) = [n(n-1)]^{-1} \sum_{i \neq j} H_n^*(D_i, D_j)$, where $H_n^*(D_i, D_j) = W_n^*(D_i, D_j) + W_n^*(D_j, D_i)$ with

$$\begin{aligned} W_n^*(D_i, D_j) &= nh^{-1}\gamma_n[I(\epsilon_i + \Delta_i(u_0) \leq \epsilon_j + \Delta_j(u_0)) - I(\epsilon_i \leq \epsilon_j)] \\ &\quad \times (\mathbf{X}_j - \mathbf{X}_i)K\left(\frac{U_i - u_0}{h}\right)K\left(\frac{U_j - u_0}{h}\right). \end{aligned}$$

Note that

$$\begin{aligned} &\Delta_j(u_0) - \Delta_i(u_0) \\ &= \frac{1}{2}[(U_j - u_0)^2\mathbf{X}_j^T - (U_i - u_0)^2\mathbf{X}_i^T]\mathbf{a}''(u_0) \\ &\quad + \frac{1}{2}[(U_j - u_0)^2 - (U_i - u_0)^2]a_0''(u_0) \\ &\quad + o((U_i - u_0)^2) + o((U_j - u_0)^2). \end{aligned}$$

By applying Lemma A.1, it can be shown that $S_{nb}(\mathbf{0}, 0) = E[H_n^*(D_i, D_j)] + o_p(1)$. It follows by using the same arguments as those in the proof of Lemma A.2 that

$$\begin{aligned} &E[H_n^*(D_i, D_j)] \\ &= 2nh^{-1}\gamma_n E\left\{ \int [G(\epsilon + \Delta_j(u_0) - \Delta_i(u_0)) - G(\epsilon)]g(\epsilon) d\epsilon \right. \\ &\quad \times (\mathbf{X}_j - \mathbf{X}_i)K\left(\frac{U_i - u_0}{h}\right)K\left(\frac{U_j - u_0}{h}\right) \left. \right\} \\ &= 2nh^{-1}\gamma_n[\tau + O(h)]E\left[(\Delta_j(u_0) - \Delta_i(u_0))(\mathbf{X}_j - \mathbf{X}_i) \right. \\ &\quad \times K\left(\frac{U_i - u_0}{h}\right)K\left(\frac{U_j - u_0}{h}\right) \left. \right] (1 + o(1)) \\ &= \frac{2h^2}{\gamma_n}[\tau f^2(u_0)\mu_2\Sigma(u_0)\mathbf{a}''(u_0) + o(1)]. \end{aligned}$$

This proves (A.4). By combining (A.3) and (A.4) and using the approximation given in (A.1), we obtain (10).

Proof of Theorem 3.3

A result of Hodges and Lehmann (1956) indicates that the ARE has a lower bound $0.864^{4/5} = 0.8896$, with this lower bound being attained at the density $f(t) = \frac{3}{20\sqrt{5}}(5 - x^2)I(|x| \leq 5)$.

Proof of Theorem 3.4

Let

$$\begin{aligned} & V_n(\alpha_1^*, \xi_1, \xi_2, \xi_3) \\ &= (nh)^{-1} \sum_{i=1}^n |\epsilon_i - \gamma_n \alpha_1^* - \xi_1(U_i - u_0) - \xi_2^T \mathbf{X}_i - \xi_3^T (U_i - u_0) \mathbf{X}_i \\ & \quad + \Delta_i(u_0)| K\left(\frac{U_i - u_0}{h}\right), \end{aligned}$$

where $\alpha_1^* = \gamma_n^{-1}(\alpha_1 - a_0(u_0))$, $\xi_1 \in \mathbb{R}$, $\xi_2 \in \mathbb{R}^p$, and $\xi_3 \in \mathbb{R}^p$. The subgradient of $V_n(\alpha_1^*, \xi_1, \xi_2, \xi_3)$ with respect to α_1^* is

$$\begin{aligned} & S_n^*(\alpha_1^*, \xi_1, \xi_2, \xi_3) \\ &= \frac{2\gamma_n}{nh} \sum_{i=1}^n [I(\epsilon_i \leq \gamma_n \alpha_1^* + \xi_1(U_i - u_0) + \xi_2^T \mathbf{X}_i + \xi_3^T (U_i - u_0) \mathbf{X}_i \\ & \quad - \Delta_i(u_0)) - 1/2] K\left(\frac{U_i - u_0}{h}\right). \end{aligned}$$

We have $S_n^*(0, 0, \mathbf{0}, \mathbf{0}) = 2\gamma_n(nh)^{-1} \sum_{i=1}^n [I(\epsilon_i \leq \Delta_i(u_0)) - 1/2] \times K\left(\frac{U_i - u_0}{h}\right)$, which is the same as the S_{n0} defined in (18). Let $U_n(\alpha_1^*, \xi_1, \xi_2, \xi_3) = \gamma_n^{-1}[S_n^*(\alpha_1^*, \xi_1, \xi_2, \xi_3) - S_n^*(0, 0, \mathbf{0}, \mathbf{0})]$, then

$$\begin{aligned} & U_n(\alpha_1^*, \xi_1, \xi_2, \xi_3) \\ &= 2(nh)^{-1} \sum_{i=1}^n [I(\epsilon_i \leq \gamma_n \alpha_1^* + \xi_1(U_i - u_0) + \xi_2^T \mathbf{X}_i \\ & \quad + \xi_3^T (U_i - u_0) \mathbf{X}_i - \Delta_i(u_0)) - I(\epsilon_i \leq \Delta_i(u_0))] K\left(\frac{U_i - u_0}{h}\right). \end{aligned}$$

For any positive constants c_i , $i = 1, 2, 3$ and $\forall \xi_1, \xi_2, \xi_3$ such that $\|\xi_1\| \leq c_1 h^{-1} \gamma_n$, $\|\xi_2\| \leq c_2 \gamma_n$, and $\|\xi_3\| \leq c_3 h^{-1} \gamma_n$, we have

$$U_n(\alpha_1^*, \xi_1, \xi_2, \xi_3) = 2\gamma_n g(0) f(u_0) \alpha_1^* + o_p(1). \quad (\text{A.5})$$

This can be proved by directly checking the mean and variance. More specifically,

$$\begin{aligned} & E[U_n(\alpha_1^*, \xi_1, \xi_2, \xi_3)] \\ &= 2h^{-1} E \left\{ [G(\gamma_n \alpha_1^* + \xi_1(U_i - u_0) + \xi_2^T \mathbf{X}_i + \xi_3^T (U_i - u_0) \mathbf{X}_i \right. \\ & \quad \left. - \Delta_i(u_0)) - G(-\Delta_i(u_0))] K\left(\frac{U_i - u_0}{h}\right) \right\} \\ &= 2h^{-1} g(0) E \left\{ [\gamma_n \alpha_1^* + \xi_1(U_i - u_0) + \xi_2^T \mathbf{X}_i + \xi_3^T (U_i - u_0) \mathbf{X}_i] \right. \\ & \quad \left. \times K\left(\frac{U_i - u_0}{h}\right) \right\} (1 + O(h)) \\ &= 2\gamma_n g(0) f(u_0) \alpha_1^* (1 + O(h)). \end{aligned}$$

And

$$\begin{aligned} & \text{Var}[U_n(\alpha_1^*, \xi_1, \xi_2, \xi_3)] \\ & \leq 4n^{-1} h^{-2} E \left\{ [I(\epsilon_i \leq \gamma_n \alpha_1^* + \xi_1(U_i - u_0) + \xi_2^T \mathbf{X}_i \right. \\ & \quad \left. + \xi_3^T (U_i - u_0) \mathbf{X}_i - \Delta_i(u_0)) \right. \\ & \quad \left. - I(\epsilon_i \leq \Delta_i(u_0))]^2 K^2\left(\frac{U_i - u_0}{h}\right) \right\} \\ & \leq 4n^{-1} h^{-2} E \left\{ K^2\left(\frac{U_i - u_0}{h}\right) \right\} = O(n^{-1} h^{-1}) = o(1). \end{aligned}$$

By (A.5) and similar proof as that for Lemma 3.1, we have

$$\gamma_n^{-1} V_n(\alpha_1^*, \xi_1, \xi_2, \xi_3) = V_n^*(\alpha_1^*) + o_p(1), \quad (\text{A.6})$$

where $V_n^*(\alpha_1^*) = \gamma_n^{-1} S_n^*(0, 0, \mathbf{0}, \mathbf{0}) \alpha_1^* + \gamma_n g(0) f(u_0) \alpha_1^{*2} + \gamma_n^{-1} V_n(0, 0, \mathbf{0}, \mathbf{0})$. Because the function $V_n(\alpha_1^*, \xi_1, \xi_2, \xi_3)$ is convex in its arguments, (A.6) can be strengthened to uniform convergence (convexity lemma, see Pollard 1991), that is,

$$\sup_{\substack{\alpha_1^* \in \mathbb{C}, \|\xi_1\| \leq c_1 h^{-1} \gamma_n \\ \|\xi_2\| \leq c_2 \gamma_n, \|\xi_3\| \leq c_3 h^{-1} \gamma_n}} |\gamma_n^{-1} V_n(\alpha_1^*, \xi_1, \xi_2, \xi_3) - V_n^*(\alpha_1^*)| = o_p(1),$$

where \mathbb{C} is a compact set in \mathbb{R} . By Theorem 3.2, $\widehat{\alpha}_2 - a'_0(u_0) = O_p(h^{-1} \gamma_n)$, $\widehat{\mathbf{a}}(u_0) - \mathbf{a}(u_0) = O_p(\gamma_n)$, and $\widehat{\mathbf{a}}'(u_0) - \mathbf{a}'(u_0) = O_p(h^{-1} \gamma_n)$, we thus have

$$\sup_{\alpha_1^* \in \mathbb{C}} |\gamma_n^{-1} V_n(\alpha_1^*, \widehat{\alpha}_2 - a'_0(u_0), \widehat{\mathbf{a}}(u_0) - \mathbf{a}(u_0), \widehat{\mathbf{a}}'(u_0) - \mathbf{a}'(u_0)) - V_n^*(\alpha_1^*)| = o_p(1).$$

Note that $V_n(\alpha_1^*, \widehat{\alpha}_2 - a'_0(u_0), \widehat{\mathbf{a}}(u_0) - \mathbf{a}(u_0), \widehat{\mathbf{a}}'(u_0) - \mathbf{a}'(u_0)) = Q_{n0}^*(\alpha_1^*, \widehat{\alpha}_2, \widehat{\boldsymbol{\beta}})$, $S_n^*(0, 0, \mathbf{0}, \mathbf{0}) = S_{n0}$, where Q_{n0}^* and S_{n0} are defined in Section 3.3. The quadratic function $V_n^*(\alpha_1^*)$ is minimized by $\widetilde{\alpha}_{1n}^* = \frac{1}{2} \gamma_n^{-2} [g(0) f(u_0)]^{-1} S_{n0}$. Similar argument as that for Theorem 3.2 shows that $\widehat{\alpha}_{1n}^* = \widetilde{\alpha}_{1n}^* + o_p(1)$. Thus we have (19). We can write $\gamma_n^{-2} S_{n0} = T_{1n} + T_{2n}$, where

$$\begin{aligned} T_{1n} &= \frac{2\gamma_n^{-1}}{nh} \sum_{i=1}^n [I(\epsilon_i \leq 0) - 1/2] K\left(\frac{U_i - u_0}{h}\right), \\ T_{2n} &= \frac{2\gamma_n^{-1}}{nh} \sum_{i=1}^n [I(\epsilon_i \leq -\Delta_i(u_0)) - I(\epsilon_i \leq 0)] K\left(\frac{U_i - u_0}{h}\right). \end{aligned}$$

By the Lindeberg–Feller central limit theorem, $T_{1n} \rightarrow \mathbf{N}(0, f(u_0) \nu_0/3)$ in distribution. By checking mean and variance, we have

$$T_{2n} = -\frac{h^2}{\gamma_n} g(0) f(u_0) a''_0(u_0) \mu_2 (1 + o(1)) + o_p(1).$$

Combining the above results and using (19), the proof is completed.

To prove Theorem 3.5, we first extend Lemma A.1 to almost sure convergence.

Lemma A.4. If $E[\|H_n(D_i, D_j)\|^2] = O(h^{-2})$, then $U_n - \widehat{U}_n = o(1)$ almost surely and $U_n = \bar{\tau}_n + o(1)$ a.s.

Proof. The proof of Powell, Stock, and Stoker (1989) for Lemma A.1 suggests that $E[\|U_n - \widehat{U}_n\|^2] = O(n^{-2} h^{-2})$. By theorem 1.3.5 of Serfling (1980), $\sum_{i=1}^n E[\|U_n - \widehat{U}_n\|^2] = O(n^{-1} h^{-2}) < \infty$. This implies that $U_n - \widehat{U}_n = o(1)$ almost surely. The second result follows by an application of the strong law of large numbers to \widehat{U}_n .

Proof of Theorem 3.5

Let $\boldsymbol{\beta}^*$ and α_2^* be defined the same as before. We introduce the reparametrized objective function $\overline{Q}_n^*(\boldsymbol{\beta}^*, \alpha_2^*)$. Let $\overline{S}_n(\boldsymbol{\beta}^*, \alpha_2^*) = (\overline{S}_{n1}^T(\boldsymbol{\beta}^*, \alpha_2^*), \overline{S}_{n2}(\boldsymbol{\beta}^*, \alpha_2^*))^T$ denote the gradient function of $\overline{Q}_n^*(\boldsymbol{\beta}^*, \alpha_2^*)$, which is defined similarly as in Section 3.1. We first show that $\overline{S}_n(\boldsymbol{\beta}^*, \alpha_2^*)$ has a similar local linear approximation as stated in Lemma A.3. To make the proof concise, we prove this for $\overline{S}_{n1}(\boldsymbol{\beta}^*, \alpha_2^*)$, where

$$\begin{aligned} & \overline{S}_{n1}(\boldsymbol{\beta}^*, \alpha_2^*) \\ &= 2\gamma_n [n(n-1)]^{-1} \end{aligned}$$

$$\begin{aligned} & \times \sum_{i \neq j} (V_i + V_j) [I(\epsilon_i - \gamma_n \alpha_2^*(U_i - u_0)/h - \gamma_n \boldsymbol{\beta}^{*T} \mathbf{Z}_i + \Delta_i(u_0)) \\ & \leq \epsilon_j - \gamma_n \alpha_2^*(U_j - u_0)/h - \gamma_n \boldsymbol{\beta}^{*T} \mathbf{Z}_j + \Delta_j(u_0)) - 1/2] \\ & \times (\mathbf{Z}_i - \mathbf{Z}_j) K_h(U_i - u_0) K_h(U_j - u_0). \end{aligned}$$

Let $U_n = \gamma_n^{-1} [\bar{S}_{n1}(\boldsymbol{\beta}^*, \alpha_2^*) - \bar{S}_{n1}(\mathbf{0}, 0)] = [n(n-1)]^{-1} \sum_{i \neq j} (V_i + V_j) M_n(D_i, D_j, \boldsymbol{\beta}^*, \alpha_2^*)$, where $M_n(D_i, D_j, \boldsymbol{\beta}^*, \alpha_2^*) = \frac{1}{2} [m_n(D_i, D_j, \boldsymbol{\beta}^*, \alpha_2^*) + m_n(D_j, D_i, \boldsymbol{\beta}^*, \alpha_2^*)]$ and

$$\begin{aligned} & m_n(D_i, D_j, \boldsymbol{\beta}^*, \alpha_2^*) \\ & = 2[I(\epsilon_i - \gamma_n \alpha_2^*(U_i - u_0)/h - \gamma_n \boldsymbol{\beta}^{*T} \mathbf{Z}_i + \Delta_i(u_0)) \\ & \leq \epsilon_j - \gamma_n \alpha_2^*(U_j - u_0)/h - \gamma_n \boldsymbol{\beta}^{*T} \mathbf{Z}_j + \Delta_j(u_0)) - 1/2] \\ & \times (\mathbf{Z}_i - \mathbf{Z}_j) K_h(U_i - u_0) K_h(U_j - u_0). \end{aligned}$$

Note that $U_n = 2n^{-1} \sum_{i=1}^n V_i [(n-1)^{-1} \sum_{j=1, j \neq i}^n M_n(D_i, D_j, \boldsymbol{\beta}^*, \alpha_2^*)]$. Conditional on $\{D_i\}_{i=1}^n$, this is a weighted average of V_i . Note that

$$E(U_n | \{D_i\}_{i=1}^n) = [n(n-1)]^{-1} \sum_{i \neq j} M_n(D_i, D_j, \boldsymbol{\beta}^*, \alpha_2^*),$$

$$\text{Var}(U_n | \{D_i\}_{i=1}^n) = n^{-2} \sum_{i=1}^n \left[(n-1)^{-1} \sum_{j=1, j \neq i}^n M_n(d_i, d_j, \boldsymbol{\beta}^*, \alpha_2^*) \right]^2.$$

By Lemma A.4, it can be shown that $[n(n-1)]^{-1} \sum_{i \neq j} M_n(D_i, D_j, \boldsymbol{\beta}^*, \alpha_2^*) = \gamma_n A^* \boldsymbol{\beta}^* + o_p(1)$ almost surely, where $A^* = 4\tau f^2(u_0) \times \text{diag}(\mathbf{I}_p, \mu_2 \mathbf{I}_p) \otimes \Sigma(\mu_0)$. It is also easy to check that $n^{-2} \sum_{i=1}^n [(n-1)^{-1} \sum_{j=1, j \neq i}^n M_n(D_i, D_j, \boldsymbol{\beta}^*, \alpha_2^*)]^2 = o_p(1)$ almost surely. Thus for almost surely every sequence $\{D_i\}_{i=1}^n$, $U_n = \gamma_n A^* \boldsymbol{\beta}^* + o_p(1)$, where $o_p(1)$ is in the probability space generated by $\{V_i\}_{i=1}^n$. The proofs of Lemma 3.1 and the asymptotic representation in Theorem 3.2 can be similarly carried out to show that for almost surely every sequence $\{D_i\}_{i=1}^n$,

$$\begin{aligned} & \sqrt{nh}[\hat{\mathbf{a}}_n(u_0) - \mathbf{a}(u_0)] \\ & = -\gamma_n^{-2} (4\tau f^2(u_0) \Sigma(\mu_0))^{-1} \bar{S}_{na}^*(\mathbf{0}, 0) + o_p(1), \quad (\text{A.7}) \end{aligned}$$

where $o_p(1)$ is in the probability space generated by $\{V_i\}_{i=1}^n$, and

$$\begin{aligned} \bar{S}_{na}^*(\mathbf{0}, 0) & = 2\gamma_n [n(n-1)]^{-1} \sum_{i \neq j} (V_i + V_j) [I(\epsilon_i + \Delta_i(u_0)) \\ & \leq \epsilon_j + \Delta_j(u_0)) - 1/2] (\mathbf{X}_i - \mathbf{X}_j) K_h(U_i - u_0) K_h(U_j - u_0). \end{aligned}$$

The approximation (A.1) can be strengthened to almost surely convergence, that is,

$$\begin{aligned} & \sqrt{nh}[\hat{\mathbf{a}}_n(u_0) - \mathbf{a}(u_0)] \\ & = -\gamma_n^{-2} (4\tau f^2(u_0) \Sigma(\mu_0))^{-1} S_{na}^*(\mathbf{0}, 0) + o(1) \quad \text{a.s.} \quad (\text{A.8}) \end{aligned}$$

Combining (17) and (A.7), we have that for almost surely every sequence $\{D_i\}_{i=1}^n$,

$$\begin{aligned} & \sqrt{nh}[\hat{\mathbf{a}}_n(u_0) - \hat{\mathbf{a}}_n(u_0)] \\ & = -\gamma_n^{-2} (4\tau f^2(u_0) \Sigma(\mu_0))^{-1} [\bar{S}_{na}^*(\mathbf{0}, 0) - S_{na}^*(\mathbf{0}, 0)] + o_p(1). \end{aligned}$$

Note that

$$\begin{aligned} & \gamma_n^{-2} [\bar{S}_{na}^*(\mathbf{0}, 0) - S_{na}^*(\mathbf{0}, 0)] \\ & = 2\gamma_n^{-1} [n(n-1)]^{-1} \sum_{i \neq j} [(V_i - 1/2) + (V_j - 1/2)] \\ & \times [I(\epsilon_i + \Delta_i(u_0) \leq \epsilon_j + \Delta_j(u_0)) - 1/2] \end{aligned}$$

$$\begin{aligned} & \times (\mathbf{X}_i - \mathbf{X}_j) K_h(U_i - u_0) K_h(U_j - u_0) \\ & = 4\gamma_n^{-1} n^{-1} \sum_{i=1}^n (V_i - 1/2) \left\{ (n-1)^{-1} \right. \\ & \times \sum_{j=1, j \neq i}^n [I(\epsilon_i + \Delta_i(u_0) \leq \epsilon_j + \Delta_j(u_0)) - 1/2] \\ & \left. \times (\mathbf{X}_i - \mathbf{X}_j) K_h(U_i - u_0) K_h(U_j - u_0) \right\}. \end{aligned}$$

And $E\{\gamma_n^{-2} [\bar{S}_{na}^*(\mathbf{0}, 0) - S_{na}^*(\mathbf{0}, 0)] | \{D_i\}_{i=1}^n\} = \mathbf{0}$. We have

$$\begin{aligned} & \text{Var}\{\gamma_n^{-2} [\bar{S}_{na}^*(\mathbf{0}, 0) - S_{na}^*(\mathbf{0}, 0)] | \{D_i\}_{i=1}^n\} \\ & = 16\gamma_n^{-2} n^{-2} (n-1)^2 \\ & \times \sum_{i=1}^n \left\{ \sum_{j=1, j \neq i}^n [I(\epsilon_i + \Delta_i(u_0) \leq \epsilon_j + \Delta_j(u_0)) - 1/2] \right. \\ & \left. \times (\mathbf{X}_i - \mathbf{X}_j) K_h(U_i - u_0) K_h(U_j - u_0) \right\}^2 \\ & = W_1 + W_2, \end{aligned}$$

where

$$\begin{aligned} W_1 & = 16\gamma_n^{-2} n^{-2} (n-1)^2 h^{-4} \\ & \times \sum_{i=1}^n \sum_{j=1, j \neq i}^n [I(\epsilon_i + \Delta_i(u_0) \leq \epsilon_j + \Delta_j(u_0)) - 1/2]^2 \\ & \times (\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T K^2((U_i - u_0)/h) K^2((U_j - u_0)/h), \\ W_2 & = 16\gamma_n^{-2} n^{-2} (n-1)^2 h^{-4} \\ & \times \sum_{i=1}^n \sum_{j_1 \neq i, j_2 \neq i, j_1}^n [I(\epsilon_i + \Delta_i(u_0) \leq \epsilon_{j_1} + \Delta_{j_1}(u_0)) - 1/2] \\ & \times [I(\epsilon_i + \Delta_i(u_0) \leq \epsilon_{j_2} + \Delta_{j_2}(u_0)) - 1/2] \\ & \times (\mathbf{X}_i - \mathbf{X}_{j_1})(\mathbf{X}_i - \mathbf{X}_{j_2})^T \\ & \times K^2((U_i - u_0)/h) K((U_{j_1} - u_0)/h) K((U_{j_2} - u_0)/h). \end{aligned}$$

Lemma A.4 can be used to show that $W_1 = o_p(1)$ almost surely; and a minor extension of Lemma A.4 to third-order U -statistic can be used to show that $W_2 = \frac{4}{3} f^3(u_0) v_0 \Sigma(u_0) + o_p(1)$ almost surely. The asymptotic normality of $\gamma_n^{-2} [\bar{S}_{na}^*(\mathbf{0}, 0) - S_{na}^*(\mathbf{0}, 0)]$ follows by showing that the condition of Lindeberg–Feller central limit theorem for triangular arrays holds almost surely. We have, for almost surely every sequence $\{D_i\}_{i=1}^n$,

$$\gamma_n^{-2} [\bar{S}_{na}^*(\mathbf{0}, 0) - S_{na}^*(\mathbf{0}, 0)] \rightarrow \mathbf{N}(\mathbf{0}, \frac{4}{3} f^3(u_0) v_0 \Sigma(u_0))$$

in distribution. This completes the proof.

[Received January 2009. Revised June 2009.]

REFERENCES

Brumback, B., and Rice, J. A. (1998), "Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves" (with discussion), *Journal of the American Statistical Association*, 93, 961–994.
 Cai, Z., Fan, J., and Li, R. (2000), "Efficient Estimation and Inferences for Varying-Coefficient Models," *Journal of the American Statistical Association*, 95, 888–902.
 Cleveland, W. S., Grosse, E., and Shyu, W. M. (1992), "Local Regression Models," in *Statistical Models in S*, eds. J. M. Chambers and T. J. Hastie, Pacific Grove, CA: Wadsworth & Brooks, pp. 309–376.

- David, H. A. (1998), "Early Sample Measures of Variability," *Statistical Science*, 13, 368–377.
- Fan, J., and Li, R. (2006), "An Overview on Nonparametric and Semiparametric Techniques for Longitudinal Data," in *Frontiers in Statistics*, eds. J. Fan and H. Koul, London: Imperial College Press, pp. 277–303.
- Fan, J., and Zhang, W. (1999), "Statistical Estimation in Varying-Coefficient Models," *The Annals of Statistics*, 27, 1491–1518.
- (2000), "Simultaneous Confidence Bands and Hypothesis Testing in Varying-Coefficient Models," *Scandinavian Journal of Statistics*, 27, 715–731.
- (2008), "Statistical Methods With Varying Coefficient Models," *Statistics and Its Interface*, 1, 179–195.
- Hall, P., and Kang, K. H. (2001), "Bootstrapping Nonparametric Density Estimators With Empirically Chosen Bandwidths," *The Annals of Statistics*, 29, 1443–1468.
- Hastie, T. J., and Tibshirani, R. J. (1993), "Varying-Coefficient Models" (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 55, 757–796.
- Hettmansperger, T. P., and McKean, J. W. (1998), *Robust Nonparametric Statistical Methods*, London: Arnold.
- Hjort, N. L., and Pollard, D. (1993), "Asymptotics for Minimisers of Convex Processes," preprint, available at <http://citeseer.ist.psu.edu/hjort93asymptotics.html>.
- Hodges, J. L., and Lehmann, E. L. (1956), "The Efficiency of Some Nonparametric Competitors of the t -Test," *The Annals of Mathematical Statistics*, 27, 324–335.
- Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L.-P. (1998), "Nonparametric Smoothing Estimates of Time-Varying Coefficient Models With Longitudinal Data," *Biometrika*, 85, 809–822.
- Jin, Z., Ying, Z., and Wei, L. J. (2001), "A Simple Resampling Method by Disturbing the Minimax," *Biometrika*, 88, 381–390.
- Kauermann, G., and Tutz, G. (1999), "On Model Diagnostics Using Varying Coefficient Models," *Biometrika*, 86, 119–128.
- Kim, M.-O. (2007), "Quantile Regression With Varying Coefficients," *The Annals of Statistics*, 35, 92–108.
- Leung, D. H. (2005), "Cross-Validation in Nonparametric Regression With Outliers," *The Annals of Statistics*, 33, 2291–2310.
- McKean, J. W. (2004), "Robust Analysis of Linear Models," *Statistical Science*, 19, 562–570.
- Pollard, D. (1991), "Asymptotics for Least Absolute Deviation Regression Estimators," *Econometric Theory*, 7, 186–199.
- Powell, J. L., Stock, J. H., and Stoker, T. M. (1989), "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.
- Serfling, R. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley.
- Terpstra, J., and McKean, J. (2005), "Rank-Based Analysis of Linear Models Using R," *Journal of Statistical Software*, 14, 1–26.
- Wang, L., Kai, B., and Li, R. (2009), "Local Rank Inference for Varying Coefficient Models," technical report, The Methodology Center, The Pennsylvania State University.
- Wu, C. O., Chiang, C. T., and Hoover, D. R. (1998), "Asymptotic Confidence Regions for Kernel Smoothing of a Varying-Coefficient Model With Longitudinal Data," *Journal of the American Statistical Association*, 93, 1388–1402.
- Yu, K., and Jones, M. C. (1998), "Local Linear Quantile Regression," *Journal of the American Statistical Association*, 93, 228–237.
- Zhang, W., and Lee, S. Y. (2000), "Variable Bandwidth Selection in Varying-Coefficient Models," *Journal of Multivariate Analysis*, 74, 116–134.