Research Article

Least Absolute Deviation Estimate for Functional Coefficient Partially Linear Regression Models

Yanqin Feng,[1] Guoxin Zuo,[2] and Li Liu[1]

[1] *School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China*
[2] *School of Mathematics and Statistics, Central China Normal University, Wuhan 430079, China*

Correspondence should be addressed to Yanqin Feng, yanqf2008@yahoo.com.cn

The functional coefficient partially linear regression model is a useful generalization of the nonparametric model, partial linear model, and varying coefficient model. In this paper, the local linear technique and the L_1 method are employed to estimate all the functions in the functional coefficient partially linear regression model. The asymptotic properties of the proposed estimators are studied. Simulation studies are conducted to show the validity of the estimate procedure.

1. Introduction

In this paper, we are concerned with a functional coefficient partially linear regression model (FCPLR), that is,

$$Y = a_0(X) + \sum_{j=1}^{p} a_j(U)Z_j + \varepsilon, \tag{1.1}$$

where X and U are random explanatory variables, $\mathbf{Z} = (Z_1, \ldots, Z_p)'$ is a random vector, and $a_j(\cdot)$ is some measurable function from R to R for $j = 0, \ldots, p$. We call $a_0(\cdot)$ the intercept function, and $\{a_j(\cdot)\}$, $j = 1, \ldots, p$, the coefficient functions. As usual, ε denotes the errors with zero-mean and fixed variance.

The FCPLR model, first introduced by Wong et al. [1], is a generalization of the nonparametric model, partial linear model, and varying coefficient model. Zhu et al. [2] studied a similar functional coefficient model, a functional mixed model, using a new Bayesian method. Model (1.1) reduces to a varying coefficient regression mode if the intercept

function $a_0(\cdot)$ is a constant function and a partially linear regression model when the coefficient functions $\{a_j(\cdot)\}_{j=1}^{p}$ are constants. Many researchers, for example, Aneiros-Pérez and Vieu [3], made contributions to studying this kind of model. When X and U in model (1.1) coincide, the model becomes a semiparametric varying coefficient model discussed by Ahmad et al. [4]. Since the FCPLR model combines the nonparametric and functional coefficient regression model, its flexibility makes it attractable in various regression problems [1].

Statistical inference for the FCPLR model mainly includes the estimations of the intercept function $a_0(\cdot)$ and the coefficient functions $\{a_j(\cdot)\}_{j=1}^{p}$. To estimate the unknown functions in the nonparametric/semiparametric regression models, many statistical inference methods have been proposed over the past decades, such as the kernel estimate method [5–7], spline smoothing [8, 9], and two-step estimation method [10, 11]. Wong et al. [1] employed local linear regression method and integrated method to give the initial estimates of all functions in the FCPLR model. All the papers mentioned above used the least-squares technique to obtain the estimators of the unknown coefficient functions. The least-squares estimators (L_2 method), of course, have some good properties, especially for the normal random errors case. It is well known that, however, the least-squares method will perform poor when the random errors have a heavy-tailed distribution in that it is highly sensitive to extreme values and outliers. This motivates us to find more robust estimation methods instead of the aforementioned inference methods for model (1.1).

Local linear approximation is a good method for nonparametric regression problems [12], and the L_1 method based on the least absolute deviations overcomes the sensitivity caused by outliers. As noted in Wang and Scott [13] and Fan and Gijbels [12], among many robust estimation methods, the L_1 method based on the local least absolute deviations behaves quite well. In this paper, we adopt the L_1 method, accompany with the local linear technique and the integrated method to estimate all the unknown functions in model (1.1). Furthermore, the estimating problem can be reduced to a linear programming problem, and the numerical solutions are obtained quickly by some available softwares subsequently (e.g., Matlab is very useful for this kind of problems). The main difficulty encountered in the proof of the asymptotic normalities is that the L_1 estimates have no closed form. This paper shows the asymptotic normalities of L_1 estimators through a method completely different from those based on the L_2 method, and the simulation results show that the L_1 method is a robust method indeed.

The rest of this paper is organized as follows. In Section 2, we describe the estimation method and the associated bandwidth selection procedure. Section 3 gives the the asymptotic theories of the estimators. Simulation studies are conducted in Section 4. A real application is given in Section 5. Section 6 gives the proofs of the main results.

2. Least Absolute Deviation Estimate

This section gives the main idea of the proposed estimation method; that is, local linear polynomials are used to approximate the nonparametric function and the functional coefficients, and the least absolute deviation technique is used to find the best approximation. Bandwidth selection technique is also discussed in this section. Throughout this paper, we suppose that $\{(Y_i, X_i, U_i, \mathbf{Z}_i')'\}_{i=1}^{n}$ is an i.i.d. sample from model (1.1) and assume the following conditions.

Assumptions

(1) $\Omega = E(\mathbf{Z}\mathbf{Z}' \mid X = x_0, U = u_0)$ is a positive definite matrix, $E(\mathbf{Z} \mid X = x_0, U = u_0) = \omega$.

(2) Bandwidth h subjects to $h \sim n^{-1/6}$.

(3) Random error ε, with zero mean and zero median, is independent of Z conditional on (X, U). The conditional probability density $g(\cdot|x, u)$ of ε given random variables X and U is continuous in a neighborhood of the point 0, and $g(0|x, u) > 0$. $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed.

(4) The density functions $f(x, u)$, $f_1(x)$, $f_2(u)$ of (X, U), X, and U are continuous in neighborhoods of (x_0, u_0), x_0 and u_0, and $f(x_0, u_0) > 0$.

(5) All functions $a_j(u)$, $j = 1, \ldots, p$ and $a_0(x)$ are twice continuously differentiable in neighborhoods of u_0 and x_0, respectively.

(6) Kernel functions $K_k(\cdot)$, $k = 1, 2$ are bounded, nonnegative, and compactly supported.

(7) $\max_{1 \leq i \leq n} \|\mathbf{Z}_i\| = o_p(n^{1/3})$.

To simplify typesetting, we introduce the following symbols:

$$\mu_i(k) = \int v^i K_k(v)dv, \qquad \gamma_i(k) = \int v^i K_k^2(v)dv \quad \text{for } k = 1, 2. \ i = 1, 2, \ldots. \tag{2.1}$$

2.1. Local Linear Estimate Based on Least Absolute Deviation

The main idea is to approximate the functional coefficients $a_j(\cdot)$ by linear functions for $j = 0, \ldots, p$, that is, $a_0(x)$ can be approximated by

$$a_0(x) \approx a_0(x_0) + a_0'(x_0)(x - x_0) \tag{2.2}$$

for x in a neighborhood of x_0 within the closed support of X, and $a_j(u)$ by

$$a_j(u) \approx a_j(u_0) + a_j'(u_0)(u - u_0), \quad j = 1, \ldots, p \tag{2.3}$$

for u in a neighborhood of u_0 within the closed support of U. For simplicity, denoting $a_0(x_0)$, $a_0'(x_0)$ as a_0, b_0, and $a_j(u_0)$, $a_j'(u_0)$ as a_j, b_j for $j = 1, \ldots, p$. The local linear least absolute deviation estimate (L_1 estimate) of the unknown parameters $a_0, b_0, \mathbf{a} := (a_1, \ldots, a_p)'$ and $\mathbf{b} := (b_1, \ldots, b_p)'$, denoted, respectively, by $\hat{a}_0, \hat{b}_0, \hat{\mathbf{a}}, \hat{\mathbf{b}}$, is the optimal solution of the minimization problem as follows:

$$\min \sum_{i=1}^{n} |Y_i - a_0 - b_0(X_i - x_0) - \mathbf{a}'\mathbf{Z}_i - \mathbf{b}'(U_i - u_0)\mathbf{Z}_i| K_{1,h}(X_i - x_0)K_{2,h}(U_i - u_0), \tag{2.4}$$

where $K_{k,h}(\cdot) = K_k(\cdot/h)$, $k = 1,2$ are the given kernel functions and h is the chosen bandwidth. The optimization problem is equivalent to the following linear programming problem:

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{n} (e_i^+ + e_i^-) K_{1,h}(X_i - x_0) K_{2,h}(U_i - u_0) \\
\text{s.t.} \quad & a_0 + b_0(X_i - x_0) + \mathbf{a}'\mathbf{Z}_i + \mathbf{b}'(U_i - u_0)\mathbf{Z}_i + e_i^+ - e_i^- = Y_i, \\
& e_i^+ \geq 0, \quad e_i^- \geq 0, \quad i = 1, \ldots, n.
\end{aligned}
\tag{2.5}
$$

There are many algorithms available for the optimal solution of problem (2.5); for example, the feasible direction method can be directly used to compute the optimal solution [14], and the numerical solution of (2.5) can be quickly computed by a series of Matlab functions.

By the integrated method [1], the estimator of the intercept function $a_0(x)$ is defined by

$$
\tilde{a}_0(x_0) = \frac{1}{n} \sum_{k=1}^{n} \hat{a}_0(x_0, U_k),
\tag{2.6}
$$

and the estimators of the coefficient functions $a_j(u_0)$, $j = 1, \ldots, p$ are defined by

$$
\tilde{a}_j(u_0) = \frac{1}{n} \sum_{k=1}^{n} \hat{a}_j(X_k, u_0), \quad j = 1, \ldots, p.
\tag{2.7}
$$

We focus our main task in establishing the asymptotic distributions of the estimators $\tilde{a}_0(x_0)$ and $\tilde{a}_j(u_0)$ for $j = 1, \ldots, p$.

2.2. Selection of Bandwidth

It is well known that the choice of the bandwidth strongly influences the adequacy of the estimators. We use an automatic bandwidth choice procedure in this paper; that is, the absolute cross-validation (ACV) method, and the ACV bandwidth h_{ACV} is defined as

$$
h_{\text{ACV}} = \arg \min \text{ACV}(h) = \frac{1}{n} \sum_{k=1}^{n} \left| Y_k - \hat{a}_0^{-k}(X_k) - \sum_{j=1}^{p} \hat{a}_j^{-k}(U_k) Z_{kj} \right|,
\tag{2.8}
$$

where $a_0^{-k}(X_k)$ and $a_j^{-k}(U_k)$, $j = 1, \ldots, p$ are constructed based on observations with size $n - 1$ by leaving out the kth observation $((Y_k, X_k, U_k)', \mathbf{Z}_k')'$. According to Wang and Scott [13], this bandwidth is better than cross-validation (CV) bandwidth h_{CV}. The latter one is suggested by Rice and Silverman [15] and often used in curve regression as in Hoover et al. [6] and Wu et al. [7].

3. Asymptotic Theory

This section gives the asymptotic distribution theories of the estimators. Using Taylor's expansion, for $|X_i - x_0| \leq Ch$ and $|U_i - u_0| \leq Ch$, we have

$$
\begin{aligned}
a_0(X_i) &= a_0(x_0) + a_0'(x_0)(X_i - x_0) + \frac{1}{2}a_0''(\xi_i)(X_i - x_0)^2, \\
a_j(U_i) &= a_j(u_0) + a_j'(u_0)(U_i - u_0) + \frac{1}{2}a_j''(\eta_{ij})(U_i - u_0)^2, \quad \text{for } j = 1, \dots, p,
\end{aligned}
\tag{3.1}
$$

where ξ_i is between x_0 and X_i, and η_{ij} is between U_i and u_0. Let $\mathbf{a}(u_0) = (a_1(u_0), \dots, a_p(u_0))'$, $\mathbf{b}(u_0) = (a_1'(u_0), \dots, a_p'(u_0))'$, $\mathbf{c}(u_0) = (a_1''(u_0), \dots, a_p''(u_0))'$, then we have

$$
\begin{aligned}
&\left(\widehat{a}_0, \widehat{b}_0, \widehat{\mathbf{a}}', \widehat{\mathbf{b}}'\right)' \\
&= \arg\min \sum_{i=1}^{n} \left| Y_i - a_0 - b_0(X_i - x_0) - \mathbf{a}'\mathbf{Z}_i - \mathbf{b}'(U_i - u_0)\mathbf{Z}_i \right| K_{1,h}(X_i - x_0)K_{2,h}(U_i - u_0) \\
&= \arg\min \sum_{i=1}^{n} \left\{ \left| \left| \sqrt{nh^2}\left[(a_0 - a_0(x_0)) + (b_0 - a_0'(x_0))(X_i - x_0)\right] \frac{1}{\sqrt{nh^2}} \right. \right. \right. \\
&\qquad\qquad \left. + \sqrt{nh^2}\left[(\mathbf{a} - \mathbf{a}(u_0))'\mathbf{Z}_i + (\mathbf{b} - \mathbf{b}(u_0))'(U_i - u_0)\mathbf{Z}_i\right]\frac{1}{\sqrt{nh^2}} \right. \\
&\qquad\qquad \left. \left. - \left[\frac{1}{2}\left(a_0''(\xi_i)(X_i - x_0)^2 + \sum_{j=1}^{p} a_j''(\eta_{ij})(U_i - u_0)^2 Z_{ij}\right) + \varepsilon_i\right] \right| \right| \\
&\qquad\qquad \left. - \left| \frac{1}{2}\left(a_0''(\xi_i)(X_i - x_0)^2 + \sum_{j=1}^{p} a_j''(\eta_{ij})(U_i - u_0)^2 Z_{ij}\right) + \varepsilon_i \right| \right\} \\
&\quad \times K_{1,h}(X_i - x_0)K_{2,h}(U_i - u_0).
\end{aligned}
\tag{3.2}
$$

The aim of this paper is to study the asymptotic behavior of $\sqrt{nh}(\widetilde{a}_0 - a_0(x_0))$ and $\sqrt{nh}(\widetilde{\mathbf{a}} - \mathbf{a}(u_0))$. Combining some technique reasons, we first introduce the new variables $\alpha_0 = \sqrt{nh^2}(a_0 - a_0(x_0))$, $\alpha = \sqrt{nh^2}(\mathbf{a} - \mathbf{a}(u_0))$, $\beta_0 = \sqrt{nh^2}[(b_0 - a_0'(x_0))h]$, and $\beta = \sqrt{nh^2}[(\mathbf{b} - \mathbf{b}(u_0))h]$ [16] and form a new equivalent problem as follows:

$$
\begin{aligned}
\left(\widehat{\alpha}_0, \widehat{\beta}_0, \widehat{\alpha}', \widehat{\beta}'\right)' &= \arg\min \sum_{i=1}^{n} \left\{ \left| \left| \left[\alpha_0 + \beta_0\left(\frac{X_i - x_0}{h}\right)\right]\frac{1}{\sqrt{nh^2}} + \left[\alpha'\mathbf{Z}_i + \beta'\left(\frac{U_i - u_0}{h}\right)\mathbf{Z}_i\right]\frac{1}{\sqrt{nh^2}} \right. \right. \right. \\
&\qquad\qquad \left. \left. - \left[\frac{1}{2}\left(a_0''(\xi_i)(X_i - x_0)^2 + \sum_{j=1}^{p} a_j''(\eta_{ij})(U_i - u_0)^2 Z_{ij}\right) + \varepsilon_i\right] \right| \right| \\
&\qquad\qquad \left. - \left| \frac{1}{2}\left(a_0''(\xi_i)(X_i - x_0)^2 + \sum_{j=1}^{p} a_j''(\eta_{ij})(U_i - u_0)^2 Z_{ij}\right) + \varepsilon_i \right| \right\} \\
&\quad \times K_{1,h}(X_i - x_0)K_{2,h}(U_i - u_0).
\end{aligned}
\tag{3.3}
$$

Let

$$F_n = E(S_n),$$

$$R_n = S_n - F_n + \sum_{i=1}^{n} \left\{ \left[\alpha_0 + \beta_0 \left(\frac{X_i - x_0}{h} \right) \right] \frac{1}{\sqrt{nh^2}} \right.$$

$$+ \left[\alpha' \mathbf{Z}_i + \beta' \left(\frac{U_i - u_0}{h} \right) \mathbf{Z}_i \right] \frac{1}{\sqrt{nh^2}} \left. \right\} K_{1,h}(X_i - x_0) \; K_{2,h}(U_i - u_0) \, \mathrm{sgn}(\varepsilon_i)$$

$$:= S_n - F_n + \sum_{i=1}^{n} L_{ni},$$

$$(3.4)$$

where S_n is the objective function of the equality above and $\mathrm{sgn}(\cdot)$ is the sign function.

Since the L_1 estimators have no closed forms, we first give the limit form of the function F_n, which is critical to obtain the asymptotic properties of the estimators.

Theorem 3.1. *Suppose Assumptions (1)–(7) hold, and $n \to \infty$, then for any fixed α_0, β_0, α, β, F_n converges to $F(\alpha_0, \beta_0, \alpha, \beta)$, which is defined as*

$$g(0 \mid x_0, u_0) f(x_0, u_0) \left\{ \left(\alpha_0^2 + 2\alpha_0\beta_0\mu_1(1) + \beta_0^2\mu_2(1) \right) \right.$$

$$+ \left(\alpha'\Omega\alpha + 2\mu_1(2)\alpha'\Omega\beta + \mu_2(2)\beta'\Omega\beta \right) + 2\alpha_0 \left(\alpha'\omega + \mu_1(2)\beta'\omega \right)$$

$$+ 2\beta_0\mu_1(1) \left(\alpha'\omega + \mu_1(2)\beta'\omega \right) \qquad (3.5)$$

$$- a''(x_0) \left(\alpha_0\mu_2(1) + \beta_0\mu_3(1) + \mu_2(1)\alpha'\omega + \mu_2(1)\mu_1(2)\beta'\omega \right)$$

$$\left. - \mathbf{c}(u_0)' \left(\alpha_0\mu_2(2)\omega + \beta_0\mu_1(1)\mu_2(2)\omega + \mu_2(2)\Omega\alpha + \mu_3(2)\Omega\beta \right) \right\}.$$

Remark 3.2. If the kernel functions $K_1(\cdot)$, $K_2(\cdot)$ are symmetric about zero and Lipschitz continuous, the limit form of $F(\alpha_0, \beta_0, \alpha, \beta)$ can be simplified as

$$F(\alpha_0, \beta_0, \alpha, \beta)$$

$$= g(0 \mid x_0, u_0) f(x_0, u_0) \left\{ \alpha_0^2 + \mu_2(1)\beta_0^2 + \alpha'\Omega\alpha + \mu_2(2)\beta'\Omega\beta \right.$$

$$\left. + 2\alpha_0\alpha'\omega - \mu_2(1)a''(x_0)(\alpha_0 + \alpha'\omega) - \mu_2(2)\mathbf{c}(u_0)'(\alpha_0\omega + \Omega\alpha) \right\}.$$

$$(3.6)$$

Now we are in the position to state the asymptotic properties of the estimators.

Theorem 3.3. *Suppose Assumptions (1)–(7) hold and $n \to \infty$, then one has*

$$\sqrt{nh}\left(\tilde{a}_0(x_0) - a_0(x_0) - \frac{h^2\left[\mu_2^2(1) - \mu_3(1)\mu_1(1)\right]a''(x_0)}{2\left[\mu_2(1) - \mu_1^2(1)\right]}\right) \xrightarrow{d} N\left(0, \sigma_{\alpha_0}^2\right), \tag{3.7}$$

$$\sqrt{nh}\left(\tilde{\mathbf{a}}(u_0) - \mathbf{a}(u_0) - \frac{h^2\left[\mu_2^2(2) - \mu_1(2)\mu_3(2)\right]\mathbf{c}(u_0)}{2\left[\mu_2(2) - \mu_1^2(2)\right]}\right) \xrightarrow{d} N_p(\mathbf{0}, \Sigma_\alpha), \tag{3.8}$$

where

$$\sigma_{\alpha_0}^2 = \left\{ \frac{(1 - \omega\prime\Omega^{-1}\omega)\gamma_0(1)r_2^2(x_0)}{4} + \frac{\mu_1^2(1)\left[\mu_1^2(1)\gamma_0(1) - 2\mu_1(1)\gamma_1(1) + \gamma_2(1)\right]r_2^2(x_0)}{4g^2(0 \mid x_0, u_0)f^2(x_0, u_0)\left[\mu_1^2(1) - \mu_2(1)\right]^2} \right.$$

$$\left. - \frac{(1 - \omega\prime\Omega^{-1}\omega)\mu_1(1)\left[\mu_1(1)\gamma_0(1) - \gamma_1(1)\right]r_2^2(x_0)}{2g(0 \mid x_0, u_0)f(x_0, u_0)\left[\mu_1^2(1) - \mu_2(1)\right]} \right\},$$

$$\Sigma_\alpha = \left\{ \frac{(\Omega - \omega\omega')^{-1}\gamma_0(2)r_1^2(u_0)}{4} + \frac{\mu_1^2(2)\left[\mu_1^2(2)\gamma_0(2) - 2\mu_1(2)\gamma_1(2) + \gamma_2(2)\right]r_1^2(u_0)\Omega^{-1}}{4g^2(0 \mid x_0, u_0)f^2(x_0, u_0)\left[\mu_1^2(2) - \mu_2(2)\right]^2} \right.$$

$$\left. - \frac{\mu_1(2)\left[\mu_1(2)\gamma_1(2) - \gamma_1(2)\right]r_1^2(u_0)}{2g(0 \mid x_0, u_0)f(x_0, u_0)\left[\mu_1^2(2) - \mu_2(2)\right]} \right\}$$

$$\tag{3.9}$$

with $r_1^2(u_0) = \int f_1^2(u)f(u, u_0)du$ and $r_2^2(x_0) = \int f_2^2(u)f(x_0, u)du$.

Remark 3.4. If the kernel functions $K_1(\cdot), K_2(\cdot)$ are symmetric about zero and Lipschitz continuous, the results of Theorem 3.3 can be simplified as

$$\sqrt{nh}\left(\tilde{a}_0(x_0) - a_0(x_0) - \frac{h^2\mu_2(1)a''(x_0)}{2}\right) \xrightarrow{d} N\left(0, \frac{(1 - \omega\prime\Omega^{-1}\omega)\ \gamma_0(1)r_2^2(x_0)}{4}\right),$$

$$\sqrt{nh}\left(\tilde{\mathbf{a}}(u_0) - \mathbf{a}(u_0) - \frac{h^2\mu_2(2)\mathbf{c}(u_0)}{2}\right) \xrightarrow{d} N_p\left(\mathbf{0}, \frac{(\Omega - \omega\omega')^{-1}\gamma_0(2)r_1^2(u_0)}{4}\right). \tag{3.10}$$

Remark 3.5. Here, we have considered estimation method and asymptotic distributions for the case that two bandwidths are same. It is important to note that similar asymptotic theories can be obtained for the case that two bandwidths with the same order are different.

Remark 3.6. If we consider different bandwidth h_1, h_2 for kernel functions K_1 and K_2. Furthermore, suppose the Assumptions (2) and (7) are replaced by $h_1^5 h_2 = o(1/n)$ and $h_1 h_2^5 = o(1/n)$, for example, $h_1 \sim n^{-1/6}, h_2 \sim n^{-1/5}$, and $\max \|\mathbf{Z}_i\| = o_p(\min\{h_2^{-2}, \sqrt{nh_1h_2}\})$, respectively. Similar results also will be obtained except that all the second-order derivatives in the results will disappear.

Remark 3.7. This paper restricts the study to one-dimensional variable X. The ideas used here can be adapted to higher dimensional variable X, for example, consider d-dimensional

variable X for the case of same bandwidths. Similar asymptotic distribution results for $\sqrt{nh^d}(\tilde{a}(x_0) - a(x_0))$ and $\sqrt{nh}(\tilde{\mathbf{a}}(u_0) - \mathbf{a}(u_0))$ can be obtained under the assumptions with Assumptions (2) and (7) being replaced by $h \sim n^{-1/(5+d)}$ and $\max\|\mathbf{Z}_i\| = o_p(n^{2/(5+d)})$, respectively.

4. Simulations

In this section, we carry out some simulations to illustrate the performance of L_1-method, and compare the performance of our L_1-method with that of the L_2-method. All the following simulations are conducted for sample size $n = 100$.

The following example is considered:

$$Y = a_0(X) + a_1(U)Z_1 + a_2(U)Z_2 + \varepsilon, \qquad (4.1)$$

where $a_0(x) = x + 4\exp(-8x^2)$, $a_1(u) = 2\sin(1.5\pi u + u)$, $a_2(u) = 4u(1-u)$, $(X,U)' \sim U([-0.5,0.5]^2)$, Z_1, Z_2 are normally distributed with correlation coefficient $1/\sqrt{2}$, the marginal distributions of Z_1 and Z_2 are standard normal, $\varepsilon \sim N(0,0.2^2)$, and ε, (X,U) and Z_1, Z_2 are mutually independent.

In each simulation, the L_1 estimators of $a_0(x)$, $a_1(u)$, $a_2(u)$ were computed by solving the minimization problem (2.5) and using the integrated method described in (2.6) and (2.7). We use the Epanechnikov kernel, $K(u) = (3/4)(1 - u^2)I_{|u|<1}$, for every $K_l(\cdot)$, $l = 1,2$. All bandwidths in a model are selected by the method proposed in Section 2.

To evaluate the asymptotic results given in Theorem 3.3, the quantile-quantile plots of the estimators are constructed. Figure 1 presents the quantile-quantile plots for $\tilde{a}_0(0), \tilde{a}_1(0), \tilde{a}_2(0)$ with sample size $n = 100$ and 100 replications, respectively, and these plots reveal that the asymptotic approximation is reasonable.

Figure 2 displays the true function curves of $a_0(x), a_1(u)$, and $a_2(u)$ and their estimated curves with sample size $n = 100$ and one replication. We can see from the figure that the L_1 estimates perform well.

In order to illustrate that the L_1 method is a robust method. Figure 3 displays the estimated curves with four outliers, that is, $(-0.4538, -0.4402, -10)$, $(0.13377, 0.1393, 20)$, $(0.2703, 0.3085, 20)$, $(0.4174, 0.4398, 18)$ for the three element array (x, u, y). From Figure 3, we can see that the L_1 estimate also has a good performance even in the presence of four large singular points of Y. The fact that outliers have little influence on the L_1 estimates is displayed in Figure 3, so it is a robust method.

By solving the following minimization problem:

$$\min \sum_{i=1}^{n} \left(Y_i - a_0 - b_0(X_i - x_0) - \mathbf{a}'\mathbf{Z}_i - \mathbf{b}'(U_i - u_0)\mathbf{Z}_i\right)^2 K_{1,h}(X_i - x_0)K_{2,h}(U_i - u_0), \qquad (4.2)$$

we can obtain similarly the L_2 estimators of the functions $a_0(x), a_1(u), a_2(u)$ by the equations (2.6) and (2.7). For comparing the L_1-method with the L_2 method. We simulated the function $a_0(x)$ by L_2 method and display the fitted curves with (without) outliers for sample sizes $n = 100$ and 1000 replications in Figure 4. We can see that L_2 estimate cannot perform well for the data sparsity and singularity, and three small outlying data points make the estimated curve by the L_2 method deviate from the true curve significantly. Combining Figures 2, 3, and 4,
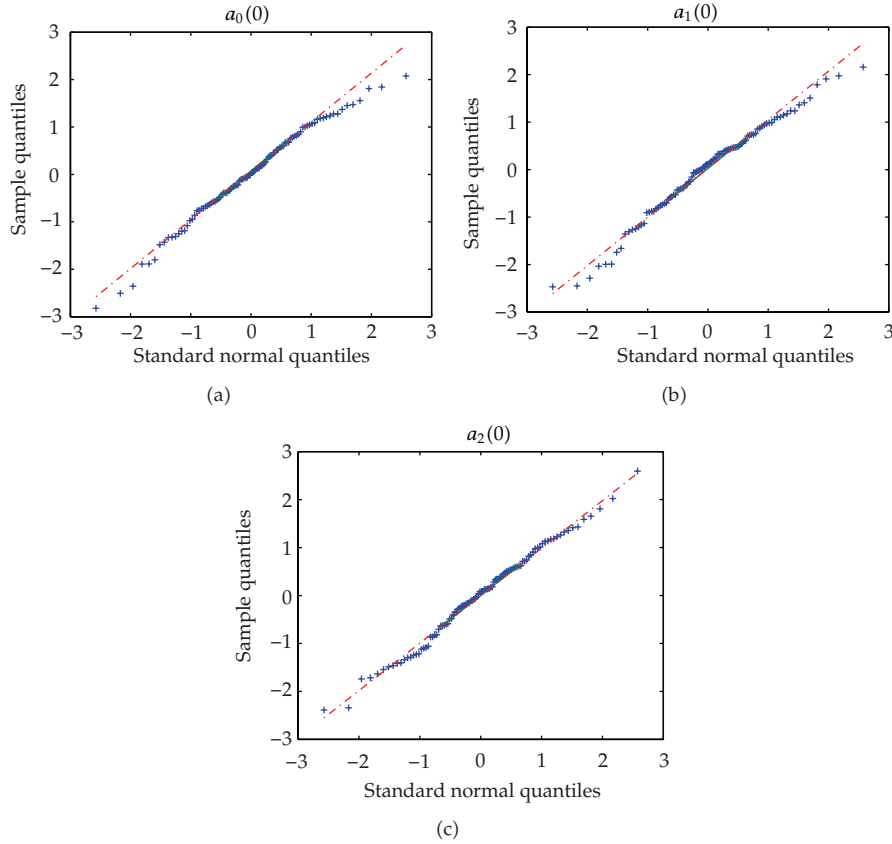
Figure 1: Normal QQ-plot for $\tilde{a}_0(0), \tilde{a}_1(0)$ and $\tilde{a}_2(0)$ $(n = 100)$.

we conclude that the L_1 method performs better than the L_2 method, the L_1 method is a robust method.

Finally, for further comparing the L_1-estimate with the L_2-estimate method, we also assess their performance via the weighted average squared error (WASE), which is defined as

$$\text{WASE} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{(a_0(x_i) - \hat{a}_0(x_i))^2}{(\text{range}(a_0))^2} + \frac{(a_1(u_i) - \hat{a}_1(u_i))^2}{(\text{range}(a_1))^2} + \frac{(a_2(u_i) - \hat{a}_2(u_i))^2}{(\text{range}(a_2))^2} \right), \tag{4.3}$$

where range (a_k) for $k = 0, 1, 2$ are the ranges of the functions $a_0(x)$, $a_1(u)$, and $a_2(u)$. The weights are introduced to account for the different scales of the functions. We conducted 200 replications with sample size $n = 100$. For the bandwidths and the Epanechnikov kernels used in the simulations, we obtain the mean and standard deviation of the WASE are 0.1201 and 0.0183 for the L_1 method, and 1.6613 and 0.8936 for the L_2 method. We can see that L_1 method outperforms L_2 method.
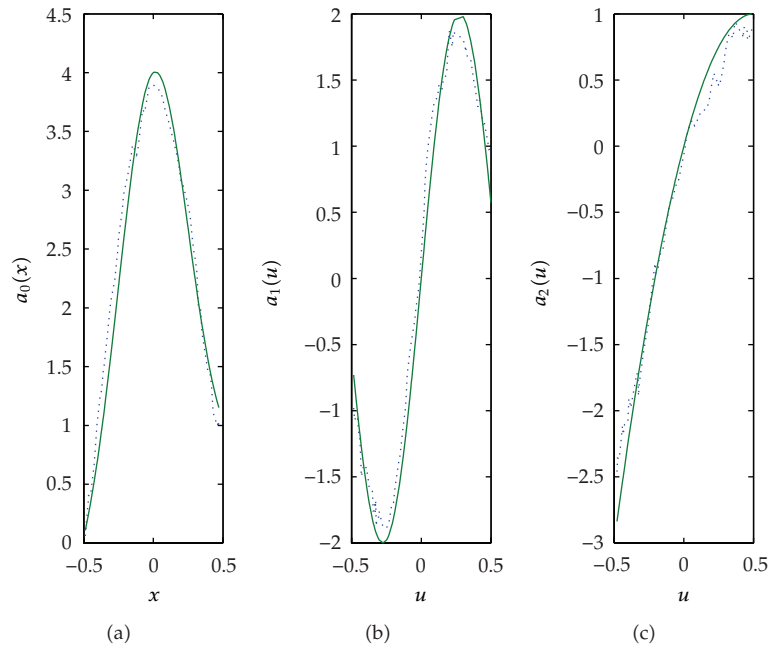
(a)   (b)   (c)

Figure 2: Simulation results, for example. Dotted curves are L_1 estimators; solid curve are the true curves.

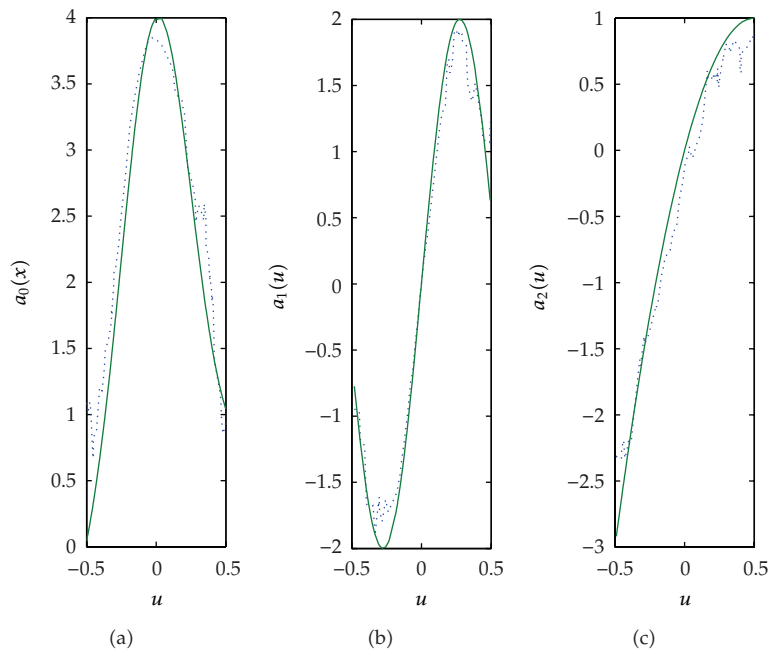(a)   (b)   (c)

Figure 3: Simulation results, for example, with outliers. Dotted curves are L_1 estimators; solid curve are the true curves.
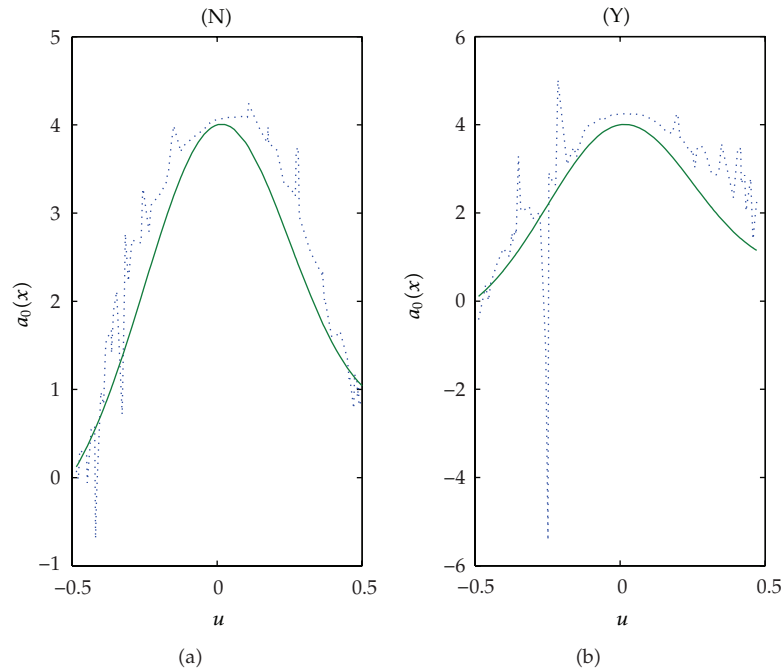
Figure 4: Simulation results for $\hat{a}_0(x)$. Dotted curves are L_2 estimators; solid curve are the true curves. (N) plot without outliers, (Y) plot with three outliers $(x, y) = (-0.4538, -5)$, $(0.1948, 8)$, and $(0.4649, 5)$.

5. Application

A real data is analyzed by the proposed L_1-method in this section. The classic gas furnace data was studied recently by Wong et al. [1]. The data set includes 296 samples (I_t, O_t), $(t = 1, \ldots, 296)$ measured at a fixed interval of 9 seconds, where I_t's represent the input gas rate in cubic feet per minute, and O_t's represent the concentration of carbon dioxide in the gas out of the furnace. Similar to the procedures of Wong et al. [1], the original data are transformed as $x_t = (I_t + 2.716)/5.55$, $y_t = (O_t - 45.6)/14.9$ for $t = 1, \ldots, 296$ such that both x's and y's are limited in the interval $[0, 1]$, and the model

$$y_t = a_0(x_{t-3}) + \sum_{j=1}^{4} a_j(y_{t-3}) y_{t-j} + e_t \tag{5.1}$$

is used to fit the data. The first 250 samples are used to establish the model, and the remained 46 samples are used for prediction.

In the proposed method, Epanechnikov kernel is used and all the bandwidths $h_j(j = 0, \ldots, 4)$ are selected as 0.14 via cross validation for simplicity. The mean absolute error (MAE) and the mean squared error (MSE) for fitting and forecasting are listed as follows.

Fitting: MAE = 0.3189, MSE = 0.1584; forecasting: MAE = 0.3705, MSE = 0.2127.

Since the model is chosen based on the L_2 errors, the results are not perfect as that in Wong et al. [1]. Moreover, the data set does not contain obvious outliers, so the advantage of the L_1 estimation method is not apparent. Compared to the results showed in Wong et al. [1], the difference between the fitting MAE/MSE and the forecasting MAE/MSE is small.
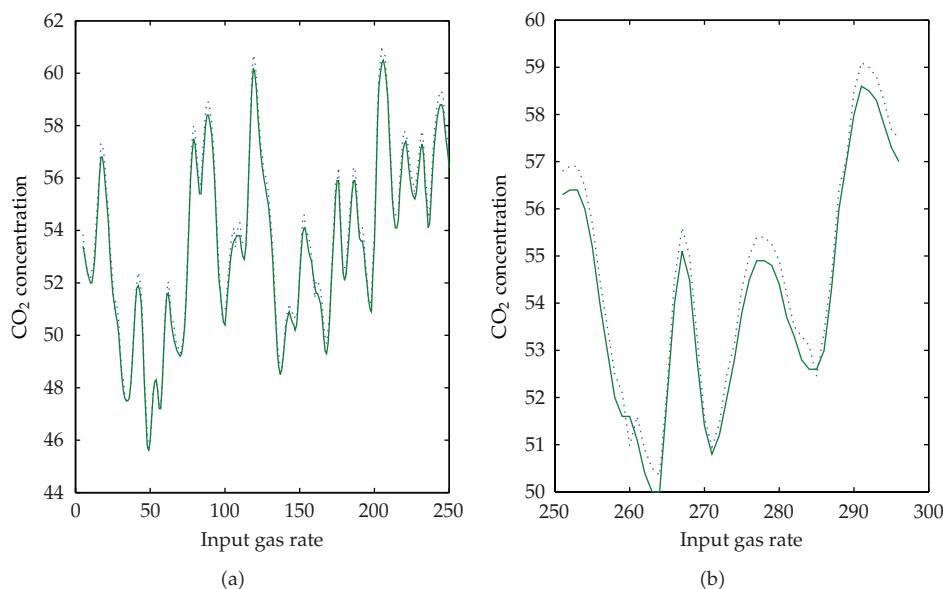
(a)   (b)

Figure 5: (a) Fitted values of $O_t(t = 5,\ldots,250)$ for gas data; (b) Predictive values $O_t(t = 251,\ldots,296)$ for gas data. Dotted lines are fitted or predicted; solid lines are observed.

The reason is that L_1-methods are employed in our method. The fitted values $\{\widehat{O}_t\}_{t=5}^{250}$ and predictive values $\{\widehat{O}_t\}_{t=251}^{296}$ are shown in Figure 5. These results indicate that the estimated results are reasonable.

6. Proofs of the Main Results

Before completing the proofs of the main results, we give the following useful lemma first.

Lemma 6.1. *Suppose Assumptions (1)–(7) hold, then for any fixed $\alpha_0, \beta_0, \alpha, \beta$, R_n converges to 0 in probability, that is, $R_n = o_p(1)$ as $n \to \infty$.*

Proof of Lemma 6.1. Let

$$
d_n = \max_{1 \le i \le n} \left\{ \left| \left[\alpha_0 + \beta_0 \left(\frac{X_i - x_0}{h} \right) \right] \frac{1}{\sqrt{nh^2}} + \left[\alpha' \mathbf{Z}_i + \beta' \left(\frac{U_i - u_0}{h} \right) \mathbf{Z}_i \right] \frac{1}{\sqrt{nh^2}} \right| \right.
$$

$$
\left. + \frac{1}{2} \left| a_0''(\xi_i)(X_i - x_0)^2 + \sum_{j=1}^{p} a_j''(\eta_{ij})(U_i - u_0)^2 Z_{ij} \right| \right\}.
\tag{6.1}
$$

Note that $|X_i - x_0| \le Ch$, $|U_i - u_0| \le Ch$ and Assumptions (2), (5), and (7), for any fixed $\alpha_0, \beta_0, \alpha$ and β, we have that

$$
d_n = o_p(1) \quad \text{as } n \longrightarrow \infty.
\tag{6.2}
$$

Let

$$
\begin{aligned}
T_{ni} = L_{ni} + &\left\{ \left| \left| \left[\alpha_0 + \beta_0 \left(\frac{X_i - x_0}{h} \right) \right] \frac{1}{\sqrt{nh^2}} + \left[\alpha' \mathbf{Z}_i + \beta' \left(\frac{U_i - u_0}{h} \right) \mathbf{Z}_i \right] \frac{1}{\sqrt{nh^2}} \right. \right. \right. \\
&\left. \left. - \left[\frac{1}{2} \left(a_0''(\xi_i)(X_i - x_0)^2 + \sum_{j=1}^{p} a_j''(\eta_{ij})(U_i - u_0)^2 Z_{ij} \right) + \varepsilon_i \right] \right| \right. \\
&\left. - \left| \frac{1}{2} \left(a_0''(\xi_i)(X_i - x_0)^2 + \sum_{j=1}^{p} a_j''(\eta_{ij})(U_i - u_0)^2 Z_{ij} \right) + \varepsilon_i \right| \right\} \\
&\times K_{1,h}(X_i - x_0) K_{2,h}(U_i - u_0).
\end{aligned}
\tag{6.3}
$$

It can be easily seen that $T_{ni} = 0$ if $|\varepsilon_i| \geq d_n$. Hence we have

$$
\begin{aligned}
|T_{ni}| \leq 2 &\left| \left\{ \left[\alpha_0 + \beta_0 \left(\frac{X_i - x_0}{h} \right) \right] + \left[\alpha' \mathbf{Z}_i + \beta' \left(\frac{U_i - u_0}{h} \right) \mathbf{Z}_i \right] \right\} \right| \frac{1}{\sqrt{nh^2}} \\
&\times K_{1,h}(X_i - x_0) K_{2,h}(U_i - u_0) I_{\{|\varepsilon_i| < d_n\}},
\end{aligned}
\tag{6.4}
$$

where $I(\cdot)$ is the indicator function. For any $\epsilon > 0, \delta > 0$, we have

$$
P\{|R_n| > \epsilon\} \leq P\{d_n \geq \delta\} + P\{I_{\{d_n < \delta\}} |R_n| > \epsilon\}.
\tag{6.5}
$$

Since $E(\text{sgn}(\varepsilon_i)) = 0$, we have

$$
E\left(I_{\{d_n < \delta\}} R_n^2 \right) = E\left\{ I_{\{d_n < \delta\}} (S_n - F_n + L_n)^2 \right\} = E\left\{ I_{\{d_n < \delta\}} \left(\sum_{i=1}^{n} (T_{ni} - E(T_{ni})) \right)^2 \right\},
\tag{6.6}
$$

here $L_n = \sum_{i=1}^{n} L_{ni}$. Combining (6.2) and (6.4), we have

$$
\begin{aligned}
\lim_{n \to \infty} E\left(I_{\{d_n < \delta\}} R_n^2\right) &= \lim_{n \to \infty} E\left\{ I_{\{d_n < \delta\}} \left[\sum_{i=1}^{n} E\left((T_{ni} - E(T_{ni}))^2 \mid \mathbf{X}, \mathbf{U}, \mathbf{Z} \right) \right.\right. \\
&\qquad\qquad \left.\left. + \sum_{i \neq j} E\left((T_{ni} - E(T_{ni}))(T_{nj} - E(T_{nj})) \mid \mathbf{X}, \mathbf{U}, \mathbf{Z} \right) \right] \right\} \\
&\leq 8 \lim_{n \to \infty} E\left\{ \left[\left(\alpha_0 + \beta_0 \left(\frac{X_i - x_0}{h} \right) \right)^2 + \left(\alpha' \mathbf{Z}_i + \beta' \left(\frac{U_i - u_0}{h} \right) \mathbf{Z_i} \right)^2 \right] \right. \\
&\qquad\qquad \left. \times \frac{1}{h^2} K_{1,h}^2(X_i - x_0) K_{2,h}^2(U_i - u_0) P\{|\varepsilon_i| \leq \delta \mid \mathbf{X}, \mathbf{U}\} \right\} \\
&\leq 16\delta g(0 \mid x_0, u_0)\, f(x_0, u_0) \Big[\gamma_0(2) \left(\alpha_0^2 \gamma_0(1) + 2\alpha_0 \beta_0 \gamma_1(1) + \beta_0^2 \gamma_2(1) \right) \\
&\qquad\qquad + \gamma_0(1)\left(\gamma_0(2) \alpha' \Omega \alpha \right. \\
&\qquad\qquad\qquad \left. + 2\gamma_2(1) \alpha' \Omega \beta + \gamma_2(2) \beta' \Omega \beta \right) + o(1) \Big] \\
&\longrightarrow 0,
\end{aligned}
\tag{6.7}
$$

when $\delta \to 0$. Combining (6.2), (6.5), and the argument $E(I_{\{d_n < \delta\}} R_n^2) \to 0$ as $\delta \to 0$ and $n \to \infty$, the desired conclusion follows by using the Chebyshev's inequality. This completes the proof of Lemma 6.1. $\qquad\square$

Proof of Theorem 3.1. Set

$$
\begin{aligned}
A(X_i, U_i, \mathbf{Z}_i) &= \left(\alpha_0 + \beta_0 \left(\frac{X_i - x_0}{h} \right) \right) \frac{1}{\sqrt{nh^2}} + \left(\alpha' \mathbf{Z}_i + \beta' \left(\frac{U_i - u_0}{h} \right) \mathbf{Z}_i \right) \frac{1}{\sqrt{nh^2}}, \\
B(X_i, U_i, \mathbf{Z}_i) &= \frac{1}{2} \left(a_0''(\xi_i)\, (X_i - x_0)^2 + \sum_{j=1}^{p} a_j''(\eta_{ij})(U_i - u_0)^2 Z_{ij} \right).
\end{aligned}
\tag{6.8}
$$

By Lemma 6.1, under Assumptions (1) – (7), $A(X_i, U_i, \mathbf{Z}_i)$ and $B(X_i, U_i, \mathbf{Z}_i)$ converge to zero as $n \to \infty$.

Start from the equality,

$$
F_n = I_{\{d_n < \delta\}} F_n + I_{\{d_n \geq \delta\}} F_n.
\tag{6.9}
$$

We first give the limit form of F_n, for fixed $\alpha_0, \beta_0, \alpha, \beta$, we have

$$
I_{\{d_n<\delta\}}F_n
$$

$$
= I_{\{d_n<\delta\}}E\left\{\sum_{i=1}^{n}|A(X_i,U_i,\mathbf{Z}_i) - B(X_i,U_i,\mathbf{Z}_i) - \varepsilon_i| - |B(X_i,U_i,\mathbf{Z}_i) + \varepsilon_i|\right\}
$$

$$
\times K_{1,h}(X_i - x_0)K_{2,h}(U_i - u_0)
$$

$$
= I_{\{d_n<\delta\}}nE\{|A(X_1,U_1,\mathbf{Z}_1) - B(X_1,U_1,\mathbf{Z}_1) - \varepsilon_1| - |B(X_1,U_1,\mathbf{Z}_1) + \varepsilon_1|\}
$$

$$
\times K_{1,h}(X_1 - x_0)K_{2,h}(U_1 - u_0)
$$

$$
= I_{\{d_n<\delta\}}nE\{E\{[|A(X_1,U_1,\mathbf{Z}_1) - B(X_1,U_1,\mathbf{Z}_1) - \varepsilon_1| - |B(X_1,U_1,\mathbf{Z}_1) + \varepsilon_1|]
$$

$$
\times K_{1,h}(X_1 - x_0)K_{2,h}(U_1 - u_0) \mid \mathbf{X},\mathbf{U},\mathbf{Z}\}\}
$$

$$
= I_{\{d_n<\delta\}}nE\left\{\int_{\delta>w>A(X_1,U_1,\mathbf{Z}_1)-B(X_1,U_1,\mathbf{Z}_1)}[w+B(X_1,U_1,\mathbf{Z}_1)-A(X_1,U_1,\mathbf{Z}_1)]g(w \mid X_1,U_1)dw\right.
$$

$$
+ \int_{-\delta<w<A(X_1,U_1,\mathbf{Z}_1)-B(X_1,U_1,\mathbf{Z}_1)}[A(X_1,U_1,\mathbf{Z}_1)-B(X_1,U_1,\mathbf{Z}_1)-w]g(w \mid X_1,U_1)dw
$$

$$
- \int_{\delta>w>-B(X_1,U_1,\mathbf{Z}_1)}[w + B(X_1,U_1,\mathbf{Z}_1)]g(w \mid X_1,U_1)dw
$$

$$
+ \int_{-\delta<w<-B(X_1,U_1,\mathbf{Z}_1)}[B(X_1,U_1,\mathbf{Z}_1) + w]g(w \mid X_1,U_1)\ dw
$$

$$
+ \int_{w>\delta}-g(w \mid X_1,U_1)A(X_1,U_1,\mathbf{Z}_1)dw
$$

$$
\left.+ \int_{w<-\delta}g(w \mid X_1,U_1)A(X_1,U_1,\mathbf{Z}_1)dw\right\}K_{1,h}(X_1 - x_0)K_{2,h}(U_1 - u_0).
$$

$$(6.10)$$

By the Integral Mean Value Theorem (refer to the Appendix), we have

$$
I_{\{d_n<\delta\}}F_n
$$

$$
= nI_{\{d_n<\delta\}}E\left\{g(\tau_1 \mid X_1,U_1)\right.
$$

$$
\times \left[\frac{\delta^2}{2}+(B(X_1,U_1,\mathbf{Z}_1)-A(X_1,U_1,\mathbf{Z}_1))\delta+\frac{(B(X_1,U_1,\mathbf{Z}_1) - A(X_1,U_1,\mathbf{Z}_1))^2}{2}\right]
$$

$$
+ g(\tau_2 \mid X_1,U_1)
$$

$$\times \left[\frac{\delta^2}{2} + (A(X_1, U_1, \mathbf{Z}_1) - B(X_1, U_1, \mathbf{Z}_1))\delta + \frac{(B(X_1, U_1, \mathbf{Z}_1) - A(X_1, U_1, \mathbf{Z}_1))^2}{2}\right]$$

$$+ A(X_1, U_1, \mathbf{Z}_1)[G(\delta \mid X_1, U_1) + G(-\delta \mid X_1, U_1) - 1]$$

$$- g(\tau_3 \mid X_1, U_1)\left[\frac{\delta^2}{2} + B(X_1, U_1, \mathbf{Z}_1)\delta + \frac{B^2(X_1, U_1, \mathbf{Z}_1)}{2}\right]$$

$$- g(\tau_4 \mid X_1, U_1)\left[\frac{\delta^2}{2} - B(X_1, U_1, \mathbf{Z}_1)\delta + \frac{B^2(X_1, U_1, \mathbf{Z}_1)}{2}\right]\Bigg\}$$

$$\times K_{1,h}(X_1 - x_0)K_{2,h}(U_1 - u_0),$$

$$(6.11)$$

where $G(\cdot | \cdot)$ is the conditional probability distribution function of ε, and $\tau_1, \tau_2, \tau_3, \tau_4$ converge to zero as $\delta \to 0$ and $n \to \infty$. Then by Assumptions (1)–(5), for any small enough $\delta > 0$, we have

$$I_{\{d_n < \delta\}} F_n = I_{\{d_n < \delta\}} n E\left\{\left(g(0 \mid X_1, U_1) + o(1)\right)\left(A^2(X_1, U_1, \mathbf{Z}_1) - 2A(X_1, U_1, \mathbf{Z}_1)B(X_1, U_1, \mathbf{Z}_1)\right)\right\}$$

$$\times K_{1,h}(X_1 - x_0)K_{2,h}(U_1 - u_0)$$

$$= I_{\{d_n < \delta\}}\Big(g(0 \mid x_0, u_0)f(x_0, u_0)$$

$$\times \left\{\left(\alpha_0^2 + 2\alpha_0\beta_0\mu_1(1) + \beta_0^2\mu_2(1)\right)\right.$$

$$+ (\alpha'\Omega\alpha + 2\mu_1(2)\alpha'\Omega\beta + \mu_2(2)\beta'\Omega\beta)$$

$$+ 2\alpha_0(\alpha'\omega + \mu_1(2)\beta'\omega) + 2\beta_0\mu_1(1)(\alpha'\omega + \mu_1(2)\beta'\omega)$$

$$- a''(x_0)(\alpha_0\mu_2(1) + \beta_0\mu_3(1) + \mu_2(1)\alpha'\omega + \mu_2(1)\mu_1(2)\beta'\omega)$$

$$- \mathbf{c}(u_0)'(\alpha_0\mu_2(2)\omega + \beta_0\mu_1(1)\mu_2(2)\omega$$

$$\left.+ \mu_2(2)\Omega\alpha + \mu_3(2)\Omega\beta\right\} + o(1)\Big)$$

$$= I_{\{d_n < \delta\}}\left(F(\alpha_0, \beta_0, \alpha, \beta) + o(1)\right),$$

$$F_n = F(\alpha_0, \beta_0, \alpha, \beta) + o(1) - \left[F(\alpha_0, \beta_0, \alpha, \beta) + o(1) - F_n\right]I_{\{d_n \geq \delta\}}.$$

$$(6.12)$$

Note that, for any fixed $\alpha_0, \beta_0, \alpha, \beta, d_n \to 0$ as $n \to \infty$, we draw the desired conclusion. This completes the proof of Theorem 3.1.   □

Proof of Theorem 3.3. Since the proofs of (3.7) and (3.8) are quite similar, we only give the proof of (3.8).

Start from the equality

$$L_n = \sum_{i=1}^{n} \left\{ \left[\alpha_0 + \beta_0 \left(\frac{X_i - x_0}{h} \right) \right] + \left[\alpha' \mathbf{Z}_i + \beta' \left(\frac{U_i - u_0}{h} \right) \mathbf{Z}_i \right] \right\} \frac{1}{\sqrt{nh^2}}$$

$$\times K_{1,h}(X_i - x_0) K_{2,h}(U_i - u_0) \, \text{sgn}(\varepsilon_i). \tag{6.13}$$

By Lemma 6.1 and Theorem 3.1, for fixed α_0, β_0, α, β, we have

$$S_n = F(\alpha_0, \beta_0, \alpha, \beta) - L_n + R_n + o(1) \triangleq F(\alpha_0, \beta_0, \alpha, \beta) - L_n + \overline{R}_n, \tag{6.14}$$

hence

$$S_n + L_n = F(\alpha_0, \beta_0, \alpha, \beta) + \overline{R}_n. \tag{6.15}$$

Note that

$$EL_n^2 = \sum_{i=1}^{n} EL_{ni}^2$$

$$\leq 2nE \left\{ \left(\alpha_0 + \beta_0 \left(\frac{X_i - x_0}{h} \right) \right)^2 \frac{1}{nh^2} K_{1,h}^2(X_i - x_0) K_{2,h}^2(U_i - u_0) \right.$$

$$\left. + \left(\alpha' \mathbf{Z}_i + \beta' \left(\frac{U_i - u_0}{h} \right) \mathbf{Z}_i \right)^2 \frac{1}{nh^2} K_{1,h}^2(X_i - x_0) \, K_{2,h}^2(U_i - u_0) \right\}$$

$$= 2 \left[\gamma_0(2) \left(\alpha_0^2 \gamma_0(1) + 2\alpha_0 \beta_0 \gamma_1(1) + 2\beta_0^2 \gamma_2(1) \right) + \gamma_0(1) \left(\alpha' \Omega \alpha + 2\gamma_1(1) \alpha' \Omega \beta + \gamma_2(2) \beta' \Omega \beta \right) \right]. \tag{6.16}$$

We obtain that L_n is bounded in probability for any fixed α_0, β_0, α, β. Thus, for any fixed $\alpha_0, \beta_0, \alpha, \beta$, the random convex function $S_n + L_n$ converges to $F(\alpha_0, \beta_0, \alpha, \beta)$. According to the *convexity lemma* [17], we can deduce that, for any compact set $\mathbb{K}$,

$$\sup_{(\alpha_0, \beta_0, \alpha', \beta')' \in \mathbb{K}} \left| \overline{R}_n \right| = o_p(1), \tag{6.17}$$

when $\delta \to 0$. By the proof of Theorem 2 in Wang [18], we obtain that the "limit" here is not only in the sense of the limit of a sequence of random variables, but also in the sense of the limit of a sequence of stochastic process, and the minimizer of S_n converges to the minimizers of $F(\alpha_0, \beta_0, \alpha, \beta) - L_n$.

By the convexity of the function $F(\alpha_0, \beta_0, \alpha, \beta)$, we have

$$
\widehat{\beta}_0 = \frac{\sum_{i=1}^{n} \left[\mu_1(1) - ((X_i - x_0)/h)\right] \left(1/\sqrt{nh^2}\right) K_{1,h}(X_i - x_0) K_{2,h}(U_i - u_0) \operatorname{sgn}(\varepsilon_i)}{2g(0 \mid x_0, u_0) f(x_0, u_0) \left[\mu_1^2(1) - \mu_2(1)\right]}
$$

$$
+ \frac{\left[\mu_2(1)\mu_1(1) - \mu_3(1)\right] a''(x_0)}{2\left[\mu_1^2(1) - \mu_2(1)\right]},
$$

$$
\widehat{\beta} = \frac{\Omega^{-1} \sum_{i=1}^{n} \left[\mu_1(2) - ((U_i - u_0)/h)\right] \left(\mathbf{Z}_i/\sqrt{nh^2}\right) K_{1,h}(X_i - x_0) K_{2,h}(U_i - u_0) \operatorname{sgn}(\varepsilon_i)}{2g(0 \mid x_0, u_0) f(x_0, u_0) \left[\mu_1^2(2) - \mu_2(2)\right]}
$$

$$
+ \frac{\left[\mu_2(2)\mu_1(2) - \mu_3(2)\right] \mathbf{c}(u_0)}{2\left[\mu_1^2(2) - \mu_2(2)\right]},
$$

$$
\widehat{\alpha} = \frac{\left[\mu_2^2(2) - \mu_1(2)\mu_3(2)\right] \mathbf{c}(u_0)}{2\left[\mu_2(2) - \mu_1^2(2)\right]}
$$

$$
+ \frac{1}{2\sqrt{nh^2}} \sum_{i=1}^{n} (\Omega - \omega\omega')^{-1} (\mathbf{Z}_i - \omega) K_{1,h}(X_i - x_0) K_{2,h}(U_i - u_0) \operatorname{sgn}(\varepsilon_i)
$$

$$
- \frac{\mu_1(2)\Omega^{-1} \sum_{i=1}^{n} \left[\mu_1(2) - ((U_i - u_0)/h)\right] \mathbf{Z}_i \left(1/\sqrt{nh^2}\right) K_{1,h}(X_i - x_0) K_{2,h}(U_i - u_0) \operatorname{sgn}(\varepsilon_i)}{2g(0 \mid x_0, u_0) f(x_0, u_0) \left[\mu_1^2(2) - \mu_2(2)\right]}
$$

$$
\widehat{\alpha}_0 = \frac{\left[\mu_2^2(1) - \mu_3(1)\mu_1(1)\right] a''(x_0)}{2\left[\mu_2(1) - \mu_1^2(1)\right]} + \frac{1}{2\sqrt{nh^2}} \sum_{i=1}^{n} \left(1 - \omega'\Omega^{-1}\mathbf{Z}_i\right) K_{1,h}(X_i - x_0) K_{2,h}(U_i - u_0) \operatorname{sgn}(\varepsilon_i)
$$

$$
- \frac{\mu_1(1) \sum_{i=1}^{n} \left[\mu_1(1) - ((X_i - x_0)/h)\right] \left(1/\sqrt{nh^2}\right) K_{1,h}(X_i - x_0) K_{2,h}(U_i - u_0) \operatorname{sgn}(\varepsilon_i)}{2g(0 \mid x_0, u_0) f(x_0, u_0) \left[\mu_1^2(1) - \mu_2(1)\right]}.
$$

$$(6.18)$$

Thus we obtain

$$
\frac{\widetilde{\alpha}(u_0)}{\sqrt{h}} = \frac{1}{n\sqrt{h}} \sum_{k=1}^{n} \widehat{\alpha}(X_k, u_0)
$$

$$
= \frac{1}{2nh\sqrt{nh}} \sum_{k=1}^{n} \sum_{i=1}^{n} (\Omega - \omega\omega')^{-1} (\mathbf{Z}_i - \omega) K_{1,h}(X_i - X_k) K_{2,h}(U_i - u_0) \operatorname{sgn}(\varepsilon_i)
$$

$$
- \frac{\mu_1(2)\Omega^{-1} \sum_{k=1}^{n} \sum_{i=1}^{n} \left[\mu_1(2) - ((U_i - u_0)/h)\right] \mathbf{Z}_i K_{1,h}(X_i - X_k) K_{2,h}(U_i - u_0) \operatorname{sgn}(\varepsilon_i)}{2nh\sqrt{nh}g(0 \mid x_0, u_0) f(x_0, u_0) \left[\mu_1^2(2) - \mu_2(2)\right]}
$$

$$
+ \frac{\left[\mu_2^2(2) - \mu_1(2)\mu_3(2)\right] \mathbf{c}(u_0)}{2\sqrt{h}\left[\mu_2(2) - \mu_1^2(2)\right]}
$$

$$
:= A_1 + \frac{\left[\mu_2^2(2) - \mu_1(2)\mu_3(2)\right] \mathbf{c}(u_0)}{2\sqrt{h}\left[\mu_2(2) - \mu_1^2(2)\right]}.
$$

$$(6.19)$$

By interchanging summation signs and noting that

$$\frac{1}{n}\sum_{k=1}^{n}K_{1,h}(X_i - X_k) = hf_1(X_i)\{1 + o_p(1)\} \tag{6.20}$$

A_1 can be rewritten as

$$A_1 = \frac{1}{2\sqrt{nh}}\sum_{i=1}^{n}(\Omega - \omega\omega')^{-1}(\mathbf{Z}_i - \omega)K_{2,h}(U_i - u_0)f_1(X_i)\operatorname{sgn}(\varepsilon_i)\{1 + o_p(1)\}$$

$$- \frac{\mu_1(2)\Omega^{-1}\sum_{i=1}^{n}\left[\mu_1(2) - ((U_i - u_0)/h)\right]\mathbf{Z}_i K_{2,h}(U_i - u_0)f_1(X_i)\operatorname{sgn}(\varepsilon_i)\{1 + o_p(1)\}}{2\sqrt{nh}g(0 \mid x_0, u_0)f(x_0, u_0)\left[\mu_1^2(2) - \mu_2(2)\right]}. \tag{6.21}$$

For all $\mathbf{t} \in R^p$, by some straightforward computations, we get that the mean and variance of $\mathbf{t}'A_1$ are

$$E(\mathbf{t}'A_1) = 0$$

$$E\left((\mathbf{t}'A_1)^2\right) = \mathbf{t}'\left\{\frac{(\Omega - \omega\omega')^{-1}\gamma_0(2)r_2^2(u_0)}{4}\right.$$

$$+ \frac{\mu_1^2(2)\left[\mu_1^2(2)\gamma_0(2) - 2\mu_1(2)\gamma_1(2) + \gamma_2(2)\right]r_1^2(u_0)\Omega^{-1}}{4g^2(0 \mid x_0, u_0)f^2(x_0, u_0)\left[\mu_1^2(2) - \mu_2(2)\right]^2} \tag{6.22}$$

$$\left. - \frac{\mu_1(2)\left[\mu_1(2)\gamma_1(2) - \gamma_1(2)\right]r_1^2(u_0)}{2g(0 \mid x_0, u_0)f(x_0, u_0)\left[\mu_1^2(2) - \mu_2(2)\right]}\right\}\mathbf{t}\{1 + o_p(1)\}.$$

By Assumptions (1)–(7), and using the methods used in the proof of Theorem 1 in Tang and Wang [19], we can easily check that the Lindeberg-Feller's condition for $\mathbf{t}'A_1$ is held. So Cramer-Wold theorem tells us (3.8) follows. This completes the proof of Theorem 3.3. □

Appendix

Integral Mean Value Theorem

If f and g are continuous functions defined on $[a, b]$, and $g(x) \geq 0$ (or $g(x) \leq 0$), then there exists a number $\xi \in [a, b]$ such that
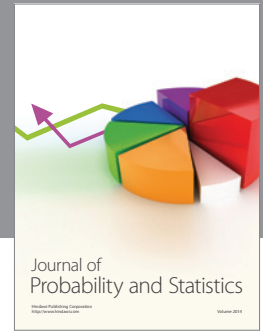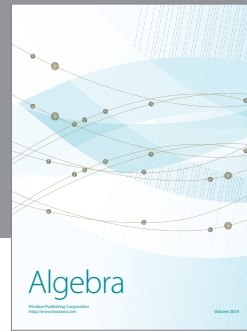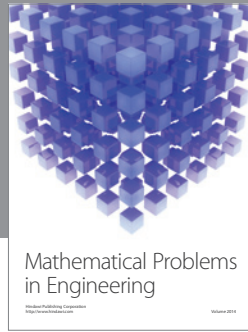
$$\int_a^b f(x)g(x)\mathrm{d}x = f(\xi)\int_a^b g(x)\mathrm{d}x. \tag{A.1}$$

Acknowledgments

References

[1] H. Wong, R. Zhang, W.-c. Ip, and G. Li, "Functional-coefficient partially linear regression model," *Journal of Multivariate Analysis*, vol. 99, no. 2, pp. 278–305, 2008.

[2] H. Zhu, P. J. Brown, and J. S. Morris, "Robust, adaptive functional regression in functional mixed model framework," *Journal of the American Statistical Association*, vol. 106, no. 495, pp. 1167–1179, 2011.

[3] G. Aneiros-Pérez and P. Vieu, "Nonparametric time series prediction: a semi-functional partial linear modeling," *Journal of Multivariate Analysis*, vol. 99, no. 5, pp. 834–857, 2008.

[4] I. Ahmad, S. Leelahanon, and Q. Li, "Efficient estimation of a semiparametric partially linear varying coefficient model," *The Annals of Statistics*, vol. 33, no. 1, pp. 258–283, 2005.

[5] W. S. Cleveland, E. Grosse, and W. M. Shyu, "Local regression models," in *Statistical Models in S*, J. M. Chambers and T. J. Hastie, Eds., pp. 309–376, Wadsworth and Brooks, Pacific Grove, Calif, USA, 1992.

[6] D. R. Hoover, J. A. Rice, C. O. Wu, and L.-P. Yang, "Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data," *Biometrika*, vol. 85, no. 4, pp. 809–822, 1998.

[7] C. O. Wu, C.-T. Chiang, and D. R. Hoover, "Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data," *Journal of the American Statistical Association*, vol. 93, no. 444, pp. 1388–1402, 1998.

[8] T. Hastie and R. Tibshirani, "Varying-coefficient models," *Journal of the Royal Statistical Society Series B*, vol. 55, no. 4, pp. 757–796, 1993, With discussion and a reply by the authors.

[9] C.-T. Chiang, J. A. Rice, and C. O. Wu, "Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables," *Journal of the American Statistical Association*, vol. 96, no. 454, pp. 605–619, 2001.

[10] J. Fan and W. Zhang, "Statistical estimation in varying coefficient models," *The Annals of Statistics*, vol. 27, no. 5, pp. 1491–1518, 1999.

[11] J. Fan and J.-T. Zhang, "Two-step estimation of functional linear models with applications to longitudinal data," *Journal of the Royal Statistical Society Series B*, vol. 62, no. 2, pp. 303–322, 2000.

[12] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*, vol. 66 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London, UK, 1996.

[13] F. T. Wang and D. W. Scott, "The L_1 method for robust nonparametric regression," *Journal of the American Statistical Association*, vol. 89, no. 425, pp. 65–76, 1994.

[14] M. S. Bazaraa and C. M. Shetty, *Nonlinear Programming*, John Wiley & Sons, New York, NY, USA, 1979, Theory and Algorithms.

[15] J. A. Rice and B. W. Silverman, "Estimating the mean and covariance structure nonparametrically when the data are curves," *Journal of the Royal Statistical Society Series B*, vol. 53, no. 1, pp. 233–243, 1991.

[16] B. L. S. Prakasa Rao, *Asymptotic Theory of Statistical Inference*, John Wiley & Sons, New York, NY, USA, 1987.

[17] D. Pollard, "Asymptotics for least absolute deviation regression estimators," *Econometric Theory*, vol. 7, no. 2, pp. 186–199, 1991.

[18] J. D. Wang, "Asymptotic normality of L_1-estimators in nonlinear regression," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 227–238, 1995.

[19] Q. Tang and J. Wang, "L_1-estimation for varying coefficient models," *Statistics*, vol. 39, no. 5, pp. 389–404, 2005.