Functional-coefficient quantile regression with nonstationary time series

Han-Ying Liang[a], Yu Shen[a], Qiying Wang[b]

[a]School of Mathematical Science, Tongji University, Shanghai 200092, P. R. China

E-mail: hyliang@tongji.edu.cn, ipqnojug@gmail.com

[b]School of Mathematics and Statistics, The University of Sydney, Sydney, NSW 2006, Australia

E-mail: qiying@maths.usyd.edu.au

Abstract. This paper explores kernel and local linear quantile estimation for a functional-coefficient regression model with nonstationary time series as regressor. Our main results are established by allowing for the heavy-tailed distributional assumption in the error term, which enables the quantile approach applicable in econometrics and many other fields where outliers and aberrant observations are at present. Our main results further indicate that the linear term in kernel quantile estimator can not be eliminated from the asymptotic bias. This feature is different from the previous researches on nonlinear regression with nonstationary time series, where the conventional kernel estimator is shown to have the same limit distribution (to the second order including bias) as the local linear nonparametric estimator. Simulation result shows good performance for the proposed estimators as predicted by our asymptotic theory. An empirical application for the monthly road casualties in Great Britain has also been considered.

Key words and phrases: Functional-coefficient regression, Kernel quantile smoothing, local linear quantile smoothing, nonstationary time series.

AMS Subject Classifications: Primary: 62M10, 62G07; Secondary 60F05.

JEL Classification: C14, C22.

1 Introduction

Since the initial work by Koenker and Bassett (1978), quantile estimation in a regression model has been successfully and widely used in finance and economics. Estimation of conditional quantiles nowadays is a common practice in risk management, portfolio optimization, and asset pricing. Asymptotic theory underlying quantile estimators for many commonly used models has been well established for independent and identically distributed (iid) data as well as for weakly dependent data. We refer to Koenker (2005), Cai, Gu and Li (2009) and articles therein for current development. In comparison with the extensive researches on quantile estimation with stationary data, little is known about the behaviors with nonstationary time series. The early contributions on quantile estimation with nonstationary time series include Xiao (2009a, 2009b), Chen, Li and Zhang (2010) and Honda (2013). Xiao (2009a) considered quantile cointegrating regression, while the others investigated the least absolute deviation estimation for nonlinear regression with nonstationary time series. More currently, Li and Li (2015) considered local composite quantile regression smoothing for Harries Recurrent Markov processes.

This paper considers quantile estimation in a more general model with nonstationary time series. Explicitly, we focus on the varying coefficient regression model having the form:

$$y_t = x_t^T \beta_0(z_t) + \sigma(x_t, z_t)\, \epsilon_t, \tag{1.1}$$

where y_t, z_t and ϵ_t are all scalars, x_t is of dimension d, $\beta_0(\cdot)$ is a $d \times 1$ vector of unknown smooth function and A^T denotes the transpose of a vector or a matrix A. We will investigate the quantile estimator of $\beta_0(\cdot)$ under the conditions:

- x_t is stationary and z_t is an $I(1)$ process.

There are extensive researches for the quantile estimator of $\beta_0(\cdot)$ under the assumption that both x_t and z_t are stationary processes. See, for instance, Honda (2004), Kim (2007), Cai and Xu (2008), Cai, Gu and Li (2009) and references therein. Xiao (2009a) considered the situation that x_t is an $I(1)$ process, $\beta_0(z_t) \equiv \beta_0$ and $\sigma(x_t, z_t) \equiv 1$. The situation for both x_t and z_t being non-stationary time series seems to be difficult and requires very different techniques. We will leave the topic for future work.

Model (1.1) under the setting in this paper is becoming increasedly popular due to its flexibility. The proposed model includes the nonlinear cointegrating regression model which was currently developed by Wang and Phillips (2009a, 2009b, 2011, 2012, 2015), Wang (2014) and Wang (2015). As in the regression model with stationary data, $\beta_0(\cdot)$ can be estimated by using standard kernel and local linear method. When $E\epsilon_t = 0$ and ϵ_t satisfies certain moment conditions, the asymptotics for the local linear estimator of $\beta_0(\cdot)$ has been considered in Cai, Li and Park (2009). Also see Xiao (2009b), Gao and Phillips (2013) and Sun, Cai and Li (2013, 2015). Unlike the mean regression method in existing literature that relies only on the central tendency of the data, the quantile approach in this paper allows to estimate the functional dependence between variables for all portions of the conditional distribution of the response variable, which extends the framework of estimating only the behavior of the central part of a cloud of data points onto all parts of the conditional distribution. As a consequence, the quantile approach provides a more complete view of relationships between variables of interest. Furthermore, our asymptotic results allow for modeling data with heterogeneous conditional distributions and makes no distributional assumption about the error term in the model. These features enable the quantile approach in this paper useful since outliers or aberrant observations are common in nonstationary time series data from econometrics and many other applied areas, and heavy-tailed distribution is an important feature of some nonstationary time series data from finance.

This paper is organized as follows. In Section 2, we construct kernel quantile smoothing estimators and local linear (LL) quantile estimators of $\beta_0(\cdot)$ in the model (1.1), and investigate asymptotic distributions of the proposed estimators. Simulation result showing good performance for the proposed estimators as predicted by asymptotic theory is presented in Section 3. Section 4 provides an empirical application. Proofs of the main results are put in Section 5. Some preliminary lemmas and proof of auxiliary results are collected in Appendix (Section 6).

Throughout the paper, we make use of the following notation: for $x \in R^d$, $||x|| = \sum_{j=1}^{d} |x_j|$. We denote constants by $C, C_1, ...$, which may be different at each appearance.

2 Quantile regression estimator

Let $\tau \in (0, 1)$ and $z \in R$ be fixed. Throughout the paper, we assume that the τth conditional quantile of ϵ_t given $\mathcal{F}_t = \sigma(x_s, z_s, s \le t)$ equals zero in model (1.1), namely,

A1. $P(\epsilon_t < 0 \mid \mathcal{F}_t) = \tau$ for all $t \ge 1$.

As in standard kernel estimation where one usually assumes that $E(\epsilon_t \mid \mathcal{F}_t) = 0$, condition A1 is crucial for the construction of an unbiased quantile estimator in model (1.1). Let

$$Q_{y_t}(\tau|x_t, z_t) = \inf\{y : F_0(y|x_t, z_t) \ge \tau\}$$

be the τth conditional quantile of y_t given x_t and z_t, where $F_0(y|x_t, z_t)$ is the conditional distribution function of y_t given x_t and z_t. Due to Condition A1, $Q_{y_t}(\tau|x_t, z_t) = x_t^T \beta_0(z_t)$. Supposing that $\beta_0(x)$ is locally approximated by a constant vector $\beta \equiv \beta_0(z) \in R^d$ for x in a neighborhood of z, a kernel quantile estimator of $\beta_0(z)$ in model (1.1) can be obtained by solving the problem

$$\widehat{\beta}_n(z) = \arg\min_{\beta \in R^d} \sum_{t=1}^{n} \rho_\tau(y_t - x_t^T \beta) K\Big(\frac{z_t - z}{h}\Big), \tag{2.1}$$

where $\rho_\tau(t) = t[\tau - I(t < 0)]$ is called the "check" (loss) function, $I(A)$ is an indicator function of set A, $0 < h \equiv h_n \to 0$ is a bandwidth and $K(x)$ is a positive kernel function. Similarly, if $\beta_0(x)$ can be locally approximated by

$$\beta_0(x) \approx \beta_0(z) + \beta_0'(z)(x - z) \equiv \alpha_0 + \alpha_1(x - z)$$

for x in a neighborhood of z, an estimator $\widehat{\beta}_L(z)$ of $\beta_0(z)$ is $\widehat{\alpha}_0$, where $(\widehat{\alpha}_0, \widehat{\alpha}_1)$ is the minimizer of

$$\sum_{t=1}^{n} \rho_\tau \Big\{ y_t - x_t^T [\alpha_0 + \alpha_1 (z_t - z)] \Big\} K\Big(\frac{z_t - z}{h}\Big). \tag{2.2}$$

$\widehat{\beta}_L(z)$ is called a local linear quantile estimator of $\beta_0(z)$.

This paper will investigate the asymptotic normalities of $\widehat{\beta}_n(z)$ and $\widehat{\beta}_L(z)$. To this end, let $\eta_j, j = 0, \pm 1, \pm 2, \cdots$ be a sequence of i.i.d. random variables with $\mathrm{E}\eta_0 = 0$, $\mathrm{E}\eta_0^2 = 1$ and $|Ee^{it\eta_0}| \le t^{-\delta}$ for some $\delta > 0$. Let $\xi_j, j \ge 1$, be a linear process defined by

$$\xi_j = \sum_{k=0}^{\infty} \phi_k \, \eta_{j-k},$$

where the coefficients $\phi_k, k \ge 0$, satisfy one of the following conditions:

LM. $\phi_k \sim k^{-\mu} \rho(k)$, where $1/2 < \mu < 1$ and $\rho(k)$ is a function slowly varying at ∞.

SM. $\sum_{k=0}^{\infty} |\phi_k| < \infty$ and $\phi \equiv \sum_{k=0}^{\infty} \phi_k \ne 0$.

To establish the asymptotics of $\widehat{\beta}(z)$ and $\widehat{\beta}_L(z)$, we need the following restrictions on x_t, z_t, ϵ_t, $\beta_0(z)$, $\sigma(x, z)$ and the kernel function $K(x)$, where z is a given constant on R.

A2. (i) $z_t = \gamma \, z_{t-1} + \xi_t$, where $c \ge 0$ is a constant, $\gamma = 1 - c/n$ and $z_0 = o_P(\sqrt{n})$;

(ii) x_t is a stationary random vector independent of η_s for all $s \le t - m_0$ and some $m_0 \ge 0$.

A3. Let $f_t(x) = F_t'(x)$, where $F_t(x) = P(\epsilon_t < x | \mathcal{F}_t)$. Assume that $f_t(0) \equiv f(0) > 0$, $f_t'(0) \equiv f'(0)$ and

(i) $|f_t(s) - f_t(0)| \le C \, \min\{|s|^\lambda, 1\}$; or

(ii) $|f_t(s) - f_t(0) - f_t'(0) s| \le C \, \min\{|s|^{1+\lambda}, 1\}$

for some $0 < \lambda \le 1$ and for all $s \in R$.

A4. For x in a neighborhood of z,

(i) $||\beta_0(x) - \beta_0(z)|| \le C|x - z|$; or

(ii) $||\beta_0(x) - \beta_0(z) - \beta_0'(z)(x - z)|| \le C|x - z|^2$; or

(iii) $||\beta_0(x) - \beta_0(z) - \beta_0'(z)(x - z) - \frac{1}{2}\beta_0''(z)(x - z)^2|| \le C|x - z|^{2+\lambda}$, for some $0 < \lambda \le 1$.

A5. There exist $k_0 > 0$ and $0 < \lambda \leq 1$ such that, for t in a neighborhood of z and all $x \in R^d$,

(i) $|\sigma^{-1}(x, z+t) - \sigma^{-1}(x, z)| \leq C(1 + ||x||^{k_0})\, |t|^{\lambda}$; or

(ii) $|\sigma^{-1}(x, z+t) - \sigma^{-1}(x, z) - \sigma_1(x, z)t| \leq C(1 + ||x||^{k_0})\, |t|^{1+\lambda}$, where $|\sigma_1(x, z)| \leq C(1 + ||x||^{k_0})$.

A6. (i) $K(\cdot)$ is a bounded kernel function with $\int_{-\infty}^{\infty} K(x)dx = 1$ and a compact support;

(ii) $\int_{-\infty}^{\infty} xK(x)dx = 0$.

A7. $Ex_1x_1^T > 0$, $E\big[\sigma^{-1}(x_1, z)\, x_1x_1^T\big] > 0$ and

$$E\Big\{\big[\sigma^{-1}(x_1, z) + ||x_1||^{k_0} + 1\big]\,(||x_1||^2 + 1)\Big\}^3 < \infty$$

where k_0 is given as in condition **A5**.

Remark 2.1 *Condition* **A2** *is quite general, which allows for nearly integrated long and short memory regressors. The* m_0 *in* **A2** *(ii) can be chosen as large (but not depending on* n*) as required and the independence between* x_t *and* $\eta_s, s \leq t$*, can be eliminated if* x_t *has certain structure. More details can be found in Remark 2.2. Due to the model (1.1), condition* **A3** *on the distribution function of* ϵ_t *is natural. In many application,* ϵ_t *is independent of* x_t *and* z_t*, implying* $F_t(x) = P(\epsilon_t \leq x)$*. As a consequence, condition* **A3** *is satisfied if only* ϵ_t *is stationary, together with certain smoothing conditions on the distribution function of* ϵ_t*. It should be mentioned that no moments are imposed on the distribution of* ϵ_t*, which makes the quantile regression a big advantage. Conditions* **A4**–**A6** *are minor smooth conditions on the regression and kernel functions. In particular, if* $\sigma(x, z) = \sigma(x)$ *as in most of practical applications,* **A5** *holds automatically with* $\sigma_1(x, z) = 0$*.* **A6** *can be extended to include the normal kernel, but requiring more other restriction on* x_t *and* z_t*. We omit the details.*

Remark 2.2 *We do not impose extra restrictions on* x_t *except stationarity and the independence between* x_t *and* $\eta_s, s \leq t - m_0$ *for some* $m_0 \geq 0$*. When* x_t *has certain structure,* m_0 *may be chosen to be zero. As an illustration, let* (η_t, ν_t) *be a sequence of iid random vectors. If* $z_t = z_{t-1} + \eta_t$ *and* $x_t = \sum_{j=0}^{\infty} \phi_j \nu_{t-j}$*, where* $\sum_{j=0}^{\infty} |\phi_j| < \infty$*, rough calculations show that it is possible to establish a result without* **A3** *(i). However, due to its complexity, detailed development requires new limit theorems, and hence leaves for future work.*

We next state our main result. For the convenience of notation, write

$$K_j(x) = x^j K(x), \quad \mu_j = \int_{-\infty}^{\infty} K_j(x)dx, \quad \text{for } j \geq 0,$$

$c_0 = \tau(1-\tau)\int_{-\infty}^{\infty} K^2(x)dx$, $d_n^2 = \mathrm{Var}\big(\sum_{j=1}^n \xi_j\big)$ and we further write

$$\begin{aligned}\Lambda &= f(0)E\big[\sigma^{-1}(x_1, z)\, x_1 x_1^T\big], \quad \Lambda_1 = f'(0)E\big\{\sigma^{-2}(x_1, z)\, x_1 \left[x_1^T \beta_0'(z)\right]^2\big\}, \\ \Lambda_2 &= f(0)E\big[\sigma_1(x_1, z)\, x_1 x_1^T\big].\end{aligned}$$

Note that Λ and Λ_2 are $d \times d$ matrixes and Λ_1 ia a $d \times 1$ vector.

Theorem 2.1 *Suppose that A1–A2 and A7 hold and $h \equiv h_n > 0$ satisfying $nh/d_n \to \infty$.*

(a) If in addition part (i) in A3–A6 and $nh^3/d_n \to 0$, then

$$\Big[\sum_{t=1}^n K\Big(\frac{z_t - z}{h}\Big)\Big]^{1/2} \Big[\widehat{\beta}_n(z) - \beta_0(z)\Big] \to_D c_0 \Lambda^{-1}\, \mathbb{N}, \tag{2.3}$$

where $\mathbb{N}$ is a d-dimensional normal vector with mean zero and covariance $\Omega = E x_1 x_1^T$.

(b) If in addition A3(i), A4(ii), A5(i) and A6, result (2.3) holds whenever $nh^5/d_n \to 0$.

(c) If in addition A3(ii), A4(iii), A5(ii), A6 and $nh^{5+\delta}/d_n \to 0$ for some $\delta > 0$, then

$$\Big[\sum_{t=1}^n K\Big(\frac{z_t - z}{h}\Big)\Big]^{1/2} \Big[\widehat{\beta}_n(z) - \beta_0(z) - \frac{\mu_2 h^2}{2}\, \Lambda^{-1}\, \alpha\Big] \to_D c_0 \Lambda^{-1}\, \mathbb{N} \tag{2.4}$$

where $\alpha = \Lambda\, \beta_0''(z) - \Lambda_1 + 2\Lambda_2 \beta_0'(z)$.

Remark 2.3 *Theorem 2.1 indicates that better asymptotic results can be achieved if strong smooth conditions on $\beta_0(x)$, $\sigma(x, z)$ and $K(x)$ are used, which is matching with the empirical applications. Theorem 2.1 allows for the x_t to be a sequence of deterministic constants. In particular, when $d = 1$ and $x_t \equiv 1$, (1.1) reduces to the nonlinear cointegrating regression model considered in Wang and Phillips (2009a, 2009b, 2011, 2015) and Wang (2014, 2015), where authors investigated the asymptotics of the conventional kernel estimator and the local linear estimator for $\beta_0(z)$. In comparison with these existing papers, the quantile regression approach in Theorem 2.1 allows for the heavy-tailed distributional assumption in the error term, which is important in econometrics since outliers or aberrant observations are common in nonstationary time series data.*

Similar results exist for the local linear quantile estimator generated from (2.2).

Theorem 2.2 *Suppose that A1–A2 and A6–A7 hold and $h \equiv h_n > 0$ satisfying $nh/d_n \to \infty$.*

(a) If in addition part (i) in A3–A5 and $nh^3/d_n \to 0$, then

$$\Big[\sum_{t=1}^n K\Big(\frac{z_t - z}{h}\Big)\Big]^{1/2} \Big[\widehat{\beta}_L(z) - \beta_0(z)\Big] \to_D c_0 \Lambda^{-1}\, \mathbb{N}, \tag{2.5}$$

where $\mathbb{N}$ *is a d-dimensional normal vector with mean zero and covariance* $\Omega = Ex_1x_1^T$.

(b) If in addition A3(i), A4(ii) and A5(i), result (2.5) holds whenever $nh^5/d_n \to 0$.

(c) If in addition A3(i), A4(iii) and A5(i), $nh^{5+\delta}/d_n \to 0$ *for some* $\delta > 0$*, then*

$$\Big[\sum_{t=1}^{n} K\Big(\frac{z_t - z}{h}\Big)\Big]^{1/2}\Big[\widehat{\beta}_L(z) - \beta_0(z) - \frac{\mu_2\, h^2}{2}\,\beta_0''(z)\Big] \to_D c_0\Lambda^{-1}\,\mathbb{N}. \tag{2.6}$$

Remark 2.4 *It follows from (2.4) and (2.6) that, in comparison with the local linear quantile estimator* $\widehat{\beta}_L(z)$*, the limit distribution of the kernel quantile estimator* $\widehat{\beta}_n(z)$ *has an extra asymptotic bias term*

$$\frac{1}{2}\big[-\Lambda_1 + 2\Lambda_2\beta_0'(z)\big]\,\Lambda^{-1}\mu_2 h^2.$$

As a consequence, as in stationary situation, the local linear quantile estimator $\widehat{\beta}_L(z)$ *of* $\beta_0(z)$ *generated from (2.2) has its advantage in reducing bias. This feature is different from nonlinear regression with nonstationary time series, where previous researches shown that the conventional kernel estimator has the same limit distribution (to the second order including bias) as the local linear nonparametric estimator.*

3 Simulation Study

In this section, we investigate the finite sample performance of the proposed estimators $\widehat{\beta}_n(z)$ and $\widehat{\beta}_L(z)$ of $\beta_0(z)$ through Monte Carlo simulation. The observed data are generated by the following varying-coefficients model:

$$Y_t = b_1(Z_t)X_{t1} + b_2(Z_t)X_{t2} + \sigma(X_t, Z_t)\epsilon_t,\ \ t = 1, 2, \ldots, n, \tag{3.1}$$

where $b_1(Z_t) = \sin(\pi Z_t)$, $b_2(Z_t) = \exp(-Z_t^2) + 1$, $\sigma(X_t, Z_t) = 0.5[1 + 0.5\sin(\pi X_{t1}Z_t)]$, and X_{t1} and X_{t2} are independent and from $U[0, 1]$. For non-stationary time series Z_t, let

$$Z_t = Z_{t-1} + \xi_t \quad \text{with} \quad \xi_t = \rho\xi_{t-1} + e_t$$

and e_t be a sequence of i.i.d. standard normal random variables. Thus $\beta_0(z) = (b_1(z), b_2(z))^T$. For simplicity, we just take $Z_0 = 0$ and $\xi_0 = 0$. We choose the following four different distributions for the random error ϵ_t:

(1) standard normal distribution: $\epsilon_t \sim N(0, 1)$;

(2) t distribution with degree 3: $\epsilon_t \sim t(3)$;

(3) mixture normal distribution: $\epsilon_t \sim 0.9N(0, 1) + 0.1N(0, 100)$;

Table 1: Minimum AMSE's of $\widehat{\beta}_n$ and $\widehat{\beta}_L$ and corresponding optimal bandwidths for different random errors and sample sizes with $\rho = 0.2$

n	Error	$\widehat{\beta}_n$		$\widehat{\beta}_L$	
		h_{op}	AMSE	AMSE	h_{op}
100	$N(0,1)$	0.43	0.4190	0.3879	0.41
	$t(3)$	0.41	0.4660	0.4011	0.40
	$0.9N(0,1) + 0.1N(0,100)$	0.43	0.4715	0.4088	0.45
	$C(0,1)$	0.38	0.5624	0.4360	0.37
500	$N(0,1)$	0.25	0.2549	0.2430	0.31
	$t(3)$	0.19	0.2564	0.2511	0.20
	$0.9N(0,1) + 0.1N(0,100)$	0.20	0.2672	0.2629	0.24
	$C(0,1)$	0.30	0.2952	0.2898	0.29

(4) standard Cauchy distribution: $\epsilon_t \sim C(0,1)$.

For the proposed estimators, we take $\tau = 0.5$ and employ the kernel $K(z) = \frac{3}{4}(1-z^2)I(|z| \leq 1)$. The average mean squared error (AMSE) for the estimators $\widehat{\beta}(\cdot)$ of $\beta_0(\cdot)$ based on the estimators $\widehat{b}_i(\cdot)$ of $b_i(\cdot)$ along $M = 200$ Monte Carlo trials is defined as

$$\text{AMSE}(h) = \frac{1}{2Mn_{grid}} \sum_{d=1}^{n_{grid}} \sum_{j=1}^{M} \sum_{i=1}^{2} \left[\widehat{b}_i^j(z_d) - b_i(z_d)\right]^2,$$

where $\{z_d : d = 1, 2, \ldots, n_{grid}\}$ is a sequence of grid points of z. Here, we set $\{z_d : d = 1, 2, \ldots, n_{grid}\}$ is a sequence from -1 to 1 with step 0.02. The minimal values of AMSE(h) along the grid, and the corresponding optimal bandwidths h_{op} minimizing the errors, are reported in Tables 1 and 2 for different sample sizes and the four different random errors.

From Tables 1 and 2, it can be seen that the minimum AMSEs of the estimators decrease as the sample size increases or the dependence of the observations increases, that is, the value of ρ increases. More interestingly, we can appreciate how the local linear estimator $\widehat{\beta}_L(z)$ outperforms the kernel estimator $\widehat{\beta}_n(z)$ of $\beta_0(\cdot)$ in all the considered situations.

In Figs 1-2, we plot the curves of $b_1(z)$ and $b_2(z)$ and their estimators based on $\widehat{\beta}_n(z)$ and $\widehat{\beta}_L(z)$ from $z = 0$ to 1. The Fig 1 shows that the plots get better as the sample size increase. From Fig 2, it seems that the plots become worse as the dependence of the observations increases. Fig 3 gives the plots of the AMSE vs the bandwidth h with $\rho = 0.2$ and $n = 500$. The left picture is with $\epsilon_t \sim N(0,1)$ and the right one is with $\epsilon_t \sim t(3)$. We see that for either $\widehat{\beta}_n(z)$ or $\widehat{\beta}_L(z)$, the AMSEs varies little for $h \in [0.1, 0.5]$, and the AMSEs of $\widehat{\beta}_L(z)$ are smaller than those of $\widehat{\beta}_n(z)$.

Table 2: Minimum AMSE's of $\widehat{\beta}_n$ and $\widehat{\beta}_L$ and corresponding optimal bandwidths for different random errors and sample sizes with $\rho = 0.8$

n	Error	$\widehat{\beta}_n$		$\widehat{\beta}_L$	
		h_{op}	AMSE	AMSE	h_{op}
100	$N(0,1)$	0.37	0.7302	0.6423	0.43
	$t(3)$	0.39	0.9065	0.7719	0.39
	$0.9N(0,1) + 0.1N(0,100)$	0.43	0.9552	0.7958	0.45
	$C(0,1)$	0.41	1.1733	0.9089	0.35
500	$N(0,1)$	0.23	0.5349	0.5303	0.24
	$t(3)$	0.26	0.5975	0.5864	0.25
	$0.9N(0,1) + 0.1N(0,100)$	0.26	0.6590	0.6438	0.24
	$C(0,1)$	0.25	0.7339	0.7036	0.21

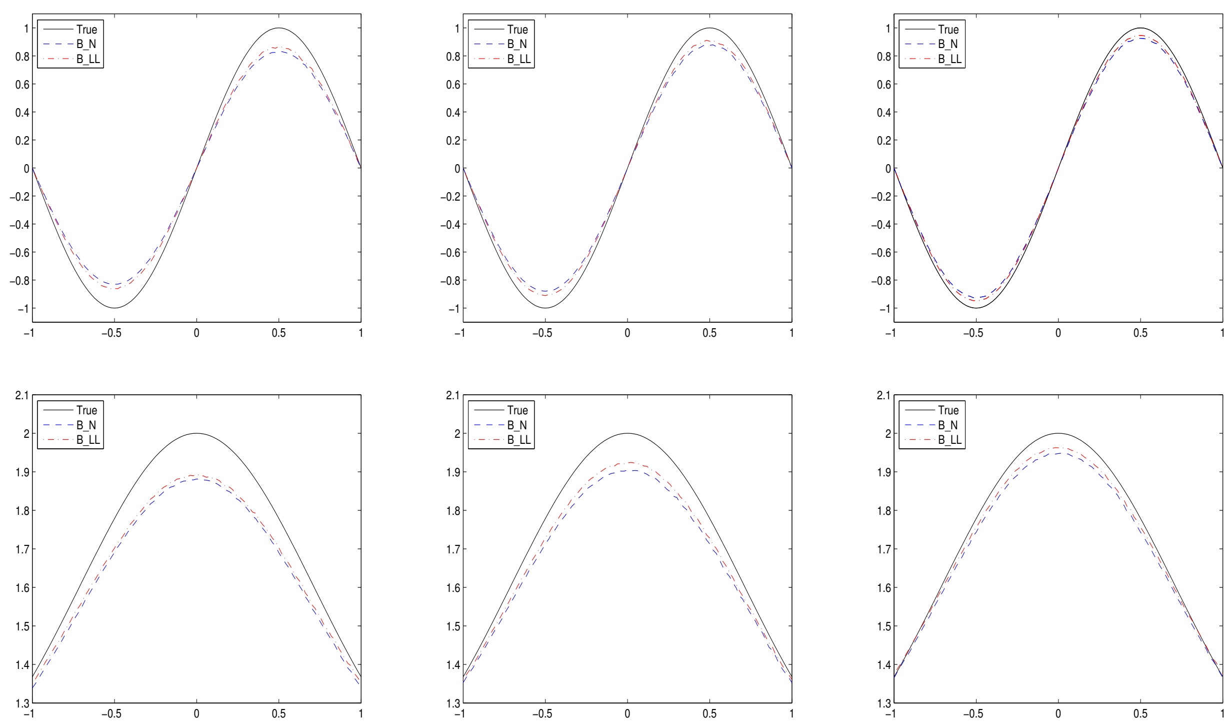

Fig 1. Plots of $\widehat{\beta}_n(z)$ and $\widehat{\beta}_L(z)$ with $\rho = 0.2$ and $\epsilon_t \sim N(0,1)$. From left to right, sample size $n = 100, 300, 500$. From up to town, $b_1(z)$ and $b_2(z)$.

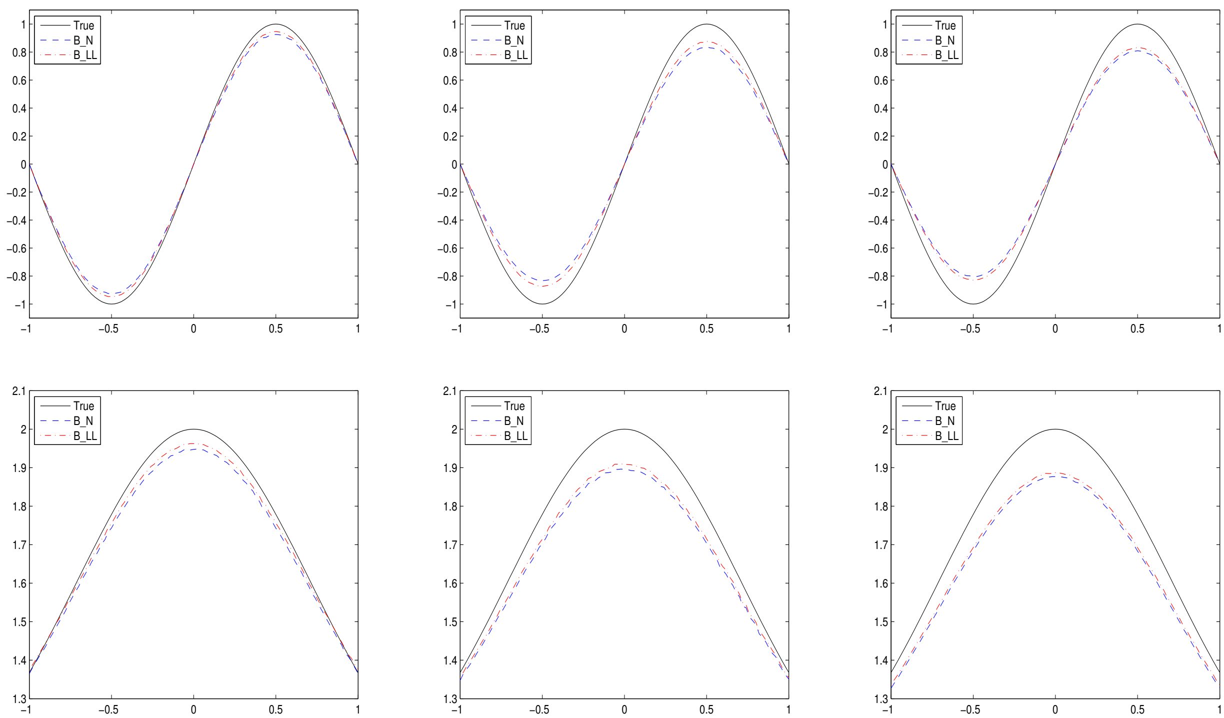

Fig 2. Plots of $\widehat{\beta}_n(z)$ and $\widehat{\beta}_L(z)$ with $n = 500$ and $\epsilon_t \sim N(0, 1)$. From left to right, $\rho = 0.2, 0.5, 0.8$. From up to town, $b_1(z)$ and $b_2(z)$.

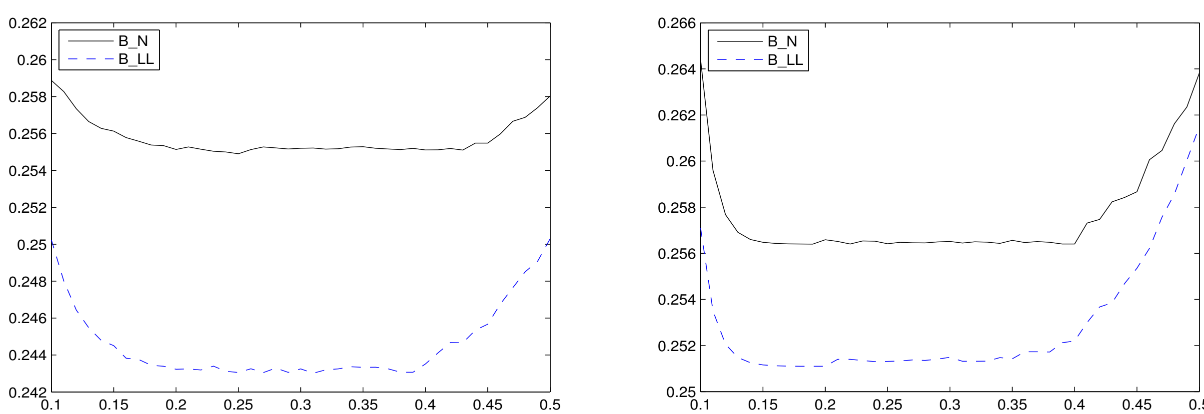

Fig 3. Plots of AMSE vs h with $\rho = 0.2$ and $n = 500$. The solid line is the plot of $\widehat{\beta}_n(z)$ and the dashed line is the plot of $\widehat{\beta}_L(z)$. The left picture is with $\epsilon_t \sim N(0, 1)$ and the right one is with $\epsilon_t \sim t(3)$.

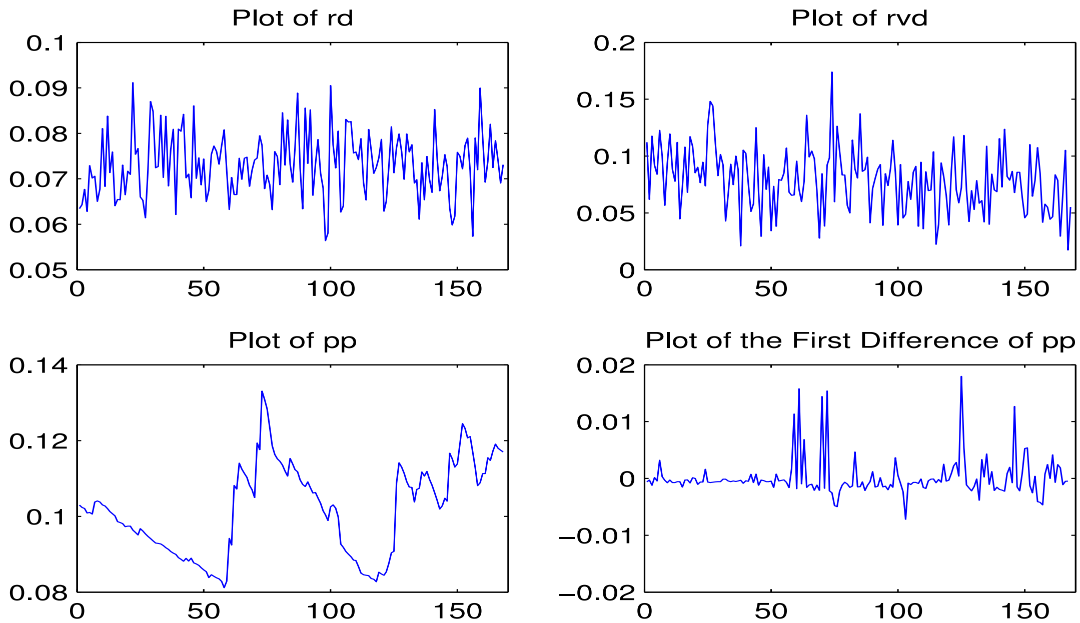

Fig 4. Plots of rd, rvd, pp and the first difference of pp.

4 Empirical applications

In this section, we apply the proposed methods to analyze UKDriverDeaths data in R language. UKDriverDeaths data gives the monthly road casualties in Great Britain from Jan. 1969 to Dec. 1984. Compulsory wearing of seat belts was introduced in Jan. 1983. Therefore in practise we just consider the data from Jan 1969 to Dec 1982, which means the data consists of 168 observations. We take rpd (ratio of the death number of passengers and the death number of drivers) as response y_t, rd (rate of the death number of drivers and the total number of drivers), rvd (rate of the death number of van car drivers and the death number of drivers) and pp (monthly petrol price) as covariates.

Firstly, it is necessary to test the stationary of covariates. Fig 4 shows the plots of rd, rvd, pp and the first difference of pp. We also use Box-Pierce's test to investigate the autocorrelation of these three covariates. Table 3 gives the p-values of Box-Pierce's test. Fig 4 and Table 3 both indicate that rd, rvd and the first difference of pp are stationary, while pp is non-stationary. Then we take rd as x_{t1}, rvd as x_{t2} and pp as z_t $(t = 1, 2, \cdots, 168)$.

Consider the following model

$$y_t = x_{t1}b_1(z_t) + x_{t2}b_2(z_t) + \sigma(x_t, z_t)\epsilon_t.$$

In practice, we set $\tau = 0.5$ and the kernel function as Gauss kernel. Since the minimum of

Table 3: P-values of Box-Pierce's test on rd, rvd, pp and $\Delta(pp)$ (the first difference of pp)

	rd	rvd	pp	$\Delta(pp)$
p-value	0.2418	0.3831	0.0000	0.6559

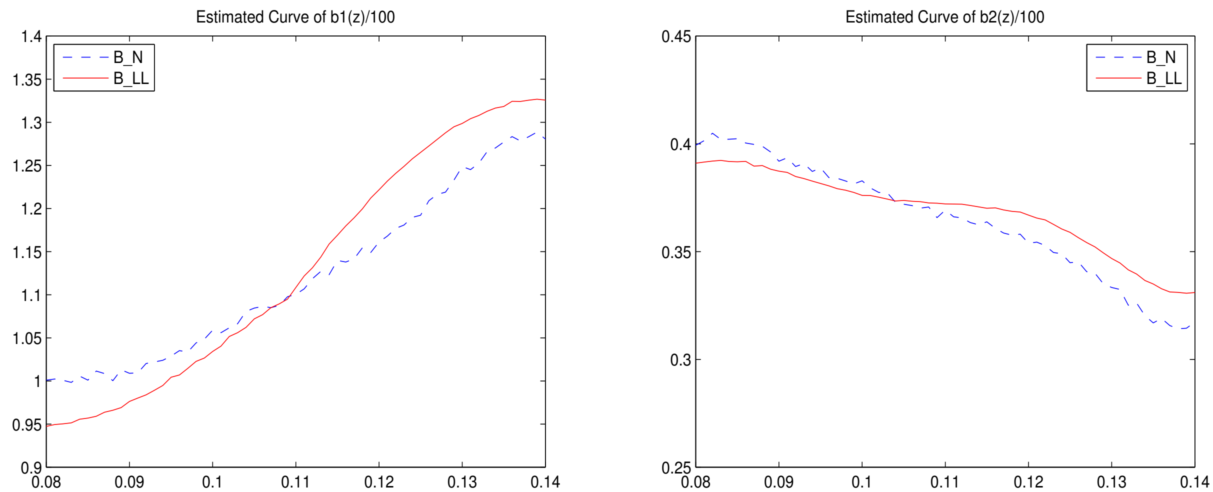

Fig 5. Curves of $\widehat{b}(\cdot)/100$ with z from 0.08 to 0.14. The left one is the curve of $\widehat{b}_1(\cdot)/100$, and the right one is the curve of $\widehat{b}_2(\cdot)/100$. The dashed line is $\widehat{\beta}_n(\cdot)$ and the solid line is $\widehat{\beta}_L(\cdot)$.

pp is 0.081 and the maximum of pp is 0.133, we take z from 0.08 to 0.14 with step 0.001. To select bandwidth, for a fixed point $z = z_0$ and a fixed bandwidth h, we first use the proposed methods to estimate $\widehat{b}_1^h(z_0)$ and $\widehat{b}_2^h(z_0)$, and then compute

$$R^2(h, z_0) = \frac{1}{168}\sum_{t=1}^{168}(\widehat{y}_t - y_t)^2,$$

where $\widehat{y}_t$ is the estimated value of y_t with $\widehat{b}_1^h(z_0)$ and $\widehat{b}_2^h(z_0)$. Define the optimal bandwidth *hop* at $z = z_0$ as the one minimizes $R^2(h, z_0)$, and take the estimated values of varying coefficients as $\widehat{b}_1(z_0) = \widehat{b}_1^{hop}(z_0)$ and $\widehat{b}_2(z_0) = \widehat{b}_2^{hop}(z_0)$.

Fig 5 shows the curves of $\widehat{b}_1(\cdot)/100$ and $\widehat{b}_2(\cdot)/100$ based on the kernel estimator $\widehat{\beta}_n(\cdot)$ and the local linear estimator $\widehat{\beta}_L(\cdot)$. From Fig 5, we can see that (1) Both rd and rvd have positive impact on rpd, while rd has bigger influence on rpd than rvd does. For example, at $z = 0.08$, based on the $\widehat{\beta}_n(\cdot)$, rpd increases about 1 as rd increases 1% and rpd increases about 0.4 as rvd increases 1%; (2) $\widehat{b}_1(\cdot)/100$ increases as pp increases. The reason may be that when petrol price is high, people are more willing to go out by car together. Every driver takes more passengers, which makes the death number of passengers increases when car accident happens. It is also seen that $\widehat{b}_2(\cdot)/100$ decreases as pp increases, but the differences are comparatively smaller; (3) the curves of $\widehat{\beta}_L(\cdot)$ are smoother that those of $\widehat{\beta}_n(\cdot)$.

5 Proofs of Main Results

We start with some preliminaries. Except mentioned explicitly, the notation used in this section is the same as in Section 2. Recall $d_n^2 = \mathrm{Var}(\sum_{j=1}^n \xi_j)$. Wang, Lin and Gulati (2003) proved that

$$d_n^2 = \begin{cases} \nu_r\, n^{3-2\mu}\, \rho^2(n), & \text{under } \mathbf{LM}, \\ \phi^2\, n, & \text{under } \mathbf{SM}, \end{cases} \tag{5.1}$$

where $\nu_r = \frac{1}{(1-r)(3-2r)} \int_0^\infty x^{-r}(x+1)^{-r}dx$. Denote by $B_H = \{B_H(t)\}_{t\geq 0}$ a fractional Brownian motion and write

$$Z_t = W(t) + \tau \int_0^t e^{-\tau(t-s)} W(s)ds,$$

where

$$W(t) = \begin{cases} B_{3/2-u}(t), & \text{under } \mathbf{LM}, \\ B_{1/2}(t), & \text{under } \mathbf{SM}. \end{cases}$$

Note that Z_t is an Ornstein-Uhlenbeck process, having a continuous local time $L_Z(t,x)$. The definition of a local time process can be found in Chapter 2 of Wang (2015).

We have the following lemma, which is crucial in the proof of our main results. Let $m(s)$, $s \in R^d$, be a measurable real function of its components and $\psi_\tau(u) = \tau - I(u<0)$.

Lemma 5.1 *Suppose that A2 and A6(i) hold, and $E|m(x_1)|^{2+\delta} < \infty$ for some $\delta > 0$. Then, for any $h = O(1)$ and $nh/d_n \to \infty$, we have*

$$\sup_{z\in R} \sum_{t=1}^n E\Big\{|m(x_t)|^{2+\delta}\, K\Big(\frac{z_t - z}{h}\Big)\Big\} = O(nh/d_n), \tag{5.2}$$

$$\sup_{z\in R} E\Big|\sum_{k=1}^{[nt]} \big[m(x_k) - Em(x_k)\big]\, K\Big(\frac{z_k - z}{h}\Big)\Big|^2$$

$$= O(nh/d_n) \begin{cases} 1+h, & \textit{under } \mathbf{LM}, \\ 1 + h\log n, & \textit{under } \mathbf{SM}, \end{cases} \tag{5.3}$$

uniformly for $0 \leq t \leq 1$, and

$$\Big\{\frac{d_n}{nh} \sum_{t=1}^n K\Big(\frac{z_t - z}{h}\Big),\ \big(\frac{d_n}{nh}\big)^{1/2} \sum_{t=1}^n m(x_t)\, K\Big(\frac{z_t - z}{h}\Big)\, \psi_\tau(\epsilon_t)\Big\}$$

$$\to_D \big\{L_Z(1,0),\ a_0\, L_Z^{1/2}(1,0)\, N\big\}, \tag{5.4}$$

where N is a standard normal variate independent of $L_Z(1,0)$ and

$$a_0^2 = \tau(1-\tau)Em^2(x_1)\int_{-\infty}^{\infty} K^2(x)dx.$$

If $\int_{-\infty}^{\infty} K(x)dx = 0$, then

$$\sum_{k=1}^n m(x_k)\, K\Big(\frac{z_k - z}{h}\Big) \;=\; O_P\big[(nh/d_n)^{1/2}(1 + h\log n)\big]. \tag{5.5}$$

Proof. For results (5.2) and (5.3), we refer to Lemma 2.2 of Wang (2015). Result (5.5) follows from (5.3) and Theorem 3.18 of Wang (2015). To prove (5.4), write

$$x_{nk} = \big(\frac{d_n}{nh}\big)^{1/2} m(x_k)\, K\Big(\frac{z_k - z}{h}\Big) := f_n(\eta_k, \eta_{k-1}, ...; x_k, x_{k-1}, ..., x_1).$$

Result (5.3), together with Theorem 2.21 of Wang (2015, page 39), implies that

$$\begin{aligned}
&\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\eta_j,\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\eta_{-j},\ \sum_{j=1}^{[nt]} x_{nj}^2\Big)\\
=\;& \Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\eta_j,\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\eta_{-j},\ Em^2(x_1)\,\frac{d_n}{nh}\sum_{j=1}^{[nt]} K^2\Big(\frac{z_k - z}{h}\Big)\Big) + o_P(1)\\
\Rightarrow\;& \big(B_{1t},\ B_{2t},\ \widetilde{a}_0^2\, Z_t\big),
\end{aligned} \tag{5.6}$$

on $D_{R^3}[0,\infty)$, where $\widetilde{a}_0^2 = Em^2(x_1)\ \int_{-\infty}^{\infty} K^2(x)dx$, $B = (B_{1t}, B_{2t})_{t\geq 0}$ is a standard 2-dimensional Brownian motion and Z_t is a functional of B. On the other hand, by recalling $\mathcal{F}_t = \sigma(x_j, z_j, j \leq t)$, it is readily seen that $\{(\eta_{t+1}, \psi_\tau(\epsilon_t)), \mathcal{F}_t\}_{t\geq 1}$ forms a sequence of martingale difference with $|\psi_\tau(\epsilon_t)| \leq \tau + 1$ and

$$E(\psi_\tau^2(\epsilon_t)|\mathcal{F}_t) = \tau(1-\tau).$$

Result (5.4) now follows from an application of Theorem 3.14 in Wang (2015, page 106). We omit the details. □

Lemma 5.2 *Suppose that A2 and A6(i) hold, and $E||x_1||^{2+\delta} < \infty$ for some $\delta > 0$. Then, for any $h = O(1)$ and $nh/d_n \to \infty$, we have*

$$\begin{aligned}
&\Big\{\frac{d_n}{nh}\sum_{t=1}^n K\Big(\frac{z_t - z}{h}\Big),\ \big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n x_t\, K\Big(\frac{z_t - z}{h}\Big)\,\psi_\tau(\epsilon_t)\Big\}\\
&\qquad \to_D \big\{L_Z(1,0),\ c_0\, L_Z^{1/2}(1,0)\,\mathbb{N}\big\},
\end{aligned} \tag{5.7}$$

where $\mathbb{N}$ is a d-dimensional normal vector independent of $L_Z(1,0)$ with mean zero and covariance $\Omega = Ex_1x_1^T$.

Proof. For any $\alpha \in R^d$, we have $E(\alpha' x_1)^2 = \alpha' E x_1 x_1^T \alpha$. The result follows from (5.4) with $m(x_t) = \alpha' x_t$ and the Cramér-Wold theorem. □

The following lemma is called the Quadratic approximation lemma, which can be found in Fan and Gijbel (1996).

Lemma 5.3 *Let $V_n(\theta)$ be a sequence of random convex function defined on a convex open subset Θ of R^d. Let F be no random positive matrix and U_n a sequence of random vectors that is stochastically bounded. Write*

$$V_n(\theta) \;\; = \;\; -\theta^T U_n + \frac{1}{2}\theta^T F\theta + f_n(\theta).$$

If for each $\theta \in \Theta$, $f_n(\theta) = o_P(1)$, then

$$\widehat{\theta}_n = F^{-1} U_n + o_P(1), \tag{5.8}$$

where $\widehat{\theta}_n$ (assumed to exist) minimizes $V_n(\theta)$.

We are now ready to prove our main results.

Proof of Theorem 2.1. We only prove (2.4). Others are similar except simpler. Let

$$v_n = (nh/d_n)^{1/2}, \quad \theta(z) = v_n\,[\beta - \beta_0(z)], \quad \widehat{\theta}_n(z) = v_n\,[\widehat{\beta}_n(z) - \beta_0(z)]$$

and $\epsilon_t^* = \sigma(x_t, z_t)\epsilon_t + x_t^T\big[\beta_0(z_t) - \beta_0(z)\big]$. Recalling (1.1), we have

$$\rho_\tau\big[y_t - x_t^T\widehat{\beta}_n(z)\big] K\Big(\frac{z_t - z}{h}\Big) = \rho_\tau\big[\epsilon_t^* - v_n^{-1} x_t^T \widehat{\theta}_n(z)\big]\, K\Big(\frac{z_t - z}{h}\Big),$$

Hence (2.1) is equivalent to

$$\widehat{\theta}_n(z) = \arg\min_\theta \sum_{t=1}^n \big[\rho_\tau\big(\epsilon_t^* - v_n^{-1} x_t^T\theta\big) - \rho_\tau(\epsilon_t^*)\big] K\Big(\frac{z_t - z}{h}\Big) := \arg\min_\theta V_n(\theta).$$

Note that, for $u \neq 0$,

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v \big[I(u \le z) - I(u \le 0)\big]\, dz, \tag{5.9}$$

where $\psi_\tau(u) = \tau - I(u < 0)$. Since $\psi_\tau[\sigma(x_t, z_t)\epsilon_t] = \psi_\tau(\epsilon_t)$, we may write

$$\begin{aligned} V_n(\theta) =& \sum_{t=1}^n \big[\rho_\tau\big(\epsilon_t^* - v_n^{-1} x_t^T\theta\big) - \rho_\tau(\epsilon_t^*)\big]\, K\Big(\frac{z_t - z}{h}\Big) \\ =& -\theta^T\Big(\frac{d_n}{nh}\Big)^{1/2} \sum_{t=1}^n x_t K\Big(\frac{z_t - z}{h}\Big)\psi_\tau(\epsilon_t^*) + \sum_{t=1}^n \xi_t(\theta) \end{aligned}$$

$$:= -\theta^T V_{n1} + V_{n2}, \tag{5.10}$$

where $\xi_t(\theta) = K(\frac{z_t - z}{h}) \int_0^{v_n^{-1} x_t^T \theta} \big[I(\epsilon_t^* \le u) - I(\epsilon_t^* \le 0) \big] du$.

Let $A_n = \frac{d_n}{nh} \sum_{t=1}^n K\Big(\frac{z_t - z}{h}\Big)$. In Appendix, for each $\theta \in R^d$, we will prove

$$V_{n2} = \frac{1}{2} \theta^T \Lambda \theta A_n + o_p(1). \tag{5.11}$$

Recalling that $A_n \to_D L_Z(1, 0)$ by (5.4) and $P(L_Z(1, 0) = 0) = 0$, it follows from (5.10) and (5.11) that

$$A_n^{-1} V_n(\theta) = -A_n^{-1} \theta^T V_{n1} + \frac{1}{2} \theta^T \Lambda \theta + f_n(\theta),$$

where $V_{n1} A_n^{-1}$ is stochastically bounded and $f_n(\theta) = o_P(1)$ for each $\theta \in \Theta$. Now, by noting

$$\arg\min_\theta V_n(\theta) = \arg\min_\theta A_n^{-1} V_n(\theta)$$

and using Lemma 5.3, we have

$$\widehat{\theta}_n(z) = A_n^{-1} \Lambda^{-1} V_{n1} + o_P(1). \tag{5.12}$$

Hence (2.4) will follow if we prove

$$\begin{aligned} V_{n1} &= \Big(\frac{d_n}{nh}\Big)^{1/2} \sum_{t=1}^n x_t K\Big(\frac{z_t - z}{h}\Big) \psi_\tau(\epsilon_t) \\ &\quad + \frac{h^2}{2} \big[\Lambda \beta_0''(z) - \Lambda_1 + 2\Lambda_2 \beta_0'(z) \big] \big(\frac{d_n}{nh}\big)^{1/2} \sum_{t=1}^n K_2\Big(\frac{z_t - z}{h}\Big) + o_P(1). \end{aligned} \tag{5.13}$$

Indeed, by noting $\int_{-\infty}^{\infty} \big[K_2(x) - \mu_2 K(x) \big] dx = 0$, (5.4) and (5.5) imply that

$$\begin{aligned} \Big| A_n^{-1} \frac{d_n}{nh} \sum_{t=1}^n K_2\Big(\frac{z_t - z}{h}\Big) - \mu_2 \Big| &= \frac{\sum_{t=1}^n \big[K_2\big(\frac{z_t - z}{h}\big) - \mu_2 K\big(\frac{z_t - z}{h}\big) \big]}{\sum_{t=1}^n K\big(\frac{z_t - z}{h}\big)} \\ &= O_P\big[\big(\frac{d_n}{nh}\big)^{1/2} (1 + h \log n) \big]. \end{aligned} \tag{5.14}$$

This, together with (5.12) and (5.13), yields that

$$\widehat{\theta}_n(z) = \Lambda^{-1} A_n^{-1} \Big(\frac{d_n}{nh}\Big)^{1/2} \sum_{t=1}^n x_t K\Big(\frac{z_t - z}{h}\Big) \psi_\tau(\epsilon_t) + \big(\frac{nh}{d_n}\big)^{1/2} \frac{\mu_2 h^2}{2} \Lambda^{-1} \alpha + o_P(1).$$

Now, by recalling $\widehat{\theta}_n(z) = \big(\frac{nh}{d_n}\big)^{1/2} [\widehat{\beta}_n(z) - \beta_0(z)]$, result (2.4) follows from Lemma 5.2 and the continuous mapping theorem .

The proof of (5.13) will be given in Appendix. The proof of Theorem 2.1 is complete. □

Proof of Theorem 2.2. The proof is similar to that of Theorem 2.1 and we only prove (2.6). Let

$$v_n = (nh/d_n)^{1/2}, \quad x_{tz} = \Big\{x_t^T, x_t^T(z_t - z)/h\Big\}^T,$$

$$\pi_0(z) = v_n[\alpha_0 - \beta_0(z)], \quad \pi_1(z) = v_n h[\alpha_1 - \beta_0'(z)], \quad \pi(z) = \{\pi_0(z)^T, \pi_1(z)^T\}^T,$$

$$\widehat{\pi}_0(z) = v_n[\widehat{\beta}_L(z) - \beta_0(z)], \quad \widehat{\pi}_1(z) = v_n h[\widehat{\beta}_L'(z) - \beta_0'(z)], \quad \widehat{\pi}_n(z) = \{\widehat{\pi}_0(z)^T, \widehat{\pi}_1(z)^T\}^T,$$

and $\widetilde{\epsilon}_t = \sigma(x_t, z_t)\epsilon_t + x_t^T[\beta_0(z_t) - \beta_0(z) - \beta_0'(z)(z_t - z)]$. It is easy to see that

$$\rho_\tau\big(y_t - x_t^T[\widehat{\beta}_L(z) + \widehat{\beta}_L'(z)(z_t - z)]\big) = \rho_\tau\big(\widetilde{\epsilon}_t - v_n^{-1}x_{tz}^T\widehat{\pi}_n(z)\big).$$

Then the minimizer of (2.2) is equivalent to

$$\widehat{\pi}_n(z) = \arg\min_\pi \sum_{t=1}^n \big[\rho_\tau\big(\widetilde{\epsilon}_t - v_n^{-1}x_{tz}^T\pi\big) - \rho_\tau(\widetilde{\epsilon}_t)\big] K\Big(\frac{z_t - z}{h}\Big) := \arg\min_\pi H_n(\pi).$$

From (5.9) we write

$$\begin{aligned} H_n(\pi) &= -\,\pi^T\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n x_{tz}K\Big(\frac{z_t - z}{h}\Big)\psi_\tau(\widetilde{\epsilon}_t) + \sum_{t=1}^n \tilde{\xi}_t(\pi) \\ &:= -\,\pi^T H_{n1} + H_{n2}, \end{aligned}$$

where $\tilde{\xi}_t(\pi) = K(\frac{z_t - z}{h})\int_0^{v_n^{-1}x_{tz}^T\pi}\big[I(\widetilde{\epsilon}_t \le u) - I(\widetilde{\epsilon}_t \le 0)\big]du$.

Write $A_n = \frac{d_n}{nh}\sum_{t=1}^n K\Big(\frac{z_t - z}{h}\Big)$ and $B_n = \frac{d_n}{nh}\sum_{t=1}^n K_2\Big(\frac{z_t - z}{h}\Big)$. In Appendix, we will prove

$$H_{n2} = \frac{1}{2}\pi^T\ \mathrm{diag}\{A_n\Lambda, B_n\Lambda\}\ \pi + o_P(1), \tag{5.15}$$

Since $\mathrm{diag}\{A_n\Lambda, B_n\Lambda\}$ is a quasi-diagonal matrix, by simple calculation of block matrix and following the same argument in the proof of (5.12), we have

$$\widehat{\pi}_0(z) = A_n^{-1}\Lambda^{-1}\, H_{n11} + o_P(1), \tag{5.16}$$

where $H_{n11} = \Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n x_tK\Big(\frac{z_t - z}{h}\Big)\psi_\tau(\widetilde{\epsilon}_t)$. In Appendix we will prove

$$H_{n11} = \Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n x_tK\Big(\frac{z_t - z}{h}\Big)\psi_\tau(\epsilon_t) + \frac{h^2}{2}\Lambda\,\beta_0''(z)\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n K_2\Big(\frac{z_t - z}{h}\Big) + o_P(1). \tag{5.17}$$

Combining with (5.14), (5.16) and (5.17), it yields that

$$\widehat{\pi}_0(z) = \Lambda^{-1}A_n^{-1}\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n x_tK\Big(\frac{z_t - z}{h}\Big)\psi_\tau(\epsilon_t) + \big(\frac{nh}{d_n}\big)^{1/2}\frac{\mu_2 h^2}{2}\,\beta_0''(z) + o_P(1).$$

Recalling that $\widehat{\pi}_0(z) = \big(\frac{nh}{d_n}\big)^{1/2}[\widehat{\beta}_L(z) - \beta_0(z)]$, then result (2.6) follows from Lemma 5.2 and the continuous mapping theorem. Thus, Theorem 2.2 is proved. □

6 Appendix

Proof of (5.11). Set $Y_t = \sigma^{-1}(x_t, z_t) K(\frac{z_t - z}{h})$. It follows from A3 that

$$0 \le \xi_t(\theta) \le v_n^{-1} |x_t^T \theta| \, K\Big(\frac{z_t - z}{h}\Big) \tag{6.1}$$

and

$$\begin{aligned} E\big[\xi_t(\theta)|\mathcal{F}_t\big] =& Y_t \int_0^{v_n^{-1}|x_t^T\theta|} \int_0^r f_t\Big\{\sigma^{-1}(x_t, z_t)\left(s - x_t^T\big[\beta_0(z_t) - \beta_0(z)\big]\right)\Big\} ds\, dr && (6.2)\\ \le& C\, v_n^{-2}\, Y_t\, (x_t^T\theta)^2 \,. && (6.3) \end{aligned}$$

From (6.2), we further have

$$E\big[\xi_t(\theta)|\mathcal{F}_t\big] = \frac{f(0)}{2} \frac{d_n}{nh} \sigma^{-1}(x_t, z)\, (x_t^T\theta)^2 \, K\Big(\frac{z_t - z}{h}\Big) + R_{t1} + R_{t2} \tag{6.4}$$

where

$$\begin{aligned} R_{t1} =& Y_t \int_0^{v_n^{-1}|x_t^T\theta|} \int_0^r \Big[f_t\{\sigma^{-1}(x_t, z_t)(s - x_t^T\big[\beta_0(z_t) - \beta_0(z)\big])\} - f(0)\Big] ds\, dr,\\ R_{t2} =& \frac{1}{2} f(0) v_n^{-2} [\sigma^{-1}(x_t, z_t) - \sigma^{-1}(x_t, z)] (x_t^T\theta)^2 \, K\Big(\frac{z_t - z}{h}\Big). \end{aligned}$$

Using A3(i) and A5(i) it can be obtained that

$$\begin{aligned} |R_{t1}| \;&\le\; C Y_t \int_0^{v_n^{-1}|x_t^T\theta|} \int_0^r \Big|\sigma^{-1}(x_t, z_t)(s - x_t^T\big[\beta_0(z_t) - \beta_0(z)\big])\Big|^\lambda ds\, dr\\ &\le\; C v_n^{-(2+\lambda)} \sigma^{-(1+\lambda)}(x_t, z_t)\, |x_t^T\theta|^{2+\lambda} \, K\Big(\frac{z_t - z}{h}\Big)\\ &\quad + C v_n^{-2} \sigma^{-(1+\lambda)}(x_t, z_t)\, (x_t^T\theta)^2 \, |x_t^T\big[\beta(z_t) - \beta(z)\big]|^\lambda K\Big(\frac{z_t - z}{h}\Big),\\ |R_{t2}| \;&\le\; C |h|^\delta \, v_n^{-2}\, (1 + ||x_t||^{k_0}) (x_t^T\theta)^2 \, K\Big(\frac{z_t - z}{h}\Big). \end{aligned}$$

Note that conditions A5 (i), A7 and $h \to 0$ imply that

$$\sigma^{-1}(x_t, z_t) K\Big(\frac{z_t - z}{h}\Big) \le C\big[\sigma^{-1}(x_t, z) + |h|^\delta (1 + ||x_t||^{k_0})\big] K\Big(\frac{z_t - z}{h}\Big),$$

$$\big[\beta_0(z_t) - \beta_0(z)\big] K\Big(\frac{z_t - z}{h}\Big) \le ChK\Big(\frac{z_t - z}{h}\Big),$$

uniformly for $1 \le t \le n$. A simple application of (5.2) yields $\sum_{t=1}^n (R_{t1} + R_{t2}) = o_P(1)$ due to $h \to 0$. This, together with (5.3) and (6.4), implies that

$$\sum_{t=1}^n E(\xi_t(\theta)|\mathcal{F}_t) \;=\; \frac{f(0)}{2} \frac{d_n}{nh} \sum_{t=1}^n \sigma^{-1}(x_t, z)\, (x_t^T\theta)^2 \, K\Big(\frac{z_t - z}{h}\Big) + o_P(1)$$

$$= \frac{f(0)}{2} E\big[\sigma^{-1}(x_1,z)\,(x_1^T\theta)^2\big]\,\frac{d_n}{nh}\sum_{t=1}^n K\Big(\frac{z_t-z}{h}\Big)+o_P(1)$$
$$= \frac{1}{2}\theta^T\Lambda\,\theta\,\frac{d_n}{nh}\sum_{t=1}^n K\Big(\frac{z_t-z}{h}\Big)+o_P(1).$$

Hence (5.11) will follow if we prove

$$\Delta_n := \sum_{t=1}^n \big[\xi_t(\theta)-E(\xi_t(\theta)|\mathcal{F}_t)\big] = o_P(1). \tag{6.5}$$

In fact, by noting that $\{\xi_t(\theta)-E(\xi_t(\theta)|\mathcal{F}_t)\}_{i\ge1}$ forms a martingale difference sequence, it follows from (6.1), (6.3) and (5.2) that

$$E\Delta_n^2 \le 2\sum_{t=1}^n E\xi_t^2(\theta) \le C\,v_n^{-1}\sum_{t=1}^n E\big[|x_t^T\theta|\,E(\xi_t(\theta)|\mathcal{F}_t)\big]$$
$$\le C\,\big(\frac{d_n}{nh}\big)^{3/2}\sum_{t=1}^n E\Big\{|x_t^T\theta|^3\,\sigma^{-1}(x_t,z_t)K\Big(\frac{z_t-z}{h}\Big)\Big\} = o(1),$$

due to $nh/d_n\to\infty$, which yields (6.5). □

Proof of (5.13). The idea is similar to the proof of (5.11), but requiring more detailed calculations. We only provide a outline.

Let $\eta_t = x_tK\Big(\frac{z_t-z}{h}\Big)\big[\psi_\tau(\epsilon_t^*)-\psi_\tau(\epsilon_t)\big]$. Then

$$V_{n1} = \Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n x_tK\Big(\frac{z_t-z}{h}\Big)\psi_\tau(\epsilon_t)+\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n\eta_t. \tag{6.6}$$

Write $\lambda_t = -\sigma^{-1}(x_t,z_t)x_t^T[\beta_0(z_t)-\beta_0(z)]$. By noting that, for some $\delta>0$,

$$|F_t(x)-F_t(0)-f(0)x-\frac{1}{2}f'(0)x^2| \le \int_0^{|x|}|f_t(s)-f_t(0)-f_t'(0)\,s|ds \le C\min\{|x|^{2+\delta},1\},$$
$$\big|\big|\beta_0(z_t)-\beta_0(z)-\beta_0'(z)(z_t-z)-\frac{1}{2}\beta_0''(z)(z_t-z)^2\big|\big|\,K\Big(\frac{z_t-z}{h}\Big) \le C|h|^{2+\delta}K\Big(\frac{z_t-z}{h}\Big).$$

Hence we have

$$E(\eta_t|\mathcal{F}_t) = x_tK\Big(\frac{z_t-z}{h}\Big)\,\big[F_t(0)-F_t(\lambda_t)\big]$$
$$= -x_tK\Big(\frac{z_t-z}{h}\Big)\,\Big[f(0)\lambda_t+\frac{1}{2}f'(0)\lambda_t^2\Big]+x_tK\Big(\frac{z_t-z}{h}\Big)O(|\lambda_t|^{2+\delta})$$
$$= f(0)\,\sigma^{-1}(x_t,z_t)x_tx_t^T\Big[h\,\beta_0'(z)\,K_1\Big(\frac{z_t-z}{h}\Big)+\frac{h^2}{2}\beta_0''(z)K_2\Big(\frac{z_t-z}{h}\Big)\Big]$$
$$-\frac{h^2}{2}f'(0)\sigma^{-2}(x_t,z_t)x_t\big[x_t^T\beta_0'(z)\big]^2\,K_2\Big(\frac{z_t-z}{h}\Big)$$

$$+ x_t \Big[\big(\sigma^{-1}(x_t, z) + ||x_t||^{k_0} + 1\big)\,\big(||x_t|| + 1\big)\Big]^3 K\Big(\frac{z_t - z}{h}\Big)\, O(h^{2+\delta}),$$

where $K_i(x) = x^i K(x), i = 1, 2$. As a consequence, it follows from Lemma 5.1 that

$$\begin{aligned}
\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n} E(\eta_t|\mathcal{F}_t) &= h\, f(0)\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n}\sigma^{-1}(x_t, z_t)x_t x_t^T \beta_0'(z)\, K_1\Big(\frac{z_t - z}{h}\Big) \\
&+ \frac{h^2 f(0)}{2}\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n}\sigma^{-1}(x_t, z_t)x_t x_t^T \beta_0''(z)\, K_2\Big(\frac{z_t - z}{h}\Big) \\
&- \frac{h^2 f'(0)}{2}\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n}\sigma^{-2}(x_t, z_t)x_t [x_t^T \beta_0'(z)]^2\, K_2\Big(\frac{z_t - z}{h}\Big) \\
&+ O_P\Big[\big(\frac{nh}{d_n}\big)^{1/2} h^{2+\delta}\Big] \\
&:= R_{1n} + R_{2n} + R_{3n} + O_P\Big[\big(\frac{nh}{d_n}\big)^{1/2} h^{2+\delta}\Big]. \qquad (6.7)
\end{aligned}$$

By recalling A5(i) and using Lemma 5.1 again, we have

$$\begin{aligned}
R_{2n} &= \frac{h^2 f(0)}{2}\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n}\sigma^{-1}(x_t, z)x_t x_t^T \beta_0''(z)\, K_2\Big(\frac{z_t - z}{h}\Big) \\
&\quad + O(h^{2+\delta})\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n} x_t \Big[\big(||x_t||^{k_0} + 1\big)\,\big(||x_t|| + 1\big)\Big]\, K_2\Big(\frac{z_t - z}{h}\Big) \\
&= \frac{h^2}{2}\,\Lambda\,\beta_0''(z)\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n} K_2\Big(\frac{z_t - z}{h}\Big) + O_P\Big[\big(\frac{nh}{d_n}\big)^{1/2} h^{2+\delta} + h^2(1 + h\log n)\Big],
\end{aligned}$$

and similarly,

$$R_{3n} = -\frac{h^2}{2}\,\Lambda_1\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n} K_2\Big(\frac{z_t - z}{h}\Big) + O_P\Big[\big(\frac{nh}{d_n}\big)^{1/2} h^{2+\delta} + h^2(1 + h\log n)\Big],$$

As for R_{1n}, it follows from A5(ii) first and then using similar arguments as in the proofs above that

$$\begin{aligned}
R_{1n} &= h\, f(0)\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n}\sigma^{-1}(x_t, z)x_t x_t^T \beta_0'(z)\, K_1\Big(\frac{z_t - z}{h}\Big) \\
&\quad + h^2 f(0)\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n}\sigma_1(x_t, z)x_t x_t^T \beta_0'(z)\, K_2\Big(\frac{z_t - z}{h}\Big) + O_P\Big[\big(\frac{nh}{d_n}\big)^{1/2} h^{2+\delta}\Big] \\
&= h\,\Lambda\,\beta_0'(z)\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n} K_1\Big(\frac{z_t - z}{h}\Big) + h^2\,\Lambda_2\beta_0'(z)\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n} K_2\Big(\frac{z_t - z}{h}\Big) \\
&\quad + O_P\Big[\big(\frac{nh}{d_n}\big)^{1/2} h^{2+\delta} + h(1 + h\log n)\Big]
\end{aligned}$$

$$= \quad h^2\,\Lambda_2\beta_0'(z)\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n K_2\Big(\frac{z_t-z}{h}\Big)+O_P\Big[\big(\frac{nh}{d_n}\big)^{1/2}h^{2+\delta}+h(1+h\log n)\Big],$$

where we have used (5.5) with $m(x)=1$, due to $\int_{-\infty}^{\infty}K_1(x)dx=0$.

Taking these estimates into (6.7), we obtain

$$\begin{aligned}\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n E(\eta_t|\mathcal{F}_t) \quad &= \quad \frac{h^2}{2}\left[\Lambda\,\beta_0''(z)-\Lambda_1+2\Lambda_2\beta_0'(z)\right]\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n K_2\Big(\frac{z_t-z}{h}\Big)\\ &\quad +O_P\Big[\big(\frac{nh}{d_n}\big)^{1/2}h^{2+\delta}+h(1+h\log n)\Big].\end{aligned}$$

On the other hand, as the proof in (5.11), we have $(d_n/nh)^{1/2}\sum_{t=1}^n[\eta_t-E(\eta_t|\mathcal{F}_t)]=o_P(1)$. We therefore obtain

$$\begin{aligned}V_{n1} \quad &= \quad \Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n x_tK\Big(\frac{z_t-z}{h}\Big)\psi_\tau(\epsilon_t)\\ &\quad +\frac{h^2}{2}\left[\Lambda\,\beta_0''(z)-\Lambda_1+2\Lambda_2\beta_0'(z)\right]\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n K_2\Big(\frac{z_t-z}{h}\Big)+o_P(1),\end{aligned}$$

due to $nh^{5+\delta}/d_n\to 0$ (yields $h^2\log n\to 0$). This completes the proof of (5.13).□

Proof of (5.15). Let $Y_t=\sigma^{-1}(x_t,z_t)K(\frac{z_t-z}{h})$ and $\widetilde{\beta}=\beta_0(z_t)-\beta_0(z)-\beta_0'(z)(z_t-z)$. We have

$$\begin{aligned}\sum_{t=1}^n E(\tilde{\xi}_t(\pi)|\mathcal{F}_t) &= \sum_{t=1}^n Y_t\int_0^{v_n^{-1}|x_{tz}^T\pi|}\int_0^r f_t\big[\sigma^{-1}(x_t,z_t)(s-x_t^T\widetilde{\beta})\big]ds\,dr\\ &= \frac{f(0)}{2}\frac{d_n}{nh}\sum_{t=1}^n\sigma^{-1}(x_t,z)\,(x_{tz}^T\pi)^2\,K\Big(\frac{z_t-z}{h}\Big)\\ &\quad+\frac{f(0)}{2}\frac{d_n}{nh}\sum_{t=1}^n[\sigma^{-1}(x_t,z_t)-\sigma^{-1}(x_t,z)](x_{tz}^T\pi)^2\,K\Big(\frac{z_t-z}{h}\Big)\\ &\quad+\sum_{t=1}^n Y_t\int_0^{v_n^{-1}|x_{tz}^T\pi|}\int_0^r\Big\{f_t\big[\sigma^{-1}(x_t,z_t)(s-x_t^T\beta^*)\big]-f(0)\Big\}ds\,dr.\end{aligned}$$

Applying similar arguments as in proof of (5.11), it follows that

$$\frac{d_n}{nh}\sum_{t=1}^n[\sigma^{-1}(x_t,z_t)-\sigma^{-1}(x_t,z)](x_{tz}^T\pi)^2\,K\Big(\frac{z_t-z}{h}\Big)=o_P(1),$$

$$\sum_{t=1}^n Y_t\int_0^{v_n^{-1}|x_{tz}^T\pi|}\int_0^r\Big\{f_t\big[\sigma^{-1}(x_t,z_t)(s-x_t^T\widetilde{\beta})\big]-f(0)\Big\}ds\,dr=o_P(1),$$

$$\sum_{t=1}^n\tilde{\xi}_t(\pi)=\sum_{t=1}^n E(\tilde{\xi}_t(\pi)|\mathcal{F}_t)+\sum_{t=1}^n\left\{\tilde{\xi}_t(\pi)-E(\tilde{\xi}_t(\pi)|\mathcal{F}_t)\right\}=\sum_{t=1}^n E(\tilde{\xi}_t(\pi)|\mathcal{F}_t)+o_P(1).$$

Hence

$$H_{n2} = \frac{f(0)}{2}\frac{d_n}{nh}\sum_{t=1}^{n}\sigma^{-1}(x_t,z)\,(x_{tz}^T\pi)^2\,K\Big(\frac{z_t-z}{h}\Big) + o_P(1). \tag{6.8}$$

Recall $\pi = (\pi_0^T, \pi_1^T)^T$. Applying (5.3) we have

$$\begin{aligned}
&\frac{d_n}{nh}\sum_{t=1}^{n}\sigma^{-1}(x_t,z)\,(x_{tz}^T\pi)^2\,K\Big(\frac{z_t-z}{h}\Big)\\
&=\frac{d_n}{nh}\sum_{t=1}^{n}\sigma^{-1}(x_t,z)\Big\{(x_t^T\pi_0)^2+(x_t^T\pi_1)^2\Big(\frac{z_t-z}{h}\Big)^2+2(x_t^T\pi_0)(x_t^T\pi_1)\Big(\frac{z_t-z}{h}\Big)\Big\}K\Big(\frac{z_t-z}{h}\Big)\\
&=\pi_0^TE[\sigma^{-1}(x_t,z)x_tx_t^T]\pi_0\cdot\frac{d_n}{nh}\sum_{t=1}^{n}K\Big(\frac{z_t-z}{h}\Big)\\
&\quad+\pi_1^TE[\sigma^{-1}(x_t,z)x_tx_t^T]\pi_1\cdot\frac{d_n}{nh}\sum_{t=1}^{n}K_2\Big(\frac{z_t-z}{h}\Big)\\
&\quad+2\pi_0^TE[\sigma^{-1}(x_t,z)x_tx_t^T]\pi_1\cdot\frac{d_n}{nh}\sum_{t=1}^{n}K_1\Big(\frac{z_t-z}{h}\Big)+o_P(1),\\
&=f^{-1}(0)\,\pi^T\,\mathrm{diag}\,\{A_n\Lambda, B_n\Lambda\}\,\pi+o_P(1),
\end{aligned} \tag{6.9}$$

where $A_n = \frac{d_n}{nh}\sum_{t=1}^{n}K\Big(\frac{z_t-z}{h}\Big)$, $B_n = \frac{d_n}{nh}\sum_{t=1}^{n}K_2\Big(\frac{z_t-z}{h}\Big)$ and we have used the fact:

$$\frac{d_n}{nh}\sum_{t=1}^{n}K_1\Big(\frac{z_t-z}{h}\Big) = O_P\big[(nh/d_n)^{1/2}(1+h\log n)\big]$$

due to $\int_{-\infty}^{\infty}K_1(x)dx = 0$ and Lemma 5.1. Taking this estimate into (6.8), we obtain (5.15). □

Proof of (5.17). Let $\widetilde{\eta}_t = x_tK\Big(\frac{z_t-z}{h}\Big)\big[\psi_\tau(\widetilde{\epsilon}_t)-\psi_\tau(\epsilon_t)\big]$. Then

$$H_{n11} = \Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^{n}x_tK\Big(\frac{z_t-z}{h}\Big)\psi_\tau(\epsilon_t)+\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^{n}\widetilde{\eta}_t.$$

Write $\widetilde{\lambda}_t = -\sigma^{-1}(x_t,z_t)x_t^T[\beta_0(z_t)-\beta_0(z)-\beta_0'(z)(z_t-z)]$. For some $\delta>0$, from A3(i), A4(iii) and A5(i) we have

$$\begin{aligned}
E(\widetilde{\eta}_t|\mathcal{F}_t) &= x_tK\Big(\frac{z_t-z}{h}\Big)\big[F_t(0)-F_t(\widetilde{\lambda}_t)\big]\\
&= -x_tK\Big(\frac{z_t-z}{h}\Big)f(0)\widetilde{\lambda}_t + x_tK\Big(\frac{z_t-z}{h}\Big)O(|\widetilde{\lambda}_t|^{1+\delta})\\
&= \frac{h^2\,f(0)}{2}\,\sigma^{-1}(x_t,z_t)x_tx_t^T\,\beta_0''(z)K_2\Big(\frac{z_t-z}{h}\Big)\\
&\quad + x_t\Big[\big(\sigma^{-1}(x_t,z)+||x_t||^{k_0}+1\big)\big(||x_t||+1\big)\Big]^2K\Big(\frac{z_t-z}{h}\Big)O(h^{2+\delta}).
\end{aligned}$$

Following from Lemma 5.1, it shows that

$$\begin{aligned}\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^{n}E(\widetilde{\eta}_t|\mathcal{F}_t) &= \frac{h^2 f(0)}{2}\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^{n}\sigma^{-1}(x_t,z_t)x_tx_t^T\beta_0''(z)\,K_2\Big(\frac{z_t-z}{h}\Big)\\ &\quad +O_P\Big[\Big(\frac{nh}{d_n}\Big)^{1/2}h^{2+\delta}\Big]\\ &= \frac{h^2}{2}\Lambda\,\beta_0''(z)\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^{n}K_2\Big(\frac{z_t-z}{h}\Big)\\ &\quad +O_P\Big[\Big(\frac{nh}{d_n}\Big)^{1/2}h^{2+\delta}+h^2(1+h\log n)\Big]. \qquad (6.10)\end{aligned}$$

Furthermore, as the proof in (5.11), we have $(d_n/nh)^{1/2}\sum_{t=1}^{n}[\widetilde{\eta}_t - E(\widetilde{\eta}_t|\mathcal{F}_t)] = o_P(1)$, which yields (5.17) by $nh^{5+\delta}/d_n \to 0$ (implying $h^3\log n \to 0$). □

Acknowledgments. Liang acknowledges research support from the National Natural Science Foundation of China (11271286) and the Specialized Research Fund for the Doctor Program of Higher Education of China (20120072110007). Wang acknowledges research support from the Australian Research Council.

References

[1] Cai, Z., Gu, J. and Li, Q. (2009). *Some Recent Developments on Nonparametric Econometrics*. Advances in Econometrics, 25, 495-549. Emerald Group Publishing Limited, UK.

[2] Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *J. Amer. Statist. Assoc.* **103**, 1595-1608.

[3] Cai, Z., Li, Q., Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics* **148**, 101-113.

[4] Chen, J., Li, D. and Zhang, L. (2010). Robust estimation in a nonlinear cointegration model. *J. Multi. Anal.*, **101**, 706–717.

[5] Fan, J., Gijbels, I. (1996). *Local Polynomial Modeling and its Applications.* Chapman and Hall, London.

[6] Gao, J. and Phillips, P. C. B. (2013). Functional coefficient nonstationary regression. it Cowles fundation discussion paper No 1911.

[7] Honda, T. (2004). Quantile regresion in varying coefficient models. *J. Statist. Plan. and Infer.* **121**, 113–125.

[8] Honda, T. (2013). Nonparametric LAD cointegrating regression. *J. Multi. Anal.* **117**, 150–162.

[9] Kim, M.-O. (2007). Quantile regression with varying coefficients. *Ann. Statist.* **35**, 92-108.

[10] Koenker, R. and Bassett G. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

[11] Koenker, R. (2005). *Quantile Regression.* Cambridge university press.

[12] Li, D. and Li, R. (2015). Local composite quantile regression smoothing for Harries Recurrient Markov processes. Manuscript.

[13] Sun, Y. Cai, Z. and Li, Q. (2013). Semeparametric functional coefficient models with integrated covariates. *Econometric Theory*, **29**, 659–672.

[14] Sun, Y. Cai, Z. and Li, Q. (2015). A consistent nonparametric test on oemeparametric functional coefficient models with integrated time series. *Econometric Theory*, forthcoming.

[15] Wang, Q. (2014). Martingale limit theorems revisited and non-linear cointegrating regression. *Econometric Theory*, forthcoming.

[16] Wang, Q. (2015). *Limit theorems for nonlinear cointegrating regression.* World Scientific. Singapore.

[17] Wang, Q., Lin, Y.-X. and Gulati, C. M. (2003). Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory* **19**, 143-164.

[18] Wang, Q. and Phillips, P. C. B. (2009a) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* **25**, 710-738.

[19] Wang, Q. and Phillips, P. C. B. (2009b). Structural Nonparametric Cointegrating Regression, *Econometrica*, **77**, 1901-1948.

[20] Wang, Q. and Phillips, P. C. B. (2011). Asymptotic Theory for Zero Energy Functionals with Nonparametric Regression Applications. *Econometric Theory*, **27**, 235-259.

[21] Wang, Q. and Phillips, P. C. B. (2012). A Specification Test for Nonlinear Nonstationary Models. *Annals of Statistics*, **40**, 727-758.

[22] Wang, Q. and Phillips, P. C. B. (2015). Nonparametricregression with endogeneity and long memory. *Econometric Theory*, forthcoming

[23] Xiao, Z. (2009a). Quantile cointegrating regression. *Journal of Econometrics* **150**, 248-260.

[24] Xiao, Z. (2009b). Functional-coefficient cointegration models. *Journal of Econometrics* **152**, 81-92.