Hitotsubashi University

Global COE Hi-Stat Discussion Paper Series 207

Research Unit for Statistical and Empirical Analysis in Social Sciences (Hi-Stat)

Nonparametric LAD Cointegrating Regression

Toshio Honda

October 2011
(Revised May 2012)

Hi-Stat
Institute of Economic Research
Hitotsubashi University
2-1 Naka, Kunitatchi Tokyo, 186-8601 Japan
http://gcoe.ier.hit-u.ac.jp

Hi-Stat Discussion Paper

Nonparametric LAD cointegrating regression

Toshio Honda*

Graduate School of Economics, Hitotsubashi University, Kunitachi, Tokyo 186-8601, Japan

Abstract

We deal with nonparametric estimation in a nonlinear cointegration model whose regressor and error term can be contemporaneously correlated. The asymptotic properties of the Nadaraya-Watson estimator are already examined in the literature. In this paper, we consider nonparametric least absolute deviation (LAD) regression and derive the asymptotic distributions of the local constant and local linear estimators by appealing to the local time approach. We also present the results of a small simulation study.

Keywords: Nonlinear cointegration, Integrated process, Local time, Least absolute deviation, Local polynomial regression, Bias
2010 MSC: 62G08, 62G35, 62M10

1. Introduction

There have been a lot of papers applying nonparametric regression techniques to time series data. Nonparametric regression techniques are flexible and robust to model misspecifications. The techniques are also useful for specification testing of parametric models. See Fan and Yao [6], Gao [7], and Li and Racine [17] and the references therein for recent developments of nonparametric estimation for stationary time series data.

Recently, Karlsen and Tjøstheim [13], Karlsen et al. [14], and Wang and Phillips [21]-[23] have successfully applied nonparametric regression estimation to nonlinear cointegration models and investigated the asymptotic properties of the estimators. Since Granger [9] and Engle and Granger [5],

*Corresponding author
Email address: honda@econ.hit-u.ac.jp (Toshio Honda)

cointegartion models have been one of most popular models for nonstationary time series data. However, most researches were limited to linear models until [13], [14], and [21]-[23]. [13], [14], and [20] are based on the theory of null recurrent Markov chains and [21]-[23] exploited the theory of local time of nonstationary processes. See [1], [3], [8], [24] for specification testing and semiparametric models of nonstationary time series.

Chen et al. [4] considered robust nonparametric regression in the setup of [21] and derived the asymptotic distribution of the estimator. In [4], the regressor and the error term are assumed to be mutually independent as in Theorem 3.1 of [21]. Their robust nonparametric regression estimators include nonparametric quantile regression estimators. However, Theorem 1 below contradicts with their Theorem 3.2. The details are given after Theorem 1 below and in Remark 2 in section 4. Lin et al. [18] deals with robust nonparametric regression by using the null recurrent Markov chain approach and we cannot apply their approach to the setup of this paper because $\{X_i\}$ is not a Markov chain and X_i and u_i are correlated in this paper.

In this paper, we consider least absolute deviation (LAD) regression in the setup of [22] where the regressor and dependent variable can be contemporaneously correlated. We examine the asymptotic properties of the local constant estimator (LCE) and local linear estimator (LLE). The proof of our main result crucially depends on the results in [21] and [22]. Our results can be easily extended to general q-th quantile regression and we also give a comment on how to deal with nonparametric robust estimators in Remark 4 in section 4. We also carried out a small simulation study. In the simulation study, we compared the nonparametric LAD estimator and the nonparametric least squares estimator and investigated the effects of bandwidths.

Our nonlinear coinegartion model is given by

$$Y_i = g(X_i) + v_i, \quad i = 1, \dots, n, \tag{1}$$

where $v_i = v(X_i, u_i)$, $\{X_i\}$ is a near-integrated or integrated process, $\{u_i\}$ is a stationary process. We estimate $g(x_0)$ for a fixed x_0 under Assumptions G and V below.

Assumption G: $g(x)$ is twice continuously differentiable in a neighborhood of x_0.

We present the assumption on $v(X_i, u_i)$ here. We specify $\{X_i\}$ and $\{u_i\}$ and the other assumptions are given later in section 2.

Assumption V: $v(x, u)$ is monotone increasing in u for any x and $v(x, m_u) = 0$ for any x, where m_u is the median of u_i. In addition, $v(x, u)$ is continuously differentiable in a neighborhood of (x_0, m_u) and $\frac{\partial v}{\partial u}(x_0, m_u) \neq 0$. When we deal with the local constant estimator (LCE), $v(x, u)$ is twice continuously differentiable in a neighborhood of (x_0, m_u).

Notice that $\text{sign}(v_i) = \text{sign}(u_i - m_u)$ under Assumption V and an example of $v(x, u)$ is $\sigma(x)(u - m_u)$. Hence we have $\text{E}\{\text{sign}(v_i)\} = 0$, where $\text{sign}(v) = -1, v < 0, = 1, v \geq 0$, and we estimate $g(x_0)$ by using nonparametric LAD regression. In [22], the error term in (1) is u_i with $\text{E}\{u_i\} = 0$ and $g(x_0)$ is estimated by nonparametric mean regression estimators such as the Nadaraya-Watson estimator. In [22] and this paper, contemporaneous correlation between the regressor and the error term is allowed.

There has been a lot of interest in quantile regression since Koenker and Basset [16]. It is because quantile regression is robust to outliers and offers more information on data than mean regression. See Koenker [15] for more details on quantile regression. There are a lot of papers which deal with nonparametric quantile regression for time series data, to name only a few, Honda [11], [12], Cai [2], Hall et al. [10]. Xiao [26] considers quantile regression in linear and time-varying cointegration models.

The rest of this paper is organized as follows. We state assumptions, define the nonparametric estimators, and present the main result Theorem 1 in section 2. We rather focus on the local linear estimator (LLE) in this paper. We present the results of a simulation study in section 3. The proof of Theorem 1 and the propositions for the proof are given in section 4. The proofs of the propositions are relegated to section 5.

We denote convergence in distribution and in probability by $\stackrel{d}{\to}$ and $\stackrel{p}{\to}$, respectively and C is a generic positive constant whose value varies from place to place. When X has a normal distribution with mean μ and covariance matrix Σ, we write $X \sim \text{N}(\mu, \Sigma)$. For a vector v, v^T is the transpose of v. We write $[a]$ for the largest integer less than or equal to a. We introduce two i.i.d. processes $\{\epsilon_i | -\infty < i < \infty\}$ and $\{\lambda_i | -\infty < i < \infty\}$ later in section 2. For notational simplicity, we write $\{\epsilon_i\}$ and $\{\lambda_i\}$ for them, respectively. In addition we omit almost surely or a.s. when we consider conditional expectations or it is clear from the context. This is also for notational simplicity.

2. Estimators and asymptotic distributions

First we follow [22] to define $\{X_i\}$ and describe the limiting process $J_\kappa(t), 0 \le t \le 1$, of $X_{[nt]}/\sqrt{n}, 0 \le t \le 1$. Next we specify $\{u_i\}$ as in [22]. We borrow a lot of notation from [22] in the definitions and specifications. Then we define the LCE and LLE and present the asymptotic distributions in Theorem 1, whose proof crucially depends on the results in [21] and [22] and is postponed to section 4.

We specify $\{X_i\}$ in Assumption X below and the assumption is Assumption 1 of [22]

Assumption X: With $X_0 = 0$ and $\rho = 1+\kappa/n$ for some constant κ, we define X_i by $X_i = \rho X_{i-1} + \xi_i$. $\{\xi_i\}$ is a linear process given by $\xi_i = \sum_{k=0}^{\infty} \phi_k \epsilon_{i-k}$, where $\sum_{k=0}^{\infty} |\phi_k| < \infty$, $0 < \sum_{k=0}^{\infty} \phi_k = \phi$, and $\{\epsilon_i\}$ is an i.i.d. process. Besides, $\mathrm{E}\{\epsilon_i\} = 0$, $\mathrm{Var}\{\epsilon_i\} = 1$, and the characteristic function of ϵ_i is integrable.

Suppose that Assumption X holds throughout this paper. Then $X_{[nt]}/\sqrt{n}, 0 \le t \le 1$, converges in distribution to

$$J_\kappa(t) = \phi(W(t) + \kappa \int_0^t e^{(t-s)\kappa} W(s)ds), \quad 0 \le t \le 1, \tag{2}$$

in the Skorokhod topology on $D[0,1]$, where $W(s), 0 \le s \le 1$, is a standard Brownian motion. See Proposition 7.1 of [22] for the proof. The local time process $L(s,a)$ of $J_\kappa(t), 0 \le t \le 1$ is defined as in (3.10) of [22]. Note that $J_\kappa(t)$ in (2) is $J_\kappa(t)$ in (3.9) of [22] multiplied by ϕ.

Next we define $\{u_i\}$ in Assumption U1 below, which is essentially Assumption 2 of [22]. In the setup, X_i and u_i can be correlated.

Assumption U1: Letting $\{\lambda_i\}$ be another i.i.d. process independent of $\{\epsilon_i\}$, we have $u_i = u(\epsilon_i, \dots, \epsilon_{i-m_0}, \lambda_i, \dots, \lambda_{i-m_0})$, where m_0 is a positive integer.

We do not need any assumptions on moments of u_i. Instead we have to impose another assumption on the conditional density of u_i to deal with nonparametric LAD regression. We write $\mathcal{E}$ and $\mathcal{E}_{i-m_0}^i$ for the σ-field generated by $\{\epsilon_i\}$ and $\{\epsilon_i, \dots, \epsilon_{i-m_0}\}$, respectively. If u_i has the conditional density given $\mathcal{E}$, then we can denote it by $f_{u_i}(u|\mathcal{E}_{i-m_0}^i)$ due to Assumption U1. Recall that we denote the unconditional median of u_i by m_u.

Assumption U2: There is a fixed and nonstochastic neighborhood of m_u. In the neighborhood, the conditional distribution of u_i given $\mathcal{E}$ has the density function. Besides $f_{u_i}(u|\mathcal{E}_{i-m_0}^i)$ is uniformly bounded in $(\epsilon_i, \dots, \epsilon_{i-m_0})$

and continuously differentiable and the derivative $f'_{u_i}(u|\mathcal{E}^i_{i-m_0})$ is uniformly bounded in the neighborhood. We also have $f_u(m_u) > 0$, where $f_u(u)$ is the density function of u_i.

We assume that Assumptions U1 and U2 hold throughout this paper. Denoting the conditional density of v_i given $\mathcal{E}$ by $f_{v_i}(v|\mathcal{E})$, we have a representation of $f_{v_i}(0|\mathcal{E})$ in (3).

$$f_{v_i}(0|\mathcal{E}) = f_{u_i}(m_u|\mathcal{E}^i_{i-m_0})\Big(\frac{\partial v}{\partial u}(X_i, m_u)\Big)^{-1}. \tag{3}$$

Here we introduce another notation $f_v(v|x)$ for the density function of $v(x, u_i)$ with x fixed. Since X_i and u_i are not independent, the density function is not the conditional density function of $v(X_i, u_i)$ given $X_i = x$. We slightly abuse the standard notation for conditional density functions since it plays almost the same role as the conditional density function in the cases of stationary processes. As for the density function of $v(x, u_i)$ $f_v(v|x)$, we have

$$f_v(0|x_0) = f_u(m_u)\Big(\frac{\partial v}{\partial u}(x_0, m_u)\Big)^{-1}. \tag{4}$$

We state assumptions on the kernel function $K(s)$ and the bandwidth h. We define the Fourier transform of $f(x)$ by $\hat{f}(t) = (2\pi)^{-1/2}\int e^{\mathrm{i}tx}f(x)dx$, where $f(x)$ is an integrable function and i is the imaginary unit.

Assumption K: $K(s)$ is a nonnegative bounded continuous function with compact support and $\hat{K}(t)$ is integrable. In addition, the Fourier transforms of $sK(s)$, $s^2K(s)$, and $s^3K(s)$ are also integrable.

Assumption K above is Assumption 3 of [22] plus the last line of Assumption K. Assumption 3 is not restrictive as asserted in [22] and the last line of Assumption K is not restrictive, either because

$$\frac{d^j}{dt^j}\hat{K}(t) = \frac{\mathrm{i}^j}{\sqrt{2\pi}}\int e^{\mathrm{i}ts}s^jK(s)dx.$$

We introduce some notation related to the kernel function here.

$$K_i = K((X_i - x_0)/h) \quad \text{and} \quad \eta_i = (1, (X_i - x_0)/h)^T \tag{5}$$

$$\kappa_j = \int s^jK(s)ds \quad \text{and} \quad \nu_j = \int s^jK^2(s)ds \tag{6}$$

Assumption H: $nh^2 \to \infty$ and $nh^{10} = O(1)$.

Assumption H is a very mild condition. It is easy to see from Theorem 1 below that the asymptotically optimal bandwidth has the form of $h = C_0 n^{-1/10}$, where C_0 depends on the definition of the optimality and maybe a random variable.

We define the LLE $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)^T$ of $(g(x_0), hg'(x_0))^T$ by

$$\hat{\beta} = \operatorname*{argmin}_{\beta \in R^2} \sum_{i=1}^{n} K_i |Y_i - \eta_i^T \beta|. \tag{7}$$

The convergence rate of $\hat{\beta}$ is $(nh^2)^{-1/4}$ and we set

$$\tau_n = (nh^2)^{1/4}.$$

We use both τ_n and $(nh^2)^{1/4}$ in this paper. By normalizing $\hat{\beta}$ as

$$\hat{\theta} = \tau_n(\hat{\beta}_1 - g(x_0), \hat{\beta}_2 - hg'(x_0))^T,$$

we have from (7)

$$\hat{\theta} = \operatorname*{argmin}_{\theta \in R^2} \sum_{i=1}^{n} K_i(|v_i^* - \tau_n^{-1}\eta_i^T\theta| - |v_i^*|), \tag{8}$$

where

$$v_i^* = v_i + \frac{1}{2}\Big(\frac{X_i - x_0}{h}\Big)^2 h^2 g''(\bar{X}_i) \tag{9}$$

and $\bar{X}_i$ is defined in the second order Taylor expansion of $g(x)$ at x_0. For the LCE, we can define $\hat{\theta}$ in (8) by removing η_i and replacing v_i^* with v_i^{**} below.

$$v_i^{**} = v_i + (X_i - x_0)g'(x_0) + \frac{1}{2}\Big(\frac{X_i - x_0}{h}\Big)^2 h^2 g''(\bar{X}_i) \tag{10}$$

Here we state Theorem 1, which is the main result of this paper and will be proved in section 4. The theorem says we can estimate $g(x_0)$ without any instrumental variables as in [22]. We also give a remark on the extension to nonparametric robust regression at the end of section 4.

Theorem 1. *Suppose that Assumptions V, X, U1, U2, K, and H hold. Then we have for the LLE,*

$$\hat{\theta} - B_{1n} \xrightarrow{d} \frac{1}{2}(f_v(0|x_0)L^{1/2}(1,0))^{-1} \begin{pmatrix} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{pmatrix}^{-1} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix},$$

where

$$
\begin{aligned}
B_{1n} &= (nh^2)^{1/4}\frac{h^2}{2}\Big\{\left(\begin{array}{cc}\kappa_0 & \kappa_1\\ \kappa_1 & \kappa_2\end{array}\right)^{-1}\left(\begin{array}{c}\kappa_2\\ \kappa_3\end{array}\right)g''(x_0)+O_p\Big(\frac{1}{(nh^6)^{1/4}}\Big)\Big\},\\
\left(\begin{array}{c}Z_1\\ Z_2\end{array}\right) &\sim N\left(\left(\begin{array}{c}0\\ 0\end{array}\right),\left(\begin{array}{cc}\nu_0 & \nu_1\\ \nu_1 & \nu_2\end{array}\right)\right),
\end{aligned}
$$

Note that $(Z_1, Z_2)^T$ above is independent of $L(1,0)$.

For the LCE, we also assume that $\kappa_1 = 0$. Then we have

$$\hat{\theta} - B_{2n} \xrightarrow{d} \frac{1}{2}(f_v(0|x_0)L^{1/2}(1,0))^{-1}\kappa_0^{-1}Z_1,$$

where Z_1 is defined above,

$$
\begin{aligned}
B_{2n} &= (nh^2)^{1/4}\frac{h^2}{2}\frac{\kappa_2}{\kappa_0}\Big\{(f_v(0|x_0))^{-1}\\
&\quad\times\Big(g''(x_0)f_v(0|x_0)+2g'(x_0)\frac{\partial f_v}{\partial x}(0|x_0)-(g'(x_0))^2\frac{\partial f_v}{\partial v}(0|x_0)\Big)\\
&\quad+O_p\Big(\frac{1}{(nh^8)^{1/4}}\Big)\Big\},\\
\frac{\partial f_v}{\partial v}(0|x_0) &= \Big(\frac{\partial v}{\partial u}(x_0,m_u)\Big)^{-2}\Big\{f_u'(m_u)+f_u(m_u)\frac{\partial}{\partial u}\Big(\frac{\partial v}{\partial u}(x_0,m_u)\Big)^{-1}\Big\}.
\end{aligned}
$$

Note that $O_p(1/(nh^6)^{1/4})$ and $O_p(1/(nh^8)^{1/4})$ may affect the forms of the bias terms. However, they do not affect the asymptotic distributions since $(nh^2)^{1/4}h^2/(nh^6)^{1/4} \to 0$ and $(nh^2)^{1/4}h^2/(nh^8)^{1/4} \to 0$. The bias terms B_{1n} and B_{2n} are negligible when $nh^{10} = o(1)$. When $C_1 < nh^{10} < C_2$ for some positive constants C_1 and C_2, $O_p(1/(nh^6)^{1/4})$ and $O_p(1/(nh^8)^{1/4})$ are negligible.

We give an expression of the objective function and decompose the estimator to the stochastic part and the bias part in the proof of Theorem 1. We deal with the stochastic part by using the results in [22] and Theorem 1 allows for some endogeneity. The bias part is considered in Propositions 3 and 5.

The asymptotic distribution of the LLE will be the same as that of the Nadaraya-Watson estimator in (3.12) of [22] if $(2f_v(0|x_0))^{-1}$ is replaced with σ_u. (3.12) of [22] is derived under some restrictive assumptions. However, some of them does not seem to be used in the proof. The relaxation of the restrictive assumptions may be a topic of future research. On the other hand, the limiting distribution is $c(L(1,0))^{1/2}\xi$, where ξ is the standard normal variable independent of $L(1,0)$ and c is some constant, in Theorem 3.2 of [4]. They considered the LCE in the case of undersmoothing. The difference comes from the difference between Proposition 2 and Lemma A.1 of [4]. See Remark 2 in section 4 for the difference.

The bias term of the LCE is much more complicated than that of the LLE and that of the Nadaraya-Watson estimator in [22] and [23]. The complicated form is due to Proposition 5. The LCE also requires more technical assumptions. Thus we should use the LLE for nonparametric quantile regression with integrated covariates.

Theorem 1 implies that the asymptotically optimal bandwidth may depend on $g''(x_0)$, $L(1,0)$, and $f_v(0|x_0)$ and that larger bandwidths will be preferable. It might be difficult to estimate $f_v(0|x_0)$ from regression residuals. We will need another paper to establish the consistency even if we estimate it by standard kernel conditional density estimators. A cross-validation method as in [4] may be a promising candidate for bandwidth selection.

3. Simulation study

We carried out a small simulation study by using R to compare LAD and least squares (LS) estimators and examine the effects of bandwidths. The results are given in Tables 1-6. In the simulation study, we partly followed section 4 of [22] and set

$$Y_i = X_i + X_i^2 + v_i \quad \text{and} \quad X_i = X_{i-1} + \epsilon_i,$$

where $\epsilon_i \sim \mathrm{N}(0,1)$ and $X_0 = 0$, and

$$v_i = 0.8(\lambda_i + 0.4\epsilon_i)/\sqrt{1+(0.4)^2},$$

where $\lambda_i \sim \mathrm{N}(0,1)$, $t(3)/\sqrt{3}$, and $t(2)/\sqrt{3}$ in Tables 1-2, 3-4, and 5-6, respectively. Note that $t(j)$ means a t-distribution with d.f. j.

We estimated $g(x)$ by the local linear LAD estimator and the local linear LS estimator and denote the estimate by $\hat{g}(x)$. The results for the LAD estimator and for the LS estimator are presented in Tables 1, 3, and 5 and in Tables 2, 4, and 6, respectively. The Epanechnikov kernel was used and we employed the quantreg package for LAD regression. See [15] for the details of the quantreg package.

We took $x = 0,\ 5,\ 10$ and $h = n^{-1/\gamma}$, $\gamma = 4, \ldots, 10$ in the simulation study. The sample size was 1600 and the entries in the tables were based on 10000 replications. In the tables, Bias and MSE stand for the simulated bias and the simulated mean squared error of the estimators, respectively and SE1 means the standard error of the the simulated MSE. We also computed the approximate integrated squared error

$$\frac{1}{L+1}\sum_{l=0}^{L}(\hat{g}(x+\delta l) - g(x+\delta L))^2 \tag{11}$$

with $L = 10$ and $\delta = 0.1$. We denote the simulated MISE by MISE and SE2 means the standard error of the simulated MISE.

We have to be very careful in the simulation study of nonparametric regression for nonstationary time series since there are no or only a few observations to estimate $g(x)$ in some replications.

We used the LCE or the Nadaraya-Watson estimator to compute the estimates for numerical stability when we had only two observations on $(x - h, x + h)$. We employed the weightedMedian function in the aroma.light package of R for the LCE. The entries in the N2 rows are the numbers of the replications in which there are less than 3 observations on $(x - h, x + h)$. The entries in the N1 rows are the numbers of the replications in which there is no or only one observation on $(x - h, x + h)$. Note that we excluded the replications with no or only one observation from the computation of the MSE. When we have no observation around x, we cannot compute $\hat{g}(x)$. In addition, the simulation results were very badly affected by replications with only one observation. The numbers in the N1 rows are included in the ones in the N2 rows.

When we computed the values in the MISE rows, we used only replications which had at least three observations at each grid of (11). The numbers in the NU rows are the numbers of replications not used for computation of MISE's.

Tables 1-6 are around here.

Tables 1-6 give us the following implications.

1. When we look at MISE's, the LS estimator performs better when v_i has finite variance. On the other hand, the LAD estimator performs better in Tables 5 and 6, .

2. Larger bandwidths tend to give smaller MISE's. This means that the effects of fewer observations for $\hat{g}(x)$ are much more serious than the biases caused by larger bandwidths.
3. MSE's are larger than MISE's. Some results omitted here imply that this is partly due to replications with only two observations around x. Recall that we used such replications only for MSE's. We may need at least three observations to estimate $g(x)$. The values in the N2, N1, and NU rows and the differences between MSE's and MISE's imply that we will need a very large sample size to carry out nonparametric regression for nonstationary time series.
4. Biases are small in spite of the endogeneity.

4. Proof of Theorem 1

We give Propositions 1-5 before we prove Theorem 1. The proofs of the propositions are postponed to section 5.

Proposition 1 is essentially (3.8) combined with Proposition 7.2 of [22] and the first two elements of the random vector in Proposition 1 are related to the stochastic part of the nonparametric LAD regression estimators. Recall that $\tau_n = (nh^2)^{1/4}$.

Proposition 1. *Suppose that Assumptions X, U1, U2, and K hold and that $nh^2 \to \infty$ and $h \to 0$. Then we have*

$$\begin{aligned}
&\Big(\tau_n^{-1}\sum_{i=1}^n K_i \mathrm{sign}(u_i - m_u), \tau_n^{-1}\sum_{i=1}^n \Big(\frac{X_i - x_0}{h}\Big)K_i \mathrm{sign}(u_i - m_u),\\
&\quad \tau_n^{-2}\sum_{i=1}^n K_i, \tau_n^{-2}\sum_{i=1}^n K_i f_{u_i}(m_u|\mathcal{E}^i_{i-m_0}),\\
&\quad \tau_n^{-2}\sum_{i=1}^n \Big(\frac{X_i - x_0}{h}\Big)K_i f_{u_i}(m_u|\mathcal{E}^i_{i-m_0}), \tau_n^{-2}\sum_{i=1}^n \Big(\frac{X_i - x_0}{h}\Big)^2 K_i f_{u_i}(m_u|\mathcal{E}^i_{i-m_0}),\\
&\quad \tau_n^{-2}\sum_{i=1}^n \Big(\frac{X_i - x_0}{h}\Big)^3 K_i f_{u_i}(m_u|\mathcal{E}^i_{i-m_0})\Big)^T\\
&\stackrel{d}{\to} \ (L^{1/2}(1,0)Z_1, L^{1/2}(1,0)Z_2, \kappa_0 L(1,0), \kappa_0 f_u(m_u)L(1,0),\\
&\qquad \kappa_1 f_u(m_u)L(1,0), \kappa_2 f_u(m_u)L(1,0), \kappa_3 f_u(m_u)L(1,0))^T,
\end{aligned}$$

where $(Z_1, Z_2)^T$ is defined as in Theorem 1 and independent of $L(1,0)$.

We need to apply the almost sure representation theorem in Remark 1 to the result in Proposition 1 for technical reasons.

Remark 1. Let Ω be a σ-field generated by $\{\epsilon_i\}$ and $\{\lambda_i\}$. Addendum 1.10.5 of [25] implies that there exists a σ-field $\tilde{\Omega}$ satisfies
1. $\tilde{\Omega}$ virtually contains Ω,
2. $(Z_1, Z_2)^T$ and $L(1,0)$ can be defined on $\tilde{\Omega}$,
3. We can replace convergence in distribution with almost sure convergence in Proposition 1.
Hence we will assume that the sequence of random vectors in Proposition 1 also converges almost surely in Proposition 4 below and the proof of Theorem 1.

Proposition 2 gives the expansion of the objective function for $\hat{\theta}$. Note that θ is fixed in Proposition 2 and we consider the uniformity in θ in Proposition 4 by exploiting Proposition 2 and the convexity of the objective function.

Proposition 2. *Suppose that Assumptions V, X, U1, U2, and K hold and that $nh^2 \to \infty$ and $h \to 0$. Then for any $\theta \in R^2$, we have*

$$
\begin{aligned}
&\sum_{i=1}^n K_i(|v_i^* - \tau_n^{-1}\eta_i^T\theta| - |v_i^*|) \\
&\quad = \quad -\theta^T(\tau_n^{-1}\sum_{i=1}^n \eta_i K_i \mathrm{sign}(v_i^*)) \\
&\qquad\qquad +\theta^T\Big\{\tau_n^{-2}\sum_{i=1}^n \eta_i\eta_i^T K_i f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)\Big(\frac{\partial v}{\partial u}(x_0, m_u)\Big)^{-1}\Big\}\theta + o_p(1).
\end{aligned}
$$

Remark 2. We compare Proposition 2 and Lemma A.1 of [4]. From Proposition 1, we have

$$
\tau_n^{-2}\sum_{i=1}^n \eta_i\eta_i^T K_i f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)\Big(\frac{\partial v}{\partial u}(x_0, m_u)\Big)^{-1} \stackrel{d}{\to} \begin{pmatrix} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{pmatrix} f_v(0|x_0)L(1,0)
$$

and $L(1,0)$ is a random variable. However, $\{nh(\theta-\theta_0)^2\lambda_1(z_0)\tau\phi(0)\nu_1/(2d_n)\}$ in Lemma A.1 of [4] is a sequence of constants converging to a constant. Note that their $\theta - \theta_0$ is not normalized and that they considered the LCE. This means Proposition 2 contradict with Lemma A.1 of [4]. In the case of stationary observations, we obtain the uniformity in θ from the pointwise convergence by employing the convexity lemma for random functions given in [19] immediately. However, we cannot apply the convexity lemma in [19] as it is because of the above convergence in distribution to a random variable. Thus we need the almost sure representation in Remark 1 to obtain the uniformity in θ in Proposition 4 below.

Proposition 3 is about the bias term of the LLE.

Proposition 3. *Suppose that Assumptions V, X, U1, U2, K, and H hold. Then we have*

$$\begin{aligned} & h^{-2}\tau_n^{-2}\sum_{i=1}^{n} K_i\eta_i(\mathrm{sign}(v_i^*)-\mathrm{sign}(v_i)) \\ = \ & \tau_n^{-2}g''(x_0)\sum_{i=1}^{n}\Big(\frac{X_i-x_0}{h}\Big)^2 K_i\eta_i f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)\Big(\frac{\partial v}{\partial u}(x_0,m_u)\Big)^{-1} \\ & +O_p\Big(\frac{1}{(nh^6)^{1/4}}\Big)+o_p(1). \end{aligned}$$

Proposition 4 is a version of the convexity lemma in Pollard [19] adapted to the setup of this paper. In Proposition 4, we use the almost sure representation of the convergence in distribution in Proposition 1. See Remark 2 above. Note that the convergence in probability in Proposition 4 is from the almost sure representation and is correct.

Proposition 4. *Suppose that Assumptions V, X, U1, U2, K, and H hold. Then for any compact subset K of R^2, we have*

$$\sup_{\theta\in K}\Big|\sum_{i=1}^{n} K_i(|v_i^*-\tau_n^{-1}\eta_i^T\theta|-|v_i^*|)+\theta^T(\tau_n^{-1}\sum_{i=1}^{n}\eta_i K_i\mathrm{sign}(v_i^*))-\theta^T A\theta\Big| \xrightarrow{p} 0,$$

where

$$\begin{aligned} A &= \lim_{n\to\infty}\tau_n^{-2}\sum_{i=1}^{n} K_i\eta_i\eta_i^T f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)\Big(\frac{\partial v}{\partial u}(x_0,m_u)\Big)^{-1} \quad a.s. \\ &= \begin{pmatrix} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{pmatrix} f_v(0|x_0)L(1,0) \quad a.s. \end{aligned}$$

Proposition 5 is necessary to examine the bias term of the LCE. Recall the definition of v_i^{**} in (10).

Proposition 5. *Suppose that Assumptions V, X, U1, U2, K, and H hold. Then we have*

$$
\begin{aligned}
& h^{-2}\tau_n^{-2}\sum_{i=1}^n K_i(\mathrm{sign}(v_i^{**}) - \mathrm{sign}(v_i)) \\
&= \tau_n^{-2}\sum_{i=1}^n \Big(\frac{X_i - x_0}{h}\Big)^2 K_i\Big[g''(x_0)f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)\Big(\frac{\partial v}{\partial u}(x_0, m_u)\Big)^{-1} \\
&\quad +2g'(x_0)f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)\frac{\partial}{\partial x}\Big(\frac{\partial v}{\partial u}(x_0, m_u)\Big)^{-1} \\
&\quad -(g'(x_0))^2\Big(\frac{\partial v}{\partial u}(x_0, m_u)\Big)^{-2}\Big\{f'_{u_i}(m_u|\mathcal{E}_{i-m_0}^i) \\
&\quad +f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)\frac{\partial}{\partial u}\Big(\frac{\partial v}{\partial u}(x_0, m_u)\Big)^{-1}\Big\}\Big] + O_p\Big(\frac{1}{(nh^8)^{1/4}}\Big) + o_p(1).
\end{aligned}
$$

Remark 3. It is easy to see that Proposition 2 holds for any $\theta \in R$ with v_i^* replaced by v_i^{**} and without η_i. Proposition 4 is also true with the same changes.

We prove Theorem 1 only for the LLE by exploiting Propositions 1-4. We can deal with the LCE similarly by employing Proposition 5 instead of Proposition 3.

Proof of Theorem 1. We consider all the random variables on $\tilde{\Omega}$ given in Remark 1. Taking a compact subset K of R^2, we have from Propositions 1 and 4 that uniformly in θ on K,

$$
\begin{aligned}
& \sum_{i=1}^n K_i(|v_i^* - \tau_n^{-1}\eta_i^T\theta| - |v_i^*|) \qquad (12) \\
&= -\theta^T(\tau_n^{-1}\sum_{i=1}^n \eta_i K_i \mathrm{sign}(v_i^*)) + \theta^T \begin{pmatrix} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{pmatrix} \theta f_v(0|x_0)L(1,0) + o_p(1).
\end{aligned}
$$

We evaluate the first term of the RHS of (12) by combining Propositions 1 and 3 and get

$$
\begin{aligned}
& \tau_n^{-1}\sum_{i=1}^n K_i\eta_i \mathrm{sign}(v_i^*) \qquad (13) \\
&= \tau_n^{-1}\sum_{i=1}^n K_i\eta_i \mathrm{sign}(v_i)
\end{aligned}
$$

$$
\begin{aligned}
& +\frac{h^2 g''(x_0)}{\tau_n}\sum_{i=1}^{n}\Big(\frac{X_i-x_0}{h}\Big)^2 K_i\eta_i f_{u_i}(m_u|\mathcal{E}^i_{i-m_0})\Big(\frac{\partial v}{\partial u}(x_0,m_u)\Big)^{-1} \\
& +O_p\Big(\frac{\tau_n h^2}{(nh^6)^{1/4}}\Big)+o_p(1) \\
= & \left(\begin{array}{c} Z_1 \\ Z_2 \end{array}\right) L^{1/2}(1,0)+\tau_n h^2 g''(x_0)\left(\begin{array}{c} \kappa_2 \\ \kappa_3 \end{array}\right) f_v(0|x_0)L(1,0) \\
& +O_p\Big(\frac{\tau_n h^2}{(nh^6)^{1/4}}\Big)+o_p(1) \\
= & O_p(1).
\end{aligned}
$$

Since $L(1,0)$ is a random variable, we have to modify the standard argument about quantile regression.

We fix a small positive δ_1 and take a sufficiently small δ_2 s.t. $\mathrm{P}(\delta_2 < L(1,0) < 1/\delta_2) > 1-\delta_1$. Then setting $\tilde{\Omega}_{\delta_2} = \{\delta_2 < L(1,0) < 1/\delta_2\}$, we temporarily consider the conditional probability given $\tilde{\Omega}_{\delta_2}$.

Here we define Θ_M for a positive M by $\{\theta \in R^2 \,|\, \theta^T\theta = M^2\}$. Notice that we have (12) uniformly on Θ_M and inside Θ_M. By (13), the second term of the RHS of (12) is dominant on Θ_M with conditional probability arbitrarily close to 1 on $\tilde{\Omega}_{\delta_2}$ when we take a sufficiently large M. Hence the convexity of the objective function imply that $\hat{\theta}$ must be inside Θ_M with conditional probability arbitrarily close to 1 on $\tilde{\Omega}_{\delta_2}$.

As in [10] and [12], we can take any large M and minimize (12) inside Θ_M. Then from the uniformity of (12) and the second equation of (13), we have that given $\tilde{\Omega}_{\delta_2}$,

$$
\begin{aligned}
\hat{\theta} = & \frac{1}{2}(f_v(0|x_0)L^{1/2}(1,0))^{-1}\left(\begin{array}{cc} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{array}\right)^{-1}\left(\begin{array}{c} Z_1 \\ Z_2 \end{array}\right) \\
& +\frac{\tau_n h^2 g''(x_0)}{2}\left(\begin{array}{cc} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{array}\right)^{-1}\left(\begin{array}{c} \kappa_2 \\ \kappa_3 \end{array}\right)+O_p\Big(\frac{\tau_n h^2}{(nh^6)^{1/4}}\Big)+o_p(1).
\end{aligned} \tag{14}
$$

Since we can choose an arbitrarily small positive δ_1, we also have (14) on $\tilde{\Omega}$. Hence the proof of Theorem 1 is complete.

In Remark 4 below, we describe how to deal with a robust local linear estimator defined by a convex loss function.

Remark 4. Suppose that we define the LLE by using a convex loss function $\rho(v)$ instead of $|v|$. We assume that $\rho(0) = 0$ and $\rho(v) \geq 0$ and that $\rho(v)$

is differentiable except at the origin. In addition, we have $\mathrm{E}\{\rho'(v_i)\} = 0$. Then we have to make some changes to Propositions 2 and 3. Let ξ and δ be a generic random variable with density $f_\xi(\xi)$ and a constant tending to 0, respectively.

In Proposition 2, we deal with $\rho(\xi - \delta) - \rho(\xi) + \delta\rho'(\xi)$ and we need (15) and (16) below to establish the proposition.

$$\mathrm{E}\{|\rho(\xi - \delta) - \rho(\xi) + \delta\rho'(\xi)|^2\} = o(\delta^2) \tag{15}$$

$$\mathrm{E}\{\rho(\xi - \delta) - \rho(\xi) + \delta\rho'(\xi)\} = \delta^2 s_1(f_\xi) + o(\delta^2), \tag{16}$$

where $s_1(f_\xi)$ is a functional of a density function and satisfies the regularity conditions necessary in the proof of Proposition 2 given in section 5.

In Proposition 3, we consider $\rho'(\xi + \delta) - \rho'(\xi)$ and we need (17) and (18) below to establish the proposition.

$$\mathrm{E}\{|\rho'(\xi + \delta) - \rho'(\xi)|^2\} = O(\delta) \tag{17}$$

$$\mathrm{E}\{\rho'(\xi + \delta) - \rho'(\xi)\} = \delta s_2(f_\xi) + o(\delta), \tag{18}$$

where $s_2(f_\xi)$ is a functional of a density function and satisfies the regularity conditions necessary in the proof of Proposition 3 given in section 5.

When we have (15)-(18) for $\rho(v)$, we can establish the same result as in Theorem 1. However, $f_\xi(\xi)$ is $f_{v_i}(v|\mathcal{E})$ in the propositions and $f_{v_i}(v|\mathcal{E})$ depends on X_i and $\mathcal{E}^i_{i-m_0}$ in a complicated way. Therefore we have to impose much more restrictive assumptions on $f_{v_i}(v|\mathcal{E})$ or $f_{u_i}(u|\mathcal{E}^i_{i-m_0})$ to obtain the same results for a general $\rho(v)$ than for a specific $\rho(v)$ such as $|v|$. Thus we decided to focus on LAD regression in this paper.

When $\rho(v) = |v|^q$ for some $1 < q < 2$, it is easy to verify (15) and (17). We also have

$$\begin{aligned} s_1(f_\xi) &= \frac{1}{2}\int |\xi|^q f''_\xi(\xi)d\xi = -\frac{q}{2}\int |\xi|^{q-1} f'_\xi(\xi)d\xi, \\ s_2(f_\xi) &= -q\int |\xi|^{q-1} f'_\xi(\xi)d\xi \end{aligned}$$

with some conditions on $f_\xi(\xi)$. We will also need some assumptions on $f_{v_i}(v|\mathcal{E})$ or $f_{u_i}(u|\mathcal{E}^i_{i-m_0})$ to get the same results as in Propositions 2 and 3 and those assumptions will depend on q.

5. Proofs of propositions

In this section, we give the proofs of Propositions 1-5.

Proof of Proposition 1. First note the Fourier transforms of $s^j K(s)$, $j = 1, 2, 3$, are integrable from Assumption K. Besides, $f_{u_i}(m_u|\mathcal{E}^i_{i-m_0})$ satisfies Assumption 2 of [22] and we obtain the same result as in Proposition 7.2 of [22] for $\{(X_i - x_0)/h\}^j K_i f_{u_i}(m_u|\mathcal{E}^i_{i-m_0})$, $j = 0, 1, 2, 3$.

Applying the arguments on pp.1922-1924 and Proposition 7.2 of [22] at the same time to

$$\tau_n^{-1} \sum_{i=1}^n \Big\{ aK_i + b\Big(\frac{X_i - x_0}{h}\Big) K_i \Big\} \mathrm{sign}(u_i - m_u), \tag{19}$$

where a and b are arbitrary constants, and $\tau_n^{-2} \sum_{i=1}^n K_i$, $\tau_n^{-2} \sum_{i=1}^n \{(X_i - x_0)/h\}^j K_i f_{u_i}(m_u|\mathcal{E}^i_{i-m_0})$, $j = 0, 1, 2, 3$, we have the same result as in Proposition 1 with the first two elements of the both sides replaced with (19) and

$$(a^2\nu_0 + 2ab\nu_1 + b^2\nu_2)^{1/2} L^{1/2}(1, 0) Z, \tag{20}$$

respectively. Note that Z in (20) has the standard normal distribution and is independent of $L(1, 0)$. Since a and b are arbitrary constants, the desired result follows from the Cramér-Wold device. Hence the proof of Proposition 1 is complete.

Proof of Proposition 2. Set

$$B_{2i}(\theta) = |v_i^* - \tau_n^{-1}\eta_i^T\theta| - |v_i^*| + \tau_n^{-1}\eta_i^T\theta\,\mathrm{sign}(v_i^*)$$

and notice

$$|B_{2i}(\theta)| \le C\tau_n^{-1}|\eta_i^T\theta| I(|v_i^*| \le C\tau_n^{-1}|\eta_i^T\theta|). \tag{21}$$

We also set

$$D_{2i}(\theta) = B_{2i}(\theta) - \mathrm{E}\{B_{2i}(\theta)|\mathcal{E}\}.$$

First we evaluate $\sum_{i=1}^n K_i D_{2i}(\theta)$. From (21) and Assumption U2, we have

$$\mathrm{E}\{D_{2i}^2(\theta)|\mathcal{E}\} \le C\tau_n^{-2}\mathrm{E}\{I(|v_i^*| \le C\tau_n^{-1}|\eta_i^T\theta|)|\mathcal{E}\} \le C\tau_n^{-3}. \tag{22}$$

Assumption U1, (22), and (5.19) of [21] imply

$$\begin{aligned}
&\mathrm{E}\Big[\Big\{\sum_{i=1}^{n} K_i D_{2i}(\theta)\Big\}^2\Big] \qquad (23)\\
&\le \mathrm{E}\Big[\sum_{i=1}^{n} K_i^2 \mathrm{E}\{D_{2i}^2(\theta)|\mathcal{E}\}\Big] + \mathrm{E}\Big[\sum_{|i-i'|\le m_0} K_i K_{i'} \mathrm{E}\{D_{2i}(\theta)D_{2i'}(\theta)|\mathcal{E}\}\Big]\\
&\le C\mathrm{E}\Big\{\tau_n^{-3}\sum_{i=1}^{n} K_i^2\Big\} = O(\tau_n^{-1}).
\end{aligned}$$

Next we evaluate $\sum_{i=1}^n K_i \mathrm{E}\{B_{2i}(\theta)|\mathcal{E}\}$. From Assumption U2 and the standard calculation, we obtain uniformly in i,

$$\mathrm{E}\{B_{2i}(\theta)|\mathcal{E}\} = \tau_n^{-2}(\eta_i^T\theta)^2 f_{u_i}(m_u|\mathcal{E}^i_{i-m_0})\Big(\frac{\partial v}{\partial u}(X_i, m_u)\Big)^{-1} + o_p(\tau_n^{-2}). \qquad (24)$$

The desired result follows from (23), (24), Assumption V, and Proposition 1. Hence the proof of Proposition 2 is complete.

Proof of Proposition 3. We can establish Proposition 3 almost in the same way as Proposition 2. Set

$$B_{3i} = \mathrm{sign}(v_i^*) - \mathrm{sign}(v_i) \quad \text{and} \quad D_{3i} = B_{3i} - \mathrm{E}\{B_{3i}|\mathcal{E}\}.$$

Notice that

$$|\mathrm{sign}(v_i^*) - \mathrm{sign}(v_i)| \le CI(|v_i| \le Ch^2).$$

Hence we have

$$\mathrm{E}\{|D_{3i}|^2|\mathcal{E}\} \le Ch^2.$$

The above inequality and the same argument for (23) yield

$$h^{-2}\tau_n^{-2}\sum_{i=1}^{n} K_i \eta_i D_{3i} = O_p((nh^6)^{-1/4}). \qquad (25)$$

By some calculation, we get uniformly in i,

$$\begin{aligned}
&h^{-2}\mathrm{E}\{B_{3i}|\mathcal{E}\} \qquad (26)\\
&= \Big(\frac{X_i - x_0}{h}\Big)^2 g''(\bar{X}_i) f_{u_i}(m_u|\mathcal{E}^i_{i-m_0})\Big(\frac{\partial v}{\partial u}(X_i, m_u)\Big)^{-1} + o_p(1).
\end{aligned}$$

The desired result follows from (25), (26), the continuity of $g''(x)$ at x_0, Assumption V, and Proposition 1. Hence the proof of Proposition 3 is complete.

Proof of Proposition 4. We verify this proposition by modifying the proof of the convex lemma in Pollard [19].

From Propositions 1 and 2, we have for any fixed $\theta \in K$,

$$\Big|\sum_{i=1}^{n} K_i B_{2i}(\theta) - \theta^T A\theta\Big| \xrightarrow{p} 0. \tag{27}$$

As in the proof of Theorem 1, choose a small positive δ_3 and take δ_4 s.t. $\mathrm{P}(\delta_4 < L(1,0) < 1/\delta_4) > 1-\delta_3$. Then we set $\tilde{\Omega}_{\delta_4} = \{\delta_4 < L(1,0) < 1/\delta_4\}$.

On $\tilde{\Omega}_{\delta_4}$, we can take δ-cubes on p.197 of [19] for any small positive ϵ. Then $\theta^T A\theta$ varies by less than ϵ in each of the δ-cubes. Since we have δ-cubes, we can proceed exactly in the same way as on pp.197-198 of [19]. Thus from (27) and the convexity of $\sum_{i=1}^n K_i B_{2i}(\theta)$ and $\theta^T A\theta$, we have that given $\tilde{\Omega}_{\delta_4}$,

$$\sup_{\theta\in K}\Big|\sum_{i=1}^{n} K_i B_{2i}(\theta) - \theta^T A\theta\Big| \xrightarrow{p} 0. \tag{28}$$

Since we can choose any small δ_3, we have (28) on $\tilde{\Omega}$. Hence the proof of Proposition 4 is complete.

Proof of Proposition 5. Set

$$\delta_i^{**} = v_i^{**} - v_i = (X_i - x_0)g'(x_0) + \frac{1}{2}\Big(\frac{X_i - x_0}{h}\Big)^2 h^2 g''(\bar{X}_i), \tag{29}$$
$$B_{4i} = \mathrm{sign}(v_i^{**}) - \mathrm{sign}(v_i), \text{ and } D_{4i} = B_{4i} - \mathrm{E}\{B_{4i}|\mathcal{E}\}.$$

Since

$$|B_{4i}| \le CI(|v_i| \le Ch),$$

we have

$$\mathrm{E}\{|D_{4i}|^2|\mathcal{E}\} \le Ch. \tag{30}$$

From (30) and the same argument as in the proofs of Propositions 2 and 3, we obtain

$$h^{-2}\tau_n^{-2}\sum_{i=1}^{n} K_i D_{4i} = O_p((nh^8)^{-1/4}). \tag{31}$$

Next we consider $\mathrm{E}\{B_{4i}|\mathcal{E}\}$. By some calculation, we have uniformly in i,

$$h^{-2}\mathrm{E}\{B_{4i}|\mathcal{E}\} = 2h^{-2}\delta_i^{**} f_{v_i}(0|\mathcal{E}) - h^{-2}(\delta_i^{**})^2 f'_{v_i}(0|\mathcal{E}) + o_p(1) \tag{32}$$

We evaluate the first and second terms of the RHS of (32). By some calculation, we obtain

$$
\begin{aligned}
& 2h^{-2}\tau_n^{-2}\sum_{i=1}^{n} K_i\delta_i^{**} f_{v_i}(0|\mathcal{E}) && (33)\\
= \ & \tau_n^{-2}\sum_{i=1}^{n}\Big(\frac{X_i - x_0}{h}\Big)^2 K_i g''(x_0) f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)\Big(\frac{\partial v}{\partial u}(x_0, m_u)\Big)^{-1} \\
& +2\tau_n^{-2}\sum_{i=1}^{n}\Big(\frac{X_i - x_0}{h}\Big)^2 K_i g'(x_0) f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)\frac{\partial}{\partial x}\Big(\frac{\partial v}{\partial u}(x_0, m_u)\Big)^{-1} \\
& +o_p(1)
\end{aligned}
$$

We used Theorem 2.1 of [23] to evaluate $\sum_{i=1}^n \{(X_i - x_0)/h\}K_i$ here.

We give a representation of $f'_{v_i}(0|\mathcal{E})$ by Assumptions V and U2 and some calculation before we evaluate the second term of (32).

$$
\begin{aligned}
& f'_{v_i}(0|\mathcal{E}) && (34)\\
= \ & \Big(\frac{\partial v}{\partial u}(X_i, m_u)\Big)^{-2}\Big(f'_{u_i}(m_u|\mathcal{E}_{i-m_0}^i) + f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)\frac{\partial}{\partial u}\Big(\frac{\partial v}{\partial u}(X_i, m_u)\Big)^{-1}\Big).
\end{aligned}
$$

From (29), (34), and Assumption V, we have

$$
\begin{aligned}
& h^{-2}\tau_n^{-2}\sum_{i=1}^{n} K_i(\delta_i^{**})^2 f'_{v_i}(0|\mathcal{E}) && (35)\\
= \ & \tau_n^{-2}\sum_{i=1}^{n}\Big(\frac{X_i - x_0}{h}\Big)^2 K_i\Big(\frac{\partial v}{\partial u}(x_0, m_u)\Big)^{-2}\Big\{f'_{u_i}(m_u|\mathcal{E}_{i-m_0}^i) \\
& +f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)\frac{\partial}{\partial u}\Big(\frac{\partial v}{\partial u}(x_0, m_u)\Big)^{-1}\Big\} + o_p(1).
\end{aligned}
$$

Proposition 5 follows from (31), (32), (33), (35), and Proposition 1. Hence the proof of Proposition 5 is complete.

Acknowledgments

This research is partially supported by the Global COE Program Research Unit for Statistical and Empirical Analysis in Social Sciences at Hitotsubashi University, Japan.

References

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

[2] Z. Cai, Regression quantiles for time series, Econom. Theory 18 (2002) 169-192.

[3] J. Chen, J. Gao, D. Li, Semiparametric regression estimation in null recurrent nonlinear time series, Working paper available at http://www.adelaide.edu.au/directory/jiti.gao#Publications, 2008.

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

[5] R. Engle, C.W.J. Granger, Cointegration and error correction: Representation, estimation and testing, Econometrica 55 (1987) 251-276.

[6] J. Fan, Q. Yao, Nonlinear Time Series: Nonparametric and Parametric Methods, Springer, New York, 2003.

[7] J. Gao, Nonlinear Time Series: Semiparametric and Nonparametric Methods, Chapman & Hall/CRC, London, 2007.

[8] J. Gao, M. King, Z. Lu, D. Tjøstheim, Specification testing in nonlinear and nonstationary time series autoregression, Ann. Statist. 37 (2009) 3893-3928.

[9] C.W.J. Granger, Some properties of time series data and their use in econometric model specification, J. Econometrics 16 (1981) 121-130.

[10] P. Hall, L. Peng, Q. Yao, Prediction and nonparametric estimation for time series with heavy tails, J. Time Series Anal. 23 (2002) 313-331.

[11] T. Honda, Nonparametric estimation of a conditional quantile for α-mixing processes, Ann. Inst. Statist. Math. 52 (2000) 459-470.

[12] T. Honda, Nonparametric estimation of conditional medians for linear and related processes, Ann. Inst. Statist. Math. 62 (2010) 995-1021.

[13] H.A. Karlsen, D. Tjøstheim, Nonparametric estimation in null recurrent time series, Ann. Statist. 29 (2001) 372-416.

[14] H.A. Karlsen, T. Mykelbust, D. Tjøstheim, Nonparametric estimation in a nonlinear cointegration type model, Ann. Statist. 35 (2007) 252-299.

[15] R. Koenker, Quantile Regression, Cambridge University Press, New York, 2005.

[16] R. Koenker, G. Basset, Regression quantiles, Econometrica 46 (1978) 33-50.

[17] Q. Li, J. Racine, Nonparametric Econometrics: Theory and Practice, Princeton University Press, Princeton, 2007.

[18] Z. Lin, D. Li, J. Chen, Local linear M-estimators in null recurrent time series, Statist. Sin. 19 (2009) 1683-1703.

[19] D. Pollard, Asymptotics for least absolute deviation regression estimates, Econom. Theory 7 (1991) 186-198.

[20] M. Schienle, Nonparametric nonstationary regression, Unpublished Ph.D. thesis, Mannheim University, 2008.

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

[22] Q.Y. Wang, P.C.B. Phillips, Structural nonparametric cointegrating regression, Econometrica 77 (2009) 1901-1948.

[23] Q.Y. Wang, P.C.B. Phillips, Asymptotic theory for zero energy functionals with nonparametric regression applications, Econom. Theory 27 (2011) 235-259.

[24] Q.Y. Wang, P.C.B. Phillips, Specification testing for nonlinear cointegrating regression, Cowles Foundation Discussion Paper No.1779, 2011.

[25] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, Springer, New York, 1996.

[26] Z. Xiao, Quantile cointegrating regression, J. Econometrics 150 (2009) 248-260.

Supplemetary material

Details of (22), (24), (26), (32), and (33) are given here. This is written for reviewers. I will remove this and make this available on the internet when the paper is formaly published.

(22), (24), (26):

We can establish (22), (24), and 26) by combining the standard arguments in the literature of nonparametric quantile regerssion. First put

$$a_i = \frac{1}{2}\Big(\frac{X_i - x_0}{h}\Big)^2 h^2 g''(\bar{X}_i) \text{ and } b_i = \tau_n^{-1}\eta_i^T\theta.$$

and notice that a_i and b_i tends to 0 uniformly in i since we can assume $|X_i - x_0| \leq Ch$.

(22): v_i^* is defined in (9) as $v_i^* = v_i + a_i$. Then

$$|v_i^*| \leq C|b_i| \Rightarrow -C|b_i| - a_i \leq v_i \leq C|b_i| - a_i.$$

Recall that $a_i/\tau_n^{-1} = O(1)$ uniformly in i from Assumption H. Hence we obtain (22) from from Assumptions V and U2.

(24): When $a_i \geq 0$ and $b_i \geq 0$, $B_{2i}(\theta)$ is not 0 only when $-a_i \leq v_i \leq -a_i + b_i$. Then we have

$$B_{2i}(\theta) = -2(v_i + a_i - b_i)$$

and

$$-2\int_{-a_i}^{-a_i+b_i}(v_i + a_i - b_i)f_{v_i}(v_i|\mathcal{E})dv_i = b_i^2 f_{v_i}(0|\mathcal{E}) + o_p(b_i^2)$$

uniformly in i from Assumption V and U2. We can deal with the other cases in the same way.

(26): When $a_i > 0$, we have

$$\text{sign}(v_i^*) - \text{sign}(v_i) = 2I(-a_i < v_i < 0)$$

and

$$2(F_{v_i}(0|\mathcal{E}) - F_{v_i}(-a_i|\mathcal{E})) = 2a_i f_{v_i}(0|\mathcal{E}) + o_p(a_i)$$

uniformly in i from Assumptions V and U2. We can deal with the other case in the same way.

(32), (33):

It is not easy to establish (32) and (33).

(32): Recall that

$$v_i^{**} = v_i + \delta_i^{**} \quad \text{and} \quad \delta_i^{**} = O(h).$$

When $\delta_i^{**} > 0$, we have

$$\mathrm{sign}(v_i^{**}) - \mathrm{sign}(v_i) = 2I(-\delta_i^{**} < v_i < 0)$$

and

$$\begin{aligned} & 2h^{-2}(F_{v_i}(0|\mathcal{E}) - F_{v_i}(-\delta_i^{**}|\mathcal{E})) \\ = \;& 2h^{-2}\delta_i^{**} f_{v_i}(0|\mathcal{E}) - h^{-2}(\delta_i^{**})^2 f'_{v_i}(0|\mathcal{E}) + O(|f'_{v_i}(0|\mathcal{E}) - f'_{v_i}(\bar{\delta}_i^{**}|\mathcal{E})|), \end{aligned}$$

where $\bar{\delta}_i^{**}$ is between 0 and δ_i^{**}.

Assumption V and (34) imply that

$$\begin{aligned} & |f'_{v_i}(0|\mathcal{E}) - f'_{v_i}(\bar{\delta}_i^{**}|\mathcal{E})| \\ \leq \;& C(\sup_{|m_u - m| \leq Ch} |f'_{u_i}(m_u|\mathcal{E}^i_{i-m_0}) - f'_{u_i}(m|\mathcal{E}^i_{i-m_0})| + o(1)) \end{aligned}$$

uniformly in i. Since

$$\lim_{h \to 0} \mathrm{E}\{\sup_{|m_u - m| \leq Ch} |f'_{u_i}(m_u|\mathcal{E}^i_{i-m_0}) - f'_{u_i}(m|\mathcal{E}^i_{i-m_0})|\} = 0,$$

we get from [22] (for example, see Proposition 1 of this paper) that

$$\tau_n^{-2} \sum_{i=1}^n K_i |f'_{v_i}(0|\mathcal{E}) - f'_{v_i}(\bar{\delta}_i^{**}|\mathcal{E})| = o_p(1).$$

(33): First write

$$\begin{aligned} f_{v_i}(0|\mathcal{E}) &= f_{u_i}(m_u|\mathcal{E}^i_{i-m_0})\Big(\frac{\partial v}{\partial u}(X_i, m_u)\Big)^{-1} \\ &= f_{u_i}(m_u|\mathcal{E}^i_{i-m_0})(w(x_0) + (X_i - x_0)w'(x_0) + o(|X_i - x_0|)), \end{aligned}$$

where $w(x)$ is clearly defined in the above equation. Using the above notaion, we have

$$
\begin{aligned}
&\frac{2h^{-2}}{\tau_n^2}\sum_{i=1}^{n} K_i\delta_i^{**} f_{v_i}(0|\mathcal{E}) \\
&= \frac{2h^{-2}}{\tau_n^2}\sum_{i=1}^{n} K_i\Big\{\frac{X_i-x_0}{h}hg'(x_0)+\frac{1}{2}\Big(\frac{X_i-x_0}{h}\Big)^2 h^2 g''(\bar{X}_i)\Big\} \\
&\quad \times f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)(w(x_0)+(X_i-x_0)w'(x_0)+o(|X_i-x_0|)) \\
&= 2\tau_n^{-2}\sum_{i=1}^{n} K_i f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)\Big\{\Big(\frac{X_i-x_0}{h}\Big)^2 g'(x_0)w'(x_0) \qquad (36) \\
&\quad +\frac{1}{2}\Big(\frac{X_i-x_0}{h}\Big)^2 g''(x_0)w(x_0)+o(h^2)\Big\} \\
&\quad +\frac{2h^{-1}}{\tau_n^2}\sum_{i=1}^{n} K_i\frac{X_i-x_0}{h} f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)g'(x_0)w(x_0)+o_p(1).
\end{aligned}
$$

We can handle the first term of (36) by using Proposition 1.

Finally we consider the second term of (36). Write

$$
\begin{aligned}
&\frac{h^{-1}}{\tau_n^2}\sum_{i=1}^{n} K_i\frac{X_i-x_0}{h} f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i) \\
&= \frac{\tau_n^{-1}}{(nh^6)^{1/4}}\sum_{i=1}^{n} K_i\frac{X_i-x_0}{h}\{(f_{u_i}(m_u|\mathcal{E}_{i-m_0}^i)-f_u(m_u))+f_u(m_u)\}. \qquad (37)
\end{aligned}
$$

We can use Theorem 2.1 of [23] to show $\tau_n^{-1}\sum_{i=1}^{n}\{(X_i-x_0)/h\}K_i = O_p(1)$. We can deal with the first term inside the braces of (37) by using a result similar to the second element of Proposition 1.

Table1 LAD for N(0,1)

h		$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$	$n^{-1/7}$	$n^{-1/8}$	$n^{-1/9}$	$n^{-1/10}$
x=0	Bias	0.006	0.008	0.016	0.019	0.028	0.036	0.044
	MSE	0.353	0.360	0.271	0.242	0.207	0.163	0.126
	SE1	0.050	0.062	0.047	0.047	0.037	0.031	0.009
	N2	1874	1294	1023	850	738	677	624
	N1	1263	871	672	569	508	467	428
	MISE	0.163	0.143	0.125	0.102	0.100	0.093	0.086
	SE2	0.005	0.007	0.006	0.004	0.006	0.004	0.003
	NU	3587	2415	1817	1519	1288	1154	1066
x=5	Bias	0.026	0.023	0.025	0.033	0.036	0.042	0.052
	MSE	0.415	0.368	0.264	0.258	0.229	0.230	0.222
	SE1	0.063	0.080	0.024	0.022	0.016	0.017	0.016
	N2	2663	2121	1852	1690	1577	1503	1458
	N1	2075	1714	1512	1407	1337	1287	1241
	MISE	0.173	0.157	0.122	0.104	0.209	0.097	0.087
	SE2	0.006	0.009	0.004	0.003	0.082	0.005	0.003
	NU	4329	3237	2687	2377	2185	2057	1972
x=10	Bias	0.034	0.042	0.053	0.054	0.050	0.057	0.069
	MSE	0.527	0.474	0.403	0.428	0.778	0.503	0.517
	SE1	0.055	0.040	0.026	0.036	0.263	0.080	0.054
	N2	3629	3109	2850	2674	2575	2517	2477
	N1	3072	2707	2537	2416	2348	2302	2273
	MISE	0.192	0.175	0.278	0.119	0.149	0.134	0.141
	SE2	0.015	0.030	0.151	0.008	0.031	0.028	0.029
	NU	5218	4194	3673	3374	3177	3038	2947

Table2 LS for N(0,1)

h		$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$	$n^{-1/7}$	$n^{-1/8}$	$n^{-1/9}$	$n^{-1/10}$
x=0	Bias	0.009	0.012	0.025	0.035	0.048	0.059	0.071
	MSE	0.252	0.241	0.165	0.123	0.113	0.109	0.081
	SE1	0.054	0.055	0.049	0.028	0.023	0.023	0.007
	N2	1874	1294	1023	850	738	677	624
	N1	1263	871	672	569	508	467	428
	MISE	0.090	0.080	0.070	0.058	0.058	0.053	0.051
	SE2	0.004	0.005	0.005	0.003	0.004	0.002	0.002
	NU	3587	2415	1817	1519	1288	1154	1066
x=5	Bias	0.025	0.032	0.038	0.051	0.058	0.071	0.081
	MSE	0.276	0.187	0.195	0.194	0.172	0.168	0.179
	SE1	0.033	0.017	0.018	0.015	0.013	0.013	0.014
	N2	2663	2121	1852	1690	1577	1503	1458
	N1	2075	1714	1512	1407	1337	1287	1241
	MISE	0.095	0.081	0.067	0.059	0.058	0.056	0.055
	SE2	0.004	0.004	0.003	0.002	0.002	0.002	0.002
	NU	4329	3237	2687	2377	2185	2057	1972
x=10	Bias	0.040	0.045	0.057	0.061	0.070	0.075	0.095
	MSE	0.360	0.397	0.351	0.356	0.425	0.423	0.441
	SE1	0.032	0.041	0.027	0.030	0.067	0.069	0.046
	N2	3629	3109	2850	2674	2575	2517	2477
	N1	3072	2707	2537	2416	2348	2302	2273
	MISE	0.096	0.115	0.070	0.068	0.089	0.091	0.092
	SE2	0.004	0.035	0.003	0.004	0.028	0.028	0.028
	NU	5218	4194	3673	3374	3177	3038	2947

Table3 LAD for t(3)

h		$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$	$n^{-1/7}$	$n^{-1/8}$	$n^{-1/9}$	$n^{-1/10}$
x=0	Bias	-0.009	0.006	0.022	0.027	0.035	0.047	0.047
	MSE	0.800	0.276	0.261	0.188	0.163	0.271	0.150
	SE1	0.463	0.080	0.089	0.071	0.069	0.112	0.069
	N2	1796	1274	984	812	734	657	618
	N1	1215	846	670	565	490	440	408
	MISE	0.204	0.168	0.079	0.070	0.061	0.064	0.065
	SE2	0.059	0.070	0.004	0.004	0.004	0.009	0.010
	NU	3537	2403	1853	1491	1302	1162	1069
x=5	Bias	0.019	0.036	0.035	0.042	0.044	0.049	0.054
	MSE	0.316	0.593	0.201	0.205	0.169	0.210	0.233
	SE1	0.038	0.317	0.018	0.020	0.010	0.020	0.033
	N2	2671	2175	1894	1748	1644	1555	1489
	N1	2148	1720	1550	1438	1368	1303	1267
	MISE	0.133	0.165	0.076	0.074	0.078	0.070	0.143
	SE2	0.017	0.050	0.003	0.007	0.007	0.006	0.076
	NU	4379	3298	2758	2459	2259	2110	1991
x=10	Bias	0.009	0.024	0.031	0.016	0.023	0.041	0.052
	MSE	0.363	0.402	0.401	3.104	0.554	0.648	0.725
	SE1	0.021	0.031	0.031	2.700	0.129	0.163	0.179
	N2	3625	3106	2860	2690	2601	2518	2460
	N1	3061	2711	2532	2430	2353	2311	2268
	MISE	0.156	0.105	0.085	0.070	0.088	0.084	0.079
	SE2	0.021	0.006	0.005	0.002	0.018	0.017	0.012
	NU	5250	4197	3692	3382	3184	3065	2973

Table4 LS for t(3)

h		$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$	$n^{-1/7}$	$n^{-1/8}$	$n^{-1/9}$	$n^{-1/10}$
x=0	Bias	-0.003	0.013	0.029	0.041	0.049	0.061	0.071
	MSE	0.576	0.161	0.201	0.165	0.148	0.144	0.143
	SE1	0.389	0.026	0.079	0.075	0.074	0.074	0.074
	N2	1796	1274	984	812	734	657	618
	N1	1215	846	670	565	490	440	408
	MISE	0.141	0.073	0.062	0.055	0.051	0.055	0.051
	SE2	0.048	0.005	0.003	0.003	0.003	0.008	0.005
	NU	3537	2403	1853	1491	1302	1162	1069
x=5	Bias	0.019	0.034	0.040	0.054	0.062	0.070	0.083
	MSE	0.279	0.258	0.179	0.153	0.152	0.184	0.192
	SE1	0.039	0.055	0.020	0.010	0.009	0.022	0.023
	N2	2671	2175	1894	1748	1644	1555	1489
	N1	2148	1720	1550	1438	1368	1303	1267
	MISE	0.093	0.128	0.065	0.057	0.066	0.061	0.117
	SE2	0.005	0.050	0.003	0.002	0.007	0.006	0.062
	NU	4379	3298	2758	2459	2259	2110	1991
x=10	Bias	0.019	0.031	0.036	0.055	0.047	0.063	0.080
	MSE	0.295	0.317	0.351	0.367	0.457	0.616	0.630
	SE1	0.016	0.021	0.027	0.031	0.064	0.151	0.146
	N2	3625	3106	2860	2690	2601	2518	2460
	N1	3061	2711	2532	2430	2353	2311	2268
	MISE	0.111	0.077	0.061	0.056	0.064	0.065	0.062
	SE2	0.012	0.006	0.002	0.002	0.008	0.008	0.006
	NU	5250	4197	3692	3382	3184	3065	2973

Table5 LAD for t(2)

h		$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$	$n^{-1/7}$	$n^{-1/8}$	$n^{-1/9}$	$n^{-1/10}$
x=0	Bias	0.009	0.002	0.013	0.022	0.030	0.033	0.043
	MSE	0.972	0.391	0.273	0.231	0.229	0.209	0.179
	SE1	0.326	0.075	0.042	0.033	0.035	0.028	0.024
	N2	1882	1325	1033	861	758	687	632
	N1	1273	882	683	577	508	453	413
	MISE	0.411	0.364	0.126	0.107	0.090	0.072	0.076
	SE2	0.131	0.109	0.013	0.012	0.010	0.004	0.005
	NU	3588	2433	1847	1504	1326	1185	1099
x=5	Bias	0.010	0.014	0.029	0.026	0.042	0.049	0.057
	MSE	0.962	0.940	0.675	3.003	0.736	0.670	0.638
	SE1	0.315	0.341	0.258	2.039	0.280	0.264	0.263
	N2	2741	2236	1958	1804	1693	1626	1561
	N1	2172	1817	1631	1511	1443	1381	1341
	MISE	0.279	0.169	0.167	0.155	0.155	0.162	0.152
	SE2	0.070	0.019	0.038	0.040	0.041	0.053	0.046
	NU	4399	3317	2790	2477	2274	2125	2043
x=10	Bias	0.030	0.003	0.018	0.028	0.037	0.044	0.048
	MSE	2.172	2.986	1.155	0.676	0.975	0.472	0.483
	SE1	1.109	1.746	0.315	0.111	0.437	0.045	0.047
	N2	3656	3121	2870	2685	2565	2507	2455
	N1	3077	2702	2532	2411	2354	2308	2271
	MISE	0.348	0.213	0.190	0.111	0.121	0.106	0.097
	SE2	0.095	0.060	0.055	0.010	0.026	0.013	0.009
	NU	5180	4189	3684	3370	3174	3040	2939

Table6 LS for t(2)

h		$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$	$n^{-1/7}$	$n^{-1/8}$	$n^{-1/9}$	$n^{-1/10}$
x=0	Bias	0.004	0.009	0.022	0.039	0.049	0.058	0.069
	MSE	0.579	0.384	0.290	0.232	0.206	0.194	0.177
	SE1	0.089	0.052	0.030	0.022	0.018	0.017	0.016
	N2	1882	1325	1033	861	758	687	632
	N1	1273	882	683	577	508	453	413
	MISE	0.487	0.424	0.209	0.183	0.153	0.146	0.133
	SE2	0.119	0.101	0.027	0.022	0.016	0.016	0.012
	NU	3588	2433	1847	1504	1326	1185	1099
x=5	Bias	0.012	0.017	0.035	0.042	0.060	0.066	0.083
	MSE	0.669	0.912	0.694	0.695	0.645	0.593	0.348
	SE1	0.188	0.336	0.261	0.258	0.239	0.226	0.037
	N2	2741	2236	1958	1804	1693	1626	1561
	N1	2172	1817	1631	1511	1443	1381	1341
	MISE	0.337	0.213	0.218	0.187	0.179	0.181	0.147
	SE2	0.068	0.018	0.040	0.037	0.037	0.039	0.018
	NU	4399	3317	2790	2477	2274	2125	2043
x=10	Bias	0.025	0.006	0.027	0.045	0.061	0.066	0.070
	MSE	1.217	2.277	0.943	0.710	0.704	0.686	0.806
	SE1	0.381	1.251	0.210	0.105	0.149	0.148	0.187
	N2	3656	3121	2870	2685	2565	2507	2455
	N1	3077	2702	2532	2411	2354	2308	2271
	MISE	0.376	0.378	0.231	0.200	0.194	0.192	0.212
	SE2	0.060	0.099	0.026	0.023	0.021	0.023	0.036
	NU	5180	4189	3684	3370	3174	3040	2939