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

Hi-Stat Discussion Paper

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

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 dependent variable 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.

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 dependent variable 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 contradict with their Theorem 3.2. The asymptotic distributions in Theorem 1 are similar to that in (3.12) of [22]. These implies that the result of [4] may be false. See also Remark 2 below. 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 depend 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 3.

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, and they will be specified later in section 2 as in [22]. We estimate $g(x_0)$ for a fixed x_0. We assume that $g(x)$ is twice continuously differentiable in a neighborhood of x_0.

In this paper we assume that $\mathrm{E}\{\mathrm{sign}(v_i)\} = 0$, where $\mathrm{sign}(v) = -1, v < 0, = 1, v \geq 0$, while $v_i = u_i$ and $\mathrm{E}\{v_i\} = 0$ in [22]. We can estimate $g(x_0)$ by using nonparametric LAD regression in spite of the contemporaneous correlation between X_i and v_i as in [22].

We present the assumption on $v(x, u)$ here. The other assumptions are given 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).

An example of $v(x,u)$ is $\sigma(x)(u-m_u)$.

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. The proof of Theorem 1 and the propositions for the proof are given in section 3. The proofs of the propositions are relegated to section 4.

We denote convergence in distribution and in probability by $\xrightarrow{d}$ and $\xrightarrow{p}$, respectively and C is a generic positive constant whose value vary from place to place. When X has a normal distribution with mean μ and covariance matrix Ω, we write $X \sim \mathrm{N}(\mu,\Omega)$. 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.

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, which crucially depends on the results in [21] and [22] and will be proved in section 3.

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 $X_i = \rho X_{i-1} + \eta_i$. $\{\eta_i\}$ is a linear process given by $\eta_i = \sum_{k=0}^{\infty} \phi_k \epsilon_{i-k}$, where $0 < \sum_{k=0}^{\infty} \phi_k = \phi < \infty$ 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 $X_{[nt]}, 0 \le t \le 1$, 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 $f_u(u|\mathcal{E}_{i-m_0}^i)$ due to Assumption U1. Recall that we denote the median of u_i by m_u.

Assumption U2: There exists the conditional density function of u_i given $\mathcal{E}$ in a neighborhood of m_u. The neighborhood does not depend on $(\epsilon_i, \dots, \epsilon_{i-m_0})$. Besides $f_u(u|\mathcal{E}_{i-m_0}^i)$ is uniformly bounded in $(\epsilon_i, \dots, \epsilon_{i-m_0})$ and continuously differentiable in the neighborhood and the derivative $f_u'(u|\mathcal{E}_{i-m_0}^i)$ is uniformly bounded. 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).

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

Writing $f_v(v|x)$ for the density function of $v(x, u_i)$, we have

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

Recall that $f_u(u)$ is the density function of u_i. Since X_i and u_i are not independent, $f_v(v|x)$ is not the conditional density function of $v(X_i, u_i)$ given $X_i = x$.

We state assumptions on the kernel function $K(\xi)$ 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: $K(\xi)$ is a nonnegative bounded continuous function with compact support and $\hat{K}(t)$ is integrable. In addition, the Fourier transforms of $\xi K(\xi)$, $\xi^2 K(\xi)$, and $\xi^3 K(\xi)$ are also integrable.

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

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

We introduce some notation related to the kernel function here.

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

$$\kappa_j = \int \xi^j K(\xi) d\xi \quad \text{and} \quad \nu_j = \int \xi^j K^2(\xi) d\xi \tag{6}$$

Assumption H: $nh^6 \to \infty$ and $nh^{10} = O(1)$ for the LLE and $nh^8 \to \infty$ and $nh^{10} = O(1)$ for the LCE.

Assumption H is more restrictive than that for Theorem 3.1 of [22] because we closely examine the bias term here. It is easy to see from Theorem 1 below that the asymptotically optimal bandwidth has the form of $h = cn^{-1/10}$.

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

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

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 3. 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 3.

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

$$\begin{aligned}\hat{\theta} \ &\stackrel{d}{\to} \ \frac{1}{2}(f_v(0|x_0)L^{1/2}(1,0))^{-1}\begin{pmatrix}\kappa_0 & \kappa_1\\ \kappa_1 & \kappa_2\end{pmatrix}^{-1}\begin{pmatrix}Z_1\\ Z_2\end{pmatrix}\\ &\qquad +(nh^2)^{1/4}\frac{h^2}{2}\begin{pmatrix}\kappa_0 & \kappa_1\\ \kappa_1 & \kappa_2\end{pmatrix}^{-1}\begin{pmatrix}\kappa_2\\ \kappa_3\end{pmatrix}g''(x_0),\end{aligned}$$

where $(Z_1, Z_2)^T$ below is independent of $L(1,0)$ and

$$\begin{pmatrix}Z_1\\ Z_2\end{pmatrix} \sim N\left(\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}\nu_0 & \nu_1\\ \nu_1 & \nu_2\end{pmatrix}\right).$$

For the LCE, we have

$$\begin{aligned}\hat{\theta} \ &\stackrel{d}{\to} \ \frac{1}{2}(f_v(0|x_0)L^{1/2}(1,0))^{-1}\kappa_0^{-1}Z_1 + (nh^2)^{1/4}\frac{h^2}{2}\frac{\kappa_2}{\kappa_0}(f_v(0|x_0))^{-1}\\ &\quad \times\Big(g''(x_0)f_v(0|x_0) + 2g'(x_0)\frac{\partial f_v}{\partial x}(0|x_0) - (g'(x_0))^2\frac{\partial f_v}{\partial v}(0|x_0)\Big),\end{aligned}$$

where

$$\begin{aligned}&\frac{\partial f_v}{\partial v}(0|x_0)\\ &= \Big(\frac{\partial v}{\partial u}(x_0, m_u)\Big)^{-2}\Big\{f_u'(m_u) + f_u(m_u)\frac{\partial}{\partial u}\Big(\frac{\partial v}{\partial u}(x_0, m_u)\Big)^{-1}\Big\}.\end{aligned}$$

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. Thus we should use the LLE for nonparametric quantile regression with integrated covariates to avoid the complicated bias term. The asymptotic distributions in Theorem 1 are different from that in [4]. Compare their Lemma A.1 with Proposition 2 and see Remark 2 below for more details.

Theorem 1 implies that the asymptotically optimal bandwidth depends 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. We will examine the effects of bandwidths by simulation studies in the follow-up study. A cross-validation method as in [4] may be a promising candidate for bandwidth selection.

3. Proof of Theorem 1

We give Propositions 1-5 before we prove Theorem 1. The proofs of the propositions are postponed to section 4.

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*

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

$\kappa_1 f_u(m_u)L(1,0), \kappa_2 f_u(m_u)L(1,0), \kappa_3 f_u(m_u)L(1,0))^T,$

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

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 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 $\hat{\theta}$.

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*

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

Remark 2. From Proposition 1, we have

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

This and Proposition 2 contradict with Lemma A.1 of [4].

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*

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

Proposition 4 is a version of the convex theorem in Pollard [19] adapted to the setup of this paper. We use Proposition 1 as in Remark 1.

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

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

where

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

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*

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

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 $\tilde{\Omega}$ in Remark 1. Taking a compact subset K of R^2, we have from Propositions 1 and 4 that uniformly in θ on K,

$$\begin{aligned}
&\sum_{i=1}^{n} K_i(|v_i^* - \tau_n^{-1}\eta_i^T\theta| - |v_i^*|) \qquad (11)\\
&= -\theta^T(\tau_n^{-1}\sum_{i=1}^{n}\eta_i K_i \mathrm{sign}(v_i^*)) + \theta^T \begin{pmatrix} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{pmatrix} \theta f_v(0|x_0)L(1,0) + o_p(1).
\end{aligned}$$

We evaluate the first term of the RHS of (11) by combining Propositions 1 and 3 and get

$$\begin{aligned}
&\tau_n^{-1}\sum_{i=1}^{n} K_i\eta_i \mathrm{sign}(v_i^*) \qquad (12)\\
&= \tau_n^{-1}\sum_{i=1}^{n} K_i\eta_i \mathrm{sign}(v_i) + \frac{h^2 g''(x_0)}{\tau_n}\sum_{i=1}^{n}\Big(\frac{X_i - x_0}{h}\Big)^2 K_i\eta_i f_v(0|x_0) + o_p(1)\\
&= \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} L^{1/2}(1,0) + \tau_n h^2 g''(x_0) \begin{pmatrix} \kappa_2 \\ \kappa_3 \end{pmatrix} f_v(0|x_0)L(1,0) + o_p(1)\\
&= O_p(1).
\end{aligned}$$

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 $\tilde{\Omega}_{\delta_2}$.

The uniformity of (11), (12), and the convexity of the objective function yield that $\hat{\theta} = O_p(1)$ on $\tilde{\Omega}_{\delta_2}$. Then from the uniformity of (11), Proposition 1, and the standard argument as in [10] and [12], we have that given $\tilde{\Omega}_{\delta_2}$,

$$\begin{aligned}
\hat{\theta} &= \frac{1}{2}(f_v(0|x_0)L^{1/2}(1,0))^{-1}\begin{pmatrix} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{pmatrix}^{-1}\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \qquad (13)\\
&\quad + \frac{\tau_n h^2 g''(x_0)}{2}\begin{pmatrix} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{pmatrix}^{-1}\begin{pmatrix} \kappa_2 \\ \kappa_3 \end{pmatrix} + o_p(1).
\end{aligned}$$

Since we can choose an arbitrarily small positive δ_1, we also have (13) 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 (14) and (15) below to establish the proposition.

$$\mathrm{E}\{|\rho(\xi - \delta) - \rho(\xi) + \delta\rho'(\xi)|^2\} = o(\delta^2) \tag{14}$$

$$\mathrm{E}\{\rho(\xi - \delta) - \rho(\xi) + \delta\rho'(\xi)\} = \delta^2 s_1(f_\xi) + o(\delta^2), \tag{15}$$

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 4.

In Proposition 3, we consider $\rho'(\xi + \delta) - \rho'(\xi)$ and we need (16) and (17) below to establish the proposition.

$$\mathrm{E}\{|\rho'(\xi + \delta) - \rho'(\xi)|^2\} = O(\delta) \tag{16}$$

$$\mathrm{E}\{\rho'(\xi + \delta) - \rho'(\xi)\} = \delta s_2(f_\xi) + o(\delta), \tag{17}$$

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 4.

When we have (14)-(17) 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 $f_u(u|\mathcal{E}^i_{i-m_0})$ 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 (14) and (16). 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 $f_u(u|\mathcal{E}^i_{i-m_0})$ to get the same results as in Propositions 2 and 3 and the assumptions will depend on q.

4. Proofs of propositions

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

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

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

$$\tau_n^{-1} \sum_{i=1}^{n} \Big\{ aK_i + b\Big(\frac{X_i - x_0}{h}\Big) K_i \Big\} \mathrm{sign}(u_i), \tag{18}$$

where a and b are arbitrary constants, and $\tau_n^{-2} \sum_{i=1}^n K_i$, $\tau_n^{-2} \sum_{i=1}^n \{(X_i - x_0)/h\}^j K_i f_u(m_u|\mathcal{E}^i_{i-m_0})$, $j = 0, 1, 2, 3$, we have the same result as in Proposition 1 with the first two elements of the both sides replaced with (18) and

$$(a^2\nu_0 + 2ab\nu_1 + b^2\nu_2)^{1/2} L^{1/2}(1, 0) Z, \tag{19}$$

respectively. Note that Z in (19) 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

$$|B_{2i}(\theta)| \le C\tau_n^{-1}|\eta_i^T\theta| I(|v_i^*| \le C\tau_n^{-1}|\eta_i^T\theta|). \tag{20}$$

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 (20) and Assumption U2, we have

$$\mathrm{E}\{D_{2i}^2(\theta)|\mathcal{E}\} \le C\tau_n^{-2}\mathrm{E}\{I(|v_i^*| \le C\tau_n^{-1}|\eta_i^T\theta|)|\mathcal{E}\} \le C\tau_n^{-3}. \tag{21}$$

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

$$\begin{aligned}
&\mathrm{E}\Big[\Big\{\sum_{i=1}^{n} K_i D_{2i}(\theta)\Big\}^2\Big] \qquad (22)\\
&\le \mathrm{E}\Big[\sum_{i=1}^{n} K_i^2 \mathrm{E}\{D_{2i}^2(\theta)|\mathcal{E}\}\Big] + \mathrm{E}\Big[\sum_{|i-i'|\le m_0} K_i K_{i'} \mathrm{E}\{D_{2i}(\theta)D_{2i'}(\theta)|\mathcal{E}\}\Big]\\
&\le C\mathrm{E}\Big\{\tau_n^{-3}\sum_{i=1}^{n} K_i^2\Big\} = O(\tau_n^{-1}).
\end{aligned}$$

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,

$$\mathrm{E}\{B_{2i}(\theta)|\mathcal{E}\} = \tau_n^{-2}(\eta_i^T\theta)^2 f_u(m_u|\mathcal{E}_{i-m_0}^i)\Big(\frac{\partial v}{\partial u}(X_i, m_u)\Big)^{-1} + o_p(\tau_n^{-2}). \qquad (23)$$

The desired result follows from (22), (23), 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 (22) yield

$$h^{-2}\tau_n^{-2}\sum_{i=1}^{n} K_i \eta_i D_{3i} = O_p((nh^6)^{-1/4}). \qquad (24)$$

By some calculation, we get uniformly in i,

$$\begin{aligned}
&h^{-2}\mathrm{E}\{B_{3i}|\mathcal{E}\} \qquad (25)\\
&= \Big(\frac{X_i - x_0}{h}\Big)^2 g''(\bar{X}_i) f_u(m_u|\mathcal{E}_{i-m_0}^i)\Big(\frac{\partial v}{\partial u}(X_i, m_u)\Big)^{-1} + o_p(1).
\end{aligned}$$

The desired result follows from (24), (25), 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 $\theta \in K$,

$$\Big| \sum_{i=1}^{n} K_i B_{2i}(\theta) - \theta^T A \theta \Big| \xrightarrow{p} 0. \tag{26}$$

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 (26) and the convexity of $\sum_{i=1}^{n} K_i B_{2i}(\theta)$ and $\theta^T A \theta$, we have that given $\tilde{\Omega}_{\delta_4}$,

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

Since we can choose any small δ_3, we have (26) on $\tilde{\Omega}$. Hence the proof of Proposition 4 is complete.

Proof of Proposition 5. Set

$$\delta_i^{**} = v_i^{**} - v_i = (X_i - x_0) g'(x_0) + \frac{1}{2}\Big(\frac{X_i - x_0}{h}\Big)^2 h^2 g''(\bar{X}_i), \tag{28}$$
$$B_{4i} = \mathrm{sign}(v_i^{**}) - \mathrm{sign}(v_i), \text{ and } D_{4i} = B_{4i} - \mathrm{E}\{B_{4i} | \mathcal{E}\}.$$

Since
$$|B_{4i}| \le C I(|v_i| \le Ch),$$
we have
$$\mathrm{E}\{|D_{4i}|^2 | \mathcal{E}\} \le Ch. \tag{29}$$
From (29) and the same argument as in the proofs of Propositions 2 and 3, we obtain
$$h^{-2} \tau_n^{-2} \sum_{i=1}^{n} K_i D_{4i} = O_p((nh^8)^{-1/4}). \tag{30}$$

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

$$h^{-2} \mathrm{E}\{B_{4i} | \mathcal{E}\} = 2 \delta_i^{**} f_{v_i}(0|\mathcal{E}) - (\delta_i^{**})^2 f'_{v_i}(0|\mathcal{E}) + o_p(1). \tag{31}$$

We evaluate the first and second terms of the RHS of (31). By some calculation, we obtain

$$2h^{-2}\tau_n^{-2}\sum_{i=1}^{n} K_i\delta_i^{**} f_{v_i}(0|\mathcal{E}) \tag{32}$$
$$= \tau_n^{-2}\sum_{i=1}^{n}\Big(\frac{X_i-x_0}{h}\Big)^2 K_i g''(x_0) f_u(m_u|\mathcal{E}_{i-m_0}^{i})\Big(\frac{\partial v}{\partial u}(x_0,m_u)\Big)^{-1}$$
$$+2\tau_n^{-2}\sum_{i=1}^{n}\Big(\frac{X_i-x_0}{h}\Big)^2 K_i g'(x_0) f_u(m_u|\mathcal{E}_{i-m_0}^{i})\frac{\partial}{\partial x}\Big(\frac{\partial v}{\partial u}(x_0,m_u)\Big)^{-1} + o_p(1)$$

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 (31).

$$f'_{v_i}(0|\mathcal{E}) \tag{33}$$
$$= \Big(\frac{\partial v}{\partial u}(X_i,m_u)\Big)^{-2}\Big(f'_u(m_u|\mathcal{E}_{i-m_0}^{i}) + f_u(m_u|\mathcal{E}_{i-m_0}^{i})\frac{\partial}{\partial u}\Big(\frac{\partial v}{\partial u}(X_i,m_u)\Big)^{-1}\Big).$$

From (28), (33), and Assumption V, we have

$$h^{-2}\tau_n^{-2}\sum_{i=1}^{n} K_i(\delta_i^{**})^2 f'_{v_i}(0|\mathcal{E}) \tag{34}$$
$$= \tau_n^{-2}\sum_{i=1}^{n}\Big(\frac{X_i-x_0}{h}\Big)^2 K_i\Big(\frac{\partial v}{\partial u}(x_0,m_u)\Big)^{-2}\Big\{f'_u(m_u|\mathcal{E}_{i-m_0}^{i})$$
$$+f_u(m_u|\mathcal{E}_{i-m_0}^{i})\frac{\partial}{\partial u}\Big(\frac{\partial v}{\partial u}(x_0,m_u)\Big)^{-1}\Big\} + o_p(1).$$

Proposition 5 follows from (31), (32), (34). Hence the proof of Proposition 5 is complete.

Acknowledgements

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 (2009) 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.