Haindorf Seminar 2012

Nonparametric LAD cointegrating regression

Toshio HONDA, Hitotsubashi University, JAPAN

1 Introduction

- Nonparametric quantile regression for time series

· α-mixing processes
· Long range dependent (LRD) linear processes
· Heavy-tailed linear processes

In this talk, we deal with nonstationary time series (nonlinear cointegration models).

• Nonparametric quantile regression by the check function(Chaudhuri (1991))
First we define the check function $\rho_q(z)$ by

$$\rho_q(z) = z(q - I(z < 0))$$

and consider the following stationary case.

$$Y_i = g(X_i) + v_i \quad \text{and} \quad \mathrm{E}\{\rho_q'(v_i)|X_i\} = 0,$$

where $\{X_i\}$ and $\{v_i\}$ are stationary processes and $\rho_q'(z) = q - I(z < 0)$. Then the conditional qth quantile of Y_i given $X_i = x$ is $g(x)$.

We can estimate $g(x)$ by local polynomial regression with $\rho_q(z)$ being the loss function. In this talk, we are concerned with the local constant estimator (LCE) and the local linear estimator (LLE).

When $\{(X_i, v_i)^T\}$ is an α-mixing process and the regularity conditions are satisfied, we have for the LLE that

$$\sqrt{nh}(\hat{g}(x) - g(x) - h^2 \times Bias) \xrightarrow{d} \mathrm{N}(0, \sigma^2),$$

where n is the sample size, h is the bandwidth, and $Bias$ is a constant.

- Cointegration models

· Linear models

$$Y_i = a^T X_i + v_i,$$

where $\{Y_i\}$ and $\{X_i\}$ are nonstationary (I(1)) processes and $\{v_i\}$ is a stationary process.

· I(1) processes

If $\{X_i\}$ is nonstationary and $\{X_i - X_{i-1}\}$ is stationary, then $\{X_i\}$ is an I(1) process.

· Nonlinear models

$$Y_i = g(X_i) + v_i \quad \text{and} \quad \mathrm{E}\{v_i\} = 0,$$

where $g(x)$ is an unknown function, both $\{Y_i\}$ and $\{X_i\}$ are nonstationary (I(1)) scalar processes, and $\{v_i\}$ is a stationary process.

Nonparametric estimation of $g(x)$ is considered in Karlsen and Tjøstheim (2001), Karlsen et al. (2007), Schienle (2008) by the recurrent Markov chain method and
Wang and Phillips (2009a,b), Wang and Phillips (2011a) by the local time method.

It is proved that $g(x)$ can be estimated even in the presence of endogeneity as in linear models in WP(2009b)*Econometrica*. We adopt the setup of WP(2009b). However, we assume that $\mathrm{E}\{\mathrm{sign}(v_i)\} = 0$ here and we will carry out LAD regression.

2 Setup

We omit some technical assumptions to save time and space. We have two mutually independent i.i.d. processes $\{\epsilon_i\}$ and $\{\lambda_i\}$ in this talk. In addition, $\mathrm{E}\{\epsilon_i\} = 0$ and $\mathrm{Var}\{\epsilon_i\} = 1$.

• X_i: X_i is the nonstationary regressor.
$X_i = \rho X_{i-1} + \eta_i$ with $X_0 = 0$, $\rho = 1 + \kappa/n$, and $\eta_i = \sum_{k=0}^{\infty} \phi_k \epsilon_{i-k}$, where $0 < \sum_{k=0}^{\infty} \phi_k = \phi < \infty$.
Then it is known that on $D[0,1]$, $X_{[nt]}/\sqrt{n} \xrightarrow{d}$

$$J_\kappa(t) = \phi(W(t) + \kappa \int_0^t e^{(t-s)\kappa} W(s) ds),$$

where $W(s)$ is a standard Brownian motion.

Let $L(s, a)$ be the local time of $J_\kappa(t)$. Then $L(s, a)$ is given by

$$L(s, a) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^s I\{|J_\kappa(r) - a| \leq \epsilon\} dr$$

and we have

$$\int_A L(t, a) da = \int_0^t I\{J_\kappa(r) \in A\} dr.$$

Jeganathan(2004) and WP(2009a,b) proved that

$$\frac{1}{\sqrt{n}h}\sum_{i=1}^{n} K_i \xrightarrow{d} \kappa_0 L(1,0),$$

where $K_i = K((X_i - x)/h)$, $K(\xi)$ is a kernel function, $\kappa_j = \int \xi^j K(\xi) d\xi$.
We also define ν_j, $\mathcal{E}$, $\mathcal{E}_{i-m_0}^{i}$ by $\nu_j = \int \xi^j K^2(\xi) d\xi$,
$\mathcal{E} = \sigma\{\epsilon_j | -\infty < j < \infty\}$,
$\mathcal{E}_{i-m_0}^{i} = \sigma\{\epsilon_j | i - m_0 \leq j \leq i\}$.

• v_i: $v_i = v(X_i, u_i)$. Besides, $v(x, u)$ is monotone increasing in u and $v(x, m_u) = 0$ for any x, where m_u is the median of u_i. For example, $v(x, u)$ is $\sigma(x)(u - m_u)$.

$u_i = u(\epsilon_i, \ldots, \epsilon_{i-m_0}, \lambda_i, \ldots, \lambda_{i-m_0})$, where m_0 is a unknown positive integer. Then $\{u_i\}$ is a stationary process and X_i and u_i are contemporaneously correlated.

Note that $\mathrm{sign}(v(X_i, u_i)) = \mathrm{sign}(u_i - m_u)$ because of the monotonicity of $v(x, u)$.

We define some density functions to deal with LAD regression. Recall that u_i depends on λ_j and ϵ_j, $i - m_0 \leq j \leq i$.

$f_u(u|\mathcal{E}^i_{i-m_0})$: the conditional density of u_i given $\mathcal{E}$.
$f_u(u)$: the marginal density of u_i.
$f_{v_i}(v|\mathcal{E})$: the conditional density of v_i given $\mathcal{E}$.
$f_v(v|x)$: the marginal density of $v(x, u_i)$.

They satisfy some technical assumptions and we have

$$f_{v_i}(0|\mathcal{E}) = f_u(m_u|\mathcal{E}^i_{i-m_0})\left(\frac{\partial v}{\partial u}(X_i, m_u)\right)^{-1}.$$

3 Estimators and asymptotics

The asymptotics of the LCE and the LLE are given here. We define only the LLE. The assumptions on the kernel function are omitted.

• Bandwidth 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.

We show below that the bias part $= h^2 \times Bias$ and $(nh^2)^{1/4}\times$(the stochastic part) $\stackrel{d}{\to}$ a random variable. Thus the optimal bandwidth is given by $h = C_0 n^{-1/10}$, where C_0 depends on the criterion of the optimality.

• The LLE

Set for notational convenience

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

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

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

where $\eta_i = (1, (X_i - x_0)/h)^T$.

By normalizing $\hat{\beta}$ as

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

we have

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

where

$$v_i^* = v_i + \frac{1}{2}\left(\frac{X_i - x_0}{h}\right)^2 h^2 g''(\bar{X}_i).$$

• Asymptotics

Theorem 1. We have for the LLE

$$\hat{\theta} \xrightarrow{d} V_L + (nh^2)^{1/4} \frac{h^2}{2} B_L,$$

where

$$B_L = \begin{pmatrix} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{pmatrix}^{-1} \begin{pmatrix} \kappa_2 \\ \kappa_3 \end{pmatrix} g''(x_0)$$

and

$$V_L = \frac{1}{2}(f_v(0|x_0)L^{1/2}(1,0))^{-1}\begin{pmatrix} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{pmatrix}^{-1} Z.$$

$Z = (Z_1, Z_2)^T$ above is independent of $L(1,0)$, the local time of $J_\kappa(t)$ at 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

$$\hat{\theta} \xrightarrow{d} V_C + (nh^2)^{1/4} \frac{h^2}{2} B_C,$$

where

$$V_C = \frac{1}{2} (f_v(0|x_0) L^{1/2}(1,0))^{-1} \kappa_0^{-1} Z_1,$$

Z_1 is defined as above, and

$$B_C = \frac{\kappa_2}{\kappa_0}(f_v(0|x_0))^{-1}$$
$$\times\Big(g''(x_0)f_v(0|x_0) + 2g'(x_0)\frac{\partial f_v}{\partial x}(0|x_0)$$
$$-(g'(x_0))^2\frac{\partial f_v}{\partial v}(0|x_0)\Big).$$

B_C can be smaller or larger than B_L.

1. The asymptotic distribution of the LLE will be the same as that of the Nadaraya-Watson estimator in (3.12) of WP(2009b) if $(2f_v(0|x_0))^{-1}$ is replaced with σ_u.

2. The bias term of the LCE is much more complicated than that of the LLE and that of the Nadaraya-Watson estimator in WP(2009b,2011a).

3. A similar problem in a more restrictive setup is considered in a paper. In the paper, $\{v_i\}$ and $\{X_i\}$ are independent. However, the result contradicts Theorem 1 here and the authors did not examine the bias terms, either.

4 Proof of Theorem 1

We present the outline of the proof of Theorem 1. We establish the theorem by applying the method of Pollard (1991), which is based on the convexity lemma. We need to adapt the convexity lemma to the setup here.

First we give the propositions for the proof. The first two elements of Proposition 1 below are related to the stochastic part, V_L and V_C, of the estimators. Proposition 1 is essentially established in WP(2009b).

Proposition 1.

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

$$\tau_n^{-2}\sum_{i=1}^{n} K_i, \tau_n^{-2}\sum_{i=1}^{n} K_i f_u(m_u|\mathcal{E}_{i-m_0}^{i})\Big)^T$$

$$\stackrel{d}{\to} (L^{1/2}(1,0)Z_1, L^{1/2}(1,0)Z_2,$$

$$\kappa_0 L(1,0), \kappa_0 f_u(m_u)L(1,0)),$$

where $(Z_1, Z_2)^T$ is defined as in Theorem 1 and independent of $L(1,0)$. Recall that $\tau_n = (nh^2)^{1/4}$.

By applying the almost sure representation theorem, we can replace convergence in distribution with almost sure convergence in Proposition 1. See Addendum 1.10.5 of van der Vaart and Wellner (1996). This replacement is crucial when we consider the adapted convexity lemma.

We cannot apply the martingale CLT as it is even if $\{u_i\}$ is an i.i.d. process independent of $\{X_i\}$. It is because the conditional variance converges in distribution, not in probability. See Wang (2011), Martingale limit theorems revisited and non-linear co integrating regression.

Proposition 2 gives the expansion of the objective function in θ.

Proposition 2. For any $\theta \in R^2$, we have

$$\sum_{i=1}^{n} K_i(|v_i^* - \tau_n^{-1}\eta_i^T\theta| - |v_i^*|)$$

$$= -\theta^T(\tau_n^{-1}\sum_{i=1}^{n}\eta_i K_i \mathrm{sign}(v_i^*))$$

$$+\theta^T\tau_n^{-2}\sum_{i=1}^{n}\eta_i\eta_i^T K_i f_u(m_u|\mathcal{E}_{i-m_0}^i)\left(\frac{\partial v}{\partial u}(x_0, m_u)\right)^{-1}\theta$$

$$+o_p(1).$$

Proposition 3 is about the bias term of the LLE, B_L. We omit the proposition for the LCE.

Proposition 3.

$$h^{-2}\tau_n^{-2}\sum_{i=1}^{n} K_i\eta_i(\mathrm{sign}(v_i^*)-\mathrm{sign}(v_i))$$

$$= \tau_n^{-2} g''(x_0)$$

$$\times \sum_{i=1}^{n}\left(\frac{X_i - x_0}{h}\right)^2 K_i\eta_i f_u(m_u|\mathcal{E}_{i-m_0}^{i})\left(\frac{\partial v}{\partial u}(x_0, m_u)\right)^{-1}$$

$$+ o_p(1).$$

Proposition 4 is a version of the convexity lemma in Pollard (1991) adapted to the setup of this paper.

Proposition 4. For any compact subset K of R^2, we have

$$\sup_{\theta \in K} \Big| \sum_{i=1}^{n} K_i (|v_i^* - \tau_n^{-1} \eta_i^T \theta| - |v_i^*|)$$

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

where

$$A = \lim_{n \to \infty} \tau_n^{-2} \sum_{i=1}^{n} K_i \eta_i \eta_i^T f_u(m_u | \mathcal{E}^i_{i-m_0}) \left(\frac{\partial v}{\partial u}(x_0, m_u) \right)^{-1}$$

$$= \begin{pmatrix} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{pmatrix} f_v(0|x_0) L(1, 0).$$

Note that we used the almost sure representation and that A is a random variable.

Proof of Theorem 1. For any compact subset K of R^2, we have from Propositions 1 and 4 that uniformly in θ on K,

$$\sum_{i=1}^{n} K_i(|v_i^* - \tau_n^{-1}\eta_i^T\theta| - |v_i^*|)$$

$$= -\theta^T(\tau_n^{-1}\sum_{i=1}^{n}\eta_i K_i \mathrm{sign}(v_i^*))$$

$$+\theta^T \begin{pmatrix} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{pmatrix} \theta f_v(0|x_0)L(1,0) + o_p(1).$$

By Propositions 1 and 3, we can represent the first term of the RHS of the above equation as

$$\tau_n^{-1} \sum_{i=1}^{n} K_i \eta_i \mathrm{sign}(v_i^*)$$

$$= \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} L^{1/2}(1,0)$$

$$+ \tau_n h^2 g''(x_0) \begin{pmatrix} \kappa_2 \\ \kappa_3 \end{pmatrix} f_v(0|x_0) L(1,0) + o_p(1).$$

By combining the two equations and minimizing the objective function w.r.t. θ, we obtain

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

$$+ \frac{\tau_n h^2 g''(x_0)}{2} \begin{pmatrix} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{pmatrix}^{-1} \begin{pmatrix} \kappa_2 \\ \kappa_3 \end{pmatrix} + o_p(1).$$

Hence the proof is complete.

5 Concluding remarks

1. There is no difference in asymptotics between the LCE and the LLE for nonparametric mean regression. However, there is a difference in the bias term for nonparametric LAD regression and we recommend the LLE.
2. A practical bandwidth selection rule is a topic of future research.
3. A small simulation implies that a larger bandwidth gives a smaller MSE. The asymptotic properties suggest that a lot of observations be necessary for nonparametric regression for

nonstationary time series. This is also confirmed by the simulation.

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

Thank you for your attention!