Robust Estimation in a Nonlinear Cointegration Model

Jia Chen\textsuperscript{1,2}, Degui Li\textsuperscript{1,2,*} and Lixin Zhang\textsuperscript{1}

\textsuperscript{1} Department of Mathematics, Zhejiang University, Hangzhou 310027, China
\textsuperscript{2} School of Economics, The University of Adelaide, Adelaide SA 5005, Australia

Abstract

This paper considers the nonparametric M–estimator in a nonlinear cointegration type model. The local time density argument, which was developed by Phillips and Park (1998) and Wang and Phillips (2009a), is applied to establish the asymptotic theory for the nonparametric M–estimator. The weak consistency and the asymptotic distribution of the proposed estimator are established under mild conditions. Meanwhile, the asymptotic distribution of the local least squares estimator and the local least absolute distance estimator can be obtained as applications of our main results. Furthermore, an iterated procedure for obtaining the nonparametric M–estimator and a cross–validation bandwidth selection method are discussed, and some numerical examples are provided to show that the proposed methods perform well in finite sample case.


Key words: Cointegration model, local time density, nonparametric M–estimator.

Abbreviated Title: Nonlinear Cointegration Model

* Corresponding author
E–mail: degui.li@adelaide.edu.au, ldgozju@yahoo.com (D. Li).
1. Introduction

During the past two decades, there has been much interest in various nonparametric techniques to model time series data with possible nonlinearity. Both estimation and specification testing problems have been systematically examined for the case where the observed time series satisfy a type of stationarity. For recent development of them, we refer to Green and Silverman (1994), Fan and Gijbels (1996), Fan and Yao (2003), Gao (2007), Li and Racine (2007) and the references therein.

As pointed out in the literature, the stationarity assumption seems too restrictive in practice. When tackling economic and financial issues from a time perspective, we often deal with nonstationary components. In reality, neither prices nor exchange rates follow a stationary distribution. Thus practitioners might feel more comfortable avoiding the stationary restriction. Until now, there has been large literature on nonparametric estimation of nonlinear regression and autoregression time series models with nonstationarity. The paper by Phillips and Park (1998) was among the first to study nonparametric autoregression in the context of a random walk. Karlsen and Tjøstheim (2001) and Karlsen et al (2007) independently discussed the nonparametric estimators in the framework of null recurrent Markov chains. Wang and Phillips (2009a) developed the asymptotic theory for local time density estimation and nonparametric cointegration models. For other recent development of nonparametric and semiparametric inferences in nonstationary time series, see Müller and Elliott (2003), Elliott and Müller (2006), Schienle (2008), Cai et al (2009), Chen et al (2009), Gao et al (2009), Wang and Phillips (2009b, 2009c) and the references therein.

The notion of cointegration was introduced by Granger (1981) and Engle and Granger (1987). Two time series \{x_t\} and \{y_t\} are said to be linearly cointegrated if they are both nonstationary and if there exists a linear combination \(ax_t + by_t = w_t\) such that \{w_t\} is stationary. As the relationship between two time series is not necessarily linear, in this paper, we consider a nonlinear cointegration model defined by

\[
y_t = m(x_t) + w_t,
\]

where \(m(\cdot)\) is some nonlinear function, \{x_t\} is some unit root type nonstationary
input process defined as
\[ x_t = x_{t-1} + v_t, \quad t \geq 1, \] (1.2)
\[ x_0 = O_P(1), \] and \( \{w_t\} \) and \( \{v_t\} \) are two sequences of stationary random variables satisfying some mild conditions.

A natural nonparametric estimator of \( m(z_0) \) in model (1.1) is the Nadaraya–Watson (NW) type estimator (cf. Karlsen et al, 2007 and Wang and Phillips, 2009a),
\[ \tilde{m}_n(z_0) = \frac{\sum_{t=1}^{n} y_t K \left( \frac{x_t - z_0}{h_n} \right)}{\sum_{t=1}^{n} K \left( \frac{x_t - z_0}{h_n} \right)}, \] (1.3)
where \( K(\cdot) \) is some kernel function and \( h_n \) is the bandwidth. The NW estimator is widely used in nonparametric regression since its introduction. However, it is not robust due to the fact that the NW estimator can be considered as a local least–squares estimator and the least–squares estimator is not robust. For instance, it is sensitive to outliers and does not perform well when the error distribution is heavy–tailed. However, outliers or aberrant observations are observed very often in economic time series and finance as well as in many other applied fields. A treatment of outliers or heavy–tailed errors is an important step in highlighting features of a data set. So in order to attenuate the lack of robustness of NW estimator, M–type regression estimator is a natural candidate for achieving desirable robustness properties. In Section 2, We will construct a robust version of nonparametric regression estimator for \( m(z_0) \).


In this paper, we apply the local time limit theory to establish the asymp-
totic results of the nonparametric M–estimator. The local time approach was introduced by Phillips and Park (1998) in the context of nonparametric autoregression and was developed recently by Wang and Phillips (2009a) for nonparametric cointegrating regression. The local time argument makes the approach in this paper more closely related to conventional nonparametric approaches than the null recurrent Markov chain method. In Section 2, we combine the local time argument with the nonparametric M–type smoothing technique to estimate the regression function \( m(\cdot) \). Under some mild conditions, we obtain the weak consistency as well as the asymptotic distribution of the proposed estimator. As applications of our main results, we establish the asymptotic properties of local least square estimators (LLSE) and local least absolute distance estimators (LLADE). Since the fully iterative procedure for the nonparametric M–estimator is time consuming, we apply a one–step iterative procedure to reduce the computational burden. Furthermore, we discuss the bandwidth selection based on the robust cross–validation method and give some numerical examples to show that the nonparametric M–estimator performs well in finite sample case.

The rest of the paper is organized as follows. In Section 2, we give the definition of the nonparametric M–estimator and some assumptions. In Section 3, we state the asymptotic results together with some remarks. In Section 4, we provide an iterated procedure for the M–estimator, the choice of bandwidth as well as some numerical examples. In Section 5, we conclude the paper and give some extensions. In Appendix, we provide the proofs of the main results.

2. The estimation method and assumptions

The nonparametric M–estimator \( \hat{m}_n(z_0) \) of \( m(z_0) \) in (1.1) is defined as

\[
\hat{m}_n(z_0) = \arg \min_\theta \sum_{t=1}^{n} \rho(y_t - \theta)K\left(\frac{x_t - z_0}{h_n}\right),
\]

or as the solution to the equation

\[
\sum_{t=1}^{n} \psi(y_t - \theta)K\left(\frac{x_t - z_0}{h_n}\right) = 0,
\]

where \( \rho(\cdot) \) is a convex loss function, \( \psi(\cdot) \) is any choice of the subderivatives (or subgradient) of \( \rho(\cdot) \) and \( h_n \) is the bandwidth satisfying \( h_n \to 0 \). As the convex
function $\rho(\cdot)$ may not necessarily be differentiable at all points, we let $\psi(\cdot)$ be any choice of its subderivatives. A subderivative of $\rho(\cdot)$ at a point $u_0$ is a real number $c$ such that $\rho(u) - \rho(u_0) \geq c(u - u_0)$ for all $u$. The set of all subderivatives of $\rho(\cdot)$ at $u_0$ is a nonempty closed interval $[a, b]$ with $a$ and $b$ being the one–sided limits

$$a = \lim_{u \to u_0^-} \frac{\rho(u) - \rho(u_0)}{u - u_0}, \quad b = \lim_{u \to u_0^+} \frac{\rho(u) - \rho(u_0)}{u - u_0}.$$ 

Many authors have considered the asymptotic theory for M–estimators under some assumptions on the loss function $\rho(\cdot)$ and its derivative $\psi(\cdot)$, such as Huber (1964, 1973), Bickel (1975), Bassett and Koenker (1978), Heiler and Willers (1988), Bai et al (1990). However, most of these papers discussed particular choices of $\rho(\cdot)$ and $\psi(\cdot)$ (such as $\rho(z) = |z|$), or general $\rho(\cdot)$ and $\psi(\cdot)$ under some restrictive conditions which do not cover some important special cases. Inspired by the paper by Bai et al (1992), here we allow $\rho(\cdot)$ and $\psi(\cdot)$ to include many commonly used estimators such as LLSE and LLADE.

The following assumptions will be made to establish the asymptotic properties of the nonparametric M–estimator.

**A1.** The kernel $K(\cdot)$ is nonnegative and has compact support, say $[-1, 1]$.

**A2.** $\rho(\cdot)$ is a convex function and $\psi(\cdot)$ is any choice of the subderivatives of $\rho(\cdot)$.

Let $\mathcal{D}$ be the set of discontinuity points of $\psi(\cdot)$, then $P(\mathcal{D}) = 0$.

**A3.** There exists a function $\lambda_1(\cdot)$, such that as $|u| \to 0$,

$$E[\psi(w_t + u)|x_t = x] = \lambda_1(x)u + o(|u|), \quad \text{for all } 1 \leq t \leq n,$$

where $\lambda_1(\cdot)$ is continuous at $z_0$ with $\lambda_1(z_0) > 0$.

**A4.** (i) Given $\mathcal{F}_n(v) = \sigma\{v_t, 1 \leq t \leq n\}, \{w_t, 1 \leq t \leq n\}$ is a sequence of independent and identically distributed (i.i.d.) random variables. Furthermore,

$$E[\psi^2(w_t)|x_t = x] = \sigma^2(x), \quad \text{for all } 1 \leq t \leq n,$$

where $\sigma^2(x) > 0$ and $\sigma^2(x)$ is continuous at $z_0$;

(ii) Uniformly for $x$ in a neighborhood of $z_0$,

$$\sup_n \max_{1 \leq t \leq n} E[(\psi(w_t + u) - \psi(w_t))^2|x_t = x] \leq \lambda(|u|),$$
where $\lambda(\cdot)$ is continuous at 0 with $\lambda(0) = 0$;

(iii) Uniformly for $x$ in a neighborhood of $z_0$,

$$
\sup_n \max_{1 \leq t \leq n} E[|\psi(w_t)|^r|x_t = x|] < \infty \quad \text{for some } r > 2.
$$

**A5.** The regression function $m(\cdot)$ satisfies

$$
|m(x + u) - m(x)| \leq C |u|^\beta, \quad \text{as } |u| \to 0 \quad \text{for some } \beta > 0.
$$

**A6.** Define the array $\{x_{i,n}, \, 1 \leq i \leq n, \, n \geq 1\}$ by $x_{i,n} = x_i/d_n$, where $\{d_n\}$ a sequence of positive numbers satisfying

$$
d_n \to \infty \quad \text{and} \quad d_n/n \to 0.
$$

Then, there exist a triple array of positive constants $d_{l,k,n}$ and an array of $\sigma$–fields $\mathcal{F}_{k,n}$ such that

$$
\limsup_{n \to \infty} \max_{0 \leq k < n-1} \frac{1}{n} \sum_{l=k+1}^{n} d_{l,k,n}^{-1} < \infty,
$$

$$
\frac{1}{n} \sum_{i=1}^{n} d_{i,0,n}^{-1} \to \tau, \quad \tau > 0,
$$

$x_{k,n}$ is adapted to $\mathcal{F}_{k,n}$, and given $\mathcal{F}_{k,n}$, $(x_{l,n} - x_{k,n})/d_{l,k,n}$, $0 \leq k < l \leq n$, has a density $\phi_{l,k,n}(\cdot)$ which is uniformly bounded and tends to a limit function $\phi(\cdot)$ as $n \to \infty$ and $l - k \to \infty$. Furthermore, $\phi(\cdot)$ is continuous in a neighborhood of 0.

**A7.** The bandwidth $h_n$ satisfies

$$
h_n \to 0, \quad nh_n/d_n \to \infty, \quad nh_n^{1+2\beta}/d_n \to 0, \quad \text{as } n \to \infty,
$$

where $\beta$ is defined as in A5.

**Remark 2.1.** The above assumptions are relatively mild in this kind of problems and can be justified in details. For example, A1 and A5 are quite natural and correspond to those used for the stationary time series. A2–A4 are assumed by many authors (see Lin et al., 2009b for example) to establish the asymptotic properties of nonparametric M–estimators and they cover some well–known cases such
as LLSE, LLADE and mixed LLSE and LLADE. This conditional independence assumption in A4 (i) is a little restrict and it is satisfied when \( \{w_t\} \) is a sequence of i.i.d. random variables and is independent of \( F_n(v) = \sigma \{v_t, 1 \leq t \leq n\} \). A6 and A7 are similar to Assumption 2.3 in Wang and Phillips (2009a). Next, we will give two examples in which A6 and A7 are satisfied.

**Example 2.1.** Assume that \( \{v_t\} \) is a sequence of i.i.d. random variables and \( v_t \sim N(0, 1) \). It is easy to check that A6 is satisfied if we choose \( F_{k,n} = \sigma(\eta_{-k}) \), \( d_n = \sqrt{n} \) and \( d_{l,k,n} = \sqrt{l-k}/\sqrt{n} \), \( 0 \leq k < l \leq n \). In this case, A7 becomes

\[
h_n \to 0, \quad \sqrt{n}h_n \to \infty, \quad \sqrt{n}h_n^{1+2\beta} \to 0, \quad \text{as } n \to \infty. \tag{2.4}
\]

**Example 2.2.** Let \( \{v_t\} \) be a sequence of nonlinear transforms of a linear process, i.e.

\[
v_t = H(\bar{v}_t), \quad \bar{v}_t = \sum_{k=0}^{\infty} b_k \eta_{t-k},
\]

where \( \{\eta_i\} \) is a sequence of i.i.d. random variables satisfying

\[
E\eta_1 = 0, \quad E\eta_1^2 = 1 \quad \text{and} \quad \int |\varphi(t)|dt < \infty,
\]

where \( \varphi(\cdot) \) is the characteristic function of \( \eta_1 \). The above assumption is similar to the condition of Corollary 2.2 in Wang and Phillips (2009a). Here, \( \{\bar{v}_t\} \) can be an ARMA or fractional ARIMA process. Wu (2006) established functional limit theorems for the partial sums of \( \{\bar{v}_t\} \). When \( H(\bar{v}) = v \), Wang and Phillips (2009a) showed that A6 and A7 hold under some mild conditions. Assume that

\[
b_k \sim k^{-\alpha}L(k), \quad 1/2 < \alpha < 1,
\]

where \( L(\cdot) \) is a slowly varying function and Condition 1 in Wu (2006) is satisfied.

Let \( p_0 \) be the power rank of \( H(\cdot) \), which is similar to the definition in Wu (2006). If \( p_0(2\alpha - 1) < 1 \), by Theorem 1 in Wu (2006), we can show that A6 is satisfied with

\[
\mathcal{F}_{k,n} = \sigma(\eta_i, -\infty < i \leq k), \quad d_n = C_{p_0}n^{1-p_0(\alpha-1/2)}L^{p_0}(n),
\]

\( C_{p_0} \) is some positive constant, and

\[
d_{l,k,n}^2 = E(x_l - x_k)^2 / Ex_n^2 = O \left( \frac{(l-k)^{3-2p_0\alpha}L^{2p_0}(l-k)}{n^{3-2p_0\alpha}L^{2p_0}(n)} \right).
\]
In this case, $A7$ becomes

$$h_n \to 0, \quad n^{p_0(\alpha - 1/2)}L^{-p_0}(n)h_n \to \infty, \quad \text{and} \quad n^{p_0(\alpha - 1/2)}L^{-p_0}(n)h_n^{1+2\beta} \to 0, \quad (2.5)$$

as $n \to \infty$.

If $p_0(2\alpha - 1) > 1$ or $p_0(2\alpha - 1) = 1$ with $\sum_{n=1}^{\infty} |L^{p_0}(\alpha)n| < \infty$, by Theorem 1 in Wu (2006), we can show that $A6$ is satisfied with

$$\mathcal{F}_{k,n} = \sigma(\eta_i, -\infty < i \leq k), \quad d_n = B_{p_0}\sqrt{n},$$

$B_{p_0}$ is some positive constant, and

$$d^2_{l,k,n} = E(x_k - x_l)^2/Ex_n^2 = O((l - k)/n).$$

Then, $A7$ is satisfied when (2.4) holds.

3. Asymptotic results

As our asymptotic theory relies heavily on the local time, we first introduce the local time of a continuous semimartingale $\{M(s), \ s \geq 0\}$ at a point $x$ as follows

$$L_M(t, x) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t 1\{\varepsilon < M(s) < x + \varepsilon\}d[M],$$

where $t$ and $s$ are the time parameters, $x$ is the spatial parameter, and $[M]$ is the quadratic variation process of $M(s)$ defined by

$$[M] = \lim_{\|T\| \to 0} \sum_{k=1}^{m} (X_{t_k} - X_{t_{k-1}})^2,$$

where $T$ ranges over partitions of the interval $[0, s]$ and the norm of the partition $T$ is the mesh. In (3.1), the local time is defined with respect to the measure $d[M]$. If $M(s)$ is a standard Brownian motion, then $d[M] = ds$. As can be seen from (3.1), $L_M(t, x)$ measures how much time the process $M(s)$ has spent at $x$ up to time $t$. For more details about the local time, see Geman and Horowitz (1980) and Revuz and Yor (1994).

The local time $L_M(t, x)$ satisfies the following important Occupation Time Formula:

$$\int_0^t H(M(s))d[M] = \int_{-\infty}^{+\infty} H(x)L_M(t, x)dx, \quad \text{for all} \ t \in \mathbb{R},$$
where $H(\cdot)$ is a nonnegative transformation on $\mathbb{R}$.

Next, we present the asymptotic properties of the robust estimator $\hat{m}_n(z_0)$, which is defined in (2.1). We first give the consistency of $\hat{m}_n(z_0)$ and then establish its asymptotic distribution.

**Theorem 3.1.** Assume that $A1$–$A7$ hold. Then there exists a solution $\hat{m}_n(z_0)$ for (2.1) such that

$$\hat{m}_n(z_0) - m(z_0) = o_P(1).$$

(3.2)

**Theorem 3.2.** Assume that the conditions of Theorem 3.1 are satisfied and on a suitable probability space, there exists a stochastic process $V(t)$ having a continuous local time $L_V(t, s)$ such that

$$\sup_{0 \leq t \leq 1} |x_{[nt],n} - V(t)| = o_P(1),$$

(3.3)

where $x_{i,n} = x_i/d_n$, $[nt]$ is the integer part of $nt$ and $d_n$ is defined as in $A6$. Then, we have

$$(nh_n/d_n)^{\frac{1}{2}}(\hat{m}_n(z_0) - m(z_0)) \xrightarrow{d} (\sigma_1^2 L_V(1, 0))^{1/2} \xi,$$

(3.4)

where $\sigma_1^2 = \sigma^2(z_0)\nu_2/((\lambda_1^2(z_0)\phi^2(0)\nu_1^2))$, $\nu_j = \int K^j(u)du$, $j = 1, 2$, $\xi \sim N(0,1)$ and $\xi$ is independent of $V$, $L_V(t, s)$ stands for the local time of the process $V(t)$ at the spatial point $s$ over the time interval $[0, t]$.

**Remark 3.1.** The asymptotic distribution in Theorem 3.2 is mixed normal and is different from that for stationary time series (see Härdle, 1984, Fan and Jiang, 2000 for example). (3.4) is also different from the asymptotic distribution obtained by Lin et al (2009a). By using the null recurrent Markov chain method, Lin et al (2009a) established the asymptotic distribution of the local linear M–estimators with random convergence rate. When $\{v_t\}$ is a sequence of i.i.d. random variables with $E v_1 = 0$ and $E v_1^2 = 1$, it is easy to check that $d_n = \sqrt{n}$. Hence, the convergence rate in (3.4) is $O(n^{1/4}h_n^{1/2})$, which is the same as that obtained by Karlsen et al (2007) with $\beta = 1/2$ in their paper and it is slower than the well–known convergence rate $O(\sqrt{n\bar{h}_n})$ for stationary nonparametric regression estimator.
Remark 3.2. The condition, that given $\mathcal{F}_n(v), \{w_t\}$ is an independent sequence, can be weakened. It can be seen from the proof of Theorem 3.2 that this independence restriction can be relaxed to the condition that given $\mathcal{F}_n(v), \{\psi(w_t)\}$ is a sequence of martingale differences. We conjecture that the results also hold for other dependent $\{w_t\}$ such as $\alpha$–mixing or linear processes. But to focus on essentials in our development of asymptotic theory, we simply assume that $\{w_t\}$ are independent given $\mathcal{F}_n(v)$. Furthermore, we can also establish analogous results for generalized nonparametric M–estimators such as local linear M–estimators or local polynomial M–estimators (Fan and Jiang, 2000, Jiang and Mack, 2001). Since the proofs are similar, we will not give the details here.

Remark 3.3. Letting $\rho(x) = x^2$ and $\psi(x) = 2x$, then estimator defined by (2.1) or (2.2) corresponds to the NW estimator or LLSE. Hence, we can obtain the asymptotic properties of the NW estimator as corollaries of the above two theorems. When $\rho(x) = |x|$ and $\psi(x) = \text{sign}(x)$, the estimator (2.1) corresponds to LLADE. Assume that $\{v_t\}$ is independent of $\{w_t\}$ and denote by $G(x)$ the distribution function of $w_1$. If $G(x)$ has a density function $g(x)$ around 0 and $g(0) > 0$, then we can also establish the asymptotic distribution for LLADE as a corollary of Theorem 3.2, i.e.,

$$(nh_n/d_n)^{\frac{1}{2}}(\tilde{m}_n(z_0) - m(z_0)) \xrightarrow{d} \left(\sigma_2^2LV(1,0)\right)^{1/2} \xi,$$  

(3.5)

where $\tilde{m}_n(z_0)$ is LLADE of $m(z_0)$ and $\sigma_2^2 = \nu_2/(g^2(0)\phi^2(0)\nu_2^2)$. LLADE has been discussed by many authors in the stationary case, see, for instance, papers by Basset and Koenker (1978), Bai et al. (1990). (3.5) is a new result for LLADE in nonstationary time series. Furthermore, Theorem 3.2 also holds for Huber’s $\psi$–function.

4. Examples of implementation

In this section, we discuss some critical problems such as the iterative algorithm for obtaining the nonparametric M–estimator of $m(\cdot)$ and the choice of a proper bandwidth. We also experiment with two numerical examples to illustrate the proposed method.

4.1. The iterated procedure and cross–validation bandwidth selection
The nonparametric M–estimator defined by (2.2) can be obtained by an iterative procedure. Define

$$\hat{\theta}_t(z_0) = \hat{\theta}_{t-1}(z_0) - (W_n(\hat{\theta}_{t-1}(z_0)))^{-1} \Psi_n(\hat{\theta}_{t-1}(z_0)), \quad (4.1)$$

where

$$\Psi_n(\theta) = \frac{d_n}{n h_n} \sum_{k=1}^{n} \psi(Y_k - \theta) K\left(\frac{x_k - z_0}{h_n}\right), \quad W_n(\theta) = \frac{\partial \Psi_n(\theta)}{\partial \theta},$$

The initial value of $\hat{\theta}_0(z_0)$ can be arbitrarily chosen and the above procedure is terminated at $t_0$–th iteration if $|\hat{\theta}_{t_0}(z_0) - \hat{\theta}_{t_0-1}(z_0)| < 0.0001$. Then, we let the nonparametric M–estimator of $m(z_0)$ to be $\hat{\theta}_0(z_0)$. However, this fully iterative procedure is time–consuming when the sample size is large. To overcome this disadvantage, we apply a one–step iterative procedure in our simulation as in Fan and Jiang (2000). The one–step M–estimator is defined by

$$\hat{\theta}_{OS}(z_0) = \hat{\theta}_0(z_0) - (W_n(\hat{\theta}_0(z_0)))^{-1} \Psi_n(\hat{\theta}_0(z_0)), \quad (4.2)$$

where $\hat{\theta}_0(z_0)$ is the initial value. When the initial value satisfies

$$\hat{\theta}_0(z_0) - m(z_0) = O_P(\sqrt{\frac{d_n}{n h_n}}),$$

the one–step M–estimator $\hat{\theta}_{OS}(z_0)$ has the same asymptotic properties as the M–estimator (see Fan and Jiang 2000 for details). Following the arguments in Fan and Chen (1999) and Fan and Jiang (2000), we can use the NW estimator of $m(z_0)$ as an initial value. The simulation results below show that the one–step method works well in practice.

Another difficult problem in simulation is the choice of a proper bandwidth involved in nonparametric M–estimator. In this paper, we employ a robust cross–validation method as in Lin et al (2009a). The cross–validation method is very useful in assessing the performance of estimators via estimating their prediction errors. The basic idea is to set one of the data points aside for validation of a model and use the remaining data to build the model. Define

$$CV(h_n) = \sum_{t=1}^{n} (y_t - \hat{m}_{h_n,-t}(x_t))^2, \quad (4.3)$$
where $\hat{m}_{h_{n-t}}(x_t)$ is the nonparametric M–estimator with bandwidth $h_n$ and the $t$–th observation left out. The bandwidth is selected to minimize $CV(h_n)$.

### 4.2. Numerical examples

Next, we give two numerical examples to show that the nonparametric M–estimator is more robust than the NW estimator for contaminated observations.

**Example 4.1.** Consider the linear cointegration model

$$y_t = m(x_t) + w_t, \quad m(x) = x, \quad t = 1, 2, \ldots, n,$$

where $\{x_t\}$ is generated by the unit root process defined by

$$x_t = x_{t-1} + v_t, \quad v_t = 0.5v_{t-1} + \eta_t, \quad t \geq 1, \quad x_0 = 0, \quad v_0 \sim N(0, 4/3),$$

$\{\eta_t\}$ is a sequence of i.i.d. random variables with $\eta_t \sim i.i.d. N(0, 1)$, and $\{\eta_t\}$ is independent of $\{w_t\}$. The error $\{w_t\}$ is independently taken from one of the following three distributions: (i) normal distribution $N(0, \sigma^2)$, $\sigma = 0.2$, (ii) symmetric contaminated normal distribution $0.1N(0, 5^2\sigma^2) + 0.9N(0, \sigma^2)$, and (iii) Cauchy distribution $0.1C(0, 1)$.

The data from each of the above distributions consist of 1000 replications of samples of sizes $n = 400, 800$ and $1600$. The uniform kernel $K(u) = \frac{1}{2}I(|u| \leq 1)$ and Huber’s $\psi$-function $\psi(z) = \max\{-c, \min\{c, z\}\}$ with $c = 1.25\sigma$ are applied in the simulation. The one–step iterative algorithm and the robust cross–validation bandwidth selection method defined in Section 4.1 are used to produce the nonparametric M–estimator of $m(\cdot)$. The measure of the performance of the estimators are taken to be the mean squared error (MSE) evaluated at the sample points $x_1, \ldots, x_n$,

$$MSE = \frac{1}{n} \sum_{t=1}^{n} (m_n(x_t) - m(x_t))^2,$$

where $m_n(\cdot)$ is the M–estimator or the NW estimator of $m(\cdot)$ and $m(\cdot)$ is the true regression function. The simulation results are listed in Table 1. The quantities in Table 1 are the mean MSE’s based on 1000 replications and the standard deviation of the MSE’s throughout 1000 replications.
Table 1. Simulation results for Example 4.1

<table>
<thead>
<tr>
<th></th>
<th>NW estimator</th>
<th>M–estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) $N(0, \sigma^2)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n=400$</td>
<td>0.0157</td>
<td>0.0153</td>
</tr>
<tr>
<td></td>
<td>(0.0042)</td>
<td>(0.0035)</td>
</tr>
<tr>
<td>$n=800$</td>
<td>0.0108</td>
<td>0.0108</td>
</tr>
<tr>
<td></td>
<td>(0.0027)</td>
<td>(0.0026)</td>
</tr>
<tr>
<td>$n=1600$</td>
<td>0.0081</td>
<td>0.0082</td>
</tr>
<tr>
<td></td>
<td>(0.0021)</td>
<td>(0.0021)</td>
</tr>
<tr>
<td>(ii) $0.1N(0, 0.5^2\sigma^2) + 0.9N(0, \sigma^2)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n=400$</td>
<td>0.0162</td>
<td>0.0161</td>
</tr>
<tr>
<td></td>
<td>(0.0040)</td>
<td>(0.0037)</td>
</tr>
<tr>
<td>$n=800$</td>
<td>0.0112</td>
<td>0.0111</td>
</tr>
<tr>
<td></td>
<td>(0.0026)</td>
<td>(0.0027)</td>
</tr>
<tr>
<td>$n=1600$</td>
<td>0.0086</td>
<td>0.0084</td>
</tr>
<tr>
<td></td>
<td>(0.0020)</td>
<td>(0.0020)</td>
</tr>
<tr>
<td>(iii) $0.1C(0,1)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n=400$</td>
<td>$1.1473 \times 10^3$</td>
<td>3.6341</td>
</tr>
<tr>
<td></td>
<td>(2.3899 $\times 10^4$)</td>
<td>(36.2120)</td>
</tr>
<tr>
<td>$n=800$</td>
<td>$2.0824 \times 10^3$</td>
<td>35.5750</td>
</tr>
<tr>
<td></td>
<td>(6.0893 $\times 10^4$)</td>
<td>(1.0334 $\times 10^3$)</td>
</tr>
<tr>
<td>$n=1600$</td>
<td>$3.6817 \times 10^3$</td>
<td>2.9092</td>
</tr>
<tr>
<td></td>
<td>(8.9410 $\times 10^4$)</td>
<td>(71.8292)</td>
</tr>
</tbody>
</table>

From Tables 1, we can find that, for Example 4.1, both the nonparametric M–estimator and NW estimator perform well when the error $\{w_t\}$ has the normal or contaminated norm distribution, and the performance of the two estimators improves as the sample size increases. However, when the error is heavy–tailed (Cauchy distributed), the nonparametric M–estimator behaves much better than NW estimator. The simulation results show that the nonparametric M–estimator is more robust than NW estimator.

**Example 4.2.** Consider the nonlinear cointegration model

$$y_t = m(x_t) + w_t, \quad m(x) = x^2, \quad t = 1, 2, \cdots, n$$

(4.6)
where \( \{x_t\} \) is defined as in Example 4.1. We still study the following two cases: (i) \( \{w_t\} \) has a normal distribution \( N(0, \sigma^2) \), and (ii) \( \{w_t\} \) has a symmetric contaminated normal distribution \( 0.1N(0, 5^2\sigma^2) + 0.9N(0, \sigma^2) \), and (iii) Cauchy distribution \( 0.1C(0, 1) \).

The data from each of the above distributions still consist of 1000 replications of samples of sizes \( n = 400, 800 \) and 1600. The results are reported in Table 2.

<table>
<thead>
<tr>
<th>( \text{NW estimator} )</th>
<th>( \text{M-estimator} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (i) \ N(0, \sigma^2) )</td>
<td></td>
</tr>
<tr>
<td>( n=400 )</td>
<td>16.4380</td>
</tr>
<tr>
<td></td>
<td>(16.7393)</td>
</tr>
<tr>
<td>( n=800 )</td>
<td>24.1650</td>
</tr>
<tr>
<td></td>
<td>(26.2842)</td>
</tr>
<tr>
<td>( n=1600 )</td>
<td>35.0640</td>
</tr>
<tr>
<td></td>
<td>(38.5541)</td>
</tr>
<tr>
<td>( (ii) \ 0.1N(0, 5^2\sigma^2) + 0.9N(0, \sigma^2) )</td>
<td></td>
</tr>
<tr>
<td>( n=400 )</td>
<td>16.1361</td>
</tr>
<tr>
<td></td>
<td>(17.6536)</td>
</tr>
<tr>
<td>( n=800 )</td>
<td>22.3141</td>
</tr>
<tr>
<td></td>
<td>(24.7908)</td>
</tr>
<tr>
<td>( n=1600 )</td>
<td>35.5463</td>
</tr>
<tr>
<td></td>
<td>(45.6358)</td>
</tr>
<tr>
<td>( (iii) \ 0.1C(0, 1) )</td>
<td></td>
</tr>
<tr>
<td>( n=400 )</td>
<td>( 1.6787 \times 10^4 )</td>
</tr>
<tr>
<td></td>
<td>(5.1600 \times 10^5)</td>
</tr>
<tr>
<td>( n=800 )</td>
<td>( 447.8955 )</td>
</tr>
<tr>
<td></td>
<td>(6.1013 \times 10^3)</td>
</tr>
<tr>
<td>( n=1600 )</td>
<td>( 3.6730 \times 10^5 )</td>
</tr>
<tr>
<td></td>
<td>(1.0790 \times 10^5)</td>
</tr>
</tbody>
</table>

From Table 2, we can find that, for Example 4.2, the performance of the nonparametric M-estimator is much better than that of the NW estimator no matter which of the three distributions the error \( \{w_t\} \) take. Meanwhile, the
performance of the two estimators in this example is worse than that in Example 4.1. This may due to the fact that when the regressor \( \{x_t\} \) is nonstationary, the volatility of \( \{x_t^2\} \) is much higher than that of \( \{x_t\} \) in Example 4.1.

5. Conclusions

In this paper, we establish the weak consistency as well as the asymptotic distribution for the proposed nonparametric M–estimator in a nonlinear cointegration model. We employ an iterated algorithm and a cross–validation bandwidth selection method in the simulated examples. The simulation results show that the nonparametric M–estimator works well for both linear and nonlinear cointegration models even when the error is heavy–tailed.

There are many issues left for future study. For example, as suggested by the referees, we may allow for contemporaneous correlation between \( \{x_t\} \) and \( \{w_t\} \) as in Wang and Phillips (2009b). In this case, we might have to apply a different method to establish the asymptotic theory. We may use Lemmas 7.1 and 7.2 in Wang and Phillips (2009b) to deal with the variance term in the proofs of Lemma A.1 and Theorem 3.1 in Appendix and apply the argument in the proof of Theorem 3.1 in Wang and Phillips (2009b) to obtain the asymptotic distribution of the nonparametric M–estimator. On the other hand, the limiting distribution of the proposed estimator will be different if the initial condition \( x_0 = O_P(1) \) is replaced by \( x_0 = O_P(n^{1/2}) \) and this case will be considered in our future research. Other extensions include studying Bahadur representation of the proposed M–estimator as in Lin et al (2009a). Meanwhile, testing problems based on such robust estimation procedures will also be left for our future research.

Acknowledgement

The authors would like to thank the Associate Editor and two anonymous referees for their valuable suggestions and comments, which greatly improved the former version of the paper. This work was supported by National Natural Science Foundation of China (Grant No. 10771192). The second author was also partially supported by the National Natural Science Foundation of China (Grant No. 10671176).

Appendix
Before giving the proofs of Theorems 3.1 and 3.2, we establish the following two lemmas which correspond to Theorems 2.1 and 2.3 in Bai et al (1992). However, the proofs are more difficult here because of the nonstationarity of \( \{x_t\} \). Throughout the proofs, we assume that \( x_0 = 0 \) for simplicity. The basic idea also holds for the general case of \( x_0 = O_P(1) \) and the assumption of \( x_0 = O_P(1) \) will not affect the asymptotic properties (cf. Park and Phillips, 1999).

**Lemma A.1.** Denote \( \theta_0 = m(z_0) \) and \( K_t = K \left( \frac{x_t - z_0}{h_{\lambda}} \right) \), \( t = 1, \cdots, n \). Suppose that all the conditions of Theorem 3.1 hold, then for any constant \( c > 0 \), we have

\[
\sup_{(nh_n/d_n)^{1/2} |\theta - \theta_0| \leq c} \left| \sum_{t=1}^{n} K_t \left[ \rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0) \right] \right| \leq \frac{nh_n}{\lambda (\theta - \theta_0)^2} \lambda_1(z_0) \tau_1 \nu_1 = O_P(1).
\] (A.1)

**Proof.** As \( \rho(\cdot) \) is convex, we have

\[
K_t \left[ \rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0) \right] \\
= K_t \left[ \psi(y_t - \theta_0 + \eta(\theta_0 - \theta)) - \psi(y_t - \theta_0) \right] (\theta - \theta_0) \quad \text{(A.2)}
\]

where \( |\eta| \leq 1 \). Letting

\[
z_{t,n} = K_t[\psi(y_t - \theta_0) - \psi(w_t)] = K_t[\psi(w_t + m(x_t) - \theta_0) - \psi(w_t)]
\]

and

\[
z_{t,n}^* = K_t[\psi(y_t - \theta_0 + \eta(\theta_0 - \theta)) - \psi(w_t)] = K_t[\psi(w_t + m(x_t) - \theta_0 + \eta(\theta_0 - \theta)) - \psi(w_t)],
\]

we have

\[
K_t[\rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0)] = (\theta - \theta_0)(z_{t,n} - z_{t,n}^*). \quad \text{(A.3)}
\]

Hence, it is easy to check that

\[
\text{Var} \left( \sum_{t=1}^{n} K_t \left[ \rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0) \right] \right) \\
= (\theta - \theta_0)^2 \text{Var} \left( \sum_{t=1}^{n} (z_{t,n} - z_{t,n}^*) \right) \quad \text{(A.4)}
\]

\[
\leq 2(\theta - \theta_0)^2 \text{Var} \left( \sum_{t=1}^{n} z_{t,n} \right) + 2(\theta - \theta_0)^2 \text{Var} \left( \sum_{t=1}^{n} z_{t,n}^* \right).
\]
Noting that for $|x_t - z_0| \leq h_n$ and by A5, we have

$$|m(x_t) - \theta_0| = |m(x_t) - m(z_0)| \leq Ch_n^2,$$  \hspace{1cm} (A.5)

$$|m(x_t) - \theta_0 + \eta(\theta_0 - \theta)| \leq |m(x_t) - \theta_0| + |\theta - \theta_0| \leq Ch_n^2 + C(nh_n/d_n)^{-1/2}. \hspace{1cm} (A.6)$$

By (A.5), A4(ii), A6 and A7, we know that for $n$ sufficiently large,

$$\text{Var} \left( \sum_{t=1}^{n} z_{t,n} \right) = \sum_{t=1}^{n} \text{Var}(z_{t,n}) \leq \sum_{t=1}^{n} E z_{t,n}^2$$

$$= \sum_{t=1}^{n} E \left( K_t^2 \left( \psi(w_t + m(x_t) - \theta_0) - \psi(w_t) \right)^2 \right)$$

$$\leq \sum_{t=1}^{n} E \left( K_t^2 \lambda(m(x_t) - \theta_0) \right)$$

$$= o \left( h_n d_n^{-1} \sum_{t=1}^{n} d_{t,0,n}^{-1} \int K^2(u) \phi_{t,0,n} \left( \frac{h_n u + z_0}{d_n d_{t,0,n}} \right) du \right)$$

$$= o \left( h_n d_n^{-1} \sum_{t=1}^{n} d_{t,0,n}^{-1} \right) = o(nh_n/d_n).$$  \hspace{1cm} (A.7)

By (A.6) and applying the same argument as (A.7), we can obtain

$$\text{Var} \left( \sum_{t=1}^{n} z'_{t,n} \right) = o(nh_n/d_n).$$  \hspace{1cm} (A.8)

From (A.4), (A.7) and (A.8) we know that for $(nh_n/d_n)^{1/2}|\theta - \theta_0| \leq c$

$$\text{Var} \left( \sum_{t=1}^{n} K_t \left[ \rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0) (\theta - \theta_0) \right] \right) = o(1).$$  \hspace{1cm} (A.9)

On the other hand, by A3 and the convexity of $\rho(\cdot)$, we can show that

$$E \left[ \rho(w_t + u) - \rho(w_t) | x_t = x \right] = \frac{1}{2} \lambda_1(x) u^2 + o(|u|^2).$$  \hspace{1cm} (A.10)

The proof of (A.10) is similar to Lemma 1 in Bai et al (1992). As a result,

$$E \left( \sum_{t=1}^{n} K_t \left[ \rho(y_t - \theta) - \rho(y_t - \theta_0) \right] \right)$$

$$= E \left( \sum_{t=1}^{n} \frac{1}{2} K_t \lambda_1(x_t) \left[ (m(x_t) - \theta)^2 - (m(x_t) - \theta_0)^2 \right] \right)$$

$$= \frac{1}{2} (\theta - \theta_0) \sum_{t=1}^{n} E[K_t \lambda_1(x_t)(\theta + \theta_0 - 2m(x_t))]$$

$$= \frac{1}{2} (\theta - \theta_0) \sum_{t=1}^{n} E[K_t \lambda_1(x_t)(\theta + \theta_0 - 2m(x_t))]$$
and

$$E \left( \sum_{t=1}^{n} K_t \psi(y_t - \theta_0)(\theta - \theta_0) \right) = (\theta - \theta_0) \sum_{t=1}^{n} E[K_t \lambda_1(x_t)(m(x_t) - \theta_0)].$$ \hspace{1cm} (A.12)

By (2.3), (A.11) and (A.12), we obtain

$$E \left( \sum_{t=1}^{n} K_t \left[ \rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0) \right] \right) = \frac{1}{2} (\theta - \theta_0)^2 \sum_{t=1}^{n} E(K_t \lambda_1(x_t)) $$

$$= \frac{1}{2} (\theta - \theta_0)^2 \lambda_1(z_0) h_n d_n^{-1} \sum_{t=1}^{n} d_{t,0,n}^{-1} \times \phi(0) \int K(u) du (1 + o(1)) $$

$$= \frac{1}{2} (\theta - \theta_0)^2 \lambda_1(z_0) \tau \phi(0) \nu_1 n h_n d_n^{-1} (1 + o(1)).$$ \hspace{1cm} (A.13)

By (A.9) and (A.13), we have

$$\sum_{t=1}^{n} K_t \left[ \rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0) \right] = \frac{1}{2} (\theta - \theta_0)^2 \lambda_1(z_0) \tau \phi(0) \nu_1 n h_n d_n^{-1} = o_P(1).$$

Since

$$\sum_{t=1}^{n} K_t \left[ \rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0) \right]$$

is convex in $\theta$, and $\frac{n h_n}{d_n} (\theta - \theta_0)^2 \lambda_1(z_0) \tau \phi(0) \nu_1$ is convex and continuous in $\theta$, (A.1) is proved by Theorem 10.8 in Rockafellar (1970).

**Lemma A.2.** Under the conditions of Theorem 3.1, we have, for any $c > 0$,

$$\sup_{(nh_n/d_n)^{1/2} |\theta - \theta_0| \leq c} \left| \left( \frac{nh_n}{d_n} \right)^{-\frac{1}{2}} \sum_{t=1}^{n} K_t \left[ \psi(y_t - \theta) - \psi(y_t - \theta_0) \right] + \left( \frac{nh_n}{d_n} \right)^{\frac{1}{2}} (\theta - \theta_0) \lambda_1(z_0) \nu_1 \phi(0) \right| = o_P(1).$$

**Proof.** Using the same method as that in the proof of Lemma A.1 and by Theorem 2.5.7 in Rockafellar (1970), we can prove Lemma A.2. Details are omitted here.

**Proof of Theorem 3.1.** Let $\hat{\theta}_n = \hat{m}_n(z_0)$. We first prove

$$(nh_n/d_n)^{1/2} (\hat{\theta}_n - \theta_0) = O_P(1).$$
Nonlinear Cointegration Model

It suffices for us to prove that, for any positive sequence \( \{c_n\} \) satisfying \( c_n \to \infty \),

\[
P\left(\frac{nh_n}{d_n}^{1/2}|\theta_n - \theta_0| \geq c_n\right) \to 0. \tag{A.14}
\]

By (A.1), we can choose a sequence of positive numbers \( \{c'_n\} \), such that \( c'_n \to \infty \), \( c'_n \leq c_n \) and

\[
\sup_{nh_n\{\{\theta_\theta_0\}} \left| \sum_{t=1}^n K_t [\rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0)]
- \frac{nh_n}{d_n} (\theta - \theta_0)^2 \lambda_1(z_0) \tau \phi(0) \nu_1 \right| = o_P(1). \tag{A.15}
\]

When \( (nh_n/d_n)^{1/2} |\theta - \theta_0| = c'_n \), we have

\[
\frac{nh_n}{2d_n} (\theta - \theta_0)^2 \lambda_1(z_0) \tau \phi(0) \nu_1 \geq \frac{1}{4} \lambda_1(x_0) \tau \phi(0) \nu_1 c'^2_n. \tag{A.16}
\]

Next, we will prove

\[
\sum_{t=1}^n K_t \psi(y_t - \theta_0) = O_P \left(\frac{nh_n}{d_n}^{1/2}\right). \tag{A.17}
\]

Note that

\[
E \left( \sum_{t=1}^n K_t \psi(y_t - \theta_0) \right)^2
= \sum_{t=1}^n E(K_t \psi(y_t - \theta_0))^2 + 2 \sum_{t=1}^{n-1} \sum_{j=t+1}^n E(K_t K_j \psi(y_t - \theta_0) \psi(y_j - \theta_0)) \tag{A.18}
=: J_{n1} + J_{n2}.
\]

From (A.7) and A4 (i), we have

\[
J_{n1} = \sum_{t=1}^n E \left( K_t [(\psi(y_t - \theta_0) - \psi(w_t)) + \psi(w_t)] \right)^2
\leq 2 \sum_{t=1}^n E[ K_t (\psi(y_t - \theta_0) - \psi(w_t))]^2 + 2 \sum_{t=1}^n E[ K_t \psi(w_t)]^2
= 2 \sum_{t=1}^n E z^2_{t,n} + 2 \sum_{t=1}^n E(K_t \sigma^2(x_t))
= o(nh_n/d_n) + O \left( \sum_{t=1}^n EK_t^2 \right)
= o(nh_n/d_n) + O(nh_n/d_n) = O(nh_n/d_n). \tag{A.19}
\]
By A4 (i), we have given \( x_i \) and \( x_j \), \( w_i \) and \( w_j \) are independent. Therefore, by A3, A6 and (A.5), we have

\[
J_{n2} = 2 \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} E(K_tK_jE[\psi(w_t + m(x_t) - \theta_0)\psi(w_j + m(x_j) - \theta_0)|x_t, x_j])
\]

\[
= 2 \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} E(K_tK_jE[\psi(w_t + m(x_t) - \theta_0)|x_t]E[\psi(w_j + m(x_j) - \theta_0)|x_j])
\]

\[
= 2 \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} E(K_tK_j\lambda_1(x_t)\lambda_1(x_j)(m(x_t) - \theta_0)(m(x_j) - \theta_0))
\]

\[
= O(h^{2\beta}_n \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} EK_tK_j)
\]

\[
= O(h^{2\beta}_n \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} E \left[ K_t \int K \left( \frac{d_{j,t,n}u+x_t-z_0}{h_n} \right) \phi_{j,t,n}(u)du \right] )
\]

\[
= O(h^{1+2\beta}_n \sum_{t=1}^{n-1} \sum_{j=t+1}^{n}EK_t \sum_{j=t+1}^{n} d_{j,t,n}^{-1})
\]

\[
= O \left( n^{2+2\beta}_n \right)
\]

\[
o(nh_n/d_n).
\]  \hspace{1cm} (A.20)

The last equality holds since \( nh_n^{1+2\beta}/d_n \to 0 \). In view of (A.18)–(A.20), we obtain (A.17). Therefore, when \( (nh_n/d_n)^{1/2}|\theta - \theta_0| = c'_n \)

\[
\left| (\theta - \theta_0) \sum_{t=1}^{n} K_t\psi(y_t - \theta_0) \right| = O_P(c'_n).
\]  \hspace{1cm} (A.21)

As \( c'_n \to \infty \), we know, from (A.15), (A.16) and (A.21),

\[
P \left( \inf_{(nh_n/d_n)^{1/2}|\theta - \theta_0| = c'_n} \sum_{t=1}^{n} K_t[\rho(y_t - \theta) - \rho(y_t - \theta_0)] \leq 0 \right) \to 0.
\]

By the convexity of \( \rho(\cdot) \), we have

\[
P \left( \inf_{(nh_n/d_n)^{1/2}|\theta - \theta_0| \geq c'_n} \sum_{t=1}^{n} K_t\rho(y_t - \theta) \leq \sum_{t=1}^{n} K_t\rho(y_t - \theta_0) \right) \to 0.
\]

By the definition of \( \hat{\theta}_n \) in (2.1), we get

\[
P \left( (nh_n/d_n)^{1/2} |\hat{\theta}_n - \theta_0| \geq c'_n \right) \to 0,
\]
which implies (A.14) since $c'_n \leq c_n$. So $(nh_n/d_n)^{1/2} (\hat{\theta}_n - \theta_0) = O_P(1)$ holds. Reminding $nh_n/d_n \to \infty$, we know that (3.3) is valid.

**Proof of Theorem 3.2.** We have proved, in the proof of Theorem 3.1, that

$$(nh_n/d_n)^{1/2} (\hat{\theta}_n - \theta_0) = O_P(1).$$

By Lemma A.2, we have

$$
\left( \frac{nh_n}{d_n} \right)^{-1/2} \sum_{i=1}^n K_i [\psi(y_i - \hat{\theta}_n) - \psi(y_i - \theta_0)] + \left( \frac{nh_n}{d_n} \right)^{1/2} (\hat{\theta}_n - \theta_0) \lambda_1(z_0) \nu_1 \phi(0) = o_P(1).
$$

Noting that $\sum_{i=1}^n K_i \psi(y_i - \hat{\theta}_n) = 0$ and from the above equation, we have

$$
\hat{\theta}_n - \theta_0 = (nh_n/d_n)^{-1} \lambda_1(z_0) \nu_1 \phi(0) \sum_{i=1}^n K_i \psi(y_i - \theta_0) + o_P \left( (nh_n/d_n)^{-1/2} \right).
$$

Note that

$$
\sum_{i=1}^n K_i \psi(y_i - \theta_0) = \sum_{i=1}^n K_i \psi(w_i) + \sum_{i=1}^n z_{i,n}.
$$

From (A.7), we know that $\sum_{i=1}^n E z_{i,n}^2 = o(nh_n/d_n)$. On the other hand, by taking the same lines as (A.20), we can get $\sum_{i=1}^{n-1} \sum_{j=i+1}^n E z_{i,n} z_{j,n} = o(nh_n/d_n)$. As a result,

$$
E \left( \sum_{i=1}^n z_{i,n} \right)^2 = o(nh_n/d_n),
$$

which implies

$$
\sum_{i=1}^n z_{i,n} = o_P \left( (nh_n/d_n)^{1/2} \right).
$$

By (A.22)–(A.24), we have

$$
\hat{\theta}_n - \theta_0 = (nh_n/d_n)^{-1} \lambda_1(z_0) \nu_1 \phi(0) \sum_{i=1}^n K_i \psi(w_i) + o_P \left( (nh_n/d_n)^{-1/2} \right).
$$

(A.25)
As $K(\cdot)$ is an integrable function, by Theorem 2.1 in Wang and Phillips (2009a) we get

$$
(nh_n/d_n)^{-1} \sum_{i=1}^{\lfloor nr \rfloor} K_i^2 \xrightarrow{P} L_V(1, 0) \int K^2(u)du = L_V(1, 0)\nu_2.
$$

(A.26)

On the other hand, by A4 (i), we know that, given $\mathcal{F}_n(v)$, \{\(K_i\psi(w_i), 1 \leq i \leq n\}\} is a sequence of martingale differences. Therefore, by A4 (i), A8, (A.26) and the central limit theorem for martingale differences (Hall and Heyde, 1980), we know

$$
(nh_n/d_n)^{-1/2} \sum_{i=1}^{n} K_i\psi(w_i) \xrightarrow{d} \left(\sigma^2(z_0)L_V(1, 0)\nu_2\right)^{1/2} \xi,
$$

(A.27)

where $\xi \sim N(0,1)$ and $\xi$ is independent of $V$. In view of (A.25) and (A.27), the proof of Theorem 3.2 is completed.

References


D. Revuz, M. Yor, Continuous Martingale and Brownian Motion (2nd ed.), Springer–Verlag, New York, 1994.


