Statistica Sinica **19** (2009), 1683-1703

LOCAL LINEAR M-ESTIMATORS IN NULL RECURRENT TIME SERIES

Zhengyan Lin[1], Degui Li[1,2] and Jia Chen[1,2]

[1]*Zhejiang University and* [2]*University of Adelaide*

Abstract: In this paper, we study a nonlinear cointegration type model $Y_k = m(X_k) + w_k$, where $\{Y_k\}$ and $\{X_k\}$ are observed nonstationary processes and $\{w_k\}$ is an unobserved stationary process. The process $\{X_k\}$ is assumed to be a null–recurrent Markov chain. We apply a robust version of local linear regression smoothers to estimate $m(\cdot)$. Under mild conditions, the uniform weak consistency and asymptotic normality of the local linear M-estimators are established. Furthermore, a one–step iterated procedure is introduced to obtain the local linear M-estimator and the optimal bandwidth selection is discussed. Meanwhile, some numerical examples are given to show that the proposed theory and methods perform well in practice.

Key words and phrases: Asymptotic normality, β–null recurrent Markov chain, cointegration model, consistency, local linear M-estimator.

1. Introduction

Two time series $\{Y_k\}$ and $\{X_k\}$ are said to be linearly cointegrated if they are both nonstationary and if there exists a linear combination

$$aX_k + bY_k = w_k \tag{1.1}$$

such that $\{w_k\}$ is stationary. This implies that the time series $\{Y_k, X_k\}$ move together when considered over a long period of time. Since the introduction of unit root and cointegration analysis in time series analysis, linear models have dominated empirical work in the application of these methods. This emphasis on linearity is convenient for practical application, and accords well with the linear framework of partical summation in which the integrated process and cointegration concepts have been developed. However, the linear assumption is restrictive in application and one often encounters situations where a particular parametric linear model cannot be adopted with confidence and thus a nonlinear type model is used as an alternative.

Since the time series $\{Y_k\}$ and $\{X_k\}$ may not be linearly cointegrated, this leads to a study of the nonlinear cointegration type model defined by

$$Y_k = m(X_k) + w_k, \tag{1.2}$$

where $m(\cdot)$ is some continuous function. Throughout the paper, we assume that $\{X_k\}$ is a null–recurrent Markov chain and $\{w_k\}$ is a sequence of independent and identically distributed (i.i.d.) random variables independent of $\{X_k\}$. In fact, when $Ew_1 = 0$, $m(\cdot)$ can be viewed as the conditional expectation, i.e.,

$$m(x_0) = E(Y_k|X_k = x_0).$$

The theory of nonlinear time series has been systematically examined by many authors, see Fan and Yao (2003), Gao (2007) and Li and Racine (2007) and the references therein. However, when tackling economic and financial issues from a time perspective, we often deal with nonstationary components. For example, neither prices nor exchange rates follow a stationary law over time. There is now a large literature on parametric linear and nonlinear models of nonstationary time series (see Park and Phillips (2001) for example). In nonparametric estimation of nonlinear regression and autoregression of nonstationary time series models, existing studies include Phillips and Park (1998), Karlsen and Tjøstheim (2001), Schienle (2006), Wang and Phillips (2006) and Karlsen, Mykelbust and Tjøstheim (2007).

The main goal of this paper is to investigate the estimation theory for the regression function $m(\cdot)$. For the nonlinear cointegration type model (1.2), Karlsen, Mykelbust and Tjøstheim (2007) applied the Nadaraya–Watson (NW) method to estimate $m(\cdot)$ and established the asymptotic theory of the proposed estimator. Although the NW estimator is central in most nonparametric regression methods, Fan and Gijbels (1996) showed that this method suffers from several drawbacks, such as poor boundary performance, excessive bias and low efficiency. To overcome such drawbacks, the local linear method was developed. It is defined as the solution to the weighted least squares problem

$$\sum_{k=1}^{n}(Y_k - a - (X_k - x_0)b)^2 K\left(\frac{X_k - x_0}{h}\right), \tag{1.3}$$

where $K(\cdot)$ is some kernel function and $h := h_n$ is a sequence of positive numbers which tends to zero as n tends to infinity. The local linear smoother has become popular in recent years because of its attractive statistical properties. It has advantages over the popular kernel method, in terms of design adaptation and high asymptotic efficiency. Furthermore, the local linear method is adaptive to almost all regression settings and copes well with edge effects.

Although the local linear regression estimator has many advantages, it is not robust due to the fact that the local linear regression estimator can be considered as a local weighted least–squares estimator and the least–squares estimator is sensitive to outliers and does not perform well when the error distribution is heavy–tailed. Outliers or aberrant observations are common in economic time series, finance, and many other applied fields. To attenuate the lack of robustness of the local linear estimator, M-type regression estimators are natural candidates for achieving desirable robustness properties. There is extensive literature concerning asymptotic properties of the robust nonparametric regression estimation for stationary time series. For example, Fan and Jiang (2000) studied the robust version of local linear regression smoothers augmented with variable bandwidth for i.i.d. observations; Jiang and Mack (2001) and Cai and Ould-Saïd (2003) considered local polynomial M-estimators and local linear M-estimators for stationary dependent observations. To the best of our knowledge, however, local linear M-type estimation has not been developed for nonstationary time series. We propose estimating $m(x_0)$ by a local linear M-type estimation method. That is, find a and b to minimize

$$\sum_{k=1}^{n} \rho(Y_k - a - (X_k - x_0)b) K\left(\frac{X_k - x_0}{h}\right), \tag{1.4}$$

where $\rho(\cdot)$ is a given convex function. Here and in the sequel, the local linear M-estimators of $m(x_0)$ and $m'(x_0)$ are denoted by $\widehat{m}_n(x_0)$ and $\widehat{m}'_n(x_0)$, respectively. Sometimes, the following is also applied to define the local linear M-estimators of $m(x_0)$ and $m'(x_0)$: find a and b to satisfy

$$\sum_{k=1}^{n} \psi(Y_k - a - (X_k - x_0)b) K\left(\frac{X_k - x_0}{h}\right) \begin{pmatrix} 1 \\ \frac{X_k - x_0}{h} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{1.5}$$

A natural method of obtaining (1.5) is to take the derivative of (1.4) with respect to (a, b) when $\rho(\cdot)$ is continuously differentiable, and equate it to null (cf., Fan and Jiang (2000)). In this paper, we apply a suitably chosen function $\psi(\cdot)$ to (1.5) that brings in many common cases such as the least square estimator and the least absolute distance estimator (see Section 3 for details).

In the traditional stationary time series analysis, some sort of mixing condition is assumed for $\{X_k\}$ to obtain the asymptotic properties of $\widehat{m}_n(x_0)$ and $\widehat{m}'_n(x_0)$ (cf., Jiang and Mack (2001) and Cai and Ould-Saïd (2003)). However, the mixing condition is difficult to verify and is ruled out in general. A minimal condition for undertaking the asymptotic analysis on $\widehat{m}_n(x_0)$ is that, as the number of observations increases, there must be infinitely many observations in any neighborhood of x_0, which means that $\{X_k\}$ must return to a neighborhood of x_0 infinitely often. Throughout the paper, $\{X_k\}$ is assumed to be ϕ–irreducible Harris recurrent, making asymptotics for nonparametric estimation possible. We establish the uniform weak consistency and asymptotic normality for the local linear M-estimators under mild conditions. The asymptotic properties of local linear estimators and the least absolute distance estimators can also be obtained by suitably choosing the function $\rho(\cdot)$. Furthermore, the one–step iterated procedure of M-estimators and the choice of optimal bandwidth are discussed. Finally, numerical examples are given to show that the proposed theory and methods perform well in practice.

The rest of the paper is organized as follows. Section 2 gives a brief summary of some key terminologies in the theory of Markov chains together with some necessary conditions. The main asymptotic results, with some remarks and some interesting examples, are provided in Section 3. The one–step iterated procedure of the local linear M-estimators is discussed in Section 4. The robust cross–validation method to select the optimal bandwidth, with some numerical examples, are given in Section 5. All the proofs of the main results, as well as the key lemmas, are concentrated in the Appendix.

2. Markov Theory and Assumptions

2.1. Markov theory

Let $\{X_k,\ k \geq 0\}$ be a ϕ–irreducible Markov chains on the state space $(\mathbf{E}, \mathcal{E})$ with transition probability P. This means that for any set $A \in \mathbf{E}$ with $\phi(A) > 0$, $\sum_{n=1}^{\infty} P^n(v, A) > 0$ for all $v \in \mathbf{E}$. To make asymptotics for nonparametric estimation possible, we assume that the ϕ–irreducible Markov chain $\{X_k\}$ is Harris recurrent.

Definition 1. The chain $\{X_k\}$ is Harris recurrent if, given a neighborhood N_v of v with $\phi(N_v) > 0$, $\{X_k\}$ returns to N_v with probability one, $v \in \mathbf{E}$.

The Harris recurrence of $\{X_k\}$ allows one to construct a split chain which decomposes the partial sum of functions of $\{X_k\}$ into blocks of i.i.d. parts and the negligible remaining parts (see the proofs in Appendix for example). The number of the independent parts N_n indicates how often the process regenerates. Since $N_n \leq n$ for the null recurrent case, this indicates that the convergence rate for the proposed estimator here will be slower than that for stationary time series. On the other hand, Harris recurrence only yields stochastic rates of convergence for estimators, where distribution and size of the regeneration times N_n have no a priori known structure but fully depend on the underlying process. The class of stochastic processes we are dealing with is not the general class of null recurrent Markov chains. Instead, we need to impose some restrictions on the tail behavior of the distribution of the recurrence time of the chain.

Definition 2. A Markov chain $\{X_k\}$ is β–null recurrent if there exist a small nonnegative function $f(\cdot)$, an initial measure λ, a constant $\beta \in (0, 1)$, and a slowly varying function $L_f(\cdot)$ such that

$$\mathbf{E}_\lambda\left[\sum_{i=0}^{n} f(V_i)\right] \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_f(n) \quad \text{as} \quad n \to \infty, \tag{2.1}$$

where $\mathbf{E}_\lambda$ stands for the expectation with initial distribution λ, and $\Gamma(1+\beta)$ is the Gamma function with parameter $1 + \beta$.

Assuming β–null Harris recurrence restricts the tail behavior of the recurrence time of the process to be a regular varying function. As a standard result (cf., Karlsen and Tjøstheim (2001)), the regeneration times N_n of the β–null recurrent Markov chain $\{X_k\}$ have the following asymptotic distribution

$$\frac{N_n}{n^\beta L_s(n)} \xrightarrow{d} M_\beta(1), \tag{2.2}$$

where $L_s = L_f/(\pi_s f)$, π_s is the invariant measure of the Markov chain $\{X_k\}$, and $M_\beta(1)$ is the Mittag–Leffler distribution with parameter β (cf., Kasahara (1984)). Furthermore, for a stationary or positive recurrent process $\beta = 1$, and for a univariate random walk $\beta = 1/2$.

2.2. Assumptions

The following assumptions are necessary to establish the asymptotic theory of the local linear M-estimators.

A1. The kernel function $K(\cdot)$ is continuous, symmetric, and has a compact support, say $[-1, 1]$.

A2. $m(\cdot)$ has continuous second derivative at x_0.

A3. There exists a positive constant λ_1 such that, as $|u| \to 0$, $E[\psi(w_1 + u)] = \lambda_1 u + O(u^2)$.

A4. (i) $E[\psi^2(w_1)] = \sigma^2 > 0$;
(ii) $E[(\psi(w_1 + u) - \psi(w_1))^2] \leq \lambda(|u|)$, where $\lambda(\cdot)$ is a nonnegative function continuous at 0 with $\lambda(0) = 0$.

A5. The invariant measure π_s of the β–null recurrent Markov chain $\{X_k\}$ has a continuous density function $p_s(\cdot)$ and $p_s(x_0) > 0$.

Remark 1. The above assumptions are relatively mild in this kind of problem and can be justified in details. For example, A1 and A2 are quite natural in nonparametric estimation for stationary time series (cf., Fan and Yao (2003)). A3 and A4 are some conditions on the function $\psi(\cdot)$ that are mild and cover some well–known special cases such as the least square estimate (LSE), the least absolute distance estimate (LADE), and the mixed LSE and LADE. See Section 3 for some specific examples. A5 corresponds to the continuity condition on the density function in the stationary case.

3. Large Sample Theory

3.1. Weak consistency and asymptotic normality

Next, we present the asymptotic properties of the local linear M-estimators defined in Section 1. We first give the uniform weak consistency of the proposed estimators and then establish their asymptotic distributions.

Theorem 1. *Assume that* A1−A5 *are satisfied and* $n^{\beta-\varepsilon_0}h \to \infty$ *for some* $0<\varepsilon_0<\beta$. *If* $\inf_{x\in S} p_s(x)>0$ *for some compact support* S *and*

$$E\Big[\psi^{2m}(w_1)\Big]<\infty,\quad E\Big[\Big(\psi(w_1+u)-\psi(w_1)\Big)^{2m}\Big]\le \lambda(|u|),$$

where $m>\max\{(1+\beta)/\varepsilon_0-11/8,5/4\}$ *and* $\lambda(\cdot)$ *is defined in* A4, *then we have*

$$\sup_{x\in S}\begin{pmatrix}\widehat{m}_n(x)-m(x)\\ h(\widehat{m}_n'(x)-m'(x))\end{pmatrix}=o_P(1). \tag{3.1}$$

Remark 2. Aside from the uniform weak consistency of the local linear M-estimators, we conjecture that

$$\sup_{x\in S}\begin{pmatrix}\widehat{m}_n(x)-m(x)\\ h(\widehat{m}_n'(x)-m'(x))\end{pmatrix}=o_P\Big(\frac{1}{\sqrt{n^{\beta-\varepsilon_0}h}}\Big)+O_P(h^2),$$

where $0<\varepsilon_0<\beta$. Since different methods and more technicalities will be involved, such issues are left for future research.

Let $\mu_t=\int u^tK(u)du$ and $\nu_t=\int u^tK^2(u)du$ for $t=0,1,\ldots$. We next state the asymptotic normality of local linear M-estimators for β–null recurrent processes.

Theorem 2. *Assume that* A1−A5 *are satisfied and that* $n^{\beta-\varepsilon_0}h\to\infty$ *for some* $0<\varepsilon_0<\beta$. *Then there exists a unique solution to equation* (1.5) *such that*

$$(N_nh)^{1/2}\left\{\begin{pmatrix}\widehat{m}_n(x_0)-m(x_0)\\ h(\widehat{m}_n'(x_0)-m'(x_0))\end{pmatrix}-\Big(\frac{1}{2}h^2\Gamma^{-1}(x_0)b(x_0)+o_P(h^2)\Big)\right\}$$
$$\xrightarrow{d} N\Big((0,0)^\tau,\Gamma^{-1}(x_0)\Sigma(x_0)\Gamma^{-1}(x_0)\Big), \tag{3.2}$$

where

$$\Gamma(x_0)=\lambda_1p_s(x_0)\begin{pmatrix}\mu_0&\mu_1\\ \mu_1&\mu_2\end{pmatrix},\quad \Sigma(x_0)=\sigma^2p_s(x_0)\begin{pmatrix}\nu_0&\nu_1\\ \nu_1&\nu_2\end{pmatrix},$$

and $b(x_0)=\lambda_1p_s(x_0)m''(x_0)(\mu_2,\mu_3)^\tau$.

Remark 3. From Theorem 2, the bias of local linear M-estimators is $O(h^2)$, the same as in stationary time series. Furthermore, by Theorem 3.2 in Karlsen and Tjøstheim (2001), $N_n=O_P(n^\beta L_s(n))$ which, combined with (3.2), implies that the asymptotic mean square error of the estimator is

$$O(h^4)+O\Big(\frac{1}{n^\beta L_s(n)h}\Big). \tag{3.3}$$

This leads to the optimal bandwidth $C_{k_0}(n^{\beta}L_s(n))^{-1/5}$, where C_{k_0} is some positive constant. On the other hand, since $K(\cdot)$ is symmetric, $\mu_1 = 0$. From (3.2), we know that $\widehat{m}_n(x_0)$ is asymptotically independent of $\widehat{m}_n'(x_0)$.

Remark 4. Sometimes, the regeneration times of the Markov chain N_n in (3.2) are unobservable. In practice, N_n can be replaced by

$$N_C(n) = \sum_{k=1}^{n} I_C(X_k),$$

where C is some small set and $I_C(\cdot)$ is the indicator function. Furthermore, by Lemma 3.2 in Karlsen and Tjøstheim (2001), we have $N_C(n)/N_n = \pi_s I_C + o(1)$, a.s. Hence, by (3.2), we have

$$(N_C(n)h)^{1/2}\left\{\begin{pmatrix}\widehat{m}_n(x_0) - m(x_0)\\ h(\widehat{m}_n'(x_0) - m'(x_0))\end{pmatrix} - \Big(\frac{1}{2}h^2\Gamma^{-1}(x_0)b(x_0) + o_P(h^2)\Big)\right\}$$
$$\xrightarrow{d} N\Big((0,0)^{\tau}, (\pi_s I_C)^{-1}\Gamma^{-1}(x_0)\Sigma(x_0)\Gamma^{-1}(x_0)\Big).$$

Remark 5. Using the Cramér–Wold device, Theorem 2 can be extended to the multi–dimension case by similar arguments. It seems from the proof in the Appendix that the i.i.d. assumption on $\{w_k\}$ can be relaxed. For example, we can obtain analogous results when $\{w_k\}$ is a stationary sequence of α–mixing random variables (cf., Karlsen, Mykelbust and Tjøstheim (2007)).

As an application of Theorem 2, we obtain the asymptotic distributions of local linear estimators and least absolute distance estimators in the following corollaries. First we consider local linear estimation, which corresponds to $\rho(x) = x^2$ and $\psi(x) = 2x$ in (1.4) and (1.5). Since $\psi(\cdot)$ is continuous, by an elementary calculation we have $\lambda_1 = 2$ and $\sigma^2 = 4Ew_1^2$.

Corollary 1. *Assume* $0 < Ew_1^2 < \infty$, *the conditions* A1, A2, A5, *and* $n^{\beta-\varepsilon_0}h \to \infty$ *for some* $0 < \varepsilon_0 < \beta$. *Let* $\Gamma_1(x_0)$, $\Sigma_1(x_0)$, *and* $b_1(x_0)$ *be defined as* $\Gamma(x_0)$, $\Sigma(x_0)$ *and* $b(x_0)$ *with* λ_1 *and* σ^2 *replaced by* 2 *and* $4Ew_1^2$, *respectively. Then we have*

$$(N_n h)^{1/2}\left\{\begin{pmatrix}\widehat{m}_n(x_0) - m(x_0)\\ h(\widehat{m}_n'(x_0) - m'(x_0))\end{pmatrix} - \Big(\frac{1}{2}h^2\Gamma_1^{-1}(x_0)b_1(x_0) + o(h^2)\Big)\right\}$$
$$\xrightarrow{d} N\Big((0,0)^{\tau}, \Gamma_1^{-1}(x_0)\Sigma_1(x_0)\Gamma_1^{-1}(x_0)\Big)^{1/2}.$$

Next we consider least absolute distance estimation, which corresponds to $\rho(x) = |x|$ and $\psi(x) = \text{sign}(x)$. Assume that $F(\cdot)$ has density function $f(\cdot)$ in a neighborhood of 0 and $f(0) > 0$, where $F(\cdot)$ is the distribution function of w_1. Then, as in Bai, Rao and Wu (1992), we have $\lambda_1 = 2f(0)$ and $\sigma^2 = 1$.

Corollary 2. *Assume* A1, A2, A5, it $n^{\beta-\varepsilon_0}h \to \infty$ for some $0 < \varepsilon_0 < \beta$, and $F(\cdot)$ has density function $f(\cdot)$ in a neighborhood of 0 and $f(0) > 0$. Then

$$(N_n h)^{1/2}\left\{\begin{pmatrix} \widehat{m}_n(x_0) - m(x_0) \\ h(\widehat{m}_n'(x_0) - m'(x_0)) \end{pmatrix} - \Big(\frac{1}{2}h^2\Gamma_2^{-1}(x_0)b_2(x_0) + o(h^2)\Big)\right\}$$
$$\xrightarrow{d} N\Big((0,0)^\tau, \Gamma_2^{-1}(x_0)\Sigma_2(x_0)\Gamma_2^{-1}(x_0)\Big)^{1/2},$$

where $\Gamma_2(x_0)$, $\Sigma_2(x_0)$, *and* $b_2(x_0)$ *are* $\Gamma(x_0)$, $\Sigma(x_0)$, *and* $b(x_0)$ *with* λ_1 *and* σ^2 *replaced by* $2f(0)$ *and* 1.

Remark 6. The least absolute distance estimators for stationary time series have been discussed by many authors, see Basset and Koenker (1978) and Bai, Chen, Wu and Zhao (1990) for example. Corollary 2 is a new result for nonstationary time series.

3.2. Some extensions

The technique in this paper can be extended to local polynomial M-estimation, which has been considered by many authors in the stationary case (cf., Jiang and Mack (2001)). Thus, find a and b_j, $1 \le j \le q$, to satisfy

$$\sum_{k=1}^{n}\psi\Big(Y_k - a - \sum_{j=1}^{q} b_j(X_k - x_0)^j\Big)K\Big(\frac{X_k - x_0}{h}\Big)\begin{pmatrix} 1 \\ \frac{X_k - x_0}{h} \\ \cdots \\ \Big(\frac{X_k - x_0}{h}\Big)^q \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix},$$

where $\mathbf{0}$ is a null vector of length q. Then $m(x_0)$ is estimated by the solution $\widehat{m}_n(x_0)$ and $m^{(j)}(x_0)$ is estimated by $j!\widehat{m}_n^{(j)}(x_0)$, $1 \le j \le q$. The asymptotic properties in Sections 3.1 still hold for robust local polynomial fitting. For example, we can show that under mild conditions $\sup_{x\in S} T_n(x) = o_P(1)$, and

$$(N_n h)^{1/2}\Big(T_n(x_0) - \frac{1}{(q+1)!}h^{q+1}\Gamma_3^{-1}(x_0)b_3(x_0) + o_P(h^{q+1})\Big)$$
$$\xrightarrow{d} N\Big((0, \mathbf{0}^\tau)^\tau, \Gamma_3^{-1}(x_0)\Sigma_3(x_0)\Gamma_3^{-1}(x_0)\Big),$$

where

$$T_n(x) = \left(\widehat{m}_n(x) - m(x), h(\widehat{m}_n'(x) - m'(x)), \ldots, h^q\Big(\widehat{m}_n^{(q)}(x) - \frac{m^{(q)}(x)}{q!}\Big)\right)^\tau,$$

$$\Gamma_3(x_0) = \lambda_1 p_s(x_0)(\mu_{i+j-2})_{(q+1)\times(q+1)}, \quad \Sigma_3(x_0) = \sigma^2 p_s(x_0)(\nu_{i+j-2})_{(q+1)\times(q+1)},$$
$$b_3(x_0) = \lambda_1 p_s(x_0) m^{(q+1)}(x_0)(\mu_{q+1}, \ldots, \mu_{2q+1})^\tau.$$

Besides weak consistency and asymptotic normality, the Bahadur representation of an M-estimator is another important topic, since it not only provides a kind of asymptotic representation for the estimator but also gives the order of the remainder term. There has been a large literature on Bahadur representation of parametric and nonparametric M-estimators for stationary time series, see He and Shao (1996), Hong (2003), and Cheng and Gooijer (2005) for details. Next we introduce the strong Bahadur representation for the local linear M-estimator in null recurrent time series. When $\psi(\cdot)$ in (1.5) is Lipschitz continuous of order one and

$$n^{\beta-\varepsilon_0} h \to \infty, \quad n^{\beta-\varepsilon_0} h^5 = O(1), \tag{3.4}$$

we can show that

$$\begin{pmatrix} \widehat{m}_n(x_0) - m(x_0) \\ h(\widehat{m}_n'(x_0) - m'(x_0)) \end{pmatrix} = \Gamma^{-1}(x_0) \sum_{k=1}^{n} \eta_{nk} + O\Big(\frac{1}{n^{\beta-\varepsilon_0} h}\Big) \ \text{ a.s.}, \tag{3.5}$$

where

$$\eta_{nk} = \frac{1}{N_n h} K\Big(\frac{X_k - x_0}{h}\Big) \psi\Big(Y_k - (m(x_0) + m'(x_0)(X_k - x_0))\Big) \begin{pmatrix} 1 \\ \frac{X_k - x_0}{h} \end{pmatrix}.$$

From (3.5), we can obtain the strong Bahadur representation for the nonparametric Huber estimator since the Huber ψ–function is Lipschitz continuous. However, the Lipschitz continuity on $\psi(\cdot)$ seems too restrictive in some cases. For example, the least absolute distance estimator is excluded in this case since $\psi(x) = \text{sign}(x)$ is not continuous at 0. In fact, if (3.4) is satisfied and the Lipschitz continuity on $\psi(\cdot)$ is replaced by

$$E[\psi(w_1 + u) - \psi(w_1 + v)]^p + E[\psi(w_1 + u) - \psi(w_1 + v)]^2 = O(|u - v|), \quad |u - v| \to 0, \tag{3.6}$$

where $p = [(72\beta + 3\varepsilon_0 + 60)/(120\beta - 40\varepsilon_0)] + 1$ for some $0 < \varepsilon_0 < \beta$, the following strong Bahadur representation holds:

$$\begin{pmatrix} \widehat{m}_n(x_0) - m(x_0) \\ h(\widehat{m}_n'(x_0) - m'(x_0)) \end{pmatrix} = \Gamma^{-1}(x_0) \sum_{k=1}^{n} \eta_{nk} + O\Big((n^{\beta-\varepsilon_0} h)^{-3/4}\Big) \ \text{ a.s.}. \tag{3.7}$$

From (3.7), we can obtain the strong Bahadur representation for the least absolute distance estimator. Meanwhile, the strong consistency for the local linear M-estimators can be obtained from (3.5) and (3.7). The detailed proofs of (3.5) and (3.7) are similar to those in Chen, Li and Zhang (2008).

In practice, the errors in model (1.2) might depend on the nonstationary regressor $\{X_k\}$. Thus, it is interesting to consider

$$Y_k = m(X_k) + \sigma(X_k)w_k,$$

where $\sigma(\cdot)$ is a continuous function satisfying some mild conditions. The above model with heteroscedastic errors is important in econometrics and finance. The proposed estimation procedure in this paper can be extended to the model with heteroscedastic errors. Furthermore, with some modification, we can show that the asymptotic results in this paper still hold for the case of heteroscedastic errors.

For the case of multivariate $\{X_k\}$ with $X_k^\tau = (X_{k1}, X_{k2}, \ldots, X_{kd})$, $d \geq 3$, it is known that the regression function $m(x_0)$ may not be estimated with accuracy due to "the curse of dimensionality". The curse of dimensionality problem has been clearly illustrated in several books, such as Fan and Gijbels (1996). The issue of how to avoid it is particularly important in nonlinear regression analysis. A way to solve this problem is to consider the nonlinear additive cointegration model

$$Y_k = \sum_{t=1}^{d} m_t(X_{kt}) + w_k, \tag{3.8}$$

where the functions $m_1(\cdot), \ldots, m_d(\cdot)$ are univariate. Several different methods have been proposed to deal with the additive model generated by stationary time series. For example, Fan, Härdle and Mammen (1998) considered the method of marginal integration, and Mammen, Linton and Nielsen (1999) applied the smoothed backfitting approach that has been extended to additive models generated by nonstationary time series by Schienle (2006). Recently,Lin, Li and Gao (2009) proposed a local linear M-type marginal integration method to study the additive models in the stationary case. We can extend the Lin, Li and Gao (2009) method to model (3.8). To obtain large sample theory for local linear M-smoothers in model (3.8), we need uniform weak consistency for the proposed estimators. Since more technicalities are involved, this is left for future research.

4. One-step Iterated Procedure

As pointed out in Section 1, the local linear M-estimators inherit many nice statistical properties from local linear smoothers and M-type estimators. However, it seems that it is difficult to obtain local M-estimators directly. A natural method is to apply an iterated procedure defined as follows:

$$\widehat{\mathbf{m}}_n^{(s)} = \widehat{\mathbf{m}}_n^{(s-1)} + \Big(W_n^{(s-1)}(x_0)\Big)^{-1}\Big(N_n\Psi_n(\widehat{\mathbf{m}}_n^{(s-1)}, x_0)\Big), \quad s = 1, \ldots, \tag{4.1}$$

where $\widehat{\mathbf{m}}_n^{(i)} := \widehat{\mathbf{m}}_n^{(i)}(x_0)$ is the value obtained from the i-th iteration at point x_0,

$$\Psi_n(t,x_0)=\frac{1}{N_n}\sum_{k=1}^{n}\psi\Big(Y_k-t^{\tau}\begin{pmatrix}1\\ \frac{X_k-x_0}{h}\end{pmatrix}\Big)K_{kh}\begin{pmatrix}1\\ \frac{X_k-x_0}{h}\end{pmatrix}=:\begin{pmatrix}\Psi_{n1}(t,x_0)\\ \Psi_{n2}(t,x_0)\end{pmatrix},$$

$$W_n^{(s)}(x_0)=\begin{cases}N_n\begin{pmatrix}\frac{\partial}{\partial a}\Psi_{n1}(t,x_0) & \frac{\partial}{h\partial b}\Psi_{n1}(t,x_0)\\ \frac{\partial}{\partial a}\Psi_{n2}(t,x_0) & \frac{\partial}{h\partial b}\Psi_{n2}(t,x_0)\end{pmatrix}\Bigg|_{t=\widehat{\mathbf{m}}_n^{(s)}}, & \text{if } \psi(\cdot) \text{ is differentiable},\\ N_n\widehat{\lambda}_{n1}\widehat{p}_s(x_0)\begin{pmatrix}\mu_0 & \mu_1\\ \mu_1 & \mu_2\end{pmatrix}, & \text{otherwise}.\end{cases}$$

Here $\widehat{\lambda}_{n1}$ is some consistent estimator of λ_1 and $N_n\widehat{p}_s(x_0) = (1/h)\sum_{k=1}^{n} K[(X_k - x_0)/h]$. The initial value $\widehat{\mathbf{m}}_n^{(0)}$ can be arbitrarily chosen, and the above procedure is terminated at s_0 if $\|\widehat{\mathbf{m}}_n^{(s_0)} - \widehat{\mathbf{m}}_n^{(s_0-1)}\| < 0.0001$.

The full iterated procedure is time–consuming when the sample size is large, and it may stop because of the singularity of $W_n^{(s)}(x_0)$. To overcome these difficulties, we apply the one–step iterated procedure

$$\widehat{\mathbf{m}}_n = \widehat{\mathbf{m}}_0 + \Big(W_n^{(0)}(x_0)\Big)^{-1}(N_n\Psi_n(\widehat{\mathbf{m}}_0, x_0)), \tag{4.2}$$

where the initial value $\widehat{\mathbf{m}}_0$ satisfies some mild conditions. Fan and Chen (1999) and Fan and Jiang (2000) studied this for stationary time series, and we show that it works well in theory and practice for nonstationary time series. If the initial estimators satisfy

$$\widehat{m}_{n0}(x_0) - m(x_0) = O_P\Big(h^2 + \frac{1}{\sqrt{n^{\beta}L_s(n)h}}\Big),$$
$$\widehat{m}'_{n0}(x_0) - m'(x_0) = O_P\Big(h + \frac{1}{h\sqrt{n^{\beta}L_s(n)h}}\Big), \tag{4.3}$$

then the asymptotic distribution in (3.2) holds for the one–step local linear M-estimator. The detailed proof is similar to that of Theorem 4.1 in Fan and Jiang (2000). Furthermore, (4.3) is mild and can be satisfied by many commonly used estimators such as the local linear estimators (cf., Corollary 1).

5. Examples of Implementation

5.1. Bandwidth selection

A difficult problem in simulation is the choice of optimal bandwidth. In this section, we mainly consider the special case of the random walk. It is not difficult to see that the rates for the random walk are different from that for stationary processes with n being replaced by $\sqrt{n}$. From Karlsen, Mykelbust and Tjøstheim (2007), it can be seen that the cross–validation method works well in NW estimation for null–recurrent processes. Here we apply this method for local linear M-estimators.

Cross–validation is very useful in assessing the performance of estimators via estimating their prediction errors, here defined as

$$CV(h) = \sum_{k=1}^{n} (Y_k - \widehat{m}_{h,-k}(X_k))^2, \tag{5.1}$$

where $\widehat{m}_{h,-k}(X_k)$ is the local linear M-estimator with bandwidth h and the k–th observation left out. The bandwidth is selected to minimize $CV(h)$.

For simplicity, we apply the following method in a simulation study. For a predetermined sequence of h's selected from a wide range, say from 0.01 to 0.7 with increment 0.01, we choose the bandwidth that minimizes $CV(h)$.

5.2. Numerical examples

Suppose

$$Y_k = m(X_k) + w_k, \quad m(x) = \frac{1}{3}e^{x} + \frac{2}{3}e^{-x}, \tag{5.2}$$

where $\{X_k\}$ is generated by the random walk process $X_k = X_{k-1} + x_k$. Here $\{x_k\}$ is a sequence of i.i.d. $N(0, 0.1^2)$ random variables, $\{w_k\}$ is independent of $\{x_k\}$ and is taken from one of the following distributions.

(i) Standard normal distribution $N(0, 1)$.

(ii) Symmetric contaminated normal distribution $0.8N(0, 1) + 0.2N(0, 15^2)$.

(iii) Cauchy distribution $C(0, 1)$.

The data from each of the above distributions consisted of 1,000 replications of samples of sizes $n = 500$ and $n = 1,000$. The uniform kernel $K(u) = (1/2)I(|u| \le 1)$ was applied in this example and the cross–validation method was used to choose the optimal bandwidth. The iterated procedure for M-estimator in Section 4 was applied in our simulation and the local linear estimators were chosen as the initial value. The measure for the performance of the estimators was taken to be the mean square error (MSE) of the form

$$\frac{1}{n}\sum_{k=1}^{n}(\widehat{m}_n(X_k) - m(X_k))^2 \quad \text{and} \quad \frac{1}{n}\sum_{k=1}^{n}(\widehat{m}_n'(X_k) - m'(X_k))^2. \tag{5.3}$$

The simulation results are reported in Tables 1 and 2 below (LLE: local linear estimator, LLME: local linear M-estimator).

From Tables 1 and 2, we can see that when $\{w_k\}$ was standard normal, both the local linear estimators and the local linear M-estimators performed well, when

Table 1. The average of MSE for $m(\cdot)$.

Distribution	Sample size	LLE	LLME
Standard normal	$n = 500$	0.0390	0.0412
Standard normal	$n = 1,000$	0.0345	0.0302
Contaminated normal	$n = 500$	0.7748	0.0485
Contaminated normal	$n = 1,000$	0.4127	0.0337
Cauchy distribution	$n = 500$	34.0389	0.1302
Cauchy distribution	$n = 1,000$	21.2932	0.0876

Table 2. The average of MSE for $m'(\cdot)$.

Distribution	Sample size	LLE	LLME
Standard normal	$n = 500$	0.0968	0.0811
Standard normal	$n = 1,000$	0.0471	0.0489
Contaminated normal	$n = 500$	1.1386	0.1238
Contaminated normal	$n = 1,000$	1.0264	0.1483
Cauchy distribution	$n = 500$	47.1866	0.3390
Cauchy distribution	$n = 1,000$	22.9225	0.1676

$\{w_k\}$ was contaminated or heavy–tailed ($C(0,1)$), the local linear M-estimators were much more robust.

Next, we consider the following linear cointegration model (cf., Karlsen, Mykelbust and Tjøstheim (2007) and Chen, Li and Zhang (2009))

$$Y_k = X_k + w_k, \quad k = 1, \ldots, 1000,$$

where $X_k = X_{k-1} + x_k$, $\{x_k\}$ is a sequence of i.i.d. $N(0, 0.1^2)$ random variables and $\{w_k\}$ is independent of $\{x_k\}$ and has the common symmetric contaminated normal distribution as above.

To compare the local linear M-estimator with the NW estimator, which was considered by Karlsen, Mykelbust and Tjøstheim (2007), we report the estimated curves in Figure 1. The solid line is the true regression function, the dashed line is the NW estimator and the star line is the local linear M-estimator. From it, we can see that the local linear M-estimator performed better than the NW estimator when $\{w_k\}$ was contaminated. Furthermore, the local linear M-estimator overcame the boundary effect in our study.

Acknowledgement

The authors would like to thank an associate editor and the referee for valuable suggestions and comments that greatly improved the manuscript. This work was supported by National Natural Science Foundation of China (Grant No.

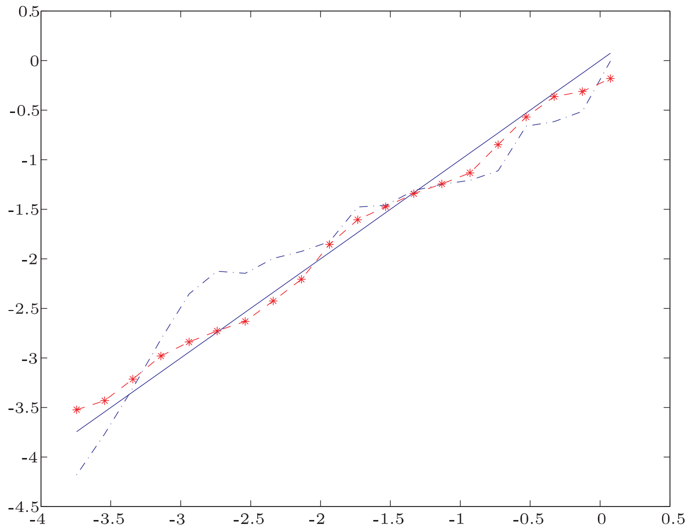

Figure 1. The estimated curves of the regression function.

10871177) and Specialized Research Fund for the Doctor Program of Higher Education (Grant No. 20060335032). The second author was also partially supported by the National Natural Science Foundation of China (Grant No. 10671176).

Appendix. Proofs of the Main Results

Let $t = (a, hb)^\tau$ and $t_x = (m(x), hm'(x))^\tau$, and write

$$\Theta_n(t,x) = \frac{1}{N_n}\sum_{k=1}^{n}\rho\left(Y_k - t^\tau\begin{pmatrix}1\\ \frac{X_k-x}{h}\end{pmatrix}\right)K_{kh}(x),$$

$$\Theta_n(t_x,x) = \frac{1}{N_n}\sum_{k=1}^{n}\rho\left(Y_k - t_x^\tau\begin{pmatrix}1\\ \frac{X_k-x}{h}\end{pmatrix}\right)K_{kh}(x),$$

$$\Psi_n(t_x,x) = \frac{1}{N_n}\sum_{k=1}^{n}\psi\left(Y_k - t_x^\tau\begin{pmatrix}1\\ \frac{X_k-x}{h}\end{pmatrix}\right)\begin{pmatrix}1\\ \frac{X_k-x}{h}\end{pmatrix}K_{kh}(x),$$

where $K_{kh}(x) = (1/h)K[(X_k - x)/h])$. Before giving the proof of Theorem 1, we establish the following two lemmas which are critical in its proof.

Lemma A.1. *Suppose that $\{X_k\}$ and $\{w_k\}$ are independent, $\{X_k\}$ is β–null recurrent, and $\{w_k\}$ is i.i.d. Then the compound process $\{X_k, w_k\}$ is β–null recurrent.*

Proof. The lemma is an immediate consequence of Lemma 3.1 in Karlsen, Mykelbust and Tjøstheim (2007).

Lemma A.2. *Suppose the conditions of Theorem* 1 *hold. Then for any constant* $c > 0$*, we have*

$$\sup_{x\in S}\sup_{\|t-t_x\|\le c}\Big|\Theta_n(t,x)-\Theta_n(t_x,x)-(t-t_x)^\tau\Psi_n(t_x,x) -\frac{1}{2}(t-t_x)^\tau\Gamma(x)(t-t_x)\Big| = o_P(1), \tag{A.1}$$

where $\|\cdot\|$ *is the* L_2*–norm in* $\mathcal{R}^2$*.*

Proof. By the convexity of $\rho(\cdot)$, it suffices to show that for $\|t-t_x\|\le c$, $x\in S$,

$$\sup_{x\in S}\Big|\Theta_n(t,x)-\Theta_n(t_x,x)-(t-t_x)^\tau\Psi_n(t_x,x)-\frac{1}{2}(t-t_x)^\tau\Gamma(x)(t-t_x)\Big| = o_P(1). \tag{A.2}$$

Note that

$$\begin{aligned}&\Theta_n(t,x)-\Theta_n(t_x,x)-(t-t_x)^\tau\Psi_n(t_x,x)\\ &\quad -E\Big[\Theta_n(t,x)-\Theta_n(t_x,x)-(t-t_x)^\tau\Psi_n(t_x,x)\Big]\\ &\quad =\sum_{k=1}^n(V_{nk}(t,x)-EV_{nk}(t,x)),\end{aligned}$$

where

$$\begin{aligned}V_{nk}(t,x)=\frac{1}{N_n}K_{kh}(x)\Bigg[&\rho\Bigg(Y_k-t^\tau\begin{pmatrix}1\\ \frac{X_k-x}{h}\end{pmatrix}\Bigg)-\rho\Bigg(Y_k-t_x^\tau\begin{pmatrix}1\\ \frac{X_k-x}{h}\end{pmatrix}\Bigg)\\ &-(t-t_x)^\tau\begin{pmatrix}1\\ \frac{X_k-x}{h}\end{pmatrix}\psi\Bigg(Y_k-t_x^\tau\begin{pmatrix}1\\ \frac{X_k-x}{h}\end{pmatrix}\Bigg)\Bigg].\end{aligned}$$

Next, we prove that for $\|t-t_x\|\le c$, $x\in S$,

$$\sup_{x\in S}\Bigg|\sum_{k=1}^n\Big[V_{nk}(t,x)-EV_{nk}(t,x)\Big]\Bigg| = o_P(1). \tag{A.3}$$

Since S is a compact set, it can be covered by a finite number of subsets $\{S_i\}$ centered at s_i with radius $(n^{\beta-\varepsilon_0/2}h)^{-(2m-1)}$, where $m>\max\{(1+\beta)/\varepsilon_0-11/8, 5/4\}$. If H_n is the number of these sets, then $H_n=O((n^{\beta-\varepsilon_0/2}h)^{2m-1})$.

Observe that

$$\begin{aligned}
&\sup_{x\in S}\left|\sum_{k=1}^{n}\Big[V_{nk}(t,x)-EV_{nk}(t,x)\Big]\right| \\
&\quad\le \max_{1\le i\le H_n}\left|\sum_{k=1}^{n}\Big[V_{nk}(t,s_i)-EV_{nk}(t,s_i)\Big]\right| \\
&\quad\quad+\max_{1\le i\le H_n}\sup_{x\in S_i}\left|\sum_{k=1}^{n}\Big[V_{nk}(t,x)-V_{nk}(t,s_i)\Big]\right| \\
&\quad\quad+\max_{1\le i\le H_n}\sup_{x\in S_i}\left|\sum_{k=1}^{n}\Big[EV_{nk}(t,x)-EV_{nk}(t,s_i)\Big]\right| \\
&\quad=: I_{n,1}+I_{n,2}+I_{n,3}.
\end{aligned}$$

By Lemma 3.4 in Karlsen and Tjøstheim (2001), we have, with probability 1,

$$n^{\beta-(\varepsilon_0/8)} \ll N_n \ll n^{\beta+(\varepsilon_0/8)}.$$

Thus, by A1, it is easy to check that

$$I_{n,3}=O_P\Big(H_n^{-1}n^{1-\beta+(\varepsilon_0/8)}h^{-2}\Big)=o_P\Big(n^{-\varepsilon_0 m+1+\beta-(11/8)\varepsilon_0}\Big)=o_P(1), \qquad (A.4)$$

since $m>(1+\beta)/\varepsilon_0-11/8$. Analogously, we can also show that

$$I_{n,2}=o_P(1). \qquad (A.5)$$

In view of (A.4) and (A.5), to show (A.3) we only need to show that

$$I_{n,1}=o_P(1). \qquad (A.6)$$

We apply the truncation method and the independence decomposition in the split Markov chain to prove (A.6). Define

$$\begin{aligned}
V'_{nk}(t,x)=K_{kh}(x)\Bigg[&\rho\bigg(Y_k-t^{\tau}\begin{pmatrix}1\\ \frac{X_k-x}{h}\end{pmatrix}\bigg)-\rho\bigg(Y_k-t_x^{\tau}\begin{pmatrix}1\\ \frac{X_k-x}{h}\end{pmatrix}\bigg)\\
&-(t-t_x)^{\tau}\begin{pmatrix}1\\ \frac{X_k-x}{h}\end{pmatrix}\psi\bigg(Y_k-t_x^{\tau}\begin{pmatrix}1\\ \frac{X_k-x}{h}\end{pmatrix}\bigg)\Bigg],
\end{aligned}$$

$$Q_l(t,x)=\begin{cases}\displaystyle\sum_{k=0}^{\tau_0}V'_{nk}(t,x), & l=0,\\ \displaystyle\sum_{k=\tau_{l-1}+1}^{\tau_l}V'_{nk}(t,x), & l=1,\ldots,N_n,\\ \displaystyle\sum_{k=\tau_{N_n}+1}^{n}V'_{nk}(t,x), & l=(n),\end{cases} \qquad (A.7)$$

where the definitions of τ_l and N_n are given in Nummelin (1984). By (1.2), $V'_{nk}(t,x)$ can be rewritten as a function of (w_k, X_k). Hence, by Lemma A.1 and the Nummelin (1984) result, we know that $\{Q_l(t,x), (\tau_l - \tau_{l-1}), l \geq 1\}$ is a sequence of i.i.d. random variables for fixed t and x. Obviously,

$$\sum_{k=1}^{n} V_{nk}(t,x) = \frac{1}{N_n}\Big(Q_0(t,x) + \sum_{l=1}^{N_n} Q_l(t,x) + Q_{(n)}(t,x)\Big). \tag{A.8}$$

By arguments similar to those used in the proof of Theorem 5.1 in Karlsen and Tjøstheim (2001), we have

$$\frac{1}{N_n} \max_{1\leq i\leq H_n} \Big|Q_l(t,s_i) - EQ_l(t,s_i)\Big| = o_P(1) \tag{A.9}$$

for $l = 0, (n)$. Thus, to show (A.6), we need only prove for any small $\epsilon > 0$

$$P\Bigg\{\max_{1\leq i\leq H_n} \Bigg|\frac{1}{N_n}\sum_{l=1}^{N_n}(Q_l(t,s_i) - EQ_l(t,s_i))\Bigg| > \epsilon\Bigg\} \to 0. \tag{A.10}$$

By Lemma 3.4 in Karlsen and Tjøstheim (2001), to show (A.10), we need only prove

$$P\Bigg\{\max_{1\leq i\leq H_n} \Bigg|\frac{1}{N_n}\sum_{l=1}^{N_n}(Q_l(t,s_i) - EQ_l(t,s_i))\Bigg| > \epsilon, n^{\beta-(\varepsilon_0/8)} \ll N_n \ll n^{\beta+(\varepsilon_0/8)}\Bigg\} \to 0. \tag{A.11}$$

Next, we apply the truncation method. Let

$$\overline{Q}_l(t,i) = Q_l(t,s_i)I\Big(|Q_l(t,s_i)| < n^{\beta-(\varepsilon_0/4)}\Big) \quad \text{and} \quad \widetilde{Q}_l(t,i) = Q_l(t,i) - \overline{Q}_l(t,i).$$

By standard arguments, and noting $EQ_l(t,i) = 0$, we have

$$\begin{aligned}
&P\Bigg\{\max_{1\leq i\leq H_n} \Bigg|\frac{1}{N_n}\sum_{l=1}^{N_n}(Q_l(t,s_i) - EQ_l(t,s_i))\Bigg| > \epsilon, n^{\beta-(\varepsilon_0/8)} \ll N_n \ll n^{\beta+(\varepsilon_0/8)}\Bigg\} \\
&\leq P\Bigg\{\max_{1\leq i\leq H_n} \Bigg|\frac{1}{N_n}\sum_{l=1}^{N_n}(\overline{Q}_l(t,i) - E\overline{Q}_l(t,i))\Bigg| > \frac{\epsilon}{2}, n^{\beta-(\varepsilon_0/8)} \ll N_n \ll n^{\beta+(\varepsilon_0/8)}\Bigg\} \\
&\quad + P\Bigg\{\max_{1\leq i\leq H_n} \Bigg|\frac{1}{N_n}\sum_{l=1}^{N_n}(\widetilde{Q}_l(t,i) - E\widetilde{Q}_l(t,i))\Bigg| > \frac{\epsilon}{2}, n^{\beta-(\varepsilon_0/8)} \ll N_n \ll n^{\beta+(\varepsilon_0/8)}\Bigg\}.
\end{aligned} \tag{A.12}$$

By Lemma 5.2 in Karlsen and Tjøstheim (2001), Lemma D.1 in Gao, King, Lu and Tjøstheim (2008), and an elementary calculation, we have

$$\max_{1\leq i\leq H_n} E|Q_l(t,i)|^{2m} \leq Ch^{-2m+1}. \tag{A.13}$$

As $m > 5/4$, we have

$$P\left\{\max_{1\le i\le H_n}\left|\frac{1}{N_n}\sum_{l=1}^{N_n}\Big(\widetilde{Q}_l(t,i)-E\widetilde{Q}_l(t,i)\Big)\right|>\frac{\epsilon}{2}, n^{\beta-(\varepsilon_0/8)}\ll N_n\ll n^{\beta+(\varepsilon_0/8)}\right\}$$
$$\le CH_nP\left\{\left|\frac{1}{N_n}\sum_{l=1}^{N_n}\Big(\widetilde{Q}_l(t,1)-E\widetilde{Q}_l(t,1)\Big)\right|>\frac{\epsilon}{2}, n^{\beta-(\varepsilon_0/8)}\ll N_n\ll n^{\beta+(\varepsilon_0/8)}\right\}$$
$$\le CH_nn^{\beta+(\varepsilon_0/8)}P\Big\{|Q_1(t,1)|>n^{\beta-(\varepsilon_0/4)}\Big\}$$
$$\le CH_nn^{\beta+(\varepsilon_0/8)}h^{1-2m}n^{-2m(\beta-(\varepsilon_0/4))}=O\Big(n^{-((\varepsilon_0 m/2)-(5/8)\varepsilon_0)}\Big)=o(1), \quad \text{(A.14)}$$

where C denotes a positive constant which may change from line to line. On the other hand, by the Bernstein Inequality, we have

$$P\left\{\max_{1\le i\le H_n}\left|\frac{1}{N_n}\sum_{l=1}^{N_n}\Big(\overline{Q}_l(t,i)-E\overline{Q}_l(t,i)\Big)\right|>\frac{\epsilon}{2}, n^{\beta-(\varepsilon_0/8)}\ll N_n\ll n^{\beta+(\varepsilon_0/8)}\right\}$$
$$\le CH_n\sum_{k=c_1n^{\beta-(\varepsilon_0/8)}}^{c_2n^{\beta+(\varepsilon_0/8)}}P\left\{\left|\frac{1}{k}\sum_{l=1}^{k}\Big(\overline{Q}_l(t,1)-E\overline{Q}_l(t,1))\Big)\right|>\frac{\epsilon}{2}\right\}$$
$$\le CH_n\sum_{k=c_1n^{\beta-(\varepsilon_0/8)}}^{c_2n^{\beta+(\varepsilon_0/8)}}\exp\Big\{-kn^{-\beta+(\varepsilon_0/4)}\Big\}\le CH_n\exp\Big\{-n^{(\varepsilon_0/8)}\Big\}=o(1), \quad \text{(A.15)}$$

where c_1 and c_2 are some positive constants. By (A.12), (A.14), and (A.15), (A.11) is proved, which implies that (A.3) and (A.6) hold. Finally, it is easy to show that

$$E\Big[\Theta_n(t,x)-\Theta_n(t_x,x)-(t-t_x)^\tau\Psi_n(t_x,x)\Big]=\frac{1}{2}(t-t_x)^\tau\Gamma(x)(t-t_x)+o(1). \quad \text{(A.16)}$$

(A.3) and (A.16) imply that (A.2) holds. Furthermore, since $\Theta_n(t,x)-\Theta_n(t_x,x)-(t-t_x)^\tau\Psi_n(t_x,x)$ is convex in t and $(t-t_x)^\tau\Gamma(x)(t-t_x)$ is continuous and convex in t, (A.1) follows from (A.2) and Theorem 10.8 in Rockafellar (1970).

Proof of Theorem 1. Let $\widehat{\mathbf{m}}_n(x)=(\widehat{m}_n(x),h\widehat{m}_n'(x))^\tau$. To show (3.1), it is enough to prove that, for any small $\epsilon>0$,

$$P\Big(\sup_{x\in S}\Big\|\widehat{\mathbf{m}}_n(x)-t_x\Big\|\ge\epsilon\Big)\to 0. \quad \text{(A.17)}$$

Taking $c=\epsilon$ in (A.1), we have

$$\sup_{x\in S}\sup_{\|t-t_x\|\le\epsilon}\Big|\Theta_n(t,x)-\Theta_n(t_x,x)+(t-t_x)^\tau\Psi_n(t_x,x)$$
$$-\frac{1}{2}(t-t_x)^\tau\Gamma(x)(t-t_x)\Big|=o_P(1). \quad \text{(A.18)}$$

Since $\inf_{x\in S} p_s(x) > 0$, then for $\|t - t_x\| = \epsilon$ we have

$$\frac{1}{2}(t - t_x)^\tau \Gamma(x)(t - t_x) \geq \frac{\epsilon^2 \kappa_0(x)}{2}, \tag{A.19}$$

where $\kappa_0(x)$ is the smallest eigenvalue of $\Gamma(x)$. On the other hand, using the decomposition technique of (A.7) in the proof of Lemma A.2, we can prove

$$\sup_{x\in S} \Psi_n(t_x) = o_P(1). \tag{A.20}$$

Hence, by (A.18)−(A.20), the convexity of $\rho(\cdot)$ and the definition of $\widehat{\mathbf{m}}_n(x)$, we can obtain (A.17). Therefore, the proof of Theorem 1 is completed.

In order to establish asymptotic normality of the local linear M-estimators, we need the following.

Lemma A.3. *Under the conditions of Theorem* 2*, we have*

$$\Big\|\Psi_n(\widehat{\mathbf{m}}_n(x_0), x_0) - \Psi_n(t_{x_0}, x_0) + \Gamma(x_0)(\widehat{\mathbf{m}}_n(x_0) - t_{x_0})\Big\| = o_P\Big(\frac{1}{\sqrt{n^\beta L_s(n) h}}\Big).$$

Proof. Using the same method as in the proof of Lemma A.2, and by Theorem 2.5.7 in Rockafellar (1970), we can prove Lemma A.3. Details are omitted here.

Proof of Theorem 2. Recall that $\Psi_n(\widehat{\mathbf{m}}_n(x_0), x_0) = (0,0)^\tau$ and, by Lemma A.3, we have

$$\widehat{\mathbf{m}}_n(x_0) - t_{x_0} = \Gamma^{-1}(x_0)\Psi_n(t_{x_0}, x_0) + o_P\left(\frac{1}{\sqrt{n^\beta L_s(n) h}}\right), \tag{A.21}$$

which is a weak Bahadur representation of the estimator $\widehat{\mathbf{m}}_n(x_0)$. Let $R(X_k) = m(X_k) - m(x_0) - m'(x_0)(X_k - x_0)$. We consider $E\Psi_n(t_{x_0}, x_0)$. Note that

$$\Psi_n(t_{x_0}, x_0) = \frac{1}{N_n}\sum_{k=1}^n \Big\{\psi(w_k) + \Big[\psi(w_k + R(X_k)) - \psi(w_k)\Big]\Big\} K_{kh}\left(\begin{matrix}1\\ \frac{X_k - x_0}{h}\end{matrix}\right). \tag{A.22}$$

By A2, A3, A5 and an elementary calculation, we have

$$E\Psi_n(t_{x_0}, x_0) = \frac{1}{2}h^2 b(x_0)(1 + o(1)). \tag{A.23}$$

Moreover, by (A.23), we have

$$\begin{aligned}&\frac{1}{N_n}\sum_{k=1}^n \Big[\psi(w_k + R(X_k)) - \psi(w_k)\Big] K_{kh}\left(\begin{matrix}1\\ \frac{X_k - x_0}{h}\end{matrix}\right)\\ &\quad = \frac{1}{2}h^2 b(x_0)(1 + o_P(1)) + o_P\left(\frac{1}{\sqrt{n^\beta L_s(n) h}}\right).\end{aligned} \tag{A.24}$$

By (A.21), (A.22) and (A.24), we need only prove

$$\sqrt{\frac{h}{N_n}}\sum_{k=1}^{n}\psi(w_k)K_{kh}\begin{pmatrix}1\\ \frac{X_k-x_0}{h}\end{pmatrix}\xrightarrow{d}N((0,0)^{\tau},\Sigma(x_0)). \qquad (A.25)$$

(A.25) can be proved using the Cramér–Wald device and Theorem 3.1 of Karlsen, Mykelbust and Tjøstheim (2007).

References

Bai, Z., Chen, X., Wu, Y. and Zhao, L. (1990). Asymptotic normality of minimum L_1-norm estimates in linear models. *Sci. China Ser. A* **33**, 449-463.

Bai, Z., Rao, C. R. and Wu, Y. (1992). M-estimation of multivariate linear regression parameters under a convex discrepancy function. *Statist. Sinica* **2**, 237-254.

Basset, G. and Koenker, R. (1978). Asymptotic theory of least absolute error regression. *J. Amer. Statist. Assoc.* **73**, 618-622.

Cai, Z. and Ould-Saïd, E. (2003). Local M-estimator for nonparametric time series. *Statist. Probab. Lett.* **65**, 433-449.

Chen, J., Li, D. and Zhang, L. (2009). Robust estimation in a nonlinear cointegration model. Revised for *J. Multivariate Anal.*

Chen, J., Li, D. and Zhang, L. (2008). Bahadur representation of nonparametric M-estimator for spatial processes. *Acta Math. Sinica, English Series* **24**, 1871-1882.

Cheng, Y. and Gooijer, J. (2005). Bahadur representation for the nonparametric M-estimator under α–mixing dependence. Tinbergen Institute Discussion Paper, Department of Quantitative Economics, Faculty of Economics and Econometrics, University of Amsterdam.

Fan, J. and Chen, J. (1999). One-step local quasi-likelihood estimation. *J. Roy. Statist. Soc. Ser. B* **61**, 927-943.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

Fan, J., Härdle, W. and Mammen, E. (1998). Direct estimation of low-dimensional components in additive models. *Ann. Statist.* **26**, 943-971.

Fan, J. and Jiang, J. (2000). Variable bandwidth and one–step local M-estimator. *Sci. China Ser. A* **43**, 65-81.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer, New York.

Gao, J. (2007). *Nonlinear Time Series: Semiparametric Methods.* Chapman and Hall, London.

Gao, J., King, M., Lu, Z. and Tjøstheim, D. (2008). Nonparametric specification testing for nonlinear time series with nonstationarity. Available at `www.adelaide.edu.au/directory/jiti.gao`.

He, X. and Shao, Q. (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *Ann. Statist.* **24**, 2608-2630.

Hong, S. (2003). Bahadur representation and its applications for local polynomial estimates in nonparametric M-estimation. *J. Nonparametr. Stat.* **52**, 237-251.

Jiang, J. and Mack, Y. P. (2001). Robust local polynomial regression for dependent data. *Statist. Sinica* **11**, 705-722.

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

Karlsen, H. A., Mykelbust, T. and Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *Ann. Statist.* **35**, 252–299.

Kasahara, Y. (1984). Limit theorems for Lévy processes and Poisson point processes and their applications to Brownian excursions. *J. Math. Kyoto Univ.* **24**, 521-538.

Li, Q. and Racine, J. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton.

Lin, Z., Li, D. and Gao, J. (2009). Local linear M-estimation in nonparametric spatial regression. *J. Time Series Anal.* **30**, 286-314.

Mammen, E., Linton, O. and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27**, 1443-1490.

Nummelin, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press.

Park, J. and Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica* **69**, 117-162.

Phillips, P. C. B. and Park, J. (1998). Nonstationary density estimation and kernel autoregression. Cowles Foundation Discussion Paper, No. 1181, Yale University.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton.

Schienle, M. (2006). Reaching for econometric generality: nonparametric nonstationary regression. Working paper available from the Department of Economics, University of Mannheim, Germany.

Wang, Q. and Phillips, P. C. B. (2006). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. Cowles Foundation Discussion Paper, No. 1594, Yale University.

Department of Mathematics, Zhejiang University, Hangzhou 310027, China.

E-mail: zlin@zju.edu.cn

School of Economics, University of Adelaide, Adelaide SA 5005, Australia.

E-mail: ldgofzju@yahoo.com, degui.li@adelaide.edu.au

School of Economics, University of Adelaide, Adelaide SA 5005, Australia

E-mail: jia.chen@adelaide.edu.au

(Received January 2008; accepted September 2008)