Local composite quantile regression smoothing for Harris recurrent Markov processes

Degui Li [a,*], Runze Li [b]

[a] *Department of Mathematics, University of York, York, YO10 5DD, UK*

[b] *Department of Statistics and the Methodology Center, Pennsylvania State University, University Park, PA 16802-2111, USA*

ARTICLE INFO

ABSTRACT

In this paper, we study the local polynomial composite quantile regression (CQR) smoothing method for the nonlinear and nonparametric models under the Harris recurrent Markov chain framework. The local polynomial CQR regression method is a robust alternative to the widely-used local polynomial method, and has been well studied in stationary time series. In this paper, we relax the stationarity restriction on the model, and allow that the regressors are generated by a general Harris recurrent Markov process which includes both the stationary (positive recurrent) and nonstationary (null recurrent) cases. Under some mild conditions, we establish the asymptotic theory for the proposed local polynomial CQR estimator of the mean regression function, and show that the convergence rate for the estimator in nonstationary case is slower than that in stationary case. Furthermore, a weighted type local polynomial CQR estimator is provided to improve the estimation efficiency, and a data-driven bandwidth selection is introduced to choose the optimal bandwidth involved in the nonparametric estimators. Finally, we give some numerical studies to examine the finite sample performance of the developed methodology and theory.

1. Introduction

During the past three decades, there has been increasing interest in nonparametric smoothing techniques to study time series data with possible nonlinearity, as the nonparametric methodology allows the time series data "speak for themselves", and thus provides a flexible tool to explore the relationship between variables and avoid some potential model misspecification problems. When the observations satisfy certain stationarity condition, there have been extensive studies on various nonparametric estimation approaches such as the local polynomial method (Fan and Gijbels, 1996) and the penalized spline method (Ruppert et al., 2003). However, as pointed out in the literature, the stationarity assumption seems too restrictive in practical applications. When tackling economic and financial issues from a time perspective, we often deal with nonstationary components. For example, neither the consumer price index nor the share price index, nor the exchange rates constitute a stationary process. Hence, in recent years, several nonparametric smoothing methods such as the Nadaraya–Watson kernel method, the local linear method and the series approximation

method have been developed to model time series data with nonstationarity. Existing studies include Park and Hahn (1999), Karlsen and Tjøstheim (2001), Karlsen et al. (2007), Cai et al. (2009), Lin et al. (2009), Wang and Phillips (2009a,b) and Chen et al. (2012).

In this paper, we consider the nonparametric regression model defined by

$$Y_t = m(X_t) + \sigma(X_t)e_t, \quad t = 1, \ldots, n, \tag{1.1}$$

where $\{X_t\}$ is generated by a Harris recurrent Markov process which will be defined in Section 2, $\{e_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with mean zero, $m(\cdot)$ is the conditional mean function, and $\sigma^2(\cdot)$ is the conditional variance (or volatility) function. For the case of stationary time series, model (1.1) has been extensively studied in the literature, see, for example, Fan and Gijbels (1996), Härdle and Tsybakov (1997) and Fan and Yao (1998). Wang and Wang (2013) studied model (1.1) when $\{X_t\}$ is a unit root process, and Cai and Tjøstheim (2015) estimated model (1.1) when $\{X_t\}$ is null recurrent. In this paper, the Harris recurrence on $\{X_t\}$ provides a general framework for our study, and both the stationary and nonstationary cases can be included. When $\{X_t\}$ is nonstationary and $\sigma(\cdot) \equiv 1$, model (1.1) reduces to the nonlinear cointegration model, which has been extensively studied

---

* Corresponding author.
E-mail addresses: degui.li@york.ac.uk (D. Li), rzli@psu.edu (R. Li).

in the econometrics literature (c.f., Karlsen et al., 2007; Wang and Phillips, 2009a,b) since the publications of Granger (1981) and Engle and Granger (1987). The main interest of this paper is to study the nonparametric estimation of $m(\cdot)$, which is also the major focus of the existing literature on nonlinear time series with nonstationarity. To estimate the conditional mean function, most of the existing literature considers the local least squares approaches such as the local linear smoothing method which is popular and convenient especially when the error follows a stationary and normal distribution. However, it is well known that the local least squares estimator is sensitive to outliers and does not perform well when the error distribution is heavy-tailed as shown by Fan et al. (1994) and Lin et al. (2009). Outliers or aberrant observations are common in nonstationary time series data from economics and many other applied fields, and heavy-tailed distribution is an important feature of some nonstationary time series data from finance. Hence, it is important to derive a robust and flexible estimation approach for the conditional mean function $m(\cdot)$.

In this paper, to estimate $m(\cdot)$, we will develop a robust local estimation procedure which attenuates the lack of robustness of the local least squares estimator. The robust estimation methodology is built on a combination of local polynomial smoothing and composite quantile regression (CQR), and thus inherits the advantages from both the local polynomial smoothing and CQR techniques. For the case of i.i.d. observations, Zou and Yuan (2008) proposed the CQR method to estimate the regression coefficients in the classical linear regression model, and Kai et al. (2010) developed local CQR smoothers to estimate the nonlinear mean regression function and its derivatives. They have shown that the CQR estimator could be much more efficient than the least squares based estimator for i.i.d observations when the model error distribution is heavy-tailed. However, the restriction of i.i.d. observations would limit the applicability of the CQR method in some applied fields, and thus it is important for us to relax such restriction. To the best of our knowledge, there is virtually no work on estimation of $m(\cdot)$ through using the local polynomial CQR method when $\{X_t\}$ is a nonstationary process. This paper aims to fill in this gap.

It is well known that a minimal condition for conducting the asymptotic analysis on local estimation of $m(x_0)$ is that, as the number of observations increases, there must be infinitely many observations in any neighborhood of x_0, which means that the process $\{X_t\}$ must return to a neighborhood of x_0 infinitely often. To achieve this, in this paper, $\{X_t\}$ is assumed to be ϕ-irreducible Harris recurrent, making asymptotics for nonparametric local estimation procedure possible. The assumption of Harris recurrence on $\{X_t\}$ indicates that the process includes both stationary (positive recurrent) and nonstationary (null recurrent) cases. Under some mild conditions, we establish the asymptotic theory for the proposed local polynomial CQR estimator of the mean regression function, and show that the convergence rate for the estimator in nonstationary case is slower than that in stationary case. A recent paper by Lin et al. (2009) considered the local linear M-smoother for model (1.1) with $\sigma(\cdot) \equiv 1$ when $\{X_t\}$ is a null recurrent Markov process. Hence, this paper provides a more general framework than that in Lin et al. (2009). On the other hand, compared with the kernel estimation method for model (1.1) studied by Wang and Wang (2013) and Cai and Tjøstheim (2015) when $\{X_t\}$ is nonstationary, the proposed local polynomial CQR estimation is more robust and more efficient when the error term has a heavy-tailed distribution. Furthermore, a weighted type local polynomial CQR estimator is provided to improve the estimation efficiency by appropriately choosing the weights, and a data-driven bandwidth selection is introduced to choose the optimal bandwidth involved in the nonparametric estimators. So far as we know, such two issues have

not been well addressed by the existing literature on nonstationary time series. Finally, we also give both the simulation study and empirical application to explore the finite sample performance of the proposed methodology and theory.

The rest of the paper is organized as follows. In Section 2, we introduce some basic theory for Harris recurrent Markov process and the methodology of local polynomial CQR smoothing. In Section 3, we establish the asymptotic theory for the proposed estimators. In Section 4, we discuss some important issues such as the weighted local polynomial CQR method and bandwidth selection. In Section 5, we give both the simulation study and empirical application of our methodology. Section 6 concludes this paper. The proofs of the main results are given in the Appendix.

2. Methodology

In this section, we first introduce some basic definitions and properties for the Harris recurrent Markov processes which generate the explanatory variables $\{X_t\}$, and then give the local q-order polynomial CQR estimators of the mean regression function $m(\cdot)$ as well as its derivatives.

2.1. Harris recurrent Markov processes

Throughout this paper, let $\{X_t, t \geq 1\}$ be a ϕ-irreducible Markov chain on the state space $(\mathbb{E}, \mathcal{E})$ with transition probability P, which means that for any set $A \in \mathcal{E}$ with $\phi(A) > 0$, we have $\sum_{t=1}^{\infty} \mathsf{P}^t(x, A) > 0$ for $x \in \mathbb{E}$. We further assume that the ϕ-irreducible Markov chain $\{X_t\}$ is Harris recurrent to make asymptotic analysis for the local polynomial CQR estimation possible. The notation used in this subsection is similar to that used by the existing literature such as Nummelin (1984) and Karlsen and Tjøstheim (2001).

A Markov process $\{X_t\}$ is Harris recurrent if, given a neighborhood $\mathbb{B}_v$ of v with $\phi(\mathbb{B}_v) > 0$, $\{X_t\}$ returns to $\mathbb{B}_v$ with probability one, $v \in \mathbb{E}$. An important feature of the Harris recurrence is to allow us to construct a split chain, which plays a critical role in the derivation of the asymptotic theory (c.f., Karlsen and Tjøstheim, 2001; Karlsen et al., 2007; Li et al., forthcoming). With the help of the split chain technique, we can decompose the partial sum of functions of $\{X_t\}$ into blocks of i.i.d. parts and two asymptotically negligible remaining parts. For the process $\{\mathcal{G}(X_t) : t \geq 1\}$, we define

$$
Z_k(\mathcal{G}) = \begin{cases} \sum_{t=1}^{\tau_0^*} \mathcal{G}(X_t), & k = 0, \\ \sum_{t=\tau_{k-1}^*+1}^{\tau_k^*} \mathcal{G}(X_t), & 1 \leq k \leq T(n), \\ \sum_{t=\tau_{T(n)}^*+1}^{n} \mathcal{G}(X_t), & k = T(n)+1, \end{cases}
$$

where $\mathcal{G}(\cdot)$ is a real function defined on $\mathbb{R}$, $\tau_k^*, k \geq 1$, are the regeneration times of the Markov process, n is the number of observations and $T(n)$ is the number of regenerations of the split chain. Usually we need to assume that the function $\mathcal{G}(\cdot)$ satisfies some integrability conditions when applying the technique of the split chain to derive asymptotic theory. For example, when $\{X_t\}$ is positive recurrent, certain moment condition needs to be imposed on $\mathcal{G}(X_t)$; and when $\{X_t\}$ is the random walk process, $\mathcal{G}(\cdot)$ is usually assumed to be an integrable function. It is easy to see that

$$
\sum_{t=1}^{n} \mathcal{G}(X_t) = Z_0(\mathcal{G}) + \sum_{k=1}^{T(n)} Z_k(\mathcal{G}) + Z_{T(n)+1}(\mathcal{G}). \tag{2.1}
$$

From the result in Nummelin (1984), $\{Z_k(\mathcal{G}), k \geq 1\}$ is a sequence of i.i.d. random variables, and the remaining two terms $Z_0(\mathcal{G})$ and $Z_{T(n+1)}(\mathcal{G})$ on the right hand side of (2.1) converge to zero almost surely when they are divided by the number of regenerations $T(n)$ by using Lemma 3.2 in Karlsen and Tjøstheim (2001).

The general Harris recurrence only yields stochastic rates of convergence in asymptotic theory, where the distribution and size of the number of regenerations $T(n)$ have no a priori known structure but fully depend on the underlying process $\{X_t\}$. For the positive Harris recurrent case, the order of $T(n)$ is the same as that of the sample size n. However, to obtain a specific rate of $T(n)$ for the null recurrent case, we need to impose some restrictions on the tail behavior of the distribution of the recurrence times of the Markov chain $\{X_t\}$.

A Markov process $\{X_t\}$ 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

$$E_\lambda \left(\sum_{i=1}^n f(X_t) \right) \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_f(n), \tag{2.2}$$

where E_λ stands for the expectation with initial distribution λ and $\Gamma(1+\beta)$ is the Gamma function with parameter $1+\beta$. The definition of a small function f in the above definition can be found in some existing literature (see, for example, pp.15 in Nummelin, 1984). Assuming the β-null recurrence restricts the tail behavior of the recurrence time of the process to be a regularly varying function. If $\{X_t\}$ is generated by the random walk process, $\beta = 1/2$. For the stationary and positive recurrent $\{X_t\}$, $\beta = 1$. Myklebust et al. (2012) and Li et al. (forthcoming) further provided an example for a general β-null recurrent Markov chain, where β could be any value between 0 and 1. The β-null recurrent process also has the so-called invariance property that if $\{X_t\}$ is β-null recurrent, then for a one-to-one transformation $\mathcal{T}(\cdot)$, $\{\mathcal{T}(X_t)\}$ is still β-null recurrent (Teräsvirta et al., 2010). Karlsen and Tjøstheim (2001) proved that the regeneration number $T(n)$ of the β-null recurrent Markov chain $\{X_t\}$ has the following asymptotic distribution

$$\frac{T(n)}{n^\beta L_s(n)} \xrightarrow{d} M_\beta(1), \tag{2.3}$$

where $L_s = L_f/(\pi_s f)$, the function $f(\cdot)$ is π_s-integrable such that $\int |f(x)| \pi_s(dx) < \infty$, $\pi_s(\cdot)$ is an invariant measure of the Markov chain $\{X_t\}$, and $M_\beta(1)$ is the Mittag-Leffler distribution with parameter β (c.f., Kasahara, 1984). As $\{X_t\}$ is assumed to be Harris recurrent throughout the paper, the invariant measure $\pi_s(\cdot)$ exists and is unique up to a constant. Section 3.2 of Karlsen and Tjøstheim (2001) provided a detailed introduction on how to construct and define the invariant measure $\pi_s(\cdot)$. In this paper, we further assume that $\pi_s(\cdot)$ is absolutely continuous with respect to the Lebesgue measure and let $p_s(\cdot)$ be the corresponding density function such that $p_s(x)dx = \pi_s(dx)$. If $\{X_t\}$ is the random walk process, $p_s(x) \equiv 1$; and if $\{X_t\}$ is stationary and positive recurrent, $p_s(\cdot)$ can be seen as a conventional density function. Since $T(n) < n$ a.s. for the null recurrent case by (2.3), it is expected that the rates of convergence for the nonparametric local estimators would be slower than those for the stationary time series case (c.f., Karlsen et al., 2007; Gao et al., 2015).

2.2. Local polynomial CQR estimators

We start with the well-known local polynomial smoothing approach. Assuming that the mean regression function $m(\cdot)$ has the $(q + 1)$th order continuous derivative function, by Taylor's expansion, we can approximate $m(X_t)$ by

$$m(X_t) \approx m(x_0) + \sum_{j=1}^q \frac{m^{(j)}(x_0)}{j!}(X_t - x_0)^j$$

when X_t is in a small neighborhood of x_0, where $m^{(j)}(\cdot)$ is the jth derivative of $m(\cdot)$. As in Fan and Gijbels (1996), define the local loss function by

$$L(b_0, b_1, \ldots, b_q) = \sum_{t=1}^n \left[Y_t - b_0 - \sum_{j=1}^q b_j(X_t - x_0)^j \right]^2 \\ \times K\left(\frac{X_t - x_0}{h} \right), \tag{2.4}$$

where $K(\cdot)$ is a kernel function and h is a bandwidth. Letting $(\widetilde{b}_0, \widetilde{b}_1, \ldots, \widetilde{b}_q)^\top$ be the minimizer to the local loss function $L(b_0, b_1, \ldots, b_q)$, then $m(x_0)$ and $m^{(j)}(x_0)$ can be estimated by $\widetilde{b}_0$ and $j!\widetilde{b}_j$, respectively, $j = 1, \ldots, q$. Such local q-order polynomial estimators can be seen as the extension of the Nadaraya–Watson kernel estimator discussed in Karlsen et al. (2007), when $\{X_t\}$ belongs to a class of Harris recurrent Markov processes. However, as mentioned in Section 1, the local polynomial estimators are not robust and would break down when the error distribution does not have certain moment conditions. To address the latter issue, we next give the local polynomial quantile regression and CQR estimation methods.

Since introduced by Koenker and Bassett (1978), the quantile regression method has been widely used in many disciplines such as economics and finance, and serves as a robust alternative to the mean regression method. Koenker (2005)'s book gives a review on various methodologies in quantile regression as well as their applications. Recently, Xiao (2009) studied the quantile regression for nonstationary time series where $\{X_t\}$ is generated by a unit root process. However, in this paper, we focus on the Harris recurrent Markov process which is quite different from the unit root process, as the latter process may be not Markovian. Furthermore, the technique developed for the unit root process relies on the linear framework, and cannot be directly applied when the underlying process is nonlinear (such as a nonlinear autoregressive Markov process).

For $0 < \tau < 1$, define $\rho_\tau(z) = \tau z - zI(z < 0)$. Let $(\bar{b}_0, \bar{b}_1, \ldots, \bar{b}_q)^\top$ be the minimizer to

$$Q(b_0, b_1, \ldots, b_q) = \sum_{t=1}^n \rho_\tau \left[Y_t - b_0 - \sum_{j=1}^q b_j(X_t - x_0)^j \right] \\ \times K\left(\frac{X_t - x_0}{h} \right). \tag{2.5}$$

Then, $m(x_0) + \kappa_\tau(x_0)$ and $m^{(j)}(x_0)$ can be estimated by $\bar{b}_0$ and $j!\bar{b}_j$, respectively, $j = 1, \ldots, q$, where $\kappa_\tau(x_0)$ is the conditional τth quantile function of $\sigma(X_t)e_t$ given $X_t = x_0$.

The local polynomial CQR estimation method is to combine the information across multiple quantile estimates, and thus improve the estimators of the mean regression function and its derivatives. For a given positive integer M, let $\tau_k = k/(M + 1)$, $k = 1, \ldots, M$. Define the local loss function for local polynomial CQR smoothing by

$$Q_C(B_0, b_1, \ldots, b_q) = \sum_{k=1}^M \sum_{t=1}^n \rho_{\tau_k} \left[Y_t - b_{0k} - \sum_{j=1}^q b_j(X_t - x_0)^j \right] \\ \times K\left(\frac{X_t - x_0}{h} \right), \tag{2.6}$$

where $B_0 = (b_{01}, \ldots, b_{0M})^\top$. Let $(\widehat{B}_0, \widehat{b}_1, \ldots, \widehat{b}_q)^\top$ be the minimizer to the local loss function $Q_C(B_0, b_1, \ldots, b_q)$ with $\widehat{B}_0 = (\widehat{b}_{01}, \ldots, \widehat{b}_{0M})^\top$. Then, the local polynomial CQR estimators of $m(x_0)$ and $m^{(j)}(x_0)$ can be defined by

$$\widehat{m}(x_0) = \frac{1}{M} \sum_{k=1}^M \widehat{b}_{0k}, \qquad \widehat{m}^{(j)}(x_0) = j!\widehat{b}_j, \quad j = 1, \ldots, q. \tag{2.7}$$

A crucial condition to ensure the unbiasedness of $\widehat{m}(x)$ is that $\{e_t\}$ is symmetric, as assumed in Assumption 3 in Section 3. In practice, M is usually chosen as a positive odd integer (say 5 or 7). In such case, the median would be included (τ_k with $k = (M + 1)/2$) and the remaining quantiles would be evenly spaced around the median. For the case of i.i.d. observations, Kai et al. (2010, 2011) have shown that the local polynomial CQR estimators enjoy nice asymptotic efficiency properties compared with the local least squares estimator. In this paper, we will further show that such nice properties hold for both stationary and nonstationary cases through theoretical analysis and numerical studies.

3. Asymptotic theory

We start with some regularity conditions to establish the asymptotic properties of the estimation method introduced in Section 2.

Assumption 1. The kernel function $K(\cdot)$ is a symmetric and continuous probability density function with compact support.

Assumption 2. Let $\pi_s(\cdot)$ be the invariant measure of the Harris recurrent Markov process $\{X_t\}$ with continuous and positive density function $p_s(\cdot)$. If $\{X_t\}$ is null recurrent, we further assume that it is β-null recurrent with $0 < \beta < 1$.

Assumption 3. Let $\{e_t\}$ be a sequence of i.i.d. random variables and independent of $\{X_t\}$. Furthermore, $\{e_t\}$ has a symmetric distribution with a cumulative distribution function $F_e(\cdot)$ and a positive and continuous density function $f_e(\cdot)$.

Assumption 4. The mean regression function $m(\cdot)$ has the $(q + 1)$th order continuous derivatives.

Assumption 5. The function $\sigma^2(\cdot)$ is continuous and positive at point x_0.

Assumption 6. The bandwidth h satisfies that $h \to 0$, and either $\frac{nh}{\log n} \to \infty$ when $\{X_t\}$ is positive recurrent or $\frac{n^\beta L_s(n)h}{\log n} \to \infty$ when $\{X_t\}$ is β-null recurrent.

Assumption 1 is a mild condition on the kernel function when local polynomial smoothing is used, and several commonly-used kernel functions such as the uniform kernel and the Epanechnikov kernel satisfy Assumption 1. Assumption 2 is comparable to some corresponding conditions in stationary case. For example, the condition on the invariant density function $p_s(\cdot)$ is similar to that on the density function in the stationary case. When $\{X_t\}$ is generated by the random walk process $X_t = X_{t-1} + x_t$ where $\{x_t\}$ is i.i.d. and $X_0 = 0$, $\{X_t\}$ corresponds to the 1/2-null recurrent Markov process with $p_s(\cdot) \equiv 1$ and Assumption 2 can be satisfied. Although we assume that $\{e_t\}$ is independent of $\{X_t\}$ in Assumption 3, there exists a heteroscedasticity structure $\sigma(X_t)e_t$ in our model, which relaxes the homoscedasticity assumption in Gao et al. (2015). Furthermore, the i.i.d. condition on $\{e_t\}$ in Assumption 3 can be replaced by the stationary and weakly dependent condition such as a stationary and α-mixing dependence (c.f., Karlsen et al., 2007). As the quantile regression is used in the developed estimation procedure, we do not need to impose any moment condition on the error $\{e_t\}$, which is an advantage over the conventional local linear smoothing method. This makes our methodology applicable to the case of heavy-tailed error distribution. It seems difficult to relax the condition that $\{e_t\}$ is independent of the regressor $\{X_t\}$, which is a standard assumption in the literature (c.f., Karlsen et al., 2007; Gao et al., 2015; Li et al., forthcoming). One possibility is to replace the mutual independence assumption by the condition that

the compound process $\{(X_t, e_t)\}$ is Harris recurrent, in which case the proofs of the asymptotic theorems in the appendix need to be substantially modified. Assumptions 4 and 5 are the same as those in Kai et al. (2010), and the bandwidth condition in Assumption 6 is also commonly used in nonparametric kernel estimators of the Harris recurrent Markov processes.

Before giving the asymptotic theory for $\widehat{m}(x_0)$ and $\widehat{m}^{(j)}(x_0), j = 1, \ldots, q$, we introduce some notation. Let $\mu_i = \int z^i K(z)dz$ and $\nu_i = \int z^i K^2(z)dz$ for $i = 0, 1, \ldots, c_k = F_e^{-1}(\tau_k)$ with $\tau_k = k/(M + 1)$, and

$$S(x_0) = \begin{bmatrix} S_{11}(x_0) & S_{12}(x_0) \\ S_{12}^\top(x_0) & S_{22}(x_0) \end{bmatrix}, \tag{3.1}$$

where $S_{11}(x_0)$ is an $M \times M$ diagonal matrix whose diagonal element is $s_{11}(k) = f_e(c_k)p_s(x_0)\mu_0/\sigma(x_0)$ for $k = 1, \ldots, M$, $S_{12}(x_0)$ is an $M \times q$ matrix whose (k, j)th element is $s_{12}(k, j) = f_e(c_k)p_s(x_0)\mu_j/\sigma(x_0)$ for $k = 1, \ldots, M$ and $j = 1, \ldots, q$, $S_{22}(x_0)$ is a $q \times q$ matrix whose (j_1, j_2)th element is $s_{22}(j_1, j_2) = \left[\sum_{k=1}^{M} f_e(c_k)\right]p_s(x_0)\mu_{j_1+j_2}/\sigma(x_0)$ for $j_1, j_2 = 1, \ldots, q$. Let

$$\Omega(x_0) = \begin{bmatrix} \Omega_{11}(x_0) & \Omega_{12}(x_0) \\ \Omega_{12}^\top(x_0) & \Omega_{22}(x_0) \end{bmatrix}, \tag{3.2}$$

where $\Omega_{11}(x_0)$ is an $M \times M$ matrix whose (k_1, k_2)th element is $\omega_{11}(k_1, k_2) = p_s(x_0)\nu_0(\tau_{k_1} \wedge \tau_{k_2} - \tau_{k_1}\tau_{k_2})$, $\Omega_{12}(x_0)$ is an $M \times q$ matrix whose (k, j)th element is $p_s(x_0)\nu_j \sum_{k_1=1}^{M}(\tau_{k_1} \wedge \tau_k - \tau_{k_1}\tau_k)$, and $\Omega_{22}(x_0)$ is a $q \times q$ matrix whose (j_1, j_2)th element is $p_s(x_0)\nu_{j_1+j_2} \sum_{k_1=1}^{M} \sum_{k_2=1}^{M}(\tau_{k_1} \wedge \tau_{k_2} - \tau_{k_1}\tau_{k_2})$. Let $e_{M+q,k}$ be an $(M + q)$-dimensional column vector with the kth element being 1 and 0 elsewhere, and

$$E_M = \left[e_{M+q,1}, \ldots, e_{M+q,M}\right]^\top.$$

We give the asymptotic distribution theory for the local polynomial CQR estimators $\widehat{m}(x_0)$ and $\widehat{m}^{(j)}(x_0), j = 1, \ldots, q$, in the following theorem.

Theorem 3.1. *Suppose that $\{X_t\}$ is a Harris recurrent Markov chain, Assumptions 1–6 are satisfied and the matrix $S(x_0)$ is non-singular.*

(i) *The local polynomial CQR estimator of $m(x_0)$ has the following asymptotic distribution:*

$$\sqrt{T(n)h}\left\{\widehat{m}(x_0) - \mathsf{E}[\widehat{m}(x_0)]\right\} \xrightarrow{d} \mathsf{N}\left[0, \sigma_0^2(x_0)\right], \tag{3.3}$$

where

$$\sigma_0^2(x_0) = \frac{1}{M^2} \sum_{k_1=1}^{M} \sum_{k_2=1}^{M} \sigma_{k_1 k_2}(x_0),$$

$\sigma_{k_1 k_2}(x_0)$ *is the (k_1, k_2)th element of the $M \times M$ matrix $E_M S^{-1}(x_0)\Omega(x_0)S^{-1}(x_0)E_M^\top$.*

(ii) *The local polynomial CQR estimator of $m^{(j)}(x_0)$ has the following asymptotic distribution:*

$$\sqrt{T(n)h^{2j+1}}\left\{\widehat{m}^{(j)}(x_0) - \mathsf{E}[\widehat{m}^{(j)}(x_0)]\right\} \xrightarrow{d} \mathsf{N}\left[0, \sigma_j^2(x_0)\right], \tag{3.4}$$

where $\sigma_j^2(x_0) = e_{M+q,M+j}^\top S^{-1}(x_0)\Omega(x_0)S^{-1}(x_0)e_{M+q,M+j}$.

From the above theorem, we find that the asymptotic variances for $\widehat{m}(x_0)$ and $h^j \widehat{m}^{(j)}(x_0)$ have the stochastic convergence rates as the number of regeneration, $T(n)$, is random, which is different from the corresponding results in existing literature and provides a more general framework. For example, when $\{X_t\}$ is null recurrent, the above asymptotic normal distribution results (3.3) and (3.4) can be seen as the nonstationary extension of Theorem 4 in Kai

et al. (2010), and our results can also be seen as the robust generalization of the corresponding results in Karlsen et al. (2007) and Gao et al. (2015) which considered the kernel and local linear estimation of the regression function.

When $q = 1$, the local polynomial CQR estimators reduce to the local linear CQR estimators. Hence, from Theorem 3.1, we have the following asymptotic theory.

Corollary 3.1. *Suppose that the conditions of Theorem 3.1 are satisfied with $\mu_0 = 1$ and let $q = 1$. The local linear CQR estimator of $m(x_0)$ has the following asymptotic distribution:*

$$\sqrt{T(n)h}\left\{\widehat{m}(x_0) - m(x_0) - \frac{1}{2}m^{(2)}(x_0)\mu_2 h^2\right\}$$

$$\xrightarrow{d} N\left[0, \sigma_*^2(x_0)/p_s(x_0)\right], \tag{3.5}$$

and

$$\sqrt{\sum_{t=1}^n K\left(\frac{X_t - x_0}{h}\right)}\left\{\widehat{m}(x_0) - m(x_0) - \frac{1}{2}m^{(2)}(x_0)\mu_2 h^2\right\}$$

$$\xrightarrow{d} N\left[0, \sigma_*^2(x_0)\right], \tag{3.6}$$

where

$$\sigma_*^2(x_0) = \frac{\nu_0\sigma^2(x_0)}{M^2}\sum_{k_1=1}^M\sum_{k_2=1}^M\frac{\tau_{k_1}\wedge\tau_{k_2} - \tau_{k_1}\tau_{k_2}}{f_e(c_{k_1})f_e(c_{k_2})}.$$

From Theorem 3.1 and Corollary 3.1, it is clear that the asymptotic variance depends on the choice of τ_ks. In general, for a single given τ, inner quantiles (i.e., τ close to 0.5) may have similar information with median and the estimation might be more accurate, while the estimation of tail quantile (i.e., τ close to 0 or 1) is harder due to less information available. In practice, it may be worth trying using more tail quantiles (1%, 5%, 95% or 99%) or more inner quantiles (25%, 40%, 60% or 75%) to examine the change of the resulting estimates. Following the literature (Zou and Yuan, 2008; Kai et al., 2010, 2011), we choose $\tau_k = k/(M+1)$ for a given positive odd integer M throughout this paper.

From the above asymptotic distribution theory, the asymptotic bias term for the local linear CQR estimator is exactly the same as that derived in Theorem 1 of Kai et al. (2010), and has the same order for both the stationary and nonstationary cases. As the stochastic order $T(n)$ is determined by the Harris recurrent process $\{X_t\}$, the asymptotic distributions in (3.3)–(3.5) belong to the self-normalized central limit theorems. Note that

$$\frac{1}{T(n)h}\sum_{t=1}^n K\left(\frac{X_t - x_0}{h}\right) = p_s(x_0) + o_P(1),$$

by Theorem 5.1 of Karlsen and Tjøstheim (2001) if $\mu_0 = 1$. Hence, the asymptotic distribution result (3.6) in Corollary 3.1 can be easily proved by using (3.5) and the asymptotic variance $\sigma_*^2(x_0)$ in (3.6) does not rely on $p_s(x_0)$.

From (3.5) in Corollary 3.1, we can also calculate the conditional asymptotic mean squared error (AMSE) for the local linear CQR estimator given $\mathcal{F}_n(X) = \sigma(X_1, \ldots, X_n)$:

$$\text{AMSE}_1[x_0, h | \mathcal{F}_n(X)] \overset{P}{\sim} \left[\frac{1}{2}m^{(2)}(x_0)\mu_2\right]^2 h^4 + \frac{\sigma_*^2(x_0)}{T(n)hp_s(x_0)}$$

$$=: \frac{1}{4}b_m^2(x_0)h^4 + \frac{\sigma_*^2(x_0)}{T(n)hp_s(x_0)}, \tag{3.7}$$

where $a_n \overset{P}{\sim} b_n$ denotes $\frac{a_n}{b_n} = 1 + o_P(1)$ and $b_m(x_0) = m^{(2)}(x_0)\mu_2$. The proof of (3.7) will be given in the Appendix. It is easy to see

that the optimal bandwidth for the local linear CQR estimator is

$$h_{opt1}(x_0) \overset{P}{\sim} \left[\frac{\sigma_*^2(x_0)}{b_m^2(x_0)p_s(x_0)}\right]^{1/5} T^{-1/5}(n) \tag{3.8}$$

and

$$\text{AMSE}_1\left[x_0, h_{opt1}(x_0)|\mathcal{F}_n(X)\right]$$

$$\overset{P}{\sim} \frac{5}{4}\left[\frac{\sigma_*^2(x_0)b_m^{1/2}(x_0)}{p_s(x_0)}\right]^{4/5} T^{-4/5}(n). \tag{3.9}$$

Similarly, for the local linear estimator, by Corollary 1 in Lin et al. (2009), we can show that the conditional AMSE for the local linear estimator given $\mathcal{F}_n(X)$ is

$$\text{AMSE}_2\left[x_0, h|\mathcal{F}_n(X)\right] \overset{P}{\sim} \frac{1}{4}b_m^2(x_0)h^4 + \frac{\sigma_\diamond^2(x_0)}{T(n)hp_s(x_0)},$$

$$\sigma_\diamond^2(x_0) = \nu_0\sigma^2(x_0). \tag{3.10}$$

Hence, the optimal bandwidth for the local linear estimator is

$$h_{opt2}(x_0) \overset{P}{\sim} \left[\frac{\sigma_\diamond^2(x_0)}{b_m^2(x_0)p_s(x_0)}\right]^{1/5} T^{-1/5}(n) \tag{3.11}$$

and

$$\text{AMSE}_2\left[x_0, h_{opt2}(x_0)|\mathcal{F}_n(X)\right]$$

$$\overset{P}{\sim} \frac{5}{4}\left[\frac{\sigma_\diamond^2(x_0)b_m^{1/2}(x_0)}{p_s(x_0)}\right]^{4/5} T^{-4/5}(n). \tag{3.12}$$

By (3.9) and (3.12), the asymptotic relative efficiency of the local linear CQR estimator with respect to the local linear estimator is

$$\text{ARE}\left[x_0|\mathcal{F}_n(X)\right] = \frac{\text{AMSE}_2\left[x_0, h_{opt2}(x_0)|\mathcal{F}_n(X)\right]}{\text{AMSE}_1\left[x_0, h_{opt1}(x_0)|\mathcal{F}_n(X)\right]} \overset{P}{\sim} \left[\frac{\sigma_\diamond^2(x_0)}{\sigma_*^2(x_0)}\right]^{4/5}$$

$$= R^{-4/5}(M), \tag{3.13}$$

where

$$R(M) = \frac{1}{M^2}\sum_{k_1=1}^M\sum_{k_2=1}^M\frac{\tau_{k_1}\wedge\tau_{k_2} - \tau_{k_1}\tau_{k_2}}{f_e(c_{k_1})f_e(c_{k_2})}.$$

Table 1 in Kai et al. (2010) calculated the values of $R^{-4/5}(M)$ corresponding to different values of M for various cases of error distribution, and it showed that for most of the heavy-tailed error distributions including the t-distribution and mixture of normal distributions, the local linear CQR estimation with $M = 5$ is significantly better than that with $M = 1$ (which reduces to the local linear absolute deviation estimation). In practical applications, the local linear CQR estimation using relatively *small* M is often suggested. Furthermore, Theorem 2 in Kai et al. (2010) proved that $R(M) \to 1$ as $M \to \infty$, which indicates that the local linear CQR estimation would not lose efficiency even if a very large M is used.

Note that the order of $T(n)$ for the stationary case is different from that in the nonstationary case, which would lead to different convergence rates for the local polynomial CQR estimators. Hence, we next give the convergence rates of the local linear CQR estimator $\widehat{m}(x_0)$ under two different scenarios: positive recurrence and β-null recurrence, which correspond to the stationary and nonstationary cases, respectively.

Corollary 3.2. *Suppose that the conditions in Theorem 3.1 are satisfied and let $q = 1$.*

(i) *When $\{X_t\}$ is positive recurrent, we have*

$$\widehat{m}(x_0) - m(x_0) = O_P\left(h^2 + \frac{1}{\sqrt{nh}}\right). \tag{3.14}$$

(ii) *When $\{X_t\}$ is β-null recurrent with $0 < \beta < 1$, we have*

$$\widehat{m}(x_0) - m(x_0) = O_P\left(h^2 + \frac{1}{\sqrt{n^\beta L_s(n)h}}\right). \tag{3.15}$$

As $0 < \beta < 1$ and $L_s(n)$ is a positive and slowly varying function, from the above corollary, we can see that the convergence rate of $\widehat{m}(x_0)$ for the nonstationary case is slower than that for the stationary case when the bandwidth h is appropriately chosen. This is mainly because in the β-null recurrent case, the amount of time spent by the time series around any particular point is of order $n^\beta L_s(n)$ rather than n for the stationary time series case.

As $T(n)$ is random, the asymptotic normal distribution results in Theorem 3.1 and Corollary 3.1 are the self-normalized central limit theorems. By using the convergence result in (2.3), we may replace the stochastic normalization rate in (3.5) by a deterministic normalization rate and obtain the following corollary.

Corollary 3.3. *Suppose that the conditions of Corollary 3.1 are satisfied. The local linear CQR estimator of $m(x_0)$ has the following asymptotic distribution:*

$$\sqrt{n^\beta L_s(n)h}\left\{\widehat{m}(x_0) - m(x_0) - \frac{1}{2}m^{(2)}(x_0)\mu_2 h^2\right\}$$

$$\xrightarrow{d} \left[\frac{\sigma_*^2(x_0)}{p_s(x_0)M_\beta(1)}\right]^{1/2} \mathsf{N}, \tag{3.16}$$

where $\sigma_*^2(x_0)$ is defined as in Corollary 3.1 and N is a standard normal distribution independent of the Mittag-Leffler distribution $M_\beta(1)$.

The above result can be proved by combining (2.3) in Section 2.1 and (3.5) in Corollary 3.1 and using the condition that $\{e_t\}$ is independent of $\{X_t\}$. Noting that the Mittag-Leffler distribution can be regarded as a scaled local time of the underlying Markov process, the result (3.16) is comparable to the asymptotic result (3.11) in Wang and Phillips (2009b) who used the local time technique.

4. Weighted local CQR smoothing and bandwidth selection

In this section, we further introduce a weighted local linear CQR smoothing technique which could be more efficient than the ordinary local linear CQR estimation studied in the previous sections. Meanwhile, a local bandwidth selection criterion is also provided to choose the optimal bandwidth in the local CQR estimation. In this section, we mainly focus on the local linear CQR smoothing estimation as the extension to the local polynomial CQR smoothing case is straightforward.

4.1. Weighted local linear CQR smoothing

Note that in (2.7), $\widehat{m}(x_0)$ is constructed by taking a simple average of $\widehat{b}_{0k}$, $k = 1, \ldots, M$, obtained by minimizing (2.6). It is well known that the efficiency of the estimator might be improved by considering a weighted average of $\widehat{b}_{0k}$ and choosing the optimal weights (c.f., Bradic et al., 2011). Motivated by this, we introduce the following weighted local linear CQR estimator:

$$\widehat{m}_w(x_0) = \sum_{k=1}^{M} w_k \widehat{b}_{0k} \quad \text{with} \quad \sum_{k=1}^{M} w_k = 1. \tag{4.1}$$

It is easy to see that $\widehat{m}_w(x_0)$ reduces to $\widehat{m}(x_0)$ when $w_k = \frac{1}{M}$. Following the proof of Theorem 3.1 in the Appendix, we may prove that

$$\sqrt{T(n)h}\left\{\widehat{m}(x_0) - m(x_0) - \frac{1}{2}m^{(2)}(x_0)\mu_2 h^2\right\}$$

$$\xrightarrow{d} \mathsf{N}\left[0, \sigma_w^2(x_0)\right] \tag{4.2}$$

where

$$\sigma_w^2(x_0) = \frac{v_0 \sigma^2(x_0)}{p_s(x_0)} \sum_{k_1=1}^{M}\sum_{k_2=1}^{M} w_{k_1} w_{k_2} \frac{\tau_{k_1} \wedge \tau_{k_2} - \tau_{k_1}\tau_{k_2}}{f_e(c_{k_1})f_e(c_{k_2})}.$$

In order to improve the estimation efficiency, we have to carefully choose the optimal $w = (w_1, \ldots, w_M)^\top$. The aim is to minimize the quantity

$$\sum_{k_1=1}^{M}\sum_{k_2=1}^{M} w_{k_1} w_{k_2} \frac{\tau_{k_1} \wedge \tau_{k_2} - \tau_{k_1}\tau_{k_2}}{f_e(c_{k_1})f_e(c_{k_2})}$$

subject to $\sum_{k=1}^{M} w_k = 1$. Letting Ω_τ be an $M \times M$ matrix whose (k_1, k_2)th element is $\frac{\tau_{k_1} \wedge \tau_{k_2} - \tau_{k_1}\tau_{k_2}}{f_e(c_{k_1})f_e(c_{k_2})}$, we have

$$\sum_{k_1=1}^{M}\sum_{k_2=1}^{M} w_{k_1} w_{k_2} \frac{\tau_{k_1} \wedge \tau_{k_2} - \tau_{k_1}\tau_{k_2}}{f_e(c_{k_1})f_e(c_{k_2})} = w^\top \Omega_\tau w. \tag{4.3}$$

Then, the optimal weights can be obtained through

$$w_{opt} = \arg\min_{w}\left\{w^\top \Omega_\tau w \,\Big|\, \sum_{k=1}^{M} w_k = 1\right\}. \tag{4.4}$$

By the Lagrange multiplier calculation as in Kim et al. (2012), we can derive the explicit form of w_{opt}:

$$w_{opt} = \frac{\Omega_\tau^{-1}\mathbf{1}_M}{\mathbf{1}_M^\top \Omega_\tau^{-1}\mathbf{1}_M}, \tag{4.5}$$

where $\mathbf{1}_M$ is an M-dimensional column vector with each element being 1. Replacing w by w_{opt}, we get the optimal weighted local linear CQR estimator of $m(x_0)$ by

$$\widehat{m}_o(x_0) = (\widehat{b}_{01}, \ldots, \widehat{b}_{0M}) \cdot w_{opt} \tag{4.6}$$

whose asymptotic variance given $\mathcal{F}_n(X)$ is

$$\frac{1}{T(n)h} \frac{v_0 \sigma^2(x_0)}{p_s(x_0)\mathbf{1}_M^\top \Omega_\tau^{-1}\mathbf{1}_M}.$$

In practice, we may obtain the feasible optimal weights by replacing Ω_τ in (4.5) by $\widehat{\Omega}_\tau$, where $\widehat{\Omega}_\tau$ is an $M \times M$ matrix whose (k_1, k_2)th element is

$$\frac{\tau_{k_1} \wedge \tau_{k_2} - \tau_{k_1}\tau_{k_2}}{\widehat{f}_e(c_{k_1})\widehat{f}_e(c_{k_2})}, \qquad \widehat{f}_e(c_0) = \frac{1}{nb}\sum_{t=1}^{n} L\left(\frac{\widehat{e}_t - c_0}{b}\right),$$

$L(\cdot)$ is a kernel function, b is a bandwidth, $\widehat{e}_t = \left[Y_t - \widehat{m}(X_t)\right]/\widehat{\sigma}(X_t)$, $\widehat{\sigma}^2(\cdot)$ is the nonparametric estimation of the conditional variance function (c.f., Fan and Yao, 1998).

4.2. Bandwidth selection

It is well known that nonparametric kernel estimators are sensitive to the choice of the bandwidth. Hence, it is important to choose the optimal data-driven bandwidth for the proposed local linear CQR estimation. A larger bandwidth would lead to an oversmoothed estimation which incurs larger bias, and a smaller bandwidth could reduce the bias but the variance of the proposed

estimator could be larger since fewer data are used in the kernel estimation. Hence, a good choice of the bandwidth should be a tradeoff between these two types of drawback. Many useful methods have been introduced to select the optimal bandwidth in nonparametric kernel estimation for stationary time series data, see, for example, Fan and Gijbels (1995), Ruppert (1997), and Yao and Tong (1998). However, there is relatively fewer literature on the bandwidth selection for nonparametric kernel estimation with possible nonstationary regressors.

It is well known that the invariant measure of the null recurrent $\{X_t\}$ does not have a compact support, which is commonly assumed in stationary case. For example, when $X_t = X_{t-1} + x_t$ with $X_0 \equiv 0$ and $\{x_t\}$ being i.i.d. standard normal, X_t follows $N(0, t)$. Thus, for this $1/2$-null recurrent Markov process, the range of X_t would be expanding as t increases. It may be not reasonable to use a universal bandwidth to handle X_t with such wide range when the unknown regression function has a complicated structure. Hence, we next propose a variable bandwidth selection method for the local linear CQR estimator.

Our bandwidth selection is based on the idea of the "leave-one-out" cross validation, which is quite useful in assessing the performance of estimators via estimating their prediction errors. Define

$$CV_{x_0}(h) = \sum_{t=1}^{n}\left[Y_t - \widehat{m}_{h,-t}(X_t)\right]^2 H\left(\frac{X_t - x_0}{b_*}\right), \qquad (4.7)$$

where $\widehat{m}_{h,-t}(X_t)$ is the local linear CQR estimated value of $m(X_t)$ with bandwidth h, by using the observations $(Y_s, X_s), s \neq t, H(\cdot)$ is a density function with a compact support and $b_* \to 0$ is pre-determined which indicates that we only assess the estimation performance for X_t in a very small neighborhood of x_0. Then, the local optimal bandwidth for the local linear CQR estimator is obtained by $\bar{h}(x_0) = \arg\min_h CV_{x_0}(h)$.

5. Numerical study

In this section, we give both the simulated and empirical examples to illustrate the proposed methodology and theory.

Example 5.1. Consider a nonlinear regression model defined by

$$Y_t = m(X_t) + e_t, \qquad m(x) = \sin(x) + 2\exp\{-16x^2\},$$
$$t = 1, \ldots, n, \qquad (5.1)$$

which was also studied by Kai et al. (2010) for the i.i.d. observations. Here, the regressor $\{X_t\}$ is generated by one of the following three Markov processes: (i) the AR(1) process: $X_t = 0.5X_{t-1} + x_t$, (ii) the random walk process: $X_t = X_{t-1} + x_t$, and (iii) the threshold AR(1) process: $X_t = 0.5X_{t-1}I(|X_{t-1}| \leq 1) + X_{t-1}I(|X_{t-1}| > 1) + x_t$. In the above three Markov processes, we let $X_0 = 0$ for simplicity and $\{x_t\}$ be a sequence of i.i.d. standard normal random variables. It is easy to see that $\{X_t\}$ is positive recurrent when generated by the AR(1) process (i), and $1/2$-null recurrent when generated by the random walk process (ii) or the threshold AR(1) process (iii) (Gao et al., 2013). Let $\{e_t\}$ be independent of $\{x_t\}$ and be taken from one of the following distributions: (i) standard normal distribution, (ii) Laplace distribution, (iii) t distribution with degree 3, and (iv) a mixture of two normal distributions $0.95N(0, 1) + 0.05N(0, 10^2)$.

In this simulation, we consider the sample size $n = 200$ with 200 replications. In the simulation, we compare the performance of three nonparametric methods to estimate the regression function $m(\cdot)$: local linear method, local linear CQR estimator and weighted local linear CQR estimator with the weights chosen as in Section 4.1. The number of quantiles in the CQR estimation method,

M, was chosen as 5 and 7, and the optimal bandwidths for the local linear and the local linear CQR estimations were selected by using the cross-validation method introduced in Section 4.2. For simplicity, we used the same bandwidth for both the local linear CQR estimator and the weighted local linear CQR estimator. The measure for the performance of the estimators was taken to be the mean squared error (MSE) of the form:

$$MSE(\overline{m}) = \frac{1}{20}\sum_{l=1}^{20}\left[\overline{m}(z_l) - m(z_l)\right]^2, \qquad (5.2)$$

where $\overline{m}(\cdot)$ is the estimated value of the true regression function $m(\cdot)$ by using one of the three nonparametric methods mentioned above, and $z_l, l = 1, \ldots, 20$, are twenty equally-spaced grid points taken between 0.1 and 0.9 quantiles of X_t in each replication. To compare the performances of the three different estimation methods, we calculated the ratio of the MSE for the local linear CQR estimates to that for the local linear estimates (denoted by R(CQR)), and also the ratio of the MSE for the weighted local linear CQR estimates to that of the local linear estimates (denoted by R(WCQR)).

In Table 1, we give the mean values of both R(CQR) and R(WCQR) over 200 replications for different cases as well as the associated standard errors (SE) when M is 5. The results in Table 1 indicate that the local linear CQR and weighted local linear CQR methods outperform the local linear method, except for the standard normal error case where the values of R(CQR) and R(WCQR) are slightly larger than 1. In particular, the improvement by using the local linear CQR and weighted local linear CQR methods is very significant when the error terms follow a mixture normal distribution. Such finding is consistent with our asymptotic analysis. Furthermore, the values of R(WCQR) indicate that the weighted local linear CQR estimation method generally outperforms the other two local estimation methods. Table 2 reports the results for the case when M is 7, and the same finding as above can be obtained.

Example 5.2. We next apply the proposed local linear CQR and weighted local linear CQR estimation methodology to study the relationship between the UK to US export data and the real exchange rates (US/UK). The data (available at https://www.uktradeinfo.com/) were collected monthly from January 1996 to August 2013, and thus the sample size is $n = 212$. The same data set was also studied by Li et al. (forthcoming) by using the parametric nonlinear modeling framework and least squares estimation method. We next consider more flexible nonparametric nonlinear modeling, and estimate it by using the three nonparametric methods as in Example 5.1.

Let $\{Y_t\}$ denote the logarithm of the export value (in million pounds) and let X_t be defined by

$$X_t = \log(E_t) + \log(p_{t,UK}) - \log(p_{t,US}),$$

where E_t is the monthly average of the nominal exchange rate (dollars per pound), and $p_{t,UK}$ and $p_{t,US}$ are the consumption price indices of UK and US, respectively. We plot the data $\{Y_t\}$ and $\{X_t\}$ in Fig. 5.1. The empirical applications in both Gao et al. (2013) and Li et al. (forthcoming) suggested that the real exchange rates $\{X_t\}$ follow the nonstationary threshold AR (1) process and belong to a $1/2$-null recurrent Markov process.

We fitted the data with model (1.1) and estimated the regression function $m(\cdot)$ using the three nonparametric approaches. The choice of the kernel function and selection of the bandwidth are the same as those in Example 5.1. The estimated plots of the regression function are provided in Fig. 5.2, where the dotted line is the local linear estimated curve, the solid line is the local linear CQR estimated curve (with $M = 5$) and the dashed line is the weighted local

Table 1
Ratios of MSEs for the simulated example with $M = 5$.

Error distribution	R(CQR)	SE(CQR)	R(WCQR)	SE(WCQR)
The regressor X_t is generated in case (i)				
N(0, 1)	1.1608	0.3217	1.1536	0.3172
Laplace	0.9062	0.3173	0.8708	0.4110
$t(3)$	0.7910	0.3058	0.7734	0.3301
$0.95N(0, 1) + 0.05N(0, 10^2)$	0.5196	0.3005	0.5178	0.2987
The regressor X_t is generated in case (ii)				
N(0, 1)	1.1200	0.4132	1.1190	0.4124
Laplace	0.9082	0.2609	0.8986	0.3834
$t(3)$	0.8863	0.4115	0.8547	0.4033
$0.95N(0, 1) + 0.05N(0, 10^2)$	0.5661	0.3061	0.5632	0.2925
The regressor X_t is generated in case (iii)				
N(0, 1)	1.1076	0.3210	1.1048	0.3136
Laplace	0.9325	0.2468	0.9217	0.3765
$t(3)$	0.9172	0.4156	0.9030	0.4392
$0.95N(0, 1) + 0.05N(0, 10^2)$	0.5929	0.2862	0.5925	0.2875

Table 2
Ratios of MSEs for the simulated example with $M = 7$.

Error distribution	R(CQR)	SE(CQR)	R(WCQR)	SE(WCQR)
The regressor X_t is generated in case (i)				
N(0, 1)	1.1045	0.2610	1.0988	0.2589
Laplace	0.9008	0.2816	0.8553	0.3778
$t(3)$	0.8027	0.2759	0.7723	0.3165
$0.95N(0, 1) + 0.05N(0, 10^2)$	0.5308	0.2886	0.5214	0.2854
The regressor X_t is generated in case (ii)				
N(0, 1)	1.0815	0.2994	1.0795	0.2979
Laplace	0.9262	0.1991	0.9042	0.3640
$t(3)$	0.8442	0.2424	0.8125	0.3034
$0.95N(0, 1) + 0.05N(0, 10^2)$	0.5748	0.2693	0.5645	0.2526
The regressor X_t is generated in case (iii)				
N(0, 1)	1.0698	0.2522	1.0670	0.2432
Laplace	0.9194	0.2466	0.9036	0.3654
$t(3)$	0.9154	0.3734	0.9002	0.4373
$0.95N(0, 1) + 0.05N(0, 10^2)$	0.6070	0.2817	0.5996	0.2806

linear CQR estimated curve. From Fig. 5.2, we can see that the three estimated curves have similar pattern with the weighted local linear CQR estimated curve slightly above the other two. To make further comparison, we also calculated the mean squared prediction error (MSPE) defined by

$$\text{MSPE}(\overline{m}) = \frac{1}{n} \sum_{t=1}^{n} [Y_t - \overline{m}(X_t)]^2, \quad n = 212,$$

where $\overline{m}(x)$ is the estimated value of the true regression function $m(x)$ by using one of the three nonparametric methods. The value of MSPE for the local linear method is 0.0859, which is larger than 0.0847, the MSPE value for the local linear CQR method. And the value of MSPE for the weighted local linear CQR method is 0.0801, which is the smallest one. This indicates that the weighted local linear CQR method is the most efficient one among the three nonparametric estimation methods, and this is also consistent with our asymptotic theory and finite sample behavior in simulation.

To further study the robust property of the CQR estimation methodology, we deliberately generated an outlier by tripling the value of Y_t in February 2012 (the value of X_t remains unchanged). The estimated plots of the regression function are provided in Fig. 5.3. From the figure, we can see that the local linear estimation of the regression function is seriously affected by the contaminated value of Y_t in February 2012. However, the local linear CQR and the weighted local linear CQR estimated curves have the pattern similar to those in Fig. 5.2. This provides a further evidence that the local linear CQR estimation or weighted local linear CQR estimation is more robust than the local linear estimation.

6. Conclusion and discussion

In this paper, we have studied the local polynomial CQR smoothing method, a robust alternative to the local polynomial smoothing method, for the nonlinear and nonparametric models under the Harris recurrent Markov chain framework which includes both the stationary and nonstationary cases. Under some mild conditions, we establish the asymptotic theory for the proposed local CQR estimator of the mean regression function, and show that the convergence rate for the estimator in nonstationary case is slower than that in stationary case. Meanwhile, we also introduce a weighted local CQR estimation method which can further improve the estimation efficiency, and provide a data-driven bandwidth selection criterion to choose the optimal bandwidth involved in the nonparametric estimators with possible nonstationary regressors.

The estimation methodology and asymptotic results developed in this paper are limited to the univariate regressor case. A possible future topic is to generalize the theory and method to the case with multivariate regressors. For example, we can consider the stationary VAR processes as the regressors (which belong to the positive recurrent Markov framework under some mild conditions), or a mixture of one nonstationary regressor and other stationary regressors. Myklebust et al. (2012) gave some examples for the latter case, which belong to the 1/2-null recurrent Markov framework. However, when the dimension of the regressors is larger than three, the local CQR estimation may suffer from the
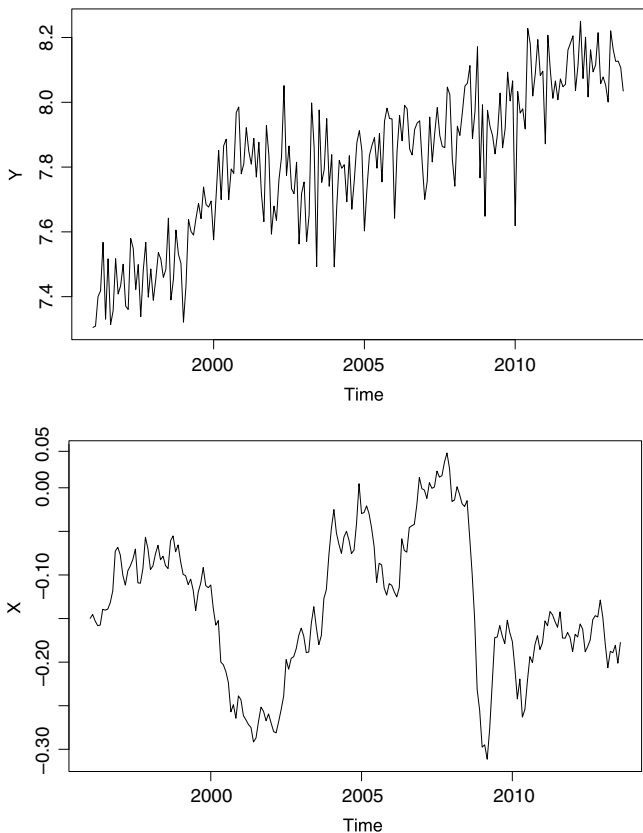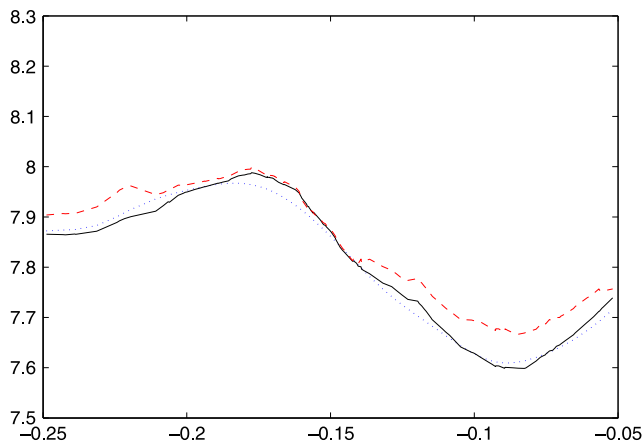
Fig. 5.1. Plots of $\{Y_t\}$ and $\{X_t\}$.

Fig. 5.2. Estimated plots of the regression function for the original series, the dotted line is the local linear estimated curve, the solid line is the local linear CQR estimated curve and the dashed line is the weighted local linear CQR estimated curve.

so-called "curse of dimensionality" issue, which indicates that an alternative modeling framework and estimation approach would be needed. This would be a good topic for future research.
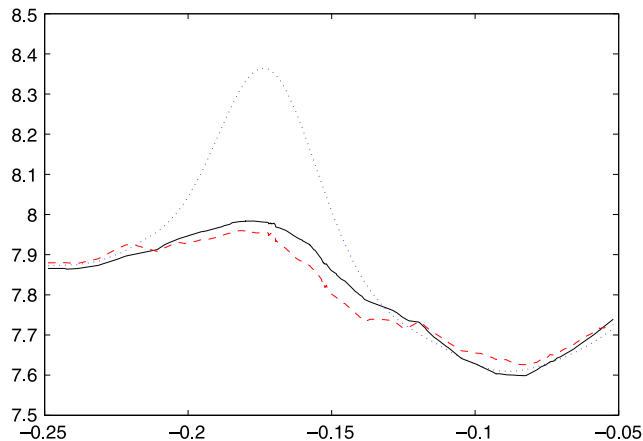
Acknowledgments

Fig. 5.3. Estimated plots of the regression function for the series with an outlier, the dotted line is the local linear estimated curve, the solid line is the local linear CQR estimated curve and the dashed line is the weighted local linear CQR estimated curve.

Appendix. Proofs of the main results

Before proving the main results, we introduce some notation to simplify the presentation. Let

$$u_k = \sqrt{T(n)h}\big[b_{0k} - m(x_0) - \sigma(x_0)c_k\big],$$
$$v_j = h^j\sqrt{T(n)h}\big[b_j - (j!)^{-1}m^{(j)}(x_0)\big]$$

for $k = 1, \ldots, M$ and $j = 1, \ldots, q$, where $c_k = F_e^{-1}(\tau_k)$ and $F_e(\cdot)$ is the cumulative distribution function of the error $\{e_t\}$. Let

$$\Delta_{t,k} = \frac{1}{\sqrt{T(n)h}}\left[u_k + \sum_{j=1}^q v_j\left(\frac{X_t - x_0}{h}\right)^j\right],$$
$$d_{t,k} = c_k\big[\sigma(X_t) - \sigma(x_0)\big] + r_t(x_0),$$
$$r_t(x_0) = m(X_t) - m(x_0) - \sum_{j=1}^q m^{(j)}(x_0)\frac{(X_t - x_0)^j}{j!},$$
$$\delta_{t,k} = \sigma(X_t)(e_t - c_k) + d_{t,k},$$
$$\eta_{t,k} = I\big(e_t \leq c_k - d_{t,k}\sigma^{-1}(X_t)\big) - \tau_k.$$

With the above notations, it is easy to see that

$$Y_t - b_{0k} - \sum_{j=1}^q b_j(X_t - x_0)^j = \sigma(X_t)(e_t - c_k) + d_{t,k} - \Delta_{t,k}. \quad \text{(A.1)}$$

Proof of Theorem 3.1. Letting

$$\overline{Q}_C(x_0) = \sum_{k=1}^M \sum_{t=1}^n \rho_{\tau_k}\big[\sigma(X_t)(e_t - c_k) + d_{t,k}\big]K\left(\frac{X_t - x_0}{h}\right),$$

it is straightforward to show that minimizing $Q_C(B_0, b_1, \ldots, b_q)$ is equivalent to minimizing

$$L_C(B_0, b_1, \ldots, b_d) := Q_C(B_0, b_1, \ldots, b_d) - \overline{Q}_C(x_0),$$
$$= \sum_{k=1}^M \sum_{t=1}^n \left\{\rho_{\tau_k}\big[\sigma(X_t)(e_t - c_k) + d_{t,k} - \Delta_{t,k}\right]$$

$$-\rho_{\tau_k}\Big[\sigma(X_t)(e_t - c_k) + d_{t,k}\Big]\Big\}K\Big(\frac{X_t - x_0}{h}\Big),\tag{A.2}$$

as $B_0, b_1, \ldots, b_d$ are not involved in $\overline{Q}_C(x_0)$.

Note that the identity result:

$$\rho_\tau(x - y) - \rho_\tau(x) = y\Big[I(x \le 0) - \tau\Big]$$
$$+ \int_0^y \Big[I(x \le z) - I(x \le 0)\Big]dz,$$

see, for example, Knight (1998). Letting $x = \sigma(X_t)(e_t - c_k) + d_{t,k} = \delta_{t,k}$ and $y = \Delta_{t,k}$ in the above equality and $K_{tj}(x_0) = \big(\frac{X_t - x_0}{h}\big)^j K\big(\frac{X_t - x_0}{h}\big)$, $L_C(B_0, b_1, \ldots, b_d)$ can be rewritten as

$$L_C(B_0, b_1, \ldots, b_d) = \sum_{k=1}^M u_k \Big[\sum_{t=1}^n \frac{\eta_{t,k}K_{t0}(x_0)}{\sqrt{T(n)h}}\Big]$$
$$+ \sum_{j=1}^q v_j \Big[\sum_{k=1}^M \sum_{t=1}^n \frac{\eta_{t,k}K_{tj}(x_0)}{\sqrt{T(n)h}}\Big]$$
$$+ \sum_{k=1}^M \sum_{t=1}^n K_{t0}(x_0) \int_0^{\Delta_{t,k}} \Big[I\Big(e_t \le c_k + \frac{z - d_{t,k}}{\sigma(X_t)}\Big)$$
$$- I\Big(e_t \le c_k - \frac{d_{t,k}}{\sigma(X_t)}\Big)\Big]dz.$$

Let $\boldsymbol{\theta}_n = (u_1, \ldots, u_M, v_1, \ldots, v_q)^\top$ and $W_n = \big(W_{n1}, \ldots, W_{nM}, W_{n1}^*, \ldots, W_{nq}^*\big)^\top$ with

$$W_{nk} = \frac{1}{\sqrt{T(n)h}} \sum_{t=1}^n \eta_{t,k}K_{t0}(x_0),$$

$$W_{nj}^* = \frac{1}{\sqrt{T(n)h}} \sum_{k=1}^M \sum_{t=1}^n \eta_{t,k}K_{tj}(x_0)$$

for $k = 1, \ldots, M$ and $j = 1, \ldots, q$. Then, we have

$$\sum_{k=1}^M u_k\Big[\sum_{t=1}^n \frac{\eta_{t,k}K_{t0}(x_0)}{\sqrt{T(n)h}}\Big]$$
$$+ \sum_{j=1}^q v_j\Big[\sum_{k=1}^M \sum_{t=1}^n \frac{\eta_{t,k}K_{tj}(x_0)}{\sqrt{T(n)h}}\Big] = W_n^\top \boldsymbol{\theta}_n.\tag{A.3}$$

We next consider $\overline{L}_C(\boldsymbol{\theta}_n)$ which is defined by

$$\overline{L}_C(\boldsymbol{\theta}_n) = \sum_{k=1}^M \sum_{t=1}^n K_{t0}(x_0)$$
$$\times \int_0^{\Delta_{t,k}} \Big[I\Big(e_t \le c_k + \frac{z - d_{t,k}}{\sigma(X_t)}\Big) - I\Big(e_t \le c_k - \frac{d_{t,k}}{\sigma(X_t)}\Big)\Big]dz.$$

Recalling that $\mathcal{F}_n(X) = \sigma\{X_1, \ldots, X_n\}$, it is easy to see $E\big[\overline{L}_C(\boldsymbol{\theta}_n)\big] = E\big\{E\big[\overline{L}_C(\boldsymbol{\theta}_n)|\mathcal{F}_n(X)\big]\big\}$. Note that $F_e(x + z) - F_e(x) = zf_e(x) + o(z)$ when z is sufficiently small, and $f_e(c_k - d_{t,k}/\sigma(X_t)) \to f_e(c_k)$ as $f_e(\cdot)$ is continuous and $d_{t,k} \to 0$ when $|X_t - x_0| = O(h) = o(1)$. By using these two facts and the condition that $\{e_t\}$ is

independent of $\{X_t\}$ in Assumption 3, we can prove that

$$E\big[\overline{L}_C(\boldsymbol{\theta}_n)|\mathcal{F}_n(X)\big] = \sum_{k=1}^M \sum_{t=1}^n K_{t0}(x_0)$$
$$\times \int_0^{\Delta_{t,k}} \Big[F_e\Big(c_k + \frac{z - d_{t,k}}{\sigma(X_t)}\Big) - F_e\Big(c_k - \frac{d_{t,k}}{\sigma(X_t)}\Big)\Big]dz$$
$$= (1 + o_P(1))\sum_{k=1}^M \sum_{t=1}^n K_{t0}(x_0) \int_0^{\Delta_{t,k}} \frac{z}{\sigma(X_t)}f_e\Big(c_k - \frac{d_{t,k}}{\sigma(X_t)}\Big)dz$$
$$= (1 + o_P(1))\sum_{k=1}^M \sum_{t=1}^n K_{t0}(x_0)\Big[\frac{\Delta_{t,k}^2}{2\sigma(X_t)}f_e(c_k)\Big]$$
$$\overset{P}{\sim} \frac{1}{2}\sum_{k=1}^M u_k^2\Big[\frac{f_e(c_k)}{T(n)h}\sum_{t=1}^n \sigma^{-1}(X_t)K_{t0}(x_0)\Big]$$
$$+ \sum_{k=1}^M \sum_{j=1}^q u_k v_j\Big[\frac{f_e(c_k)}{T(n)h}\sum_{t=1}^n \sigma^{-1}(X_t)K_{tj}(x_0)\Big]$$
$$+ \frac{1}{2}\sum_{j_1=1}^q \sum_{j_2=1}^q v_{j_1}v_{j_2}\Big[\sum_{k=1}^M \frac{f_e(c_k)}{T(n)h}\sum_{t=1}^n \sigma^{-1}(X_t)K_{t,j_1+j_2}(x_0)\Big]$$
$$=: \frac{1}{2}\boldsymbol{\theta}_n^\top S_n(x_0)\boldsymbol{\theta}_n,\tag{A.4}$$

where

$$S_n(x_0) = \begin{bmatrix} S_{n11}(x_0) & S_{n12}(x_0) \\ S_{n12}^\top(x_0) & S_{n22}(x_0) \end{bmatrix},$$

$S_{n11}(x_0)$ is an $M \times M$ diagonal matrix whose diagonal element is

$$s_{n11}(k) = \frac{f_e(c_k)}{T(n)h}\sum_{t=1}^n \sigma^{-1}(X_t)K_{t0}(x_0) \quad \text{for } k = 1, \ldots, M,$$

$S_{n12}(x_0)$ is an $M \times q$ matrix whose (k, j)th element is

$$s_{n12}(k, j) = \frac{f_e(c_k)}{T(n)h}\sum_{t=1}^n \sigma^{-1}(X_t)K_{tj}(x_0)$$

for $k = 1, \ldots, M$ and $j = 1, \ldots, q$,

and $S_{n22}(x_0)$ is a $q \times q$ matrix whose (j_1, j_2)th element is

$$s_{n22}(j_1, j_2) = \sum_{k=1}^M \frac{f_e(c_k)}{T(n)h}\sum_{t=1}^n \sigma^{-1}(X_t)K_{t,j_1+j_2}(x_0)$$

for $j_1, j_2 = 1, \ldots, q$.

As $\{X_t\}$ is Harris recurrent, we could use the split chain technique introduced in Section 2.1 to derive the limit for $S_n(x_0)$. Let

$$Z_i(\sigma, K, j) = \begin{cases} \displaystyle\sum_{t=1}^{\tau_0^*} \sigma^{-1}(X_t)K_{tj}(x_0), & k = 0, \\ \displaystyle\sum_{t=\tau_{i-1}^*+1}^{\tau_i^*} \sigma^{-1}(X_t)K_{tj}(x_0), & 1 \le i \le T(n), \\ \displaystyle\sum_{t=\tau_{N(n)}^*+1}^{n} \sigma^{-1}(X_t)K_{tj}(x_0), & i = T(n) + 1. \end{cases}$$

We next only derive the limit for $S_{n11}(x_0)$ as the proofs for $S_{n12}(x_0)$ and $S_{n22}(x_0)$ are analogous. By Lemma 3.2 in Karlsen and Tjøstheim

(2001), we have

$$
s_{n11}(k) = \frac{f_e(c_k)}{T(n)h}\left[Z_0(\sigma, K, 0) + \sum_{i=1}^{T(n)} Z_i(\sigma, K, 0) \right.
$$
$$
\left. + Z_{T(n)+1}(\sigma, K, 0) \right]
$$
$$
\overset{P}{\sim} \frac{f_e(c_k)}{T(n)h} \sum_{i=1}^{T(n)} Z_i(\sigma, K, 0). \tag{A.5}
$$

By Assumptions 1, 2 and 5, Bochner's lemma (c.f., Chapter 9 in Wheeden and Zygmund, 1977) and some standard calculation, we can prove that

$$
\mathsf{E}\left[h^{-1} Z_i(\sigma, K, 0) \right] = p_s(x_0)\mu_0/\sigma(x_0) + O(h^2). \tag{A.6}
$$

On the other hand, as mentioned in Section 2.1, $\{Z_i(\sigma, K, 0)\}$ is a sequence of i.i.d. random variables. Note that $T(n) \sim n$ for the positive recurrent case, and following the proof of Corollary 3.2, we can prove that $C_l n^\beta L_s(n) \le T(n) \le C_u n^\beta L_s(n)$ in probability for the β-null recurrent case with $0 < C_l < C_u < \infty$. By some tedious calculations, we can prove that

$$
s_{n11}(k) = f_e(c_k)p_s(x_0)\mu_0/\sigma(x_0) + o_P(1) = s_{11}(k) + o_P(1). \tag{A.7}
$$

The detailed proof of (A.7) will be given later in this appendix. Similarly, we can also prove that

$$
s_{n12}(k, j) = f_e(c_k)p_s(x_0)\mu_j/\sigma(x_0) + o_P(1) = s_{12}(k, j) + o_P(1) \tag{A.8}
$$

and

$$
s_{n22}(j_1, j_2) = \left[\sum_{k=1}^{M} f_e(c_k) \right] p_s(x_0)\mu_{j_1+j_2}/\sigma(x_0) + o_P(1)
$$
$$
= s_{22}(j_1, j_2) + o_P(1). \tag{A.9}
$$

Letting $S(x_0)$ be defined in (3.1), Eqs. (A.4) and (A.7)–(A.9) lead to

$$
\mathsf{E}\left[\bar{L}_C(\boldsymbol{\theta}_n) \right] = \frac{1}{2} \boldsymbol{\theta}_n^\top S(x_0)\boldsymbol{\theta}_n (1 + o(1)). \tag{A.10}
$$

Meanwhile, noting that $\{e_t\}$ is i.i.d. and independent of $\{X_t\}$ by Assumption 3 and following some standard calculations (c.f., the proof of Lemma 3 in Kai et al., 2010), we have

$$
\mathsf{Var}\left[\bar{L}_C(\boldsymbol{\theta}_n) | \mathcal{F}_n(X) \right] = o_P\left(\sum_{k=1}^{M}\sum_{t=1}^{n} K_{t0}^2(x_0)\Delta_{t,k}^2 \right). \tag{A.11}
$$

Furthermore, similarly to the proof of (A.7), we can prove that

$$
\mathsf{Var}\left[\bar{L}_C(\boldsymbol{\theta}_n) \right] = o(1). \tag{A.12}
$$

Eqs. (A.3), (A.10) and (A.12) imply that

$$
L_C(B_0, b_1, \ldots, b_d) = W_n^\top \boldsymbol{\theta}_n + \frac{1}{2}\boldsymbol{\theta}_n^\top S(x_0)\boldsymbol{\theta}_n + o_P(1). \tag{A.13}
$$

Note that $L_C(B_0, b_1, \ldots, b_d) - W_n^\top \boldsymbol{\theta}_n$ converges in probability to $\frac{1}{2}\boldsymbol{\theta}_n^\top S(x_0)\boldsymbol{\theta}_n$, which is a convex function. Then, by Pollard (1991)'s convexity lemma, the minimizer to $L_C(B_0, b_1, \ldots, b_d)$ satisfies

$$
\widehat{\boldsymbol{\theta}}_n \overset{P}{\sim} S^{-1}(x_0)W_n, \tag{A.14}
$$

where $\widehat{\boldsymbol{\theta}}_n$ is defined as $\boldsymbol{\theta}_n$ with b_{0k} and b_j replaced by $\widehat{b}_{0k}$ and $\widehat{b}_j$, respectively, for $k = 1, \ldots, M$ and $j = 1, \ldots, q$.

Let $W_{n\diamond}$ be defined as W_n with $\eta_{t,k}$ replaced by $\eta_{t,k\diamond} = I(e_t \le c_k) - \tau_k$. Then, we have

$$
W_n - \mathsf{E}[W_n] = W_{n\diamond} - \mathsf{E}[W_{n\diamond}] + W_n - W_{n\diamond}
$$
$$
- \mathsf{E}[W_n - W_{n\diamond}]. \tag{A.15}
$$

Note that

$$
\mathsf{Var}\left(W_n - W_{n\diamond} \right) = O\left(\mathsf{E}\left\{ \frac{1}{T(n)h}\sum_{t=1}^{n}\left[\sum_{j=0}^{q} K_{tj}^2(x_0) \right] \right.\right.
$$
$$
\left.\left. \times \mathsf{E}\left[(\eta_{t,k} - \eta_{t,k\diamond})^2 | \mathcal{F}_n(X) \right] \right\} \right)
$$
$$
= O\left(\mathsf{E}\left\{ \frac{1}{T(n)h}\sum_{t=1}^{n}\left[\sum_{j=0}^{q} K_{tj}^2(x_0) \right] \right.\right.
$$
$$
\left.\left. \times \left[F_e(c_k - d_{t,k}\sigma^{-1}(X_t)) - F_e(c_k) \right] \right\} \right)
$$
$$
= O\left(\mathsf{E}\left\{ \frac{1}{T(n)h}\sum_{t=1}^{n}\left[\sum_{j=0}^{q} K_{tj}^2(x_0) \right] f_e(c_k)\frac{d_{t,k}}{\sigma(X_t)} \right\} \right)
$$
$$
= o\left(\mathsf{E}\left\{ \frac{1}{T(n)h}\sum_{t=1}^{n}\sum_{j=0}^{q} K_{tj}^2(x_0) \right\} \right)
$$

by using the fact that $d_{t,k} = o(1)$ when $|X_t - x_0| = O(h) = o(1)$. Then, similar to the proof of (A.7), we can prove that $\mathsf{Var}(W_n - W_{n\diamond}) = o(1)$, which indicates that the leading term of $W_n - \mathsf{E}[W_n]$ is $W_{n\diamond} - \mathsf{E}[W_{n\diamond}]$. Hence, to prove (3.3) and (3.4), we only need to prove

$$
W_{n\diamond} - \mathsf{E}[W_{n\diamond}] \overset{d}{\longrightarrow} \mathsf{N}\left(\mathbf{0}, \Omega(x_0) \right), \tag{A.16}
$$

where $\Omega(x_0)$ is defined in (3.2). By the so-called Cramér–Wold device (Billingsley, 1968), to prove (A.16), it suffices to prove

$$
\boldsymbol{\alpha}^\top \left\{ W_{n\diamond} - \mathsf{E}[W_{n\diamond}] \right\} \overset{d}{\longrightarrow} \mathsf{N}\left(\mathbf{0}, \boldsymbol{\alpha}^\top \Omega(x_0)\boldsymbol{\alpha} \right), \tag{A.17}
$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{M+q})^\top$. Recall that

$$
W_{n\diamond} = \frac{1}{\sqrt{T(n)h}}\sum_{t=1}^{n} \mathbf{w}_{nt}
$$

with

$$
\mathbf{w}_{nt} = \left[\eta_{t,1\diamond}K_{t0}(x_0), \ldots, \eta_{t,M\diamond}K_{t0}(x_0), \right.
$$
$$
\left. \left(\sum_{k=1}^{M}\eta_{t,k\diamond} \right)K_{t1}(x_0), \ldots, \left(\sum_{k=1}^{M}\eta_{t,k\diamond} \right)K_{tq}(x_0) \right]^\top.
$$

Letting $\mathcal{G}_{nt} = \sigma(X_{s_1}, e_{s_2} : 1 \le s_1 \le n, s_2 \le t)$, we can find that $\{(\mathbf{w}_{nt}, \mathcal{G}_{nt})\}$ is a sequence of martingale differences as $\mathsf{E}(\mathbf{w}_{nt}|\mathcal{G}_{nt-1}) = 0$ a.s. Note that

$$
\mathsf{E}\left[\eta_{t,k_1\diamond}\eta_{t,k_2\diamond}|\mathcal{G}_{nt-1} \right] = \tau_{k_1} \wedge \tau_{k_2} \quad a.s.
$$

for $1 \le k_1, k_2 \le M$. Then, we can show that

$$
\mathsf{Var}(\boldsymbol{\alpha}^\top W_{n\diamond}|\mathcal{G}_{nt-1}) = \boldsymbol{\alpha}^\top \Omega(x_0)\boldsymbol{\alpha} + o_P(1). \tag{A.18}
$$

On the other hand, it is easy to check that

$$
\sum_{t=1}^{n}\mathsf{E}\left[\frac{1}{T(n)h}(\boldsymbol{\alpha}^\top \mathbf{w}_{nt})^2 I(|\boldsymbol{\alpha}^\top \mathbf{w}_{nt}| > \epsilon\sqrt{T(n)h})|\mathcal{G}_{nt-1} \right]
$$
$$
= O_P\left(\frac{1}{T^2(n)h^2}\sum_{t=1}^{n}\left[\sum_{j=0}^{q} K_{tj}^4(x_0) \right] \right) = o_P(1). \tag{A.19}
$$

Then, by (A.18), (A.19) and the central limit theorem for martingale differences (Hall and Heyde, 1980), we can complete the proof of (A.17). ∎

Proof of Corollary 3.1. To prove (3.5), we only need to derive the asymptotic bias term for the local linear CQR estimator. Note that $\sum_{k=1}^{M} c_k = 0$ as the distribution of e_t is symmetric,

$$\mathsf{E}\big[\eta_{t,k} - \eta_{t,k\Diamond}|\mathcal{F}_n(X)\big] = F_e(c_k - d_{t,k}\sigma^{-1}(X_t)) - F_e(c_k)$$
$$\overset{P}{\sim} f_e(c_k)d_{t,k}/\sigma(X_t),$$

and

$$S(x_0) = \mathrm{diag}$$

$$\times \left\{ \frac{f_e(c_1)p_s(x_0)\mu_0}{\sigma(x_0)}, \ldots, \frac{f_e(c_M)p_s(x_0)\mu_0}{\sigma(x_0)}, \frac{\left[\sum_{k=1}^{M} f_e(c_k)\right]p_s(x_0)\mu_2}{\sigma(x_0)} \right\}.$$

Then, by some elementary calculations, we can show that

$$\mathsf{E}[\widehat{m}(x_0)] - m(x_0) = \frac{1}{2}m^{(2)}(x_0)\mu_2 h^2(1 + o(1)), \quad (A.20)$$

which completes the proof of (3.5). The asymptotic distribution (3.6) can be proved using (3.5). ∎

Proof of Corollary 3.2. As $T(n)$ has the same order of n for the positive recurrent case, (3.14) can be proved by using Corollary 3.1. We next only consider the β-null recurrent case. The proof is similar to the proof of Corollary 3.2 in Li et al. (forthcoming). By the definition of Mittag-Leffler distribution, there exist two positive constants $0 < C_l < C_u < \infty$ such that $\mathsf{P}\big(C_l < M_\beta(1) \leq C_u\big) \geq 1 - \frac{\epsilon}{2}$ for any small $\epsilon > 0$. By (2.3) in Section 2.1, for n large enough, we have

$$\mathsf{P}\left(\frac{T(n)}{n^\beta L_s(n)} \leq C_u\right) - \mathsf{P}\left(M_\beta(1) \leq C_u\right) \geq -\frac{\epsilon}{4}, \quad (A.21)$$

$$\mathsf{P}\left(\frac{T(n)}{n^\beta L_s(n)} \leq C_l\right) - \mathsf{P}\left(M_\beta(1) \leq C_l\right) \geq -\frac{\epsilon}{4}. \quad (A.22)$$

Thus, (A.21) and (A.22) imply for large enough n

$$\mathsf{P}\{J_n(\beta)\} \geq 1 - \epsilon, \quad (A.23)$$

where $J_n(\beta)$ is the event $\big\{C_l n^\beta L_s(n) \leq N(n) \leq C_u n^\beta L_s(n)\big\}$. Let C_* be a positive constant and $J_n^c(\beta)$ be the complement of $J_n(\beta)$. Observe that

$$\mathsf{P}\left\{\big\|\widehat{m}(x_0) - m(x_0)\big\| \geq C_*[h^2 + 1/\sqrt{n^\beta L_s(n)h}]\right\}$$

$$= \mathsf{P}\left\{\big\|\widehat{m}(x_0) - m(x_0)\big\| \geq C_*[h^2 + 1/\sqrt{n^\beta L_s(n)h}], J_n(\beta)\right\}$$

$$+ \mathsf{P}\left\{\big\|\widehat{m}(x_0) - m(x_0)\big\| \geq C_*[h^2 + 1/\sqrt{n^\beta L_s(n)h}], J_n^c(\beta)\right\}$$

$$\leq \mathsf{P}\left\{\big\|\widehat{m}(x_0) - m(x_0)\big\| \geq C_*[h^2 + 1/\sqrt{n^\beta L_s(n)h}], J_n(\beta)\right\}$$

$$+ \mathsf{P}\{J_n^c(\beta)\}.$$

By letting C_* sufficiently large and using (3.5) in Corollary 3.1, we have

$$\mathsf{P}\left\{\big\|\widehat{m}(x_0) - m(x_0)\big\| \geq C_*[h^2 + 1/\sqrt{n^\beta L_s(n)h}], J_n(\beta)\right\} \to 0 \quad (A.24)$$

as $n \to \infty$. On the other hand, by (A.23) and letting $\epsilon \to 0$, we also have

$$\mathsf{P}\{J_n^c(\beta)\} \to 0. \quad (A.25)$$

By (A.24)–(A.25), we can complete the proof of (3.15). ∎

Proof of Corollary 3.3. The asymptotic distribution result in (3.16) can be proved by using (2.3) in Section 2.1, (3.5) in Corollary 3.1 as well as the condition that $\{e_t\}$ is independent of $\{X_t\}$ in Assumption 3. Details are omitted here. ∎

Proof of (A.7). Throughout this proof, we let C denote a positive constant whose value may change from line to line. We only consider the case that $\{X_t\}$ is β-null recurrent as the proof of the positive recurrent case is analogous and simpler.

Observe that for any $\epsilon > 0$

$$\left\{\left\|\frac{1}{T(n)}\sum_{i=1}^{T(n)}\left[\frac{Z_i(\sigma, K, 0)}{h} - \frac{p_s(x_0)\mu_0}{\sigma(x_0)}\right]\right\| > \epsilon\right\}$$

$$\subset \left\{\left\|\frac{1}{T(n)}\sum_{i=1}^{T(n)}\left[\frac{Z_i(\sigma, K, 0)}{h} - \frac{p_s(x_0)\mu_0}{\sigma(x_0)}\right]\right\| > \epsilon, J_n(\beta)\right\}$$

$$\cup J_n^c(\beta), \quad (A.26)$$

where $J_n(\beta)$ and $J_n^c(\beta)$ are defined as in the proof of Corollary 3.2. By (A.25), in order to prove (A.7), it suffices to show that as $n \to \infty$

$$\mathsf{P}\left\{\left\|\frac{1}{T(n)}\sum_{i=1}^{T(n)}\left[\frac{Z_i(\sigma, K, 0)}{h} - \frac{p_s(x_0)\mu_0}{\sigma(x_0)}\right]\right\| > \epsilon, J_n(\beta)\right\} \to 0. \quad (A.27)$$

As $\{Z_i(\sigma, K, 0) : i \geq 1\}$ is i.i.d., we next use Bernstein's inequality (van der Vaart and Wellner, 1996) to prove (A.27). By Assumptions 1 and 5, we have $\left|\frac{Z_i(\sigma,K,0)}{h}\right| \leq Ch^{-1}$ for some $C < \infty$. On the other hand, similarly to the proof of (A.6), we have $\mathsf{E}\left[\frac{Z_i^2(\sigma,K,0)}{h^2}\right] \leq Ch^{-1}$. Then, by Assumption 6 and using Bernstein's inequality, we can prove

$$\mathsf{P}\left\{\left\|\frac{1}{T(n)}\sum_{i=1}^{T(n)}\left[\frac{Z_i(\sigma, K, 0)}{h} - \frac{p_s(x_0)\mu_0}{\sigma(x_0)}\right]\right\| > \epsilon, J_n(\beta)\right\}$$

$$\leq \sum_{k=C_l n^\beta L_s(n)}^{C_u n^\beta L_s(n)} \mathsf{P}\left\{\left\|\frac{1}{k}\sum_{i=1}^{k}\left[\frac{Z_i(\sigma, K, 0)}{h} - \frac{p_s(x_0)\mu_0}{\sigma(x_0)}\right]\right\| > \epsilon\right\}$$

$$\leq \sum_{k=C_l n^\beta L_s(n)}^{C_u n^\beta L_s(n)} \exp\{-Ckh\}$$

$$\leq \sum_{k=C_l n^\beta L_s(n)}^{C_u n^\beta L_s(n)} \exp\{-C^* \log k\} = o(1),$$

where C^* can be taken sufficiently large. We then complete the proof of (A.27). ∎

Proof of (3.7). As in the proof of Corollary 3.1, for the local linear CQR estimation with $q = 1$, we may show that $S(x_0)$ is an $(M + 1) \times (M + 1)$ diagonal matrix, i.e.,

$$S(x_0) = \mathrm{diag}$$

$$\times \left\{ \frac{f_e(c_1)p_s(x_0)}{\sigma(x_0)}, \ldots, \frac{f_e(c_M)p_s(x_0)}{\sigma(x_0)}, \frac{\left[\sum_{k=1}^{M} f_e(c_k)\right]p_s(x_0)\mu_2}{\sigma(x_0)} \right\}$$

as $\mu_0 = 1$ and $\mu_1 = 0$. Let $\widehat{\theta}_M = (\widehat{u}_1, \ldots, \widehat{u}_M)^\top$ with $\widehat{u}_k = \sqrt{T(n)h}\big[\widehat{b}_{0k} - m(x_0) - \sigma(x_0)c_k\big]$. By (A.14) in the proof of Theorem 3.1,

$$\widehat{\boldsymbol{\theta}}_M^\top \stackrel{P}{\sim} \left\{ \left[\frac{f_e(c_1)p_s(x_0)}{\sigma(x_0)} \right]^{-1} \cdot \frac{1}{\sqrt{T(n)h}} \sum_{t=1}^n \eta_{t,1} K_{t0}(x_0), \ldots, \right.$$

$$\left. \left[\frac{f_e(c_M)p_s(x_0)}{\sigma(x_0)} \right]^{-1} \cdot \frac{1}{\sqrt{T(n)h}} \sum_{t=1}^n \eta_{t,M} K_{t0}(x_0) \right\}. \tag{A.28}$$

Noting that $\sum_{k=1}^M c_k = 0$, we have

$$\widehat{m}(x_0) - m(x_0) \stackrel{P}{\sim} \frac{1}{T(n)h} \sum_{t=1}^n \xi_t K_{t0}(x_0), \tag{A.29}$$

where

$$\xi_t = \frac{1}{M} \sum_{k=1}^M \left[\frac{f_e(c_k)p_s(x_0)}{\sigma(x_0)} \right]^{-1} \eta_{t,k}$$

$$= \frac{1}{M} \sum_{k=1}^M \left[\frac{f_e(c_k)p_s(x_0)}{\sigma(x_0)} \right]^{-1} \left[I(e_t \le c_k - d_{t,k}\sigma^{-1}(X_t)) - \tau_k \right].$$

Then we may show that

$$\mathsf{E}\left\{ [\widehat{m}(x_0) - m(x_0)]^2 \,|\, \mathscr{F}_n(X) \right\}$$

$$\stackrel{P}{\sim} \frac{1}{T(n)h} \mathsf{E}\left[\frac{1}{T(n)h} \left(\sum_{t=1}^n \{\xi_t - \mathsf{E}\left[\xi_t|\mathscr{F}_n(X)\right]\} K_{t0}(x_0) \right)^2 \,\Big|\, \mathscr{F}_n(X) \right]$$

$$+ \left[\frac{1}{T(n)h} \sum_{t=1}^n \mathsf{E}\left[\xi_t|\mathscr{F}_n(X)\right] K_{t0}(x_0) \right]^2$$

$$=: \Pi_{n1} + \Pi_{n2}. \tag{A.30}$$

Denote

$$\xi_{t\diamond} = \frac{1}{M} \sum_{k=1}^M \left[\frac{f_e(c_k)p_s(x_0)}{\sigma(x_0)} \right]^{-1} \eta_{t,k\diamond}$$

$$= \frac{1}{M} \sum_{k=1}^M \left[\frac{f_e(c_k)p_s(x_0)}{\sigma(x_0)} \right]^{-1} \left[I(e_t \le c_k) - \tau_k \right].$$

Following the proof of Corollary 3.1, we have

$$\Pi_{n2} = \left[\frac{1}{T(n)h} \sum_{t=1}^n \{\mathsf{E}\left[\xi_t|\mathscr{F}_n(X)\right] - \mathsf{E}\left[\xi_{t\diamond}|\mathscr{F}_n(X)\right]\} K_{t0}(x_0) \right]^2$$

$$\stackrel{P}{\sim} \left[\frac{1}{2} m^{(2)}(x_0)\mu_2 \right]^2 h^4 =: \frac{1}{4} b_m^2(x_0)h^4. \tag{A.31}$$

On the other hand, by the argument in the proof of Theorem 3.1, we have

$$\Pi_{n1} \stackrel{P}{\sim} \frac{1}{T(n)h} \mathsf{E}\left[\frac{1}{T(n)h} \left(\sum_{t=1}^n \{\xi_{t\diamond} - \mathsf{E}\left[\xi_{t\diamond}|\mathscr{F}_n(X)\right]\} K_{t0}(x_0) \right)^2 \right.$$

$$\left. \times |\mathscr{F}_n(X) \right] = \frac{\sigma_*^2(x_0)}{T(n)hp_s(x_0)}, \tag{A.32}$$

where $\sigma_*^2(x_0)$ is defined in Corollary 3.1. We then complete the proof of (3.7) by (A.30)–(A.32). ∎

References

Billingsley, P., 1968. Convergence of Probability Measure. Wiley, New York.

Bradic, J., Fan, J., Wang, W., 2011. Penalized composite quasi-likelihood for ultrahigh-dimensional variable selection. J. R. Stat. Soc. Ser. B Stat. Methodol. 73, 325–349.

Cai, Z., Li, Q., Park, J., 2009. Functional-coefficient models for nonstationary time series data. J. Econometrics 148, 101–113.

Cai, B., Tjøstheim, D., 2015. Nonparametric regression estimation for multivariate null recurrent processes. Econometrics 3, 265–288.

Chen, J., Gao, J., Li, D., 2012. Estimation in semi-parametric regression with non-stationary regressors. Bernoulli 18, 678–702.

Engle, R., Granger, C.W.J., 1987. Cointegration and error correction: representation, estimation and testing. Econometrica 55, 251–276.

Fan, J., Gijbels, I., 1995. Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. J. R. Stat. Soc. Ser. B Stat. Methodol. 57, 371–394.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman and Hall, London.

Fan, J., Hu, T.C., Truong, Y.K., 1994. Robust non-parametric function estimation. Scand. J. Stat. 21, 433–446.

Fan, J., Yao, Q., 1998. Efficient estimation of conditional variance functions in stochastic regression. Biometrika 85, 645–660.

Gao, J., Kanaya, S., Li, D., Tjøstheim, D., 2015. Uniform consistency for nonparametric estimators in null recurrent time series. Econometric Theory 31, 911–952.

Gao, J., Tjøstheim, D., Yin, J., 2013. Estimation in threshold autoregressive models with nonstationarity. J. Econometrics 172, 1–13.

Granger, C.W.J., 1981. Some properties of time series data and their use in econometric model specification. J. Econometrics 16, 121–130.

Härdle, W., Tsybakov, A., 1997. Local polynomial estimators of the volatility function in nonparametric autoregression. J. Econometrics 81, 223–242.

Hall, P., Heyde, C., 1980. Martingale Limit Theory and its Applications. Academic Press, New York.

Kai, B., Li, R., Zou, H., 2010. Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. J. R. Stat. Soc. Ser. B Stat. Methodol. 72, 49–69.

Kai, B., Li, R., Zou, H., 2011. New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. Ann. Statist. 39, 305–332.

Karlsen, H.A., Myklebust, T., Tjøstheim, D., 2007. Nonparametric estimation in a nonlinear cointegration type model. Ann. Statist. 35, 252–299.

Karlsen, H.A., Tjøstheim, D., 2001. Nonparametric estimation in null recurrent time series. Ann. Statist. 29, 372–416.

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.

Kim, S., Zhao, Z., Xiao, Z., 2012. Efficient Estimation for Time-varying Coefficient Longitudinal Models. Working Paper, Department of Statistics, Penn State University.

Knight, K., 1998. Limiting distributions for L_1 regression estimators under general conditions. Ann. Statist. 26, 755–770.

Koenker, R., 2005. Quantile Regression. Cambridge University Press.

Koenker, R., Bassett, G.W., 1978. Regression quantiles. Econometrica 46, 33–50.

Li, D., Tjøstheim, D., Gao, J., 2016. Estimation in nonlinear regression with Harris recurrent Markov chains. Ann. Statist. (forthcoming).

Lin, Z., Li, D., Chen, J., 2009. Local linear M-estimators in null recurrent time series. Statist. Sinica 19, 1683–1703.

Myklebust, T., Karlsen, H.A., Tjøstheim, D., 2012. Null recurrent unit root processes. Econometric Theory 28, 1–41.

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

Park, J., Hahn, S.B., 1999. Cointegrating regressions with time varying coefficients. Econometric Theory 15, 664–703.

Pollard, D., 1991. Asymptotics for least absolute deviation regression estimators. Econometric Theory 7, 186–199.

Ruppert, D., 1997. Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. J. Amer. Statist. Assoc. 92, 1042–1062.

Ruppert, D., Wand, M.P., Carroll, R.J., 2003. Semiparametric Regression. In: Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Teräsvirta, T., Tjøstheim, D., Granger, C.W.J., 2010. Modelling Nonlinear Economic Time Series. Oxford University Press.

van der Vaart, A.W., Wellner, J., 1996. Weak Convergence and Empirical Processes with Applications to Statistics. Springer.

Wang, Q.Y., Phillips, P.C.B., 2009a. Asymptotic theory for local time density and nonparametric cointegration regression. Econometric Theory 25, 710–738.

Wang, Q.Y., Phillips, P.C.B., 2009b. Structural nonparametric cointegrating regression. Econometrica 77, 1901–1948.

Wang, Q.Y., Wang, Y., 2013. Nonparametric cointegrating regression with NNH errors. Econometric Theory 29, 1–27.

Wheeden, R.L., Zygmund, A., 1977. Measure and Integral. Dekker.

Xiao, Z., 2009. Quantile cointegrating regression. J. Econometrics 150, 248–260.

Yao, Q., Tong, H., 1998. Cross-validatory bandwidth selections for regression estimation based on dependent data. J. Statist. Plann. Inference 68, 387–415.

Zou, H., Yuan, M., 2008. Composite quantile regression and the oracle model selection theory. Ann. Statist. 36, 1108–1126.