Journal of the Korean Statistical Society 45 (2016) 381–394

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Local linear estimation for regression models with locally stationary long memory errors

Lihong Wang

Department of Mathematics, Nanjing University, Nanjing, 210093, China

ARTICLE INFO

Article history:
Received 2 May 2015
Accepted 30 December 2015
Available online 18 January 2016

AMS 2000 subject classifications:
primary 62M10
secondary 62G08

Keywords:
Asymptotic behavior
Local linear regression estimation
Locally stationary long memory process

ABSTRACT

In this paper we consider the local linear regression estimation for the nonparametric regression models with locally stationary long memory errors. The asymptotic behaviors of the regression estimators are established. It is shown that there is a multiple bandwidth dichotomy for the asymptotic distribution of the estimators of the regression function and its derivatives. The finite sample performance of the estimator is discussed through simulation studies.

1. Introduction

In recent years the long range dependence is often seen in many scientific investigations. In particular, locally stationary long memory (LSLM) processes are becoming an important tool for analyzing non-stationary long range dependent time series. Data examples with time-varying long memory parameter can be found, for instance, in geophysics, oceanography, meteorology, economics, telecommunication engineering and medicine, see, e.g. Beran (2009), Beran, Sherman, Taqqu, and Willinger (1995), Falconer and Fernandez (2007), Granger and Hyung (2004), Lavielle and Ludena (2000), Palma (2010), Ray and Tsay (2002), Roueff and von Sachs (2011) and Whitcher and Jensen (2000), among others. For more details about long memory time series in general, see, e.g. Beran (1994), Beran, Feng, Ghosh, and Kulik (2013), Dobrushin and Major (1979), Doukhan, Oppenheim, and Taqqu (2003), Giraitis, Koul, and Surgailis (2012), Guégan (2005), Palma (2007) and the references therein.

For the LSLM processes, Beran (2009) proposed a maximum likelihood type method to estimate the time-varying long memory parameter $d(u)$. Roueff and von Sachs (2011) investigated the asymptotic behaviors of a local log-regression wavelet estimator of $d(u)$. Wang (2015) explored the properties of the GPH-type estimator and the Local Whittle estimator for a LSLM process characterized by a singularity at the origin of the time varying generalized spectral density. Palma (2010) studied the estimation of the mean of LSLM processes. The mean estimation is actually the special case of the nonparametric regression estimation, which is a fundamental problem in statistics. Nonparametric regression models allow empirical investigation of the data without imposing any parametric structure a priori. In this article we will study the nonparametric estimation for the regression models having the LSLM errors with a general class of time-varying long memory parameters.

Proceeding a bit more precisely, we consider the random process $(Y_t, \boldsymbol{X}_t) \in \boldsymbol{R} \times \boldsymbol{R}^p$, $p \geq 1$, $t = 1, 2, \ldots$, and the following regression model:

$$Y_t = g(\boldsymbol{X}_t) + E_t, \quad t = 1, 2, \ldots, \tag{1.1}$$

E-mail address: lhwang@nju.edu.cn.

http://dx.doi.org/10.1016/j.jkss.2015.12.005
1226-3192/

where the regression function $g(\cdot)$ has continuous second partial derivatives, the disturbances $E_t = \sigma_t(\boldsymbol{X}_t)\varepsilon_t$, where $\{\varepsilon_t,\ t = 1, 2, \ldots\}$ is a LSLM process with zero mean and finite variance σ_ε^2. We assume that $\{\varepsilon_t,\ t = 1, 2, \ldots\}$ is independent of $\{\boldsymbol{X}_t,\ t = 1, 2, \ldots\}$ and the first, and second, moments of $\sigma_t(\boldsymbol{X}_t)$ exist. Here the conditional heteroscedasticity is permitted and we do not assume the $\sigma_t^2(\cdot)$ are constant over t. We assume the availability of the data $\{Y_t, \boldsymbol{X}_t, t = 1, \ldots, n\}$, where n is the sample size. As discussed in Masry and Mielniczuk (1999) and Wang and Cai (2010), the asymptotic property of the regression estimator does not depend on whether or not the explanatory process $\boldsymbol{X}_t$ is strongly correlated. Therefore, we assume that $\{\boldsymbol{X}_t,\ t = 1, 2, \ldots\}$ is independent, however, the probability densities of $\{\boldsymbol{X}_t\}$, $f_t(\cdot)$, are allowed to vary across t. With the aim of covering more general setting of the regression models and establishing the asymptotic properties of the estimator under relatively broad conditions, we allow for conditional heteroscedasticity as well as non-identically distributed observations.

We shall use a local linear regression estimator based on the observations to estimate the function $g(\boldsymbol{x})$ and its derivatives due to the superiority of local linear fitting in function estimation. The aim of this paper is to study the asymptotic properties of the estimators for LSLM models. We establish conditions to ensure the consistency of the local linear estimator, provide uniform convergence rates and the limiting distributions under different bandwidths. For the stationary long memory processes, the nonparametric regression estimation problem has been investigated extensively, see, e.g. Ciuperca (2011), Kulik and Wichelhaus (2011) and Masry and Mielniczuk (1999), among others.

Let $\nabla g(\cdot)$ be the first partial derivative vector of the regression function $g(\cdot)$. The local linear estimators of $g(\boldsymbol{x})$ and $\nabla g(\boldsymbol{x})$, $\boldsymbol{x} \in \boldsymbol{R}^p$, are defined as

$$\begin{pmatrix} \hat{g}(\boldsymbol{x}) \\ \widehat{\nabla g}(\boldsymbol{x}) \end{pmatrix} = (\boldsymbol{X}^\tau \boldsymbol{D}\boldsymbol{X})^{-1}\boldsymbol{X}^\tau \boldsymbol{D}\boldsymbol{y}, \tag{1.2}$$

where

$$\boldsymbol{X} = \begin{pmatrix} 1 & (\boldsymbol{X_1} - \boldsymbol{x})^\tau \\ \vdots & \vdots \\ 1 & (\boldsymbol{X_n} - \boldsymbol{x})^\tau \end{pmatrix}, \qquad \boldsymbol{y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix},$$

and

$$\boldsymbol{D} = \mathrm{diag}(K_h(\boldsymbol{X_1} - \boldsymbol{x}), \ldots, K_h(\boldsymbol{X_1} - \boldsymbol{x})),$$

where $\boldsymbol{X}^\tau$ denotes the transpose of $\boldsymbol{X}$, h is the bandwidth, K is a kernel with $K_h(\boldsymbol{x}) = K(\boldsymbol{x}/h)/h^p$, where $\boldsymbol{x}/h = (x_1/h, \ldots, x_p/h)^\tau$.

In the following section, we will show the asymptotic properties of the local linear regression estimators $\hat{g}(\boldsymbol{x})$ and $\widehat{\nabla g}(\boldsymbol{x})$ for locally stationary long memory setup. We see that, as in the stationary long memory case (see Masry & Mielniczuk, 1999), there is a multiple bandwidth dichotomy for the asymptotic distribution of the estimators of g and its derivatives. Section 3 illustrates the estimation method for LSLM data with simulation studies. The proofs of the theorems are in Section 4.

2. Asymptotic properties of the estimator

Throughout this paper, we assume the following regularity conditions:

Assumption (A). (A1) $\{\varepsilon_t,\ t = 1, 2, \ldots, n\}$ is a zero mean locally stationary process with the time-varying covariance function satisfying

$$\gamma(s, t) = cov(\varepsilon_s, \varepsilon_t) \sim G\left(\frac{s}{n}, \frac{t}{n}\right)|s - t|^{d(s/n)+d(t/n)-1}$$

for large $|s - t| > 0$, where $0 < d(u) \le d_0 < \frac{1}{2}$ for $u \in [0, 1]$ and G is a continuous function over $[0, 1] \times [0, 1]$ with $G(u, u) > 0$ for all $u \in [0, 1]$.

(A2) The function $d(\cdot)$ reaches its maximum value, d_0, at u_0 with the second derivative $d''(u_0) < 0$ and continuous third derivative.

(A3) The marginal density functions of $\boldsymbol{X}_t$, $f_t(\cdot)$, are continuous at $\boldsymbol{x}$ uniformly over t, $\sup_t f_t(\boldsymbol{x}) < \infty$, and there exists a function $\bar{f}(\boldsymbol{x})$ such that $\lim_{n\to\infty} n^{-1}\sum_{t=1}^n f_t(\boldsymbol{x}) = \bar{f}(\boldsymbol{x})$.

(A4) The functions $\sigma_t^2(\cdot)$ are continuous at $\boldsymbol{x}$ uniformly over t, $\sup_t \sigma_t^2(\boldsymbol{x}) < \infty$, and there exist functions $\bar{\varsigma}(\boldsymbol{x})$ and $\bar{\omega}(\boldsymbol{x})$ such that $\lim_{n\to\infty} n^{-1}\sum_{t=1}^n \sigma_t^2(\boldsymbol{x})f_t(\boldsymbol{x}) = \bar{\varsigma}(\boldsymbol{x})$ and

$$\lim_{n\to\infty} \frac{\sum_{s,t=1;s\neq t}^{n} \sigma_s(\boldsymbol{x})\sigma_t(\boldsymbol{x})f_s(\boldsymbol{x})f_t(\boldsymbol{x})\gamma(s, t)}{\sum_{s,t=1;s\neq t}^{n} \gamma(s, t)} = \bar{\omega}(\boldsymbol{x}).$$