Journal of Statistical Planning and Inference 139 (2009) 900 – 915

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

On parameter estimation for locally stationary long-memory processes

Jan Beran*

Department of Mathematics and Statistics, University of Konstanz, P.O. Box 5560, 78457 Konstanz, Germany

ARTICLE INFO

Article history:
Received 24 May 2007
Received in revised form
20 December 2007
Accepted 26 May 2008
Available online 4 June 2008

Keywords:
Long memory
Fractional ARIMA process
Local stationarity
Bandwidth selection

ABSTRACT

We consider parameter estimation for time-dependent locally stationary long-memory processes. The asymptotic distribution of an estimator based on the local infinite autoregressive representation is derived, and asymptotic formulas for the mean squared error of the estimator, and the asymptotically optimal bandwidth are obtained. In spite of long memory, the optimal bandwidth turns out to be of the order $n^{-1/5}$ and inversely proportional to the square of the second derivative of d. In this sense, local estimation of d is comparable to regression smoothing with iid residuals.

1. Introduction

The usefulness of stationary long-memory processes for modelling time series has been demonstrated in the literature by numerous examples, including applications in hydrology, geophysics, economics, finance, climatology, physics, biology, medicine, music and telecommunications engineering among others (see e.g Mandelbrot, 1977; Beran, 1994, 2003; Lowen and Teich, 2005). Long memory of a second order stationary process X_t is characterized by slowly decaying nonsummable autocovariances

$$\gamma(k) = cov(X_t, X_{t+k}) \sim c_\gamma |k|^{2d-1} \quad (|k| \to \infty) \tag{1}$$

where $d \in (0, \frac{1}{2})$, and a pole of the spectral density at the origin,

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\lambda} \sim c_f |\lambda|^{-2d} \quad (|\lambda| \to 0) \tag{2}$$

Here "$\sim$" means that the ratio of both sides tends to one. For some data sets, however, it has been observed that the assumption of stationarity is too restrictive, even after trends in the mean are removed. In particular, the long-memory parameter d, as well as other parameters characterizing the spectrum of the process, may change as a function of time. Data examples with time-varying d can be found, for instance, in geophysics, oceanography, meteorology, economics, telecommunication engineering, medicine and other areas of statistical applications (see e.g. Beran et al., 1995; Vesilo and Chan, 1996; Whitcher and Jensen, 2000; Lavielle and Ludena, 2000; Ray and Tsay, 2002; Whitcher et al., 2002; Granger and Hyung, 2004; Falconer and Fernandez, 2007). This motivates introducing locally stationary processes with long-range dependence. For locally stationary processes with short-range dependence see e.g. Subba Rao (1970), Hallin (1978), Priestley (1981), Dahlhaus (1996, 1997), Dahlhaus and Giraitis (1998), and Moulines et al. (2005). Jensen and Whitcher (2000) define locally stationary fractional ARIMA (FARIMA) processes (Granger and Joyeux, 1980; Hosking, 1981), and estimate parameters using wavelets. Alternatively, given a specific linear model such as an

* Tel.: +49 7531 88 2653.
E-mail address: jan.beran@uni-konstanz.de.

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2008.05.047

FARIMA, one may consider local estimation based on estimated innovations. This is the approach taken here. For related estimates for stationary long-memory processes, see e.g. Fox and Taqqu (1986), Yajima (1985), Giraitis and Surgailis (1990) and Beran (1995). Modelling time series by locally stationary long-memory processes is closely related to change point detection in the spectral domain. For spectral change point detection in the long-memory context, see e.g. Giraitis and Leipus (1990, 1992), Horváth and Shao (1999), Lavielle and Ludena (2000), Ray and Tsay (2002), and Ben Hariz et al. (2007), also see Kokoszka and Leipus (2003) for a review. It should also be noted that shifts in the mean can also give rise to long-memory type dependence (see e.g. Granger and Ding, 1996; Diebold and Inoue, 2001). Distinguishing nonconstant mean from stationary long memory is possible either under regularity assumptions on a trend function (see e.g. Hall and Hart, 1990; Csörgö and Mielniczuk, 1995; Ray and Tsay, 1997; Beran and Feng 2002a, b) or in the presence of a finite number of change points (see e.g. Horváth and Kokoszka, 1997; Kuan and Hsu, 1998; Wright, 1998; Ray and Tsay, 2002; Sibbertsen, 2004; Berkes et al., 2006). In this paper, we assume the mean to be constant. The methods proposed here may be extended to situations with nonconstant mean by combining them with suitable algorithms for nonparametric regression smoothing (Beran and Feng, 2002b) or change point estimation (Horváth and Kokoszka, 1997).

Specifically, we consider a sequence of processes $X_{t,n}$ having a time-varying infinite autoregressive representation

$$X_{t,n} = \sum_{j=1}^{\infty} b_{j,n} X_{t-j,n} + \varepsilon_t \tag{3}$$

where ε_t are iid zero-mean random variables with finite variance $\sigma_\varepsilon^2 = \sigma_\varepsilon^2(t/n)$ and $b_{j,n} = b_j(\theta(t/n))$. Here $\sigma_\varepsilon^2(u)$ and $\theta(u) = (d(u), \theta_2(u), \ldots, \theta_k(u))^{\mathrm{T}}$ $(u \in [0,1])$ are sufficiently smooth functions of rescaled time. Moreover, for fixed $u = t/n$, the value of $d(u) \in (0, \frac{1}{2})$ is assumed to be such that

$$0 < \lim_{j\to\infty} j^{d+1} b_j(\theta(u)) = c_b < \infty \tag{4}$$

and

$$0 < \lim_{\lambda\to 0} 2\pi\sigma_\varepsilon^{-2} \lambda^{-2d} \left| 1 - \sum_{j=1}^{\infty} b_j e^{-\mathrm{i}j\lambda} \right|^2 = c_f^{-1} < \infty \tag{5}$$

where c_b, c_f are positive constants. In the case of a fractional ARIMA(p,d,q) process, we have $c_f = \sigma_\varepsilon^2/(2\pi)$ and for $z \in \mathbb{C}$, with $|z| \leqslant 1$ and $z \neq 1$,

$$1 - \sum_{j=1}^{\infty} b_j(d) z^j = \varphi(z)\psi^{-1}(z)(1-z)^d \tag{6}$$

where

$$\varphi(z) = 1 - \varphi_1 z - \cdots - \varphi_p z^p \neq 0 \quad (|z| \leqslant 1) \tag{7}$$

$$\psi(z) = 1 - \psi_1 z - \cdots - \psi_q z^q \neq 0 \quad (|z| \leqslant 1) \tag{8}$$

The time-varying parameters are then $\sigma_\varepsilon^2(t/n) = var(\varepsilon_t)$ and $\theta(t/n) = [d(t/n), \varphi_1(t/n), \ldots, \varphi_p(t/n), \psi_1(t/n), \ldots, \psi_q(t/n)]^{\mathrm{T}}$. Note that, $d(u) > 0$ means that locally the process has (approximately) a spectral density with a pole at the origin proportional to $|\lambda|^{-2d(u)}$, and, in the course of time, the rate of divergence of the pole changes slowly.

In this paper, estimation of $\theta(.)$ based on the autoregressive representation (3) is considered. For Gaussian innovations ε_t, this corresponds to an approximate maximum likelihood estimator. Two questions are addressed: (1) asymptotic distribution of $\hat{\theta}(u)$ and (2) the choice of a suitable bandwidth that determines which observations in the neighbourhood of u (or nu on the original time scale) are used for the local estimate. The paper is organized as follows. The asymptotic distribution of $\hat{\theta}$ is derived in Section 2. Section 3 addresses the issue of bandwidth choice. In particular, an asymptotic expression for the mean squared error of $\hat{d}$ is obtained. The asymptotically optimal bandwidth turns out to be proportional to $n^{-1/5}$ and inversely proportional to $\{d''\}^2$. In spite of long-range dependence, the formula are similar to results in the context of regression smoothing with iid errors. For the case of short-memory AR(p) processes also see Dahlhaus and Giraitis (1998). Simulations and data examples in Section 3 illustrate the approximate validity of the asymptotic results for finite samples. Moreover, a simple iterative plug-in algorithm for data driven bandwidth choice is proposed. General comments in Section 4 conclude the paper. Proofs are given in Appendix A.

2. Estimation, asymptotic distribution

Denote by $\theta^0(u)$ the true parameter curve. We consider estimation of $\theta^0(u)$ for a fixed rescaled time point $u_0 \in (0,1)$. Let $t_0(n) = [nu_0]$, $u_{t,n} = t(n)/n$, and denote by $K: \mathbb{R} \to \mathbb{R}_+$ a nonnegative kernel function with $K(-x) = K(x)$, $K(x) = 0\,(|x| > 1)$