AISM (2006) 58: 53–71
DOI 10.1007/s10463-005-0009-5

Sucharita Ghosh · Jan Beran

On estimating the cumulant generating function of linear processes

Received: 23 February 2004 / Revised: 22 November 2004 / Published online: 7 March 2006

Abstract We compare two estimates of the cumulant generating function of a stationary linear process. The first estimate is based on the empirical moment generating function. The second estimate uses the linear representation of the process and the empirical moment generating function of the innovations. Asymptotic expressions for the mean square errors are derived under short- and long-range dependence. For long-memory processes, the estimate based on the linear representation turns out to have a better rate of convergence. Thus, exploiting the linear structure of the process leads to an infinite gain in asymptotic efficiency.

Keywords Empirical moment generating function · Long-range dependence · Short-range dependence.

1 Introduction

Consider the problem of estimating the marginal distribution function $F_X(x) = P(X_i \leq x)$ of a stationary linear stochastic process

$$X_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j} \tag{1}$$

S. Ghosh (✉)
Landscape Department, Swiss Federal Research Institute WSL,
Zürcherstrasse 111, 8903 Birmensdorf, Switzerland
E-mail: rita.ghosh@wsl.ch

J. Beran
Department of Mathematics and Statistics, University of Konstanz,
78457 Konstanz, Germany
E-mail: jan.beran@uni-konstanz.de

where $\sum a_j^2 < \infty$, ε_i are iid with distribution F_ε, $E(\varepsilon_i) = 0$, $\sigma_\varepsilon^2 = \mathrm{var}(\epsilon_i) < \infty$ and such that all cumulants of ε_i exist. Suppose moreover that the cumulant generating function

$$\ell_X(t) = \log\, m_X(t) \tag{2}$$

is finite for $t \in (-T, T)$ where $T > 0$ is some constant and $m_X(t) = E(e^{tX})$ is the moment generating function. Then F_X is fully specified by ℓ_X. The standard estimates of $m_X(t)$ and $\ell_X(t)$ are

$$m_{X,n}(t) = \frac{1}{n}\sum_{i=1}^{n} e^{tX_i} \tag{3}$$

and

$$\ell_{X,n}(t) = \log\, m_{X,n}(t), \tag{4}$$

respectively (see e.g. Csörgő, 1982; Ghosh, 1996, 1999; Ghosh and Beran, 2000 and references therein). On the other hand, if X_i is a linear process, then

$$m_X(t) = \prod_{j=0}^{\infty} m_\varepsilon(a_j t) \tag{5}$$

and

$$\ell_X(t) = \sum_{j=0}^{\infty} \log\, m_\varepsilon(a_j t) = \sum_{j=0}^{\infty} \ell_\varepsilon(a_j t). \tag{6}$$

Thus, denoting by

$$m_{\varepsilon,n}(t) = \frac{1}{n}\sum_{i=1}^{n} e^{t\varepsilon_i} \tag{7}$$

the empirical moment generating function of ε_i, alternative estimates of $m_X(t)$ and $\ell_X(t)$ can be defined by

$$\hat{m}_X(t) = \prod_{j=0}^{N_n} m_{\varepsilon,n}(a_j t) \tag{8}$$

and

$$\hat{\ell}_X(t) = \sum_{j=0}^{N_n} \log m_{\varepsilon,n}(a_j t) = \sum_{j=0}^{N_n} \ell_{\varepsilon,n}(a_j t), \tag{9}$$

respectively. Here, N_n is a sequence of integers such that $1 \le N_n \le \infty$ and $N_n \to \infty$ as $n \to \infty$. Note that an analogous approach can also be based on the characteristic function (see e.g. Feuerverger and Mureika, 1977; Csörgő, 1981, 1986; Murota and Takeuchi, 1981; Ghosh and Ruymgaart, 1992 for asymptotic results and ideas based on the empirical characteristic function).

Note that in practice, observations consist of $X_1, \ldots, X_n$ whereas the sequence ε_i is not known directly. Thus, to make $\hat{\ell}_X(t)$ applicable, it must be possible to estimate the ε_i's from $X_1, \ldots, X_n$ with sufficient accuracy. In particular, invertibility conditions on the coefficients a_j need to be imposed to obtain an autoregressive representation $\varepsilon_i = \sum_{j=0}^{\infty} b_j X_{i-j}$. Estimated residuals may then be defined, for instance, by $\hat{\varepsilon}_i = \sum_{j=0}^{i-1} b_j X_{i-j}$ $(i = 1, \ldots, n-1)$.

For more sophisticated residual estimates based on finite past predictions see e.g. Haslett and Raftery, 1989; Brockwell and Davis, 1987. In this paper, the innovations ε_i $(i = 1, \ldots, n)$ are assumed to be known or estimated with sufficient accuracy. The case where innovations are estimated is discussed briefly in Sect. 5. A detailed study of the effect of different estimation techniques on $\ell_X(t)$ would be beyond the scope of this paper and will be discussed elsewhere.

The main focus here is on the comparison of the two estimates $\ell_{X,n}(t)$ and $\hat{\ell}_X(t)$ with respect to the mean squared error. In contrast to $\ell_{X,n}(t)$, the alternative estimate $\hat{\ell}_X(t)$ exploits the additional information of linearity. It may therefore be conjectured that $\hat{\ell}_X(t)$ could be more efficient than $\ell_{X,n}(t)$ for an appropriate choice of N_n. This question is investigated as follows: In Sect. 2, expressions for the asymptotic bias, variance and mean squared error of $\ell_{X,n}$ are given. Analogous results for $\hat{\ell}_X(t)$ are derived in Sect. 3. These results lead, in Sect. 4, to an asymptotically optimal choice of N_n and a comparison of the mean squared errors (MSEs)

$$\mathrm{MSE}(\ell_{X,n}(t)) = E[(\ell_{X,n}(t) - \ell_X(t))^2] \tag{10}$$

and

$$\mathrm{MSE}(\hat{\ell}_X(t)) = E[(\hat{\ell}_X(t) - \ell_X(t))^2]. \tag{11}$$

In particular, we focus on the case of long-memory processes, since there the difference between the two estimates becomes most pronounced. As it turns out, for long-memory processes with

$$\lim_{j\to\infty} \frac{a_j}{Cj^{d-1}} = 1 \tag{12}$$

for some $C > 0$ and $0 < d < \frac{1}{2}$, exploiting the linear structure of the process leads asymptotically to an infinite increase in efficiency. In other words, if N_n tends to infinity at a rate that is neither too fast nor too slow, then $\mathrm{MSE}(\ell_{X,n}(t))/\mathrm{MSE}(\hat{\ell}_X(t))$ tends to infinity as $n \to \infty$. The optimal rate of N_n is obtained by balancing the bias of $\hat{\ell}_X(t)$, which is due to truncation of the sum in Eq. (6) at N_n and the variance, which is due to summing up an increasing number N_n of random variables.

Nonparametric kernel estimation of the marginal density functions of a long-memory process is considered in recent papers by Wu and Mielniczuk (2002), Honda (2000), Csörgő and Mielniczuk (1995) and Hall and Hart (1990). Also see Dehling and Taqqu (1989) and Giraitis and Surgailis (1999) for related results on the empirical distribution function. For kernel estimates, the optimal bandwidth depends on the unknown long-memory parameter and the density function in a complex manner. The significance of the results presented here is that alternative and more efficient estimates may be obtained by estimating the cumulant generating function of the innovation process and subsequent back-transformation. This may open up the possibility of applying kernel density estimation to the innovations ε_i using well-known bandwidth selection procedures for independent data, (see e.g. Silverman, 1986 and references therein; Hall, 1993; Engel et al. 1994), and plugging the corresponding estimate of $m_\epsilon(t)$ into Eq. (6). From this, an estimated density of the observed process X_i may be calculated by inverse Laplace transformation. This is of particular interest for long-memory data where it is not known at present how to obtain an optimal bandwidth for a direct kernel density estimate.

2 Asymptotic results for $\ell_{X,n}(t)$

2.1 Asymptotic bias

Throughout the paper, $m_X(t)$ will be assumed to be twice continuously differentiable and X_i is a linear process defined by Eq. (1). By definition, the empirical moment generating function $m_{X,n}$ is unbiased. The asymptotic bias of $\ell_{X,n}$ follows directly by Taylor expansion under fairly general conditions:

Lemma 1 *Suppose that $m_{X,n}(t)$ converges in the L^2-norm to $m_X(t)$. Then*

$$\ell_{X,n}(t) = \ell(t) + \frac{m_{X,n}(t) - m_X(t)}{m_X(t)} - \frac{1}{2} m_X^{-2}(t)\{m_{X,n}(t) - m_X(t)\}^2 + r_1 \qquad (13)$$

with $r_1 = o_p\{(m_{X,n}(t) - m_X(t))^2\}$, and

$$\mathrm{Bias}(\ell_{X,n}(t)) = E[\ell_{X,n}(t)] - \ell_X(t) = \frac{1}{2} m_X^{-2}(t) \mathrm{var}(m_{X,n}(t)) + r_2 \qquad (14)$$

with $r_2 = o\{\mathrm{var}(m_{X,n}(t))\}$.

The proof is obvious and is therefore omitted.

2.2 Asymptotic variance and MSE

To evaluate the asymptotic variance of $m_{X,n}$ and $\ell_{X,n}(t)$, two different situtations are characterized by the following assumptions:

A1 (short memory): There exist constants $0 < C < \infty$, $0 < \varphi < 1$ such that

$$|a_j| \leq C\varphi^j \qquad (15)$$

A2 (long memory): There exist constants $0 < C < \infty, 0 < d < \frac{1}{2}$ such that

$$\lim_{j\to\infty} \frac{a_j}{Cj^{d-1}} = 1 \tag{16}$$

Assumption A1 implies short memory with $|\gamma(k)| = |\mathrm{cov}(X_i, X_{i+k})| \le A\varphi^{|k|}$ for some constant $0 < A < \infty$, and a spectral density

$$f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi}\left|\sum_{j=0}^{\infty} a_j \exp(ij\lambda)\right|^2$$

that converges to a finite positive constant at the origin. The latter can also be expressed by $\lim_{\lambda\to 0} f(\lambda)\lambda^{2d} = c_\mathrm{f}$ with $d = 0$ and $0 < c_\mathrm{f} < \infty$. These properties hold, for instance, for stationary ARMA processes. Assumption A2 implies long memory with slowly decaying aucotocovariances characterized by

$$\lim_{|k|\to\infty} \frac{\gamma(k)}{C^2|k|^{2d-1}} = 1. \tag{17}$$

In particular, $\sum_{k=-\infty}^{\infty} \gamma(k) = \infty$ and for the spectral density we have $\lim_{\lambda\to 0} f_X(\lambda)\lambda^{2d} = c_\mathrm{f}$, i.e. f has a hyperbolic pole at the origin. Best known examples are fractional ARIMA (or FARIMA) models (Granger and Joyeux, 1980; Hosking, 1981) and fractional Gaussian noise (Mandelbrot and Wallis, 1969). For an overview on long-memory processes see e.g. Beran (1994).

The asymptotic variance of $m_{X,n}(t)$ and $\ell_{X,n}(t)$ is given by the following two propositions.

Proposition 1 *Under A1, we have, as $n \to \infty$,*

$$\lim_{n\to\infty} n\,\mathrm{var}(m_{X,n}(t)) = v(t) \tag{18}$$

with

$$v(t) = \sum_{k=-\infty}^{\infty}\left\{\prod_{j=0}^{\infty} m_\varepsilon(t(a_j + a_{j+k}))\prod_{j=0}^{k-1} m_\varepsilon(ta_j) - m_X^2(t)\right\} \tag{19}$$

and

$$\lim_{n\to\infty} n\,\mathrm{var}(\ell_{X,n}(t)) = m_X^{-2}(t)v(t) \tag{20}$$

Proposition 2 *Under A2, we have, as $n \to \infty$,*

$$\lim_{n\to\infty} n^{1-2d}\,\mathrm{var}(m_{X,n}(t)) = w(t) \tag{21}$$

with

$$w(t) = \frac{C^2\sigma_\varepsilon^2 t^2 m_X^2(t)\{\int_0^\infty x^{d-1}(1+x)^{d-1}\mathrm{d}x - [2(1-2d)]^{-1}\}}{d(2d+1)} \tag{22}$$

and

$$\lim_{n\to\infty} n^{1-2d}\,\mathrm{var}(\ell_{X,n}(t)) = m_X^{-2}(t)w(t). \tag{23}$$

The rate of convergence of $\mathrm{var}(\ell_{X,n})$ is discussed in a proceedings paper by Ghosh (2003). Propositions 1 and 2 provide explicit expressions for the asymptotic variance. Detailed proofs are given in Appendix. Note, in particular, that under long-range dependence the variance converges to zero at a rate slower than n^{-1}. This phenomenon is well known for long-memory processes (see e.g. Taqqu, 1975; Giraitis and Surgailis, 1985, 1999; Beran, 1991, 1994). Together with Lemma 1, we also obtain the rate of convergence for the bias of $\ell_{X,n}(t)$.

Corollary 1 *As $n \to \infty$ the following holds*

1. *Under A1,*

$$\mathrm{Bias}(\ell_{X,n}(t)) = E[\ell_{X,n}(t)] - \ell_X(t) = O(n^{-1}) \tag{24}$$

2. *Under A2,*

$$\mathrm{Bias}(\ell_{X,n}(t)) = E[\ell_{X,n}(t)] - \ell_X(t) = O(n^{2d-1}). \tag{25}$$

Finally, the asymptotic expression for the $MSE(\ell_{X,n}(t)) = E[(\ell_{X,n}(t) - \ell_X(t))^2]$ follows from the decomposition $MSE = \mathrm{Bias}^2(\ell_{X,n}(t)) + \mathrm{Var}(\ell_{X,n}(t))$. Lemma 1 and Propositions 1 and 2 imply that the MSE is dominated asymptotically by the variance. Thus, we have

Proposition 3 *As $n \to \infty$, the following holds:*

1. *Under A1,*

$$\lim_{n\to\infty} n\,\mathrm{MSE}(m_{X,n}(t)) = v(t) \tag{26}$$

and

$$\lim_{n\to\infty} n\,\mathrm{MSE}(\ell_{X,n}(t)) = m_X^{-2}(t)v(t) \tag{27}$$

with $v(t)$ defined in Proposition 1;

2. *Under A2,*

$$\lim_{n\to\infty} n^{1-2d}\,\mathrm{MSE}(m_{X,n}(t)) = w(t) \tag{28}$$

and

$$\lim_{n\to\infty} n^{1-2d}\,\mathrm{MSE}(\ell_{X,n}(t)) = m_X^{-2}(t)w(t) \tag{29}$$

with $w(t)$ defined in Proposition 2.

3 Asymptotic results for $\hat{\ell}_X(t)$

3.1 Asymptotic bias

The bias of $m_{\varepsilon,n}$ is zero and the bias of $\ell_{\varepsilon,n}(t)$ is of the order of the variance of $\ell_{\varepsilon,n}(t)$ (see Lemma 1). The bias of $\hat{\ell}_X(t)$ depends on the rate of decay of the coefficients a_j.

Theorem 1 *Under A1 we have, as $n \to \infty$,*

$$\mathrm{Bias}(\hat{\ell}_X(t)) = E[\hat{\ell}_X(t)] - \ell_X(t) = O(\max(n^{-1}, \varphi^{2N_n})) \tag{30}$$

A different result is obtained for slowly decaying correlations.

Theorem 2 *Under A2 we have, as $n \to \infty$,*

$$\lim_{n\to\infty} N_n^{1-2d}\mathrm{Bias}(\hat{\ell}_X(t)) = \lim_{n\to\infty} N_n^{1-2d}\{E[\hat{\ell}_X(t)] - \ell_X(t)\} = B(t) \tag{31}$$

with

$$B(t) = -\frac{C^2\sigma_\varepsilon^2 t^2}{2(1-2d)}. \tag{32}$$

3.2 Asymptotic variance

The asymptotic variance of $\hat{\ell}_X(t)$ follows from the covariances of $m_{\varepsilon,n}(t)$ and $\ell_{\varepsilon,n}(t)$ together with the asymptotic behaviour of the coefficients a_j and the properties of $m_\varepsilon(t)$ near the origin. Asymptotic expressions for the covariances of $m_{\varepsilon,n}(t)$ and $\ell_{\varepsilon,n}(t)$, respectively are well known (see e.g. Csörgő, 1982; Ghosh, 1996, 1999; Ghosh and Beran, 2000 and references therein):

Lemma 2 *If ε_i are iid with existing finite moment generating function $m_\varepsilon(t)$, then*

$$\mathrm{cov}(m_{\varepsilon,n}(t), m_{\varepsilon,n}(s)) = n^{-1}\{m_\varepsilon(t+s) - m_\varepsilon(t)m_\varepsilon(s)\} \tag{33}$$

and

$$g(t,s) = \mathrm{cov}(\ell_{\varepsilon,n}(t), \ell_{\varepsilon,n}(s)) = n^{-1}\left\{\frac{m_\varepsilon(t+s)}{m_\varepsilon(t)m_\varepsilon(s)} - 1\right\} + r_n(t,s) \tag{34}$$

where $r_n(t,s) = o(g(t,s))$.

This implies

Theorem 3 *As $n \to \infty$, we have*

$$v_n = \mathrm{var}(\hat{\ell}_X(t)) = A_n(t) + r(t, N_n) \tag{35}$$

with $r(t,N_n) = o(A_n)$ and

$$A_n(t) = n^{-1}\sum_{i,j=0}^{N_n}\left\{\frac{m_\varepsilon(t(a_i+a_j))}{m_\varepsilon(ta_i)m_\varepsilon(ta_j)} - 1\right\}. \tag{36}$$

Since A_n may or may not converge to a finite constant, this result does not yet tell us the rate of convergence of v_n.

For short-memory processes we then have

Theorem 4 *Under A1,*

$$\lim_{n\to\infty} n v_n = A(t) \tag{37}$$

with

$$A(t) = \sum_{i,j=0}^{\infty} \left\{ \frac{m_\varepsilon(t(a_i + a_j))}{m_\varepsilon(ta_i)m_\varepsilon(ta_j)} - 1 \right\},\ 0 < A(t) < \infty. \tag{38}$$

The analogous result for long-memory processes is

Theorem 5 *Under A2,*

$$\lim_{n\to\infty} n N_n^{-2d} v_n = D(t) \tag{39}$$

with

$$D(t) = \frac{C^2 t^2 \sigma_\epsilon^2}{d^2} \tag{40}$$

These results, together with the expressions for the bias, imply that the role of the bias in the MSE is fundamentally different for the cases of long and short memory, respectively: For short memory, the square of the bias is of smaller order than the variance so that the MSE is asymptotically equal to the variance. In contrast, under long memory, the square of the bias may be asymptotically of the same order as the variance or it may be smaller or larger, depending on the order of N_n/n. Note, in particular, that under long memory, the variance is proportional to $n^{-1}N_n^{2d}$ instead of n^{1-2d}. This will make it possible to choose N_n such that the MSE is of smaller order than the one for $\ell_{X,n}$. More specifically we have

Theorem 6 *Under A1,*

$$\lim_{n\to\infty} n\,\mathrm{MSE}(\hat{\ell}_X(t)) = A(t) \tag{41}$$

with $A(t)$ as in Theorem 4.

Theorem 7 *Under A2,*

$$\mathrm{MSE}(\hat{\ell}_X(t)) = B^2(t)N_n^{4d-2} + D(t)N_n^{2d}n^{-1} + o(\max(N_n^{4d-2}, N_n^{2d}n^{-1})) \tag{42}$$

where $B(t)$ and $D(t)$ are defined in Theorems 2 and 5, respectively.

4 Asymptotically optimal N_n and MSE under long memory

Theorem 7 implies an asymptotically optimal choice of N_n under long memory. Assume that $N_n = \beta n^{\alpha}$. Then minimization of $B^2 N_n^{4d-2} + D N_n^{2d} n^{-1}$ with respect to α and β yields

Corollary 2 *Under A2, the optimal choice of N_n is*

$$N_n = \beta n^{1/(2-2d)} \tag{43}$$

with

$$\beta = \left(\frac{C^2 t^2 d}{4(1-2d)} \right)^{1/(2-2d)} \tag{44}$$

The resulting asymptotically optimal MSE is given by

$$\mathrm{MSE}_{\mathrm{opt}}(\hat{\ell}_X(t)) = M n^{(2d-1)/(1-d)} + o(n^{(2d-1)/(1-d)}) \tag{45}$$

with

$$M = B^2 \beta^{4d-2} + D\beta^{2d} \tag{46}$$

Note that the optimal rate of N_n is such that the contribution of the bias and the variance to the MSE are of the same order. This is similar to results in nonparametric smoothing.

We now can compare the asymptotic MSE of the two estimators $\ell_{X,n}(t)$ and $\hat{\ell}_X(t)$:

Corollary 3 *Let $N_n = \beta n^{\alpha}$ with $\frac{1}{2} < \alpha < 1$. Then under A2, there are constants $0 < \delta < \infty$ and $0 < q(t) < \infty$ such that*

$$\lim_{n\to\infty} n^{-\delta} \cdot \left\{ \frac{\mathrm{MSE}(\ell_{X,n}(t))}{\mathrm{MSE}_{\mathrm{opt}}(\hat{\ell}_X(t))} \right\} = q(t). \tag{47}$$

In particular, for $N_n = \beta n^{1/(2-2d)}$,

$$\delta = \frac{d(1-2d)}{1-d}. \tag{48}$$

This result means that the relative asymptotic efficiency of $\ell_{X,n}(t)$ as compared to $\hat{\ell}_X(t)$ is zero, with a hyperbolic rate of deterioration, provided that N_n diverges to infinity slower than n but faster than $\sqrt{n}$. The intuitive reason is that $\hat{\ell}_X(t)$ uses the additional information of linearity. The variance is kept low by not adding too many random terms, each of them having a square root n rate of convergence. The bias is kept low by adding a sufficient number of terms. The optimal choice of N_n keeps a balance between bias and variance such that their contribution to the MSE is of the same order. This possibility of optimization disappears for short-memory processes, since there the variance is of order n^{-1} and dominates the MSE, independently of the choice of N_n. This can also be seen in the expression for δ. As d tends to 0, δ tends to zero as well. On the other hand, for d tending to 1/2, we also have $\delta \to 0$. The reason is that for d close to 1/2, the bias due to leaving out terms in the infinite series (6) increases. The optimal number of terms is proportional to $n^{1/(2-2d)}$ which approaches n as d tends to 1/2. As a result, the improvement by using a finite slowly increasing number of terms in $\hat{\ell}_X(t)$ disappears.

5 Estimating innovations

The results above are derived under the assumption that ϵ_i and the coefficients a_j are known. For observed time series, the innovations ϵ_i are not observable directly but must be calculated from the observed series X_i, and the coefficients $a_j (j \geq 0)$ have to be estimated. Here, we discuss the general idea how to proceed in this case.

Suppose that X_i is an invertible linear process given by Eq. (1). Then there are uniquely defined coefficients b_j such that

$$\epsilon_i = \sum_{j=0}^{\infty} b_j X_{i-j}. \tag{49}$$

Estimation of b_j can be done, for instance, by applying a flexible class of parametric models for the spectral density function f_X. For example, we may consider fractional ARIMA-models given by Granger and Joyeux (1980) and Hosking (1981)

$$(1 - B)^d \phi(B) X_i = \psi(B)\varepsilon_i \tag{50}$$

where $-\frac{1}{2} < d < \frac{1}{2}$, B is the backshift operator, the polynomials $\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j$ and $\psi(z) = 1 + \sum_{j=1}^{q} \psi_j z^j$ have no roots with $|z| \leq 1$ and

$$(1 - B)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-B)^j.$$

Here, we have

$$f_X(\lambda) = f_X(\lambda; \theta) = \frac{\sigma_\epsilon^2}{2\pi} |\psi(\mathrm{e}^{-i\lambda})/\phi(\mathrm{e}^{-i\lambda})|^2 |1 - \mathrm{e}^{i\lambda}|^{-2d}$$

where $\theta = (\sigma_\epsilon^2, d, \phi_1, \ldots, \phi_p, \psi_1, \ldots, \psi_q)$. Given p and q, the unknown parameter vector θ can be estimated by Whittle's estimator or another (approximate) Gaussian maximum likelihood method. Giraitis and Surgailis (1990) showed that, even if X_t is non-Gaussian, this estimate is consistent and a central limit theorem holds under standard regularity conditions. Given estimates $\hat{b}_j = b_j(\hat{\theta})$, ϵ_i may thus be estimated by

$$\hat{\epsilon}_i = \sum_{j=0}^{i-1} \hat{b}_j X_{i-j}. \tag{51}$$

An alternative, more precise, estimate can be given by

$$\hat{\epsilon}_i = \sum_{j=0}^{i-1} \hat{b}_j X_{i-j} + \hat{\mu}_i \tag{52}$$

where $\hat{\mu}_i$ is the best linear forecast (into the past) of $\sum_{j=i}^{\infty} \hat{b}_j X_{i-j}$ given $X_1, \ldots, X_n$. A detailed study of the effect of different approximations $\hat{\epsilon}_i$ on the MSE of $\ell_X(t)$ would be beyond the scope of this paper and will be pursued elsewhere.

6 Simulations

The results are illustrated by a small simulation study. For $d = 0.1, 0.2, 0.3$ and 0.4, 2,000 replicated series of lengths $n = 25, 50, 100, 200, 400, 800, 1{,}000, 1{,}200, 1{,}400, 1{,}600, 1{,}800$ and $2{,}000$ of a stationary FARIMA(0,d,0) process (Granger and Joyeux, 1980; Hosking, 1981)

$$(1 - B)^d X_i = \varepsilon_i \tag{53}$$

were generated, with (a) ε_i iid $N(0, \sigma_\varepsilon^2)$ distributed, and (b) $\varepsilon_i = \xi_i - 1/2$ where ξ_i are iid exponential with expected value $1/2$. For each value of d, the simulated ratio of the MSEs $q_n = \mathrm{MSE}(\ell_{X,n})/\mathrm{MSE}_{\mathrm{opt}}(\hat{\ell}_X)$ was calculated. In Fig. 1a–d, the simulated values $\log q_n(t)$ for case (a) are plotted against $\log n$. The same plots for case (b) are given in Fig. 2a–d. As expected from the theoretical results, the simulated values $\log q_n$ are scattered approximately around an increasing straight line. Using sample sizes of $n = 1{,}000$ and larger, the fitted least squares slopes and the theoretical asymptotic values are reasonably close (see Tables 1, 2). Note also that all y coordinates are above zero, even for $n = 25$. This indicates that even for small sample sizes the estimate based on the linear representation is more efficient.

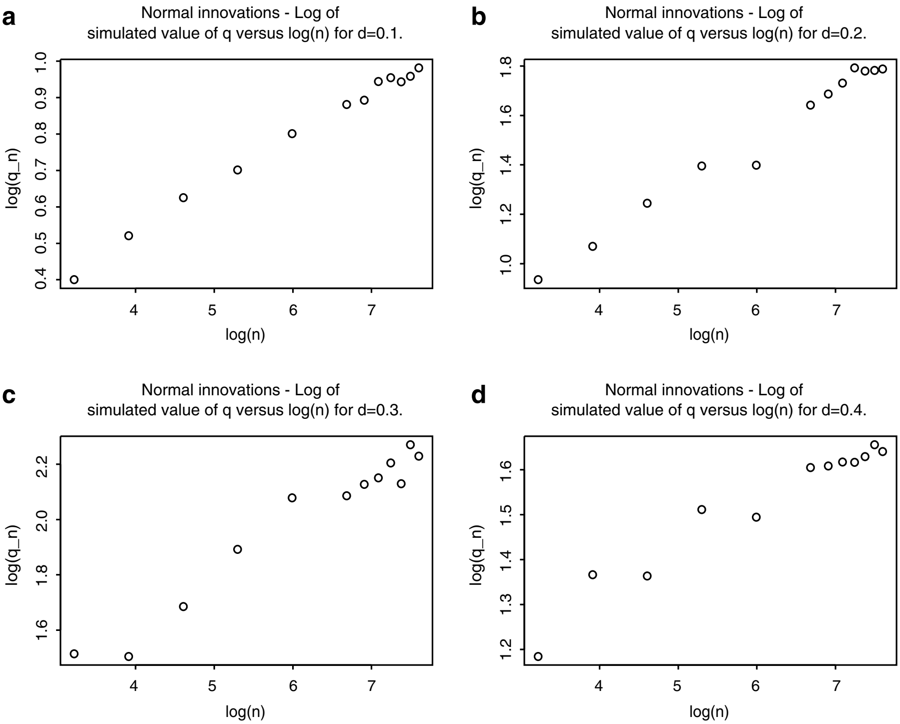

Fig. 1 Plot of $q_n = \mathrm{MSE}(\ell_{X,n})/\mathrm{MSE}(\hat{\ell}_X)$ versus $\log n$. For each value of $d = 0.1, 0.2, 0.3, 0.4$ and $n = 25, 50, 100, 200, 400, 800, 1{,}000, 1{,}200, 1{,}400, 1{,}600, 1{,}800$ and $2{,}000$, the results are based on 2,000 simulations of a fractional ARIMA$(0, d, 0)$ process with standard normal innovations ϵ_i

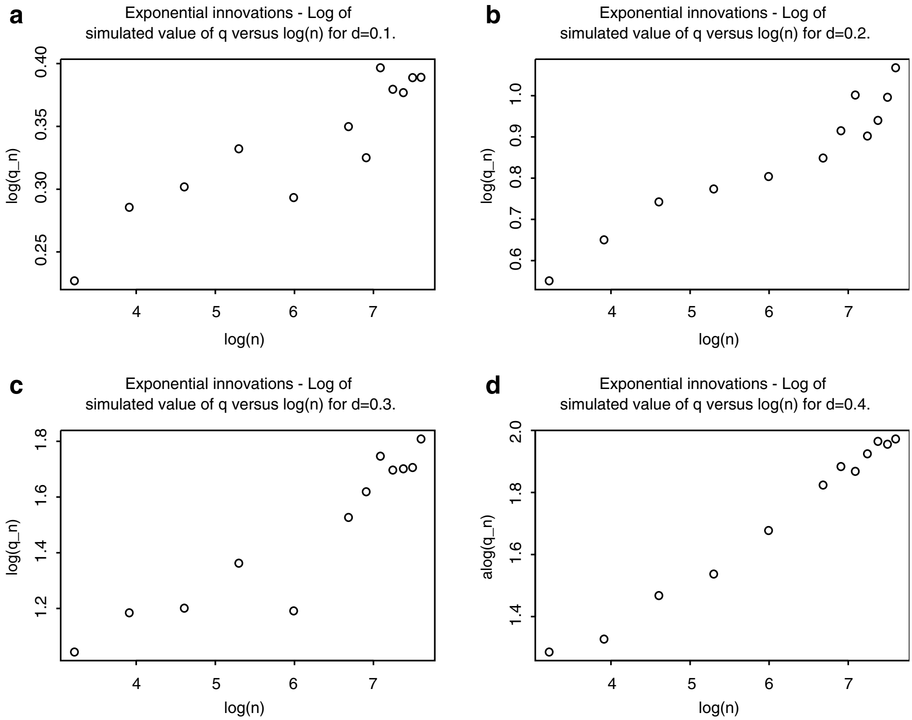

Fig. 2 Plot of $q_n = \mathrm{MSE}(\ell_{X,n})/\mathrm{MSE}(\hat{\ell}_X)$ versus $\log n$. For each value of $d = 0.1, 0.2, 0.3, 0.4$ and $n = 25, 50, 100, 200, 400, 800, 1{,}000, 1{,}200, 1{,}400, 1{,}600, 1{,}800$ and $2{,}000$, the results are based on 2,000 simulations of fractional ARIMA$(0, d, 0)$ process with innovations $\epsilon_i = \xi_i - 1/2$ and ξ_i independent exponential random variables with expected value $1/2$

Table 1 Least squares slopes δ_{sim} fitted to $\log q_n$ versus $\log n$ where $q_n = \mathrm{MSE}(\ell_{X,n})/\mathrm{MSE}(\hat{\ell}_X)$ and d = 0.1, 0.2, 0.3, 0.4. The theoretical asymptotic slopes are $\delta = d(1-2d)/(1-d)$. The simulated process is a fractional ARIMA$(0, d, 0)$ process of length n =1,000, 1,200, 1,400, 1,600, 1,800 and 2,000 respectively, with standard normal innovations ϵ_i

	δ_{sim}	δ
$d = 0.1$	0.10	0.09
$d = 0.2$	0.14	0.15
$d = 0.3$	0.16	0.17
$d = 0.4$	0.06	0.13

Table 2 Least squares slopes δ_{sim} fitted to $\log q_n$ versus $\log n$ where $q_n = \mathrm{MSE}(\ell_X)/\mathrm{MSE}(\hat{\ell}_{X,n})$ and $d = 0.1, 0.2, 0.3, 0.4$. The theoretical asymptotic slopes are $\delta = d(1-2d)/(1-d)$. The simulated process is a fractional ARIMA$(0, d, 0)$ process of length $n = 1{,}000, 1{,}200, 1{,}400, 1{,}600, 1{,}800$ and $2{,}000$ respectively, with innovations $\epsilon_i = (\xi_i - 1/2)$ and ξ_i independent exponential random variables with expected value $1/2$

	δ_{sim}	δ
$d = 0.1$	0.07	0.09
$d = 0.2$	0.15	0.15
$d = 0.3$	0.17	0.17
$d = 0.4$	0.15	0.13

Acknowledgements We would like to thank two referees for their constructive comments.

Appendix: Proofs

Proof of Proposition 1

By definition of $m_{X,n}$, we have

$$\operatorname{var}(m_{X,n}(t)) = n^{-1} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \gamma_e(k;t)$$

with $\gamma_e(k;t) = \gamma_e(-k;t)$, and for $k \geq 0$,

$$\begin{aligned}
\gamma_e(k;t) &= \operatorname{cov}(\mathrm{e}^{tX_1}, \mathrm{e}^{tX_{1+k}}) = E[\mathrm{e}^{t(X_1+X_{1+k})}] - m_X^2(t) \\
&= E\left[\exp\left\{t\sum_{j=0}^{\infty}(a_j + a_{j+k})\epsilon_{1-j}\right\}\right] E\left[\exp\left\{t\sum_{j=0}^{k-1} a_j\epsilon_{1+k-j}\right\}\right] - m_X^2(t) \\
&= \prod_{j=0}^{\infty} m_\epsilon(t(a_j + a_{j+k})) \prod_{j=0}^{k-1} m_\epsilon(ta_j) - m_X^2(t).
\end{aligned}$$

Now,

$$\begin{aligned}
\log\left\{\prod_{j=0}^{\infty} m_\epsilon(t(a_j + a_{j+k})) \prod_{j=0}^{k-1} m_\epsilon(ta_j)\right\} &= \sum_{j=0}^{\infty} \log\, m_\epsilon(t(a_j + a_{j+k})) \\
&\quad + \sum_{j=0}^{k-1} \log\, m_\epsilon(ta_j).
\end{aligned}$$

As $k \to \infty$, $a_{j+k} \to 0$ and

$$m_\epsilon(t(a_j + a_{j+k})) = m_\epsilon(ta_j) + m'_\epsilon(ta_j)ta_{j+k} + m''_\epsilon(ta_j)\frac{(ta_{j+k})^2}{2} + r_k(t)$$

with $|r_k| \leq c|ta_{j+k}|^2$ for some $c > 0$. Hence,

$$\sum_{j=0}^{\infty} \log\, m_\epsilon(t(a_j + a_{j+k})) = \sum_{j=0}^{\infty} \log\, m_\epsilon(ta_j) + \sum_{j=0}^{\infty} \frac{m'_\epsilon(ta_j)}{m_\epsilon(ta_j)} ta_{j+k} + R(t).$$

Since $a_j \to 0$ as $j \to \infty$, $|m'_\epsilon(ta_j)/m_\epsilon(ta_j)|$ is bounded from above by a constant M. Under assumption A1, it follows that

$$\left|\sum_{j=0}^{\infty} \frac{m'_\epsilon(ta_j)}{m_\epsilon(ta_j)} ta_{j+k}\right| \leq |t|M \sum_{j=0}^{\infty} \phi^{j+k} \leq D\phi^k$$

for a constant $0 < D < \infty$.

Also, noting that $m'_\epsilon(0) = 0$, we have

$$\sum_{j=0}^{k-1} \log m_\epsilon(ta_j) = \log m_X(t) - \frac{(\sigma_\epsilon t)^2}{2} \sum_{j=k}^{\infty} \{a_j^2 + u_j(t)\}$$

with $u_j(t) = o(a_j^2)$. Hence, under A1,

$$\left| \sum_{j=0}^{k-1} \log m_\epsilon(ta_j) - \log m_X(t) \right| \leq \text{const } \phi^k \frac{(\sigma_\epsilon t)^2}{2} \sum_{j=k}^{\infty} \{\phi^j + u_j(t)\}.$$

Putting these results together leads to $\gamma_e(k;t) \leq A\phi^k$ for some constant $0 < A < \infty$. It then follows by standard arguments that $v(t)$ exists, $0 < v < \infty$ and Eqs. (18), (19) and (20) hold. The result for $\ell_{X,n}$ follows by Taylor expansion.

Proof of Proposition 2

$$\begin{aligned}\gamma_e(k;t) &= \prod_{j=0}^{\infty} m_\varepsilon(t(a_j + a_{j+k})) \prod_{j=0}^{k-1} m_\varepsilon(ta_j) - m_X^2(t) \\ &= a(t,k) - m_X^2(t)\end{aligned}$$

Consider

$$b = \log a(t;k) = \sum_{j=0}^{\infty} \log m_\varepsilon(ta_j + ta_{j+k}) + \sum_{j=0}^{k-1} \log m_\varepsilon(ta_j)$$

Since $ta_{j+k} \to 0 \to$, as $k \to \infty$, and $m_\varepsilon \in C^2(-T,T)$ for some $T > 0$, we have

$$m_\varepsilon(ta_j + ta_{j+k}) = m_\varepsilon(ta_j)\left(1 + \frac{m'_\varepsilon(ta_j)}{m_\varepsilon(ta_j)} ta_{j+k} + r_{j,k}(t)\right)$$

where $|r_{j,k}(t)| \leq \sup_{|u| \leq M(t)} |m''(u)| t^2 a_j^2$ with $M(t) = \max |ta_j|$, and

$$\log m_\varepsilon(ta_j + ta_{j+k}) = \log m_\varepsilon(ta_j) + \frac{m'_\varepsilon(ta_j)}{m_\varepsilon(ta_j)} ta_{j+k} + r^*_{j,k}(t)$$

where $r_{j,k}(t) = o(a_{j+k})$ uniformly in j. Hence

$$\sum_{j=0}^{\infty} \log m_\varepsilon(ta_j + ta_{j+k}) = \sum_{j=0}^{\infty} \log m_\varepsilon(ta_j) + \sum_{j=0}^{\infty} \frac{m'_\varepsilon(ta_j)}{m_\varepsilon(ta_j)} ta_{j+k} + R_k(t)$$

where R_k is of smaller order than the other terms in the equation. Furthermore,

$$m_\varepsilon(ta_j) = 1 + \frac{1}{2} m''_\varepsilon(0)(ta_j)^2 + u_j(t) = 1 + \frac{\sigma_\varepsilon^2}{2} t^2 a_j^2 + u_j(t)$$

and

$$m'_\varepsilon(ta_j) = \sigma_\varepsilon^2 ta_j + u_j^*(t)$$

with $u_j = o(a_j^2)$ and $u_j^* = o(a_j)$. By assumption A2, $a_{j+k} = C(j+k)^{d-1} + o(k^{d-1})$, as $k \to \infty$, so that

$$\begin{aligned}
\sum_{j=0}^{\infty} \log m_\varepsilon(ta_j + ta_{j+k}) &= \log m_X(t) + \sigma_\varepsilon^2 t^2 \sum_{j=0}^{\infty} a_j a_{j+k} + R_k^*(t) \\
&= \log m_X(t) + \sigma_\varepsilon^2 t^2 C^2 \sum_{j=1}^{\infty} (j^{d-1} + o(j^{d-1}))(j+k)^{d-1} \\
&\quad + o(k^{2d-1}) \\
&= \log m_X(t) + k^{2d-1} \sigma_\varepsilon^2 t^2 C^2 \sum_{j=1}^{\infty} \left(\frac{j}{k}\right)^{d-1} \left(\frac{j}{k} + 1\right)^{d-1} \frac{1}{k} \\
&\quad + o(k^{2d-1}) \\
&= \log m_X(t) + A_1 k^{2d-1} + o(k^{2d-1})
\end{aligned}$$

with

$$A_1 = \sigma_\varepsilon^2 t^2 C^2 \int_o^\infty x^{d-1}(1+x)^{d-1} \mathrm{d}x.$$

Similarily,

$$\begin{aligned}
\sum_{j=0}^{k-1} \log m_\varepsilon(ta_j) &= \sum_{j=0}^{\infty} \log m_\varepsilon(ta_j) - \sum_{j=k}^{\infty} \log m_\varepsilon(ta_j) \\
&= \log m_X(t) - \sum_{j=k}^{\infty} \log m_\varepsilon(ta_j) \\
&= \log m_X(t) - \sum_{j=k}^{\infty} \frac{\sigma_\varepsilon^2}{2} t^2 a_j^2 + r_k \\
&= \log m_X(t) - \frac{\sigma_\varepsilon^2}{2} t^2 C^2 \sum_{j=k}^{\infty} j^{2d-2} + r_k^* \\
&= \log m_X(t) - k^{2d-1} \frac{\sigma_\varepsilon^2}{2} t^2 C^2 \int_1^\infty x^{2d-2} \mathrm{d}x + r_k^{**} \\
&= \log m_X(t) - A_2 k^{2d-1} + r_k^{**}
\end{aligned}$$

with

$$A_2 = \frac{\sigma_\varepsilon^2}{2(1-2d)} t^2 C^2$$

Putting the results together yields

$$b = 2 \log m_X(t) + (A_1 - A_2) k^{2d-1} + o(k^{2d-1}),$$

$$a = e^b = m_X^2(t)[1 + (A_1 - A_2)k^{2d-1}] + o(k^{2d-1})$$

and

$$\gamma_e(k;t) = Ak^{2d-1} + o(k^{2d-1})$$

with

$$A = C^2\sigma_\varepsilon^2 t^2 m_X^2(t)\left[\int_o^\infty x^{d-1}(1+x)^{d-1} - \frac{1}{2(1-2d)}\right]$$

As a result, we have (see e.g. Taqqu, 1975; Beran, 1994),

$$\mathrm{var}(m_{X,n}(t)) = \frac{A}{d(2d+1)}n^{2d-1} + o(n^{2d-1})$$

and

$$\mathrm{var}(\ell_{X,n}(t)) = \frac{A}{d(2d+1)}m_X^{-2}(t)n^{2d-1} + o(n^{2d-1}).$$

Proof of Theorem 1

From

$$B_{\varepsilon,n}(t) = E[\ell_{\varepsilon,n}(t)] - \ell_\varepsilon(t) = m_\varepsilon^{-2}(t)\mathrm{var}(m_{\varepsilon,n}(t)) + o\{\mathrm{var}(m_{\varepsilon,n}(t))\}$$

we have

$$\begin{aligned} E[\hat{\ell}_X(t)] &= \sum_{j=0}^{N_n} E[\ell_{\varepsilon,n}(a_j t)] = \sum_{j=0}^{N_n}\{\ell_\varepsilon(a_j t) + B_{\varepsilon,n}(a_j t)\} \\ &= \sum_{j=0}^{\infty}\ell_\varepsilon(a_j t) - \sum_{j=N_n+1}^{\infty}\ell_{\varepsilon,n}(a_j t) + \sum_{j=0}^{N_n} B_{\varepsilon,j}(a_j t) \end{aligned}$$

and

$$\begin{aligned} \mathrm{Bias} = E[\hat{\ell}_X(t)] - \ell_X(t) &= -\sum_{j=N_n+1}^{\infty}\ell_\varepsilon(a_j t) + \sum_{j=0}^{N_n} B_{\varepsilon,j}(a_j t) \\ &= -\sum_{j=N_n+1}^{\infty}\ell_\varepsilon(a_j t) + n^{-1}\sum_{j=0}^{N_n} m_\varepsilon^{-2}(a_j t)\{m_\varepsilon(2a_j t) - m_\varepsilon^2(a_j t)\} + r(N_n, n). \end{aligned}$$

Since $a_j \to 0$ ($j \to \infty$), we may replace m_ε by its Taylor expansion around zero so that

$$\begin{aligned} m_\varepsilon^{-2}(a_j t)\{m_\varepsilon(2a_j t) - m_\varepsilon^2(a_j t)\} &= 1 - 1 + \frac{4\sigma_\varepsilon^2 a_j^2 t^2}{2} - 2\frac{\sigma_\varepsilon^2 a_j^2 t^2}{2} + o(a_j^2) \\ &= \sigma_\varepsilon^2 a_j^2 t^2 + o(a_j^2 t^2) \end{aligned}$$

Since $\sum_{j=0}^{\infty} a_j^2 < \infty$, we have

$$\left| n^{-1}\sum_{j=0}^{N_n} m_\varepsilon^{-2}(a_j t)\{m_\varepsilon(2a_j t) - m_\varepsilon^2(a_j t)\}\right| \le \mathrm{const}\, n^{-1}$$

Moreover,

$$\ell_\varepsilon(a_j t) = \frac{\sigma_\varepsilon^2 t^2}{2} a_j^2 + o(a_j^2 t^2)$$

so that

$$-\sum_{j=N_n+1}^{\infty} \ell_\varepsilon(t) = -\frac{\sigma_\varepsilon^2 t^2}{2} \sum_{j=N_n+1}^{\infty} \{a_j^2 + o(a_j^2 t^2)\}.$$

Under A1, $a_j^2 \leq \varphi^{2j}$, which leads to the upper bound

$$\left| \sum_{j=N_n+1}^{\infty} \ell_\varepsilon(a_j t) \right| \leq \text{const} \sum_{j=N_n+1}^{\infty} \varphi^{2j} \sim \text{const}\ \varphi^{2N_n+1}$$

Putting these results together leads to the upper bound for the bias of the order $O(\max(n^{-1}, \varphi^{2N_n}))$.

Proof of Theorem 2
As above,

$$\left| n^{-1} \sum_{j=0}^{N_n} m_\varepsilon^{-2}(a_j t)\{m_\varepsilon(2a_j t) - m_\varepsilon^2(a_j t)\} \right| \leq const\ n^{-1}$$

However, for the first part of the bias term in Eq. (50) we obtain a different order than in Theorem 1: Under A2, $a_j \sim C j^{d-1}$. Hence

$$\begin{aligned}
\sum_{j=N_n+1}^{\infty} \ell_\varepsilon(t) &= \frac{\sigma_\varepsilon^2 t^2}{2} \sum_{j=N_n+1}^{\infty} \{a_j^2 + o(a_j^2 t^2)\} = \frac{C^2 \sigma_\varepsilon^2 t^2}{2} \sum_{j=N_n+1}^{\infty} \{j^{2d-2} + o(j^{2d-2})\} \\
&= N_n^{2d-1} \frac{C^2 \sigma_\varepsilon^2 t^2}{2} \sum_{j=N_n+1}^{\infty} \left(\frac{j}{N_n}\right)^{2d-2} \frac{1}{N_n} + o(N_n^{2d-1}) \\
&= N_n^{2d-1} \frac{C^2 \sigma_\varepsilon^2 t^2}{2} \int_1^{\infty} x^{2d-2} \mathrm{d}x + o(N_n^{2d-1}) \\
&= N_n^{2d-1} \frac{C^2 \sigma_\varepsilon^2 t^2}{2(1-2d)} + o(N_n^{2d-1})
\end{aligned}$$

As a result, we have

$$\text{Bias} = E[\hat{\ell}_X(t)] - \ell_X(t) = -N_n^{2d-1} \frac{C^2 \sigma_\varepsilon^2 t^2}{2(1-2d)} + o(N_n^{2d-1})$$

Proof of Theorem 3

Using Lemma 1 we have

$$\mathrm{var}(\hat{\ell}_X(t)) = \mathrm{var}\left(\sum_{j=0}^{N_n} \ell_{\varepsilon,n}(a_j t)\right) = \sum_{i,j=0}^{N_n} \mathrm{cov}(\ell_{\varepsilon,n}(a_i t), \ell_{\varepsilon,n}(a_j t))$$
$$= n^{-1} \sum_{i,j=0}^{N_n} \left\{ \frac{m_\varepsilon(t(a_i + a_j))}{m_\varepsilon(ta_i) m_\varepsilon(ta_j)} - 1 \right\} + r(n, N_n)$$

where $r(n, N_n)$ is of smaller order than the first term in the last expression.

Proof of Theorem 4

Under A1,

$$\left| \sum_{i,j=0}^{N_n} \left\{ \frac{m_\varepsilon(t(a_i + a_j))}{m_\varepsilon(ta_i) m_\varepsilon(ta_j)} - 1 \right\} \right| \leq t^2 \sum_{i,j=0}^{N_n} |a_i||a_j| \leq \text{const}\, t^2 \cdot \left(\sum_{j=0}^{\infty} \varphi^j \right)^2 < \infty$$

Proof of Theorem 5

Note that for $t, s \to 0$, $m_\varepsilon(t + s)/[m_\varepsilon(t) m_\varepsilon(s)] = 1 + \sigma_\epsilon^2 ts + o(ts)$. Then, under A2

$$\sum_{i,j=0}^{N_n} \left\{ \frac{m_\varepsilon(t(a_i + a_j))}{m_\varepsilon(ta_i) m_\varepsilon(ta_j)} - 1 \right\} = C^2 t^2 \sigma_\epsilon^2 \textstyle\sum_{i,j=1}^{N_n} i^{d-1} j^{d-1} + r(n, N_n)$$

and

$$C^2 t^2 \sigma_\epsilon^2 \sum_{i,j=1}^{N_n} i^{d-1} j^{d-1} = N_n^{2d} C^2 t^2 \sigma_\epsilon^2 \sum_{i,j=1}^{N_n} \left(\frac{i}{N_n}\right)^{d-1} \left(\frac{j}{N_n}\right)^{d-1} \frac{1}{N_n} \frac{1}{N_n}$$
$$= N_n^{2d} C^2 t^2 \sigma_\epsilon^2 \left(\int_o^1 x^{d-1} \mathrm{d}x \right)^2 + r^*(n, N_n)$$
$$= N_n^{2d} C^2 t^2 \sigma_\epsilon^2 d^{-2} + r^*(n, N_n)$$

Proof of Corollary 2

The optimal value of α is obtained by plugging $N_n = \beta n^\alpha$ in Theorem 7, and setting the derivative with respect to α and β equal to zero.

Proof of Corollary 3

Let $N_n = \beta n^\alpha$. Then $\mathrm{MSE}(\hat{\ell}_X) = o\{\mathrm{MSE}(\ell_{X,n})\}$ if and only if $\alpha(4d - 2) < 2d - 1$ and $2\alpha d - 1 < 2d - 1$. This is equivalent to $1/2 < \alpha < 1$.

References

Beran, J. (1991). M-estimators of location for data with slowly decaying serial correlations. *Journal of the Americal Statistical Association*, *86*, 704–708.

Beran, J. (1994). *Statistics for long-memory processes*. London: Chapman and Hall.

Brockwell, P. J., Davis, R. A. (1987). *Time series: Theory and methods*. Berlin Heidelberg New York: Springer.
Csörgő, S. (1981). Limit behaviour of the empirical characteristic function. *The Annals of Probability*, *9*, 130–144.
Csörgő, S. (1982). The empirical moment generating function. In: B. V. Gnedenko, M. L. Puri, I. Vincze (Eds.). *Nonparametric statistical inference*. London: North-Holland, 139–150.
Csörgő, S. (1986). Testing for normality in arbitrary dimension. *The Annals of Statistics*, *14*, 708–723.
Csörgő, S., Mielniczuk, J. (1995). Density estimation under long-range dependence. *The Annals of Statistics*, *23*, 990–999.
Dehling, H., Taqqu, M. S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *The Annals of Statistics*, *17*, 1767–1783.
Engel, J., Herrmann, E., Gasser, T. (1994). An iterative bandwidth selector for kernel estimation of densities and their derivatives. *Journal of Nonparametric Statistics*, *4*, 21–34.
Feuerverger, A., Mureika, R. A. (1977). The empirical characteristic function and its applications. *The Annals of Statistics*, *5*, 88–97.
Ghosh, S. (1996). A new graphical tool to detect non-normality. *Journal of the Royal Statistical Society B*, *58*, 691–702.
Ghosh, S. (1999). T3-plot. *Encyclopedia for statistical sciences*. Update Vol. 3 (pp 739–744). New York: Wiley
Ghosh, S. (2003). Estimating the moment generating function of a linear process. *Student*, *4*(3), 211–218.
Ghosh, S., Beran, J. (2000). Two sample T3-plot: A graphical comparison of two distributions. *Journal of Computational and Graphical Statistics*, *9*(1), 167–179.
Ghosh, S., Ruymgaart, F. (1992). Applications of empirical characteristic functions in some multivariate problems. *The Canadian Journal of Statistics*, *20*, 429–440.
Giraitis, L., Surgailis, D. (1985). Central limit theorems and other limit theorems for functionals of Gaussian processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *70*, 191–212.
Giraitis, L., Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and applications to asymptotical normality of Whittle's estimate. *Probability Theory and Related Fields*, *86*, 87–104.
Giraitis, L., Surgailis, D. (1999). Central limit theorem for the empirical process of a linear sequence with long memory. *Journal of Statistical Planning and Inference*, *80*, 81–93.
Granger, C. W. J., Joyeux, R. (1980). An introduction to long-range time series models and fractional differencing. *Journal of Time Series Analysis*, *1*, 15–29.
Hall, P. (1993). On plug-in rules for local smoothing of density estimators. *The Annals of Statistics*, *21*, 694–710.
Hall, P., Hart, J. (1990). Convergence rates in density estimation for data from infinite-order moving average processes. *Probability Theory and Related Fields*, *87*, 253-274.
Haslett, J., Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. Invited paper with discussion. *Applied Statistics*, *38*, 1–50.
Honda, T. (2000). Nonparametric density estimation for a long-range dependent linear process. *Annals of the Institute of Statistical Mathematics*, *52*, 599–611.
Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, *68*, 165–176.
Mandelbrot, B. B., Wallis, J. R. (1969). Computer experiments with fractional Gaussian noises. *Water Resources Research*, *5*, 228–267.
Murota, K., Takeuchi, K. (1981). The studentized empirical characteristic function and its application to test for the shape of the distribution. *Biometrika*, *68*, 55–65.
Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.
Taqqu, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *31*, 287–302.
Wu, W. B., Mielniczuk (2002). Kernel density estimation for linear processes. *The Annals of Statistics*, *30*(5), 1441–1459.