Biometrika (1997), **84**, 4, *pp.* 791–802
Printed in Great Britain

Bandwidth selection for kernel regression with long-range dependent errors

By BONNIE K. RAY
Department of Mathematics and Center for Applied Mathematics and Statistics, New Jersey Institute of Technology, Newark, New Jersey 07102, U.S.A.
e-mail: borayx@m.njit.edu

AND RUEY S. TSAY
Graduate School of Business, University of Chicago, Chicago, Illinois 60637, U.S.A.
e-mail: rst@gsbrst.uchicago.edu

Summary

We investigate the effect of long-range dependence on bandwidth selection for kernel regression with the plug-in method of Herrmann, Gasser & Kneip (1992). A new bandwidth estimator is proposed to allow for long-range dependence. Properties of the proposed estimator are investigated theoretically and via simulation. We find that the proposed estimator performs well in terms of integrated squared error of the estimated trend, allowing us to incorporate both deterministic nonlinear features having an unknown structure and long-range dependence into a single model. The method is illustrated using biweekly measurements of the volume of the Great Salt Lake.

Some key words: Kernel smoothing; Long memory; Plug-in method.

1. Introduction

We consider fixed-design regression, under which the data $x_1, \ldots, x_n$ follow the model

$$x_i = g\left(\frac{i-\frac{1}{2}}{n}\right) + \varepsilon_i \quad (i = 1, \ldots, n), \tag{1}$$

where g is a smooth function defined on $[0, 1]$ and $\{\varepsilon_i\}$ is a zero-mean, covariance stationary process. The goal here is to estimate the trend function g nonparametrically. If the error terms, ε_i, are correlated but the serial correlation decays sufficiently fast, that is the errors are short-range dependent, kernel estimates of g of the form

$$\hat{g}_h(x) = h^{-1} \sum_{i=1}^{n} x_i \int_{(i-1)/n}^{i/n} K\left(\frac{x-u}{h}\right) du \tag{2}$$

converge to the true, unknown trend function at the same rate as in the case of uncorrelated errors (Hall & Hart, 1990). Here, h is called the bandwidth and K is a kernel function having specified properties.

A major problem associated with nonparametric trend estimation in (2) involves the selection of the bandwidth h. Various data-adaptive methods for selecting h have been proposed but have been shown to fare poorly when the data are sufficiently positively correlated (Diggle & Hutchinson, 1989; Hart, 1991). Modifications have been proposed

Downloaded from https://academic.oup.com/biomet/article-abstract/84/4/791/1745390 by University Library of Paderborn user on 06 November 2019

to account for short-range dependence (Altman, 1990; Chiu, 1989; Hart, 1994; Herrmann et al., 1992).

On the other hand, many real-world series, such as those observed in oceanography, hydrology and telecommunications, exhibit serial dependence which decays so slowly that the covariance function of the errors is not summable; in the frequency domain, the spectrum is infinite at frequency zero. These processes are termed long-range dependent; see Beran (1994) for a review. When using nonparametric regression to estimate a trend in the presence of long-range dependent errors, Hall & Hart (1990) show that kernel estimates converge at a slower rate than in the case of uncorrelated or short-range dependent errors. Csörgő & Mielniczuk (1995) show the finite-dimensional limiting distribution of (2) under some proper normalisation. Cheng & Robinson (1991) study kernel density estimation for a purely stochastic process which is long-range dependent. In that case, the effect of the long-range dependence on convergence of the kernel estimate depends on the functional estimated. Ray & Tsay (1996) provide simulation results concerning an automatic bandwidth selection procedure for the fixed-design case when the error process is long-range dependent.

Nonparametric trend estimation with long-range dependent error terms is concerned with incorporating the nonstandard features of nonlinearity and long-range dependence into a single model. Both of these features have been observed in practice, especially in hydrological and meteorological applications (Haslett & Raftery, 1989; Lewis & Ray, 1993; Lall, Sangoyomi & Abarbanel, 1996). Here we modify the iterative method of bandwidth estimation, used by Herrmann et al. (1992) in the presence of short-range dependent errors. The iterative method is chosen because of its relative computational efficiency. Section 2 gives the form of the asymptotically optimal bandwidth and an algorithm for estimating it. In § 3, we state some asymptotic properties of the estimate. Section 4 gives results of a small simulation study. Section 5 illustrates the estimation procedure and resulting trend estimate for a series of biweekly measurements of the Great Salt Lake volume from 1847–1992. This series has been found to exhibit nonlinear structure, as well as behaviour suggestive of long-range dependence (Lall et al., 1996). Proofs are given in an Appendix.

2. ITERATIVE BANDWIDTH ESTIMATION

2·1. *Optimal bandwidth estimation with short-range dependent errors*

Let $\gamma_\varepsilon(k)$ denote the lag-k autocovariance of the noise component ε_t in model (1). When the noise autocovariances are summable, Hall & Hart (1990) show that the bandwidth, h_{opt}, minimising the mean integrated squared error,

$$\text{MISE}\,(h) = E\left[\int_0^1 v(x)\{g(x) - \hat{g}_h(x)\}^2\,dx\right],$$

where v is a p times continuously differentiable function with support $[\delta, 1-\delta]$ and $v(x) > 0$ for all $x \in (\delta, 1-\delta)$ and some $\delta > 0$, is, asymptotically, given by

$$h_{\text{opt}} = \left(C_1 S_\varepsilon(0) \bigg/ \left[nC_2^2 \int_0^1 v(x)\{g''(x)\}^2\,dx\right]\right)^{1/(2p+1)} \tag{3}$$

for a kernel of order p, where $S_\varepsilon(0)$ denotes the spectral density of the noise process ε_t

Downloaded from https://academic.oup.com/biomet/article-abstract/84/4/791/1745390 by University Library of Paderborn user on 06 November 2019

evaluated at frequency zero,

$$C_1 = \int_0^1 v(t)\,dt \int \{K(x)\}^2\,dx, \quad C_2 = \int x^2 K(x)\,dx$$

and $g''(x)$ denotes the second derivative of g. In practice, $p=2$ is often used. Herrmann et al. (1992) estimate h_{opt} by plugging estimates of $S_\varepsilon(0)$ and $g''(x)$ into equation (3). They estimate $S_\varepsilon(0)$ using a function of the sample autocovariances of the original data differenced an appropriate amount. The quantity $S_\varepsilon(0)$ may also be estimated using the sample autocovariances of the estimated noise component at each iteration. The function $g''(x)$ is estimated using a kernel estimator with a kernel that obeys special properties and a bandwidth that is a modification of the bandwidth estimated at the previous iteration. The method fails asymptotically for long-range dependent errors, however, because $S_\varepsilon(0)$ is infinite. In § 2·2 we propose a modified method.

2·2. *A modified plug-in bandwidth estimator*

When the noise process in model (1) has an autocovariance function which decays as $C_3 k^{-\alpha}$, where $0<\alpha<1$ and C_3 is a positive real number, Hall & Hart (1990) show that the bandwidth, h'_{opt}, minimising the asymptotic mean integrated squared error is

$$h'_{\text{opt}} = \left(C_3 \alpha C_4 \Big/ \left[n^\alpha C_2^2 \int_0^1 v(x)\{g''(x)\}^2\,dx \right] \right)^{1/(2p+\alpha)}, \tag{4}$$

where C_2 is as in (3) and

$$C_4 = \int_0^1 v(t)\,dt \iint |x-y|^{-\alpha} K(x) K(y)\,dx\,dy.$$

To estimate h'_{opt} in the presence of long-range dependent errors, we estimate the unknown terms in (4). This requires estimating α and C_3, as well as $g''(x)$, although a full parametric model for the noise component is not necessarily required. We summarise one approach below and investigate its performance in estimating α and C_3 for the noise process.

2·3. *Estimation of long-range dependence parameters*

There are several approaches one may take to estimate α and C_3. The noise component ε_i can be estimated by using a specified bandwidth, h, to obtain an initial estimate of g and then taking

$$\hat{\varepsilon}_i = x_i - \hat{g}_h\{(i-\tfrac{1}{2})/n\} \quad (i=1,\ldots,n).$$

The $\hat{\varepsilon}_i$ are then used to estimate α and C_3. Since we wish to be as nonparametric as possible, we use a method that avoids specification of any possible short-memory behaviour of ε_i.

We estimate α using the method proposed by Geweke & Porter-Hudak (1983) and modified by Robinson (1994). We compute the periodogram of the process at M initial Fourier frequencies, omitting the first L frequencies, and regress the logarithm of the periodogram against the logarithm of the frequency. The estimated regression coefficient equals $\hat{\alpha}-1$. Typically, $L=1$ or 2 and $M=[n^{0\cdot5}]$, where $[.]$ denotes the greatest integer function. The estimated regression constant c may be thought of as the evaluation at $\omega=0$ of a function $L(\omega)$ that describes the short memory behaviour of the series and is slowly

Downloaded from https://academic.oup.com/biomet/article-abstract/84/4/791/1745390 by University Library of Paderborn user on 06 November 2019

varying at low frequencies. For example, for a fractionally integrated autoregressive moving-average (ARFIMA) (r, d, q) process

$$(1-\phi_1 B - \ldots - \phi_r B^r)(1-B)^d z_t = (1-\theta_1 B - \ldots - \theta_q B^q)a_t,$$

which has long memory when $0<d<\frac{1}{2}$ $(\alpha = 1-2d)$, we have $c=\{\sigma_a^2|\theta(1)|^2\}/\{2\pi|\phi(1)|^2\}$, where σ_a^2 is the variance of the innovation a_t. The ARFIMA process has

$$\gamma_z(k) \simeq \frac{\sigma_a^2|\theta(1)|^2\Gamma(1-2d)k^{2d-1}}{|\phi(1)|^2\Gamma(d)\Gamma(1-d)} \simeq C_3 k^{2d-1}$$

for large k (Hosking, 1981). Thus

$$C_3 = 2\pi c \frac{\Gamma(\alpha)}{\Gamma(\frac{1}{2}-\alpha/2)\Gamma(\frac{1}{2}+\alpha/2)}. \tag{5}$$

For other semiparametric methods of estimating C and α, see Delgado & Robinson (1994) and Robinson (1995).

As Hart (1991) points out, using crossvalidation to obtain an initial bandwidth will result in poor estimates of g and hence in poor initial estimates of ε_i. We mitigate this problem by updating $\hat{\varepsilon}_i$ and the estimates of α and C_3 at each iteration.

2·4. *Algorithm for iterative estimation of the optimal bandwidth under long-range dependence*

ALGORITHM. *In practice, we propose the following iterative procedure.*

Step 1. *Estimate an 'optimal' bandwidth, $\hat{h}_{\text{opt}}$, assuming only short-range dependent errors. For example, use the method of Herrmann et al.* (1992) *with a relatively small number of sample autocovariances used to estimate $S_\varepsilon(0)$ in equation* (3).

Step 2. *Let $h_0' = \hat{h}_{\text{opt}}$.*

Step 3. *For $j=1, \ldots$ estimate g using h_{j-1}' and let $\hat{\varepsilon}_i = x_i - \hat{g}\{(i-\frac{1}{2})/n\}$. Estimate α and C_3 using the log periodogram regression method applied to $\hat{\varepsilon}_i$. Evaluate C_4 using the estimated α.*

Step 4. *Estimate g'' by*

$$\hat{g}_2(t, h_2) = \sum_{i=1}^{n} x_i \int_{(i-1)/n}^{i/n} \frac{1}{h_2^3} K_2\left(\frac{t-u}{h_2}\right) du,$$

where $h_2 = h_{j-1}' n^{\hat{\alpha}/\{2(2p+\hat{\alpha})\}}$, with p defined before equation (3) *and K_2 a kernel having specified properties. The multiplying factor $n^{\hat{\alpha}/\{2(2p+\hat{\alpha})\}}$ of h_{j-1}' used to obtain h_2 inflates h_{j-1}' so as to provide a consistent estimator of h; see proof of Lemma* A2, *part* (i).

Step 5. *Set*

$$h_j' = \left(\hat{C}_3\hat{\alpha}\hat{C}_4 \middle/ \left[n^{\hat{\alpha}} C_2^2 \int_0^1 v(x)\{\hat{g}_2(x, h_{j-1}' n)^{\hat{\alpha}/\{2(2p+\hat{\alpha})\}}\}^2\, dx\right]\right)^{1/(2p+\hat{\alpha})}.$$

Step 6. *Advance j by* 1 *and repeat Steps* 3–5 *until convergence is reached, at some $\hat{h}_{\text{opt}}'$.*

Alternatively, α and C_3 may be estimated using the twice-differenced data, as suggested in Hart (1991). The above algorithm is then modified by beginning at Step 2, with $h_0' = n^{-\hat{\alpha}}$. Estimates of α and C_3 are not updated at each iteration. Theoretically, the number of iterations required to obtain an asymptotically efficient estimate is then $[2(2p+\alpha)/\alpha]$.

Downloaded from https://academic.oup.com/biomet/article-abstract/84/4/791/1745390 by University Library of Paderborn user on 06 November 2019

One way to determine if long-range dependence is present is to estimate the 'optimal' bandwidth assuming short-range dependence with different lags m used to estimate $S_\varepsilon(0)$; for example $m = [n^{0.25}]$, $[n^{0.50}]$, $[n^{0.75}]$. If long-range dependence is present, the estimated optimal bandwidth should increase as m increases for fixed n. Note that, in finite samples, continuing to increase m should result in very large bandwidths using the method of Herrmann et al. (1992) because $\hat{S}_\varepsilon(0)$ will continue to increase. If, however, a fixed m is used, asymptotically the bandwidth will be of the wrong order, i.e. too small.

3. Asymptotic results

Here we show that our estimated bandwidth is asymptotically equal to the bandwidth minimising the asymptotic MISE, provided that α and C_3 are estimated consistently. The proofs are given in the Appendix. The following assumptions are used.

Assumption 1. The design is equally spaced: $t_i = (i - \frac{1}{2})/n$ $(i = 1, \ldots, n)$.

Assumption 2. The errors ε_i follow a strictly stationary time series with mean zero, $E(|\varepsilon_i|^k) < \infty$ for all $k \in \mathcal{N}$, and serial covariances $\gamma_\varepsilon(k)$ which decay as $C_3 k^{-\alpha}$, where $0 < \alpha < 1$ and C_3 is a positive real number.

Assumption 3. The regression function g is p times differentiable, having

$$\max_{0<x<1} \max_{j=0,\ldots,p} |g^{(j)}(x)| \leqslant B < \infty, \quad |g^{(p)}(x) - g^{(p)}(y)| \leqslant L|x-y|^{p'-p}$$

for all $x, y \in [0, 1]$ and some $p' > 2$, $p \in \mathcal{N}$ with $p < p' \leqslant p+1$ and $L < \infty$. For at least one $l \in \{0, \ldots, p-1\}$ the derivative $g^{(l+1)}$ does not vanish in $[\delta, 1-\delta]$ and $g^{(l)}$ achieves an absolute maximum or absolute minimum in $[\delta, 1-\delta]$. Let $q = \min(2, p'-2)$.

Assumption 4. The kernel K is a Lipschitz continuous symmetric probability density with support $[-1, 1]$. The kernel function K_2 used for estimating g'' is assumed to be Lipschitz continuous, symmetric at zero with support $[-1, 1]$ and satisfies $\int K_2(x)\,dx = 0$ and $\int K_2(x)x^2\,dx = 2$.

These assumptions are similar to those of Herrmann et al. (1992) except for Assumption 2, which allows for nonsummable serial correlations for the noise component. Assumption 1 is taken for simplicity. Assumption 3 gives smoothness conditions for the trend function and ensures that the asymptotic formula (4) holds. Assumption 4 is commonly used and holds, for example, for optimal kernels (Gasser & Müller, 1979).

Let $\hat{\alpha}$ and $\hat{c}$ be the estimators of α and c obtained from $\hat{\varepsilon}_i = x_i - \hat{g}_h(t_i)$ for some bandwidth h, using the method of Robinson (1994) with $L = [l(n)]$, $M = [m(n)]$ and $L/M \to 0$, where $m(n) = bn^\lambda$ for some positive constant b and $0 < \lambda < 1$.

Proposition 1. *Suppose Assumptions* 2 *and* 3 *are satisfied and* h *is of order* $O(n^{-\alpha/(2p+\alpha)})$. *Then* $\hat{\alpha}$ *and* $\hat{c}$ *are consistent estimators of* α *and* c.

Remark. It follows that $\hat{C}_3$ and $\hat{C}_4$ are consistent estimators of C_3 and C_4.

Let h'_{MISE} be the mean integrated squared error optimal bandwidth and h'_{opt} be the asymptotic mean integrated squared error optimal bandwidth as in (4). The next proposition gives the rate of convergence of h'_{opt} to h'_{MISE}.

Proposition 2. *Suppose Assumptions* 1–4 *hold. Then* h'_{opt} *satisfies*

$$|h'_{\text{opt}} - h'_{\text{MISE}}| = O(n^{-(q+2\alpha)/\{2(2p+\alpha)\}}).$$

Downloaded from https://academic.oup.com/biomet/article-abstract/84/4/791/745390 by University Library of Paderborn user on 06 November 2019

Thus $\hat{h}'_{\text{opt}}$ converges to h'_{MISE} more slowly than when the noise is serially uncorrelated or has only short-range dependence.

The following theorem relates the estimated optimal bandwidth to the mean integrated squared error optimal bandwidth.

THEOREM 1. *If Assumptions* 1–4 *are satisfied and* $\hat{\alpha}$ *and* $\hat{c}$ *are consistent estimators of* α *and* c, *then the estimator* $\hat{h}'_{\text{opt}}$ *of* h'_{opt} *satisfies* $\hat{h}'_{\text{opt}} = h'_{\text{MISE}}\{1 + O(n^{-q\alpha/2(2p+\alpha)})\}$.

As noted by a referee, we assume that the covariance between points which are k sample points apart is fixed, even as n increases. Technically, this is not appropriate for processes which have correlation structure determined by spatial units rather than sampling units. We use it for convenience and to obtain some asymptotic justification. In practice, the sample size will always be finite and the relevant question is how good the approximation provided by an asymptotic result is. Under some regularity condition, our asymptotic results should provide adequate approximations. See Hart (1991, pp. 175–6) for a more thorough discussion of this point. Nonparametric regression estimators other than those given in (2), such as local linear estimators, have become popular in the last few years. The main ideas presented here should extend to other types of estimator.

4. SIMULATION RESULTS

To investigate the finite-sample performance of the modified iterative method, we generated series of length $n = 500$ and $n = 1000$ with trend functions

$$g_1(t) = 2 - 5t + 5\exp\{-100(t-0{\cdot}5)^2\}, \quad g_2(t) = 2\sin(8\pi t).$$

These are similar to two of the trend functions used by Herrmann et al. (1992). Larger sample sizes than used by them are needed to estimate long-range dependence properties adequately. In small samples, short-range dependent models have been found to approximate long-range dependent behaviour adequately (Crato & Ray, 1996). To each trend function we added four different ARFIMA $(r, d, 0)$ error processes, the first three having $r = 0$, $d = 0{\cdot}3, 0{\cdot}4, 0{\cdot}45$ $(\alpha = 0{\cdot}4, 0{\cdot}2, 0{\cdot}1)$ and the fourth having $r = 1$, $\phi = 0{\cdot}5$, $d = 0{\cdot}4$ $(\alpha = 0{\cdot}2)$. The ARFIMA errors were generated using the algorithm of Hosking (1984). The variance of the errors was fixed at 1·0 for g_1 and 1·5 for g_2. The number of replications was set at 200. We used

$$K(x) = \tfrac{3}{4}(1 - x^2) \quad (-1 < x < 1), \quad K_2(x) = \tfrac{15}{4}(3x^2 - 1) \quad (-1 < x < 1)$$

as our kernel functions. We let $v(x) = I_{[0{\cdot}1, 0{\cdot}9]}(x)$. For each series, three bandwidth estimates were considered, as follows.

Bandwidth estimate 1: h_{sm}, which is the bandwidth resulting from the Herrmann et al. (1992) iterative procedure with $\hat{S}(0) = \hat{\gamma}_\varepsilon(0) + 2\sum_{i=1}^{10}\hat{\gamma}_\varepsilon(i)$ computed using iterative estimates of the errors. Boundary points were discarded from the estimated residual series.

Bandwidth estimate 2: h_{lm}, which is the bandwidth resulting from the modified iterative procedure with α and C_3 estimated. The log periodogram regression method with $L = 2$ and $M = [n^{0{\cdot}5}]$ was used to estimate α and c from the estimated noise at each iteration, and relation (5) was used to obtain C_3. However, if $\hat{\alpha} \leqslant 0$, we set $\hat{\alpha} = 0{\cdot}01$ and, if $\hat{\alpha} \geqslant 1$, we set $\hat{\alpha} = 0{\cdot}99$. In such cases, the spectrum of the short-memory component at frequency zero was estimated using the fractionally differenced noise component.

Downloaded from https://academic.oup.com/biomet/article-abstract/84/4/791/1745390 by University Library of Paderborn user on 06 November 2019

Bandwidth estimate 3: h_*, which is the bandwidth minimising

$$\sum_{i=1}^{n} [g\{(i-\tfrac{1}{2})/n\} - \hat{g}\{(i-\tfrac{1}{2})/n\}]^2$$

over a grid of 100 equally spaced points between zero and 0·5. This is the empirically optimal bandwidth, but is not necessarily equal to h'_{opt}.

The number of iterations was fixed at 11 for obtaining h_{sm}. We found that 15 iterations were typically sufficient for the modified method.

Table 1 shows the simulation results. When $n=500$, using the Herrmann et al. (1992) method with $m=10$ results in estimated bandwidths smaller than the empirical optimal for all values of α. For relatively weak long memory ($\alpha=0{\cdot}4$), h_{lm} is close to the empirical optimal bandwidth for g_1 and is larger than the empirical optimal for g_2. For stronger long-range dependence ($\alpha=0{\cdot}2, 0{\cdot}1$) and no autoregressive component ($r=0$), the

Table 1. *Median estimated bandwidths, h, integrated squared errors,* ISE, *and relative integrated squared errors,* RISE, *for trend functions* $g_1(t)=2-5t+5\exp\{-100(t-0{\cdot}5)^2\}$ *and* $g_2(t)=2\sin(8\pi t)$; *replications* = 200

		$n=500$			$n=1000$		
	Method	h	ISE	RISE	h	ISE	RISE
				Trend function $g_1(t)$			
$\alpha=0{\cdot}4$	h_{sm}	0·057	0·138	1·026	0·054	0·109	1·023
	h_{lm}	0·068	0·137	1·026	0·061	0·107	1·022
	h_*	0·067	0·128	—	0·061	0·102	—
$\alpha=0{\cdot}2$	h_{sm}	0·052	0·302	1·034	0·050	0·241	1·023
	h_{lm}	0·066	0·292	1·022	0·061	0·238	1·017
	h_*	0·067	0·281	—	0·066	0·229	—
$\alpha=0{\cdot}1$	h_{sm}	0·048	0·385	1·025	0·045	0·381	1·026
	h_{lm}	0·062	0·377	1·010	0·057	0·380	1·013
	h_*	0·067	0·367	—	0·066	0·370	—
$\alpha=0{\cdot}2$, $\phi=0{\cdot}5$	h_{sm}	0·022	0·523	1·343	0·019	0·454	1·392
	h_{lm}	0·071	0·390	1·021	0·066	0·316	1·018
	h_*	0·077	0·362	—	0·071	0·306	—
				Trend function $g_2(t)$			
$\alpha=0{\cdot}4$	h_{sm}	0·040	0·254	1·024	0·038	0·205	1·023
	h_{lm}	0·054	0·254	1·041	0·048	0·205	1·022
	h_*	0·047	0·237	—	0·046	0·191	—
$\alpha=0{\cdot}2$	h_{sm}	0·038	0·466	1·025	0·036	0·404	1·026
	h_{lm}	0·052	0·460	1·016	0·049	0·389	1·010
	h_*	0·047	0·452	—	0·046	0·392	—
$\alpha=0{\cdot}1$	h_{sm}	0·035	0·631	1·025	0·033	0·597	1·026
	h_{lm}	0·048	0·617	1·012	0·044	0·581	1·010
	h_*	0·047	0·590	—	0·046	0·580	—
$\alpha=0{\cdot}2$, $\phi=0{\cdot}5$	h_{sm}	0·022	0·779	1·203	0·019	0·680	1·231
	h_{lm}	0·053	0·682	1·018	0·050	0·572	1·014
	h_*	0·052	0·623	—	0·051	0·545	—

Downloaded from https://academic.oup.com/biomet/article-abstract/84/4/791/174539O by University Library of Paderborn user on 06 November 2019

estimator obtained using the modified method is slightly larger than the empirical optimal for g_2 and smaller than the empirical optimal for g_1. When an autoregressive component is present, h_{sm} is extremely small. When $n = 1000$, the results are similar, although the estimated bandwidths have a smaller standard deviation.

Although the modified procedure produces bandwidths closer to optimal in most cases, the different bandwidths do not produce significantly different integrated squared errors or relative integrated squared errors for series of length $n = 500$. The relative integrated squared error of bandwidth h_a relative to h_* is defined as the median of $\text{ISE}(h_a)/\text{ISE}(h_*)$ over all replications, where $a = sm, lm$. The greatest improvements in integrated squared errors and relative integrated squared errors are seen for series of length $n = 1000$ having strong long-range dependence. Improvements in integrated squared errors will vary, however, dependent on the underlying signal-to-noise ratio of the function being estimated.

We also estimated the optimal bandwidth under weak long memory ($\alpha = 0{\cdot}6$), using a series of length $n = 5000$ and conducting 50 replications. In this case, the modified bandwidth estimator and the short-memory bandwidth estimator performed similarly. When the long memory is weak, short-range dependent models may be adequate to capture the correlation structure in finite samples.

The simulation results can be better understood in light of the following results. Altman (1990) shows that, for short-range dependent processes, the method of moments estimator of the correlation at lag k,

$$\hat{\rho}(k, h, n) = \frac{\sum_{i=[nh/2]}^{n+1-[nh/2]-k} \hat{\varepsilon}(i, h, n)\hat{\varepsilon}(i+k, h, n)}{\sum_{i=[nh/2]}^{n+1-[nh/2]-k} \hat{\varepsilon}^2(i, h, n)},$$

has expectation

$$E\hat{\rho}(k) \simeq \frac{\rho(k) + h^4(C_2/2)^2\{\int g''(t)^2\,dt\}/\sigma^2 + S_\varepsilon(0)(nh)^{-1}\{C_1 - 2K(0)\}}{1 + h^4(C_2/2)^2\{\int g''(t)^2\,dt\}/\sigma^2 + S_\varepsilon(0)(nh)^{-1}\{C_1 - 2K(0)\}}$$

as $h \to 0$ and $nh \to \infty$. For long-range dependence, we have $\alpha C_3\{C_4 - 2K(0)\}/(nh)^\alpha$ in place of $S_\varepsilon(0)\{C_1 - 2K(0)\}/(nh)$ in the above expression. Thus, for kernels such that $C_4 < 2K(0)$, correlation produces a negative bias, which may be offset by a positive bias dependent on the signal-to-noise ratio of the underlying trend and noise, measured by $\int g''(t)^2\,dt/\sigma^2$. Note that C_4 depends on α. In our simulations, $C_4 > 2K(0)$ when $\alpha = 0{\cdot}6$ and $C_4 < 2K(0)$ when $\alpha = 0{\cdot}4, 0{\cdot}2, 0{\cdot}1$. Additionally, the fact that the signal-to-noise ratio is larger for g_1 in our study manifests itself in a smaller negative bias in the estimated bandwidths for g_1 than for g_2 when $\alpha = 0{\cdot}4, 0{\cdot}2, 0{\cdot}1$.

To avoid the problem of bias in the estimated correlations using the estimated noise component, we tried using the twice differenced series with a cosine bell taper to estimate α and C_3 initially. This resulted in extremely variable estimates. In most cases, the estimate of α was negatively biased, resulting in an initial bandwidth which was too large. This, in turn, resulted in oversmoothed data and a bandwidth which increased at each step of the iterative procedure. We found that starting with a small bandwidth, even as small as n^{-1}, and iteratively updating the estimated noise component and the estimate of α worked well in all cases, although the number of iterations before convergence increased in a manner inversely proportional to the size of the initial bandwidth.

Downloaded from https://academic.oup.com/biomet/article-abstract/84/4/791/1745390 by University Library of Paderborn user on 06 November 2019

5. An application

Biweekly measurements of the Great Salt Lake volume have been investigated in detail over the last few years in an effort to understand the dynamics of the precipitous rise of the lake during 1983–1987 and its consequent rapid retreat. Such behaviour is typical of nonlinear systems driven by large scale, persistent, climatic fluctuations. Hydrologists are interested in determining the structure of the nonlinear behaviour in order to examine questions related to long-term climatic change. Biweekly measurements from 1847 to 1992 ($n = 3526$) are available. Figure 1 shows the periodogram, measured in decibels, of the series. There is obviously strong, low frequency behaviour in the series. We use kernel regression estimation with optimal bandwidth estimated assuming both short- and long-range dependent errors to estimate the trend nonparametrically.

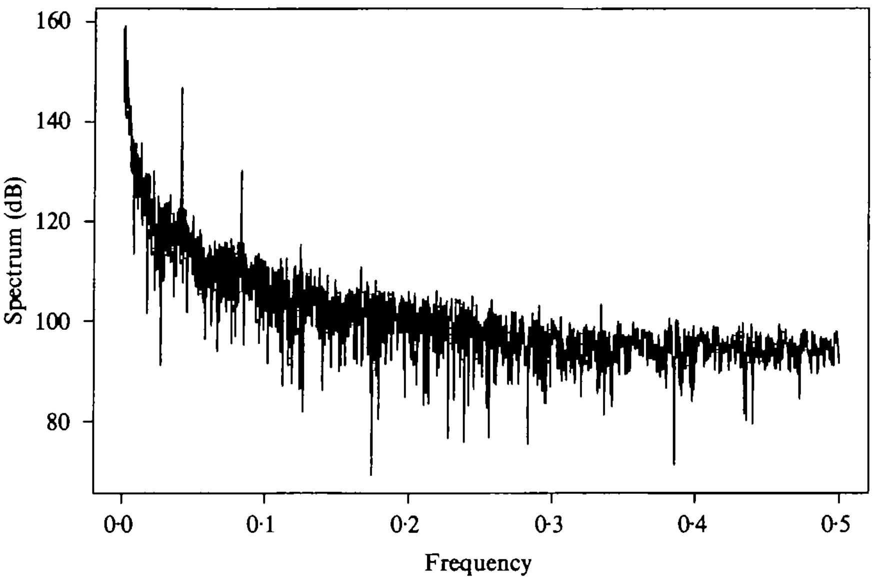

Fig. 1. Raw periodogram values (measured in dB) for biweekly measurements of the volume of the Great Salt Lake from 1847–1992.

Assuming short-range dependence and using $m = 10$ autocovariance estimates to compute $\hat{S}_\varepsilon(0)$ iteratively from the estimated errors, we obtain an estimated optimal bandwidth of approximately one month, so that almost no smoothing occurs. Using $m = 30$ and $m = 50$ autocovariance estimates, we obtain bandwidths of approximately 2 months and 19 months respectively. Thus the estimated optimal bandwidth increases as m increases, a likely indication of long-range dependence, as mentioned in § 2·4. Assuming long-range dependence and using the method proposed in § 2·4, we obtain an estimated optimal bandwidth of approximately 42 months. Figure 2(a) shows the raw data, along with the smoothed trend estimate obtained assuming long-range dependence. Since almost no smoothing is obtained under the short-memory assumption with $m = 30$, the Herrmann et al. (1992) method effectively reproduces the data. The smoothed series under the long memory assumption provides a much clearer view of the functional form of the nonlinear trend; note the presence of long-term quasi-periodic cycles in the data. The estimated value of α fluctuates between $\hat{\alpha} = 0{\cdot}9$ ($\hat{d} = 0{\cdot}05$) and $\hat{\alpha} = 0{\cdot}7$ ($\hat{d} = 0{\cdot}15$) at different iterations. This value of α is similar to the amount of persistence estimated in other hydrological

Downloaded from https://academic.oup.com/biomet/article-abstract/84/4/791/174539O by University Library of Paderborn user on 06 November 2019

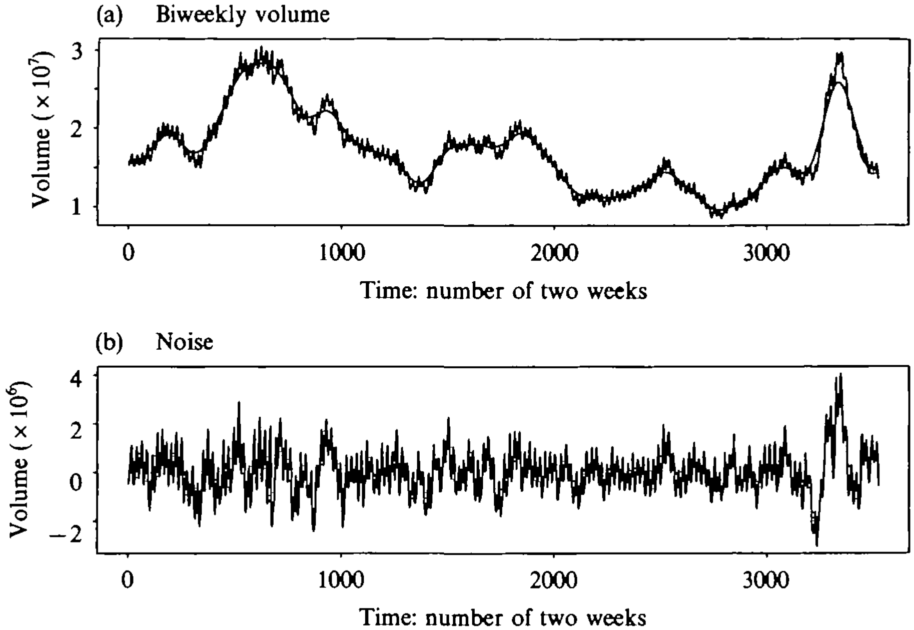

Fig. 2. (a) Raw and smoothed biweekly measurements of the volume of the Great Salt Lake from 1847–1992. The fluctuating line depicts the original data. The solid line depicts the smoothed series assuming long-range dependent errors. (b) Estimated noise component from the nonparametric fit to the data assuming long-range dependent errors.

series, such as the Nile data analysed by Hosking (1984). Figure 2(b) shows the estimated noise component obtained using the long-memory bandwidth. The series exhibits behaviour typical of a long-range dependent process, such as cycles and changes of level of all orders of magnitude (Hosking, 1984).

ACKNOWLEDGEMENT

The research of both authors was supported by the National Science Foundation. The authors wish to thank the referees for useful comments and Balaji Rajagopalan of Lamont-Doherty Earth Observatory for making the Great Salt Lake data available and for helpful discussions. The original source of the Great Salt Lake data was the United States Geological Survey.

APPENDIX

Proofs

Proof of Proposition 1. Let $\hat{\alpha}$ and $\hat{c}$ be the estimators of α and c obtained from $\hat{\varepsilon}_i = x_i - \hat{g}_h(t_i)$ for some bandwidth h, using the log periodogram regression method with $L = [l(n)]$ and $M = [m(n)]$ periodogram ordinates, where $L/M \rightarrow 0$, $m(n) = bn^{\lambda}$ for some positive constant b and $0 < \lambda < 1$. If ε_i are observed, $\hat{\alpha}$ and $\hat{c}$ are consistent estimators of α and c, having orders $O\{m(n)^{-\frac{1}{2}}\}$ and $O\{(\ln n)m(n)^{-\frac{1}{2}}\}$ respectively, using the results of Robinson (1994). From Theorem 2·2 and Remark (2) following Theorem 2·2 of Hall & Hart (1990), $\hat{\varepsilon}_i = x_i - \hat{g}_h(t_i)$ is a consistent estimator of ε_i. Thus, if $m(n)$ is selected such that $(\ln n)^{-1} n^{\lambda/2} < n^{p\alpha/(2p+\alpha)}$, $\hat{\alpha}$ and $\hat{c}$ are consistent estimators of α and c having order $O(n^{-p\alpha/(2p+\alpha)})$. □

Downloaded from https://academic.oup.com/biomet/article-abstract/84/4/791/1745390 by University Library of Paderborn user on 06 November 2019

Proof of Proposition 2. The proof follows that of Proposition 2 of Herrmann et al. (1992), requiring bounds on the bias and variance of $\hat{g}_h(t)$. The bounds on the bias are as in the short memory case. Modified upper and lower bounds for the variance may be obtained using the results of Hall & Hart (1990). From the proof of their Theorem 2·2, and using the Lipschitz continuity of K, we have

$$|\operatorname{var}\{\hat{g}_h(t)\} - C_3 C_4 (nh)^{-\alpha}| \leqslant c_1 (nh)^{-2\alpha},$$

for all $t \in [\delta, 1-\delta]$, $h \in [n^{-\alpha}, 1]$, where c_1 is a constant $<\infty$. Additionally, for a suitable constant $c_2 > 0$, we have $\operatorname{var}\{\hat{g}_h(t)\} \geqslant c_2 (nh)^{-\alpha}$. Thus MISE$(h)$ is minimised with the bandwidth $h'_{\text{MISE}} = h'_{\text{opt}} + O(n^{-(q+2\alpha)/\{2(2p+\alpha)\}})$. □

To prove Theorem 1, we need the following lemmas.

LEMMA A1. *Suppose Assumption 2 holds. Let $c(i)$, $a(i,j) \in \Re$ for all i ($j = 1, \ldots, n$) with $a(i,j) = 0$ if $|i-j| \geqslant k/2$ for some even integer k. Then*

$$E\left[\sum_{i,j} a(i,j)\{\varepsilon_i \varepsilon_j - \gamma_\varepsilon(i-j)\}\right]^{2s} \leqslant C_s n^{s\alpha} k^s \sup_{i,j} |a(i,j)|^{2s}, \quad E\left\{\sum_i \varepsilon_i c(i)\right\}^{2s} \leqslant C_s n^{2\alpha} \sup_i |c(i)|^{2s}$$

for all $n, s \in \mathcal{N}$ with a constant C_s that depends on s and the moments up to order s of the error series.

The proof follows the theory of cumulants.

LEMMA A2. *Let $H = [2(2p+\alpha)/\alpha]$ and*

$$\tilde{C} = \int_0^1 v(t)\, dt \int_{-1}^1 \{K_2(x)\}^2\, dx.$$

(i) *Suppose that $\tilde{h} = cn^{-\nu}\{1 + o_p(n^{-\beta})\}$ with $3\alpha/\{2(2p+\alpha)\} \leqslant \nu \leqslant \alpha(H-1)/\{2(2p+\alpha)\}$, $c > 0$, $0 \leqslant \beta \leqslant \alpha/\{2(2p+\alpha)\}$. Then*

$$\int_0^1 v(t)\{\hat{g}_2(t,\tilde{h})\}^2\, dt = \frac{\alpha C_3 \tilde{C}}{c^{(2p+\alpha)/\alpha}} n^{-1+(2p+\alpha)\nu/\alpha}\{1 + o_p(n^{-\beta})\}.$$

(ii) *Suppose that $\tilde{h} = cn^{-\nu}\{1 + o_p(n^{-\beta})\}$ with $\nu = \alpha/(2p+\alpha)$, $c > 0$, $0 \leqslant \beta < \alpha/\{2(2p+\alpha)\}$. Then*

$$\int_0^1 v(t)\{\hat{g}_2(t,\tilde{h})\}^2\, dt = \int_0^1 v(t)\{g''(t)\}^2\, dt + \frac{\alpha C_3 \tilde{C}}{c^{(2p+\alpha)/\alpha}} + o(1) + o_p(n^{-\beta}).$$

(iii) *Suppose that $\tilde{h} = cn^{-\nu}\{1 + o(1) + o_p(n^{-\beta})\}$ with $\nu = \alpha/\{2(2p+\alpha)\}$, $c > 0$, $\beta \geqslant 0$. Then*

$$\int_0^1 v(t)\{\hat{g}_2(t,\tilde{h})\}^2\, dt = \int_0^1 v(t)\{g''(t)\}^2\, dt + O(n^{-q\alpha/\{2(2p+\alpha)\}}) + O_p(n^{-(3+q)\alpha/\{2(2p+\alpha)\}} + n^{-\beta - q\alpha/\{2(2p+\alpha)\}}).$$

Proof. The proof follows as in Herrmann et al. (1992) by expressing $\int_0^1 v(t)\{\hat{g}_2(t,\tilde{h})\}^2\, dt$ as a sum of bias and variance components. Bounds on the components are obtained using the results of Lemma 1. □

Proof of Theorem 1. The proof of Theorem 1 follows as in Herrmann et al. (1992), with the following modifications.

(a) We start the iteration with $h_0 = n^{-\alpha}$. The first $H-2$ iteration steps reduce the order of h_i. Additional iteration steps up to $k > H - 1 + 2/q$ reduce the order of variability of h_i.
(b) We replace $\hat{S}_\varepsilon(0)$ by $\hat{C}_3 \hat{\alpha}$ and replace C_1 by $\hat{C}_4$.
(c) The orders of n are modified using the results of Lemma A2. □

Downloaded from https://academic.oup.com/biomet/article-abstract/84/4/791/1745390 by University Library of Paderborn user on 06 November 2019

REFERENCES

ALTMAN, N. S. (1990). Kernel smoothing of data with correlated errors. *J. Am. Statist. Assoc.* **85**, 749–59.

BERAN, J. (1994). *Statistics for Long Memory Processes.* New York: Chapman & Hall.

CHENG, B. & ROBINSON, P. (1991). Density estimation in strongly dependent non-linear time series. *Statist. Sinica* **1**, 335–59.

CHIU, S. T. (1989). Bandwidth selection for kernel estimate with correlated noise. *Statist. Prob. Lett.* **8**, 347–54.

CRATO, N. & RAY, B. (1996). Model selection and forecasting for long memory processes. *J. Forecasting* **15**, 107–25.

CSÖRGŐ, S. & MIELNICZUK, J. (1995). Nonparametric regression under long-range dependent normal errors. *Ann. Statist.* **23**, 1000–14.

DELGADO, M. A. & ROBINSON, P. M. (1994). New methods for the analysis of long-memory time series: Application to Spanish inflation. *J. Forecasting* **13**, 97–107.

DIGGLE, P. J. & HUTCHINSON, M. F. (1989). On spline smoothing with correlated errors. *Aust. J. Statist.* **31**, 166–82.

GASSER, T. & MÜLLER, H. G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation,* Ed. T. Gasser and M. Rosenblatt, pp. 23–68. New York: Springer.

GEWEKE, J. & PORTER-HUDAK, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.* **4**, 221–38.

HALL, P. & HART, J. (1990). Nonparametric regression with long-range dependence. *Stoch. Proces. Applic.* **36**, 339–51.

HART, J. D. (1991). Kernel regression estimation with time series errors. *J. R. Statist. Soc.* B **53**, 173–87.

HART, J. D. (1994). Automated kernel smoothing of dependent data by using time series cross-validation. *J. R. Statist. Soc.* B **56**, 529–42.

HASLETT, J. & RAFTERY, A. E. (1989). Space-time modelling with long-memory dependence: assessing Ireland's wind power resource (with Discussion). *Appl. Statist.* **38**, 1–50.

HERRMANN, E., GASSER, T. & KNEIP, A. (1992). Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika* **79**, 783–95.

HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68**, 165–76.

HOSKING, J. R. M. (1984). Modeling persistence in hydrological time series using fractional differencing. *Water Resour. Res.* **20**, 1898–908.

LALL, U., SANGOYOMI, T. & ABARBANEL, H. (1996). Nonlinear dynamics of the Great Salt Lake: Nonparametric short term forecasting. *Water Resour. Res.* **32**, 975–85.

LEWIS, P. A. W. & RAY, B. (1993). Nonlinear modeling of multivariate and categorical time series using Multivariate Adaptive Regression Splines. In *Dimension Estimation and Models*, Ed. H. Tong, pp. 136–69. Singapore: World Scientific.

RAY, B. & TSAY, R. (1996). Bandwidth selection for nonparametric regression with long-range dependent errors. In *Springer Lecture Notes in Statistics: Athens Conference on Applied Probability and Time Series,* **2**, Ed. P. Robinson and M. Rosenblatt, pp. 339–51. New York: Springer.

ROBINSON, P. (1994). Log-periodogram regression of time series with long-range dependence. *Ann. Statist.* **23**, 1048–72.

ROBINSON, P. (1995). A Gaussian semi-parametric estimator of the long-memory parameter. *Ann. Statist.* **24**, 1630–61.

[*Received January* 1996. *Revised September* 1996]

Downloaded from https://academic.oup.com/biomet/article-abstract/84/4/791/174539O by University Library of Paderborn user on 06 November 2019