Statistica Sinica 1(1991), 335-359

DENSITY ESTIMATION IN STRONGLY DEPENDENT NON-LINEAR TIME SERIES

Bing Cheng and P. M. Robinson

London School of Economics

Abstract: Smoothed nonparametric density estimates can be useful in analysing nonlinear time series. Their asymptotic properties in weakly dependent series, including limiting distributions and mean squared error, are known to be similar to those in independent series. Robinson (1987) found evidence that these properties may not hold in strongly dependent, or "long-memory" Gaussian time series. The present paper derives normal and non-normal limiting distributions in case of long-memory nonlinear series, provides a numerical comparison of integrated mean squared error, and reports estimates based on simulated series.

Key words and phrases: Density estimation, long memory time series, nonlinear time series, normal and non-normal limiting distributions, integrated mean squared error.

1. Introduction

In recent years there has been increasing interest in approaches to time series analysis that offer an alternative to linear autoregressive moving average (ARMA) modelling. Two aspects of the latter approach might be questioned.

Let y_t, $t = 1, 2, \dots$, be a real-valued, strictly stationary time series, with $E(y_1^2) < \infty$. Gaussian ARMA models assume that $E(y_t|y_s, s < t)$ is linear in $(y_s, s < t)$. Such linearity is a consequence of Gaussian y_t. But it need not be a natural assumption in case of non-Gaussian y_t. Moreover, the usual estimates of ARMA coefficients are Gaussian, in the sense that while Gaussianity need not be assumed in the asymptotic statistical theory, the estimates have the same asymptotic distribution to first order as Gaussian maximum likelihood estimates. When Gaussianity does not hold, more efficient estimates exist.

The second aspect of ARMA modelling that may be of concern is its short memory property, at least, so far as second moments are concerned. By virtue of stationarity, the autocovariance $\gamma_j = E\{(y_1 - Ey_1)(y_{1+j} - Ey_{1+j})\}$ decays exponentially as $|j| \to \infty$. A greater degree of persistence might be envisaged. In particular γ_j might follow a power law, $\gamma_j \sim |j|^{-\alpha}$, $\alpha > 0$. Of particular interest are cases when $0 < \alpha < 1$, so the γ_j are not even summable. Then y_t is sometimes said to be "strongly dependent" or "long memory".

There has been considerable activity in developing usable parametric time series models that allow explicitly either for nonlinear behaviour, or for long memory behavior. There seems to have been much less progress in developing usable, plausible models that are known to possess aspects of both properties (though see e.g. Rosenblatt (1987) and Taqqu (1987)). Indeed, much nonlinear parametric modelling has tended to emphasize processes with Markovian behavior, while much work on long memory modelling has explicitly assumed Gaussianity, and thus linearity. Evidently there exist many ways in which one might construct nonlinear models whose autocovariance structure displays long memory behavior. This very multiplicity of possibilities makes it difficult to decide which models to study. One would like some evidence that the model makes physical sense, or convincingly fits some data. Therefore an alternative approach that has some appeal in very long time series is nonparametric.

It is well known that many features of the distribution of $\{y_t, t \geq 1\}$ can be estimated nonparametrically; for example, the (possibly non-Gaussian) probability density of y_1 or the joint density of y_1, y_2 when these exist, or the (possibly non-linear) regression $E(y_2|y_1)$. Nonparametric estimates converge more slowly than estimates of correctly specified parametric models, but because of their valuable robustness property they play a useful role in exploratory analysis, with a view to possibly suggesting functional forms usable in subsequent parametric analysis. While the bulk of research on the properties of smoothed nonparametric estimates has assumed independent observations, there is also a long-standing literature which places them in dependent environments.

In particular, Rosenblatt (1970) and Roussas (1969) found that Parzen's (1962) central limit theorem for kernel probability density estimates continues to hold for many Markovian y_t. The two most interesting features are: (a) the variance in the limiting distribution is unaffected by serial dependence, in contrast to the situation with parametric estimates; (b) the multivariate central limit theorem, for density estimates at a number of fixed points, has a diagonal covariance matrix, so that estimates are asymptotically independent. The same results were subsequently shown to hold for more general schemes of serial dependence, such as many α-mixing sequences (see e.g. Robinson (1983)). Such results are practically welcome, because they justify interval estimates and test statistics of a very simple form, even in the presence of very general nonparametric serial dependence.

Analogous results are available for a popular global measure of goodness of probability density estimates, integrated mean squared error (IMSE). In particular, results of Prakasa Rao (1978) and Ahmad (1982) indicate that the results of Rosenblatt (1956) on asymptotic behavior of IMSE continue to hold for many

Markovian and α-mixing sequences, respectively. An important practical implication is that "optimal" bandwidths that minimize asymptotic IMSE in the independent case are likewise optimal under weak dependence.

There is now a large literature on smoothed nonparametric estimates under serial dependence. Given this evident interest in nonparametric treatment of nonlinear time series, and given also the interest in long-memory modelling of time series, it seems worthwhile to enquire about statistical properties of nonparametric estimates based on processes that exhibit characteristics of long-memory behaviour. In fact, nearly all the nonparametric literature appears to assume some mixing behaviour in deriving CLT and IMSE results, and often even in consistency reults. In particular, at least strong mixing is often assumed. But a Gaussian strong mixing process is not long-memory.

In Robinson (1987), some properties of kernel density estimates were investigated in long-memory environments. In case of a Gaussian process and a Gaussian kernel, some numerical calculations of exact finite sample behaviour indicated a lack of robustness of the previously described CLT to long-memory behaviour. These results were explained by a limit theorem. The Gaussianity assumption usually leads to a limiting Gaussian distribution, but the limiting variances are affected by strong dependence; and the limiting covariance matrix of estimates at several fixed points, far from being diagonal, has unit rank. A theoretical study of the asymptotic IMSE of kernel density and derivative-of-density estimates was also carried out by Robinson (1987) in the Gaussian case, with $\gamma_j \sim |j|^{-\alpha}$. It was found that only for α above a certain threshold in $(0,1)$ do the properties under independence/weak dependence continue to hold. However, this threshold diminishes as the order of derivative of interest increases (the order of magnitude of IMSE increases with derivative order). It also emerged that the IMSE-reducing benefits of using "higher-order" kernels can disappear in the presence of sufficiently long-memory behaviour, although optimality results in case of very unsmooth derivatives may remain intact. The indications are that optimal bandwidths can often, but not always, be affected by a suitable degree of long-memory dependence, as therefore can be the properties of automatic methods of bandwidth choice which approximate them, such as cross-validation. In addition, Robinson (1987) explored the role of weighted kernel estimates in the long-memory case, and a non-mixing condition that may suffice for consistency.

In this paper we substantially extend these results. The asymptotic distributional result of Robinson (1987) is suggestive, but assuming Gaussianity of the time series in nonparametric estimation is paradoxical, and one would like to know what can happen in nonlinear series. We thus consider limiting distribution theory in case y_t is a possibly non-linear, non-instantaneous function of unknown form of an underlying long-memory Gaussian sequence. The IMSE results of Robinson (1987) are all asymptotic, and convergence to the asymptotic regime, always slow in nonparametric estimates, is liable to be particularly slow for long-memory series. We thus present some finite sample numerical calculations. In addition, we also report a small comparison of density estimates based on artificial series.

Section 2 introduces the kernel estimates. Section 3 presents an asymptotic normality theorem and includes some discussion of regularity conditions. Section 4 presents a corresponding theorem to cover some cases where the limiting distribution is non-normal. Section 5 contains the IMSE calculations. Section 6 considers density estimation from artificial series. Appendix 1 discusses an important condition in Theorem 1. Appendices 2 and 3 respectively contain the proofs of our two theorems.

2. Kernel Density Estimate

In this section we introduce the kernel density estimate and summarize its known asymptotic properties under weak dependence. We shall always assume y_1 has a probability density, denoted by $f(y)$, $-\infty < y < \infty$. We wish to estimate f at various values of y, on the basis of observations $(y_t; t = 1, \dots, N)$. The kernel estimate of $f(y)$ is

$$\widehat{f}(y) = \frac{1}{Na}\sum_{t=1}^{N} K\Big(\frac{y - y_t}{a}\Big). \tag{2.1}$$

Here, a is a prescribed positive number. In asymptotic theory it is assumed, at least, that $a \to 0$ as $N \to \infty$. The real function K, the kernel, satisfies, at least,

$$\int_{-\infty}^{\infty} K(u)du = 1. \tag{2.2}$$

Under regularity conditions, we have a result of the form mentioned in the previous section

$$\begin{aligned}&(Na)^{\frac{1}{2}}\{\widehat{f}(\zeta_1) - f(\zeta_1), \dots, \widehat{f}(\zeta_s) - f(\zeta_s)\}\\ &\underset{d}{\longrightarrow} N\Big(0, \int_{-\infty}^{\infty} K^2(u)du\,\mathrm{diag}\{f(\zeta_1), \dots, f(\zeta_s)\}\Big),\end{aligned} \tag{2.3}$$

as $N \to \infty$, for distinct $\zeta_1, \dots, \zeta_s$. This is known to be true in case of many weakly dependent y_t (see e.g. Roussas (1969), Rosenblatt (1970) and Robinson (1983)).

Now consider the IMSE

$$\mathrm{IMSE}(\widehat{f}) = \int_{-\infty}^{\infty} E\{\widehat{f}(y) - f(y)\}^2 dy.$$

The order of magnitude and leading term of this varies, in a way which depends on the smoothness of f and the choice of K. If these features are fixed, then it is known that the ratio

$$\frac{\text{IMSE}(\hat{f}) : y_t \text{ independent}}{\text{IMSE}(\hat{f}) : y_t \text{ weakly dependent}}$$

converges to 1 as $N \to \infty$. This is known to be true in case of many weakly dependent y_t (see e.g. Prakasa Rao (1978) and Ahmad (1982)).

3. Asymptotic Normality under Strong Dependence

In this section we show that $\hat{f}(y)$ can sometimes still be asymptotically normal under long-memory dependence in y_t, though the precise form of the limiting distribution differs from that in (2.3). Our results extend those of Robinson (1987), who assumed y_t Gaussian. We consider the case where y_t is a nonlinear and possibly noninstantaneous function of an underlying unobservable stationary Gaussian process $\{x_t, t \geq 1\}$. Let $Ex_1 = 0$, $Ex_1^2 = 1$ and

$$\rho_j = Ex_1x_{1+j} \sim H(2H-1)j^{2H-2} \quad \text{as} \quad j \to \infty \tag{3.1}$$

where $1/2 < H < 1$. Let F be a measurable function from R^2 to R^1, and p, q be positive integers. Consider

$$y_t = F(x_{p+t-1}, x_{q+t-1}), \quad t = 1, 2, \ldots . \tag{3.2}$$

In Robinson (1987) we took F linear and $p = q$, i.e. $y_t = x_t$. The case of y_t Gaussian is a convenient and suggestive one to consider, but also a paradoxical one in view of the nonparametric density estimation. We allow F to be nonlinear, so that y_t is non-Gaussian. To gain greater generality we allow for a non-instantaneous function F; it would be straightforward but tedious to extend to a larger finite number of arguments. Under additional conditions (see Corollary in Appendix 2)

$$Ey_1y_{1+j} \sim H(2H-1)j^{2(H-1)} \text{ as } j \to \infty.$$

It is convenient to orthonormalize x_{p+t-1}, x_{q+t-1}. Let

$$u_t = cx_{p+t-1} + bx_{q+t-1}, \quad v_t = bx_{p+t-1} + cx_{q+t-1}$$

with $c = -\text{sgn}(\rho_{p-q})\{\frac{1}{2}(\Delta^2 - \Delta)\}^{\frac{1}{2}}$, $b = \{\frac{1}{2}(\Delta^2 + \Delta)\}^{\frac{1}{2}}$, and $\Delta = (1 - \rho_{p-q}^2)^{-\frac{1}{2}}$, where $|\rho_{p-q}| < 1$ is assumed. Now define $\gamma_j = E(u_1u_{1+j}) = E(v_1v_{1+j})$, $\delta_j =$

$E(u_1 v_{1+j})$. Then

$$\left.\begin{aligned}\delta_j &= 2bc\rho_j + b^2\rho_{j+p-q} + c^2\rho_{j+q-p}\\ \gamma_j &= (b^2+c^2)\rho_j + bc\rho_{j+p-q} + bc\rho_{j+q-p}\end{aligned}\right\} \tag{3.3}$$

and in particular, $\delta_0 = 0$, $\gamma_0 = 1$. Since $|\rho_{p-q}| < 1$, we have $b + c \neq 0$. So as $j \to \infty$

$$\delta_j \sim (b+c)^2 H(2H-1) j^{2(H-1)}, \quad \gamma_j \sim (b+c)^2 H(2H-1) j^{2(H-1)}. \tag{3.4}$$

We introduce the function $\widetilde{F}$ such that $y_t = \widetilde{F}(u_t, v_t)$ on substituting for x in (3.2). Expand

$$d_y(u_1, v_1) = K\Big[\frac{y - \widetilde{F}(u_1, v_1)}{a}\Big] - EK\Big[\frac{y - \widetilde{F}(u_1, v_1)}{a}\Big]$$

as

$$d_y(u_1, v_1) = \sum_{\substack{\mu,\nu=0\\ \mu+\nu\geq 1}}^{\infty} \frac{C_{\mu\nu}(y)}{\mu!\nu!} H_\mu(u_1) H_\nu(v_1)$$

where $C_{\mu\nu}(y) = E[d_y(u_t, v_t) H_\mu(u_t) H_\nu(v_t)]$, and H_μ is the μth Hermite polynomial.

Let y be a fixed point in R^1. We introduce the following assumptions.

A1. The following limits exist:

$$\lim_{a\downarrow 0} a^{-1} C_{10}(y) = \psi_1(y), \quad \lim_{a\downarrow 0} a^{-1} C_{01}(y) = \psi_2(y).$$

A2. $\psi_1(y) + \psi_2(y) \neq 0$.

A3. f has mth order continuous derivative on R^1 for some $m \geq 2$.

A4. Let y_1, y_{1+j} have a joint density function, g_j. For every compact $\Theta \subset R^2$,

$$\limsup_{|j|\to\infty} \sup_{(\tilde{u},\tilde{v})\in\Theta} |g_j(\tilde{u}, \tilde{v})| < C < \infty$$

where C only depends on Θ.

A5. We assume in (2.2) that K is bounded, has compact support and, for the same $m \geq 2$ as in A3,

$$\int_{-\infty}^{\infty} u^j K(u) du = 0, \quad j = 1, \ldots, m-1; \quad \int_{-\infty}^{\infty} u^m K(u) du \neq 0.$$

A6. As $N \to \infty$,

$$\begin{aligned} &a^m N^{1-H} + a^{-1} N^{1-2H} \to 0, \quad 1/2 < H \leq 3/4,\\ &a^m N^{1-H} + a^{-1} N^{2(H-1)} \to 0, \quad 3/4 < H < 1,\end{aligned}$$

for the same $m \geq 2$ as in A3.

Condition A1 is discussed in Appendix 1. Condition A2 is relaxed in Section 4. We could describe A3 in terms of conditions on F in view of x_t's Gaussianity. Condition A4 can be checked in some simple cases and seems mild. In A5 we introduce, when $m \geq 3$, "higher order" kernels, which are sometimes negative, in order to exploit the smoothness on f in A3. This is necessary in some cases to permit centering of F in the central limit Theorem 1. The compact support assumption on K could be relaxed at cost of some complexity in proof. Condition A6 requires at least $aN^{\frac{1}{2}} \to \infty$, a stronger upper bound on the a's rate of decay than that usually employed under weak dependence conditions. Clearly $m = 2$ suffices in case $3/4 < H < 1$, but for $1/2 < H \leq 3/4$ we need $m > (1-H)/(2H-1)$.

Theorem 1. *If* A1–A6 *hold for* $3/4 < H < 1$, $m \geq 2$, *while for* $1/2 < H \leq 3/4$, $m \geq (1-H)/(2H-1)$, *then*

$$\frac{N^{1-H}(\widehat{f}(y) - f(y))}{(b+c)(\psi_1(y) + \psi_2(y))} \xrightarrow[d]{} N(0,1).$$

The proof is in Appendix 2. The result extends the one of Robinson (1987) which covers all $y \neq 0$ in case $y_t = x_t$. In this case Robinson (1987) observed that after the norming necessary to produce a limiting $N(0,1)$ distribution density estimate, the density estimates are asymptotically perfectly correlated. The same conclusion can be drawn from the proof in Appendix 2 in the present more general case.

4. Asymptotic Non-Normality under Strong Dependence

The limiting distribution of $\widehat{f}(y)$ is not in general normal in the presence of long-memory dependence. We illustrate this fact by dropping assumption A2, and assuming that $\psi_1(y) + \psi_2(y) = 0$. This corresponds to the case of Taqqu (1975) in which the Hermite rank of a nonlinear function is greater than unity.

For the stationary Gaussian process $\{x_t\}$, there is a unique spectral measure G on $(-\pi, \pi)$ such that $\rho_j = \int e^{ij\lambda} G(d\lambda)$. Also $x_t = \int e^{itx} Z_G(dx)$, where Z_G is a random spectral measure such that $EZ_G(A)Z_G(B) = G(A \cap B)$. Define

$$h_1(x) = ce^{ix(p-1)} + be^{ix(q-1)}, \quad h_2(x) = be^{ix(p-1)} + ce^{ix(q-1)}.$$

Then $e^{ix}h_1(x)$ and $e^{ix}h_2(x)$ are orthonormal in L_G^2 and

$$u_t = \int e^{itx} h_1(x) Z_G(dx), \quad v_t = \int e^{itx} h_2(x) Z_G(dx).$$

Define $G_N(A) = N^{2(1-H)}G(N^{-1}A)$, $\forall A \in B(R^1)$. Then there exists a locally finite measure G_0 such that $\lim_{N\to\infty} G_N = G_0$, in the sense of locally weak convergence (see Dobrushin and Major (1979)). Now G_0 has the self-similar property

$$G_0(A) = t^{-2(1-H)}G_0(tA), \quad \forall A \in B(R^1), \quad t \in (0, +\infty),$$

and is determined by the relation

$$\int_{-\infty}^{\infty} e^{itx}\frac{(1-\cos x)}{x^2}G_0(dx) = \int_{-1}^{1}(1-|x|)|x+t|^{-2(1-H)}dx, \quad \forall t \in R^1.$$

We now introduce the following assumptions.

A7. For an integer $\tau \geq 2$ and for all integers μ, ν such that $\mu + \nu < \tau$, we have $C_{\mu\nu}(y) = 0$.

A8. $\lim_{a\downarrow 0} a^{-1}C_{\mu\nu}(y) = \psi_{\mu\nu}(y)$ exists for every μ, ν such that $\mu + \nu = \tau$.

A9. $\psi_\tau(y) = \sum_{\mu+\nu=\tau} \dfrac{\psi_{\mu\nu}(y)}{\mu!\nu!} \neq 0.$

A10. As $N \to \infty$,

$$a^m N^{\tau(1-H)} + a^{-1}N^{2\tau(1-H)-1} \to 0, \quad 1 - \frac{1}{2\tau} < H \leq 1 - \frac{1}{2(\tau+1)},$$

$$a^m N^{\tau(1-H)} + a^{-1}N^{2(H-1)} \to 0, \quad 1 - \frac{1}{2(\tau+1)} < H < 1,$$

for the same $m \geq 2$ as in A3.

In order to illustrate A7–A9, we consider the simplest case $y_t = u_t$ at $y = 0$. We have

$$C_{01}(y) = C_{11}(y) = C_{02}(y) = 0 \quad \text{for each } y \in R^1,$$
$$C_{10}(y) = 0, \quad \lim_{a\downarrow 0} a^{-1}C_{20}(0) = -e^{-1}/\sqrt{2\pi}.$$

Thus $\tau = 2$, $\psi_2(0) = -e^{-1}/\sqrt{2\pi}$.

Theorem 2. *If* A4, A5 *and* A7–A10 *hold, then*

$$\frac{N^{\tau(1-H)}\{\hat{f}(y) - f(y)\}}{\sqrt{\tau!}(b+c)^\tau \psi_\tau(y)} \xrightarrow[d]{} \xi,$$

where

$$\xi = \int e^{i(x_1+\cdots+x_\tau)}K_0(x_1,\ldots,x_\tau)Z_G(dx_1)\cdots Z_G(dx_\tau),$$

$$K_0(x_1,\ldots,x_\tau) = \frac{e^{i(x_1+\cdots+x_\tau)} - 1}{i(x_1+\cdots+x_\tau)}.$$

For $\tau > 1$, ξ is non-normal. The theorem is proved in Appendix 3.

5. Integrated Mean Squared Error Calculations

Although asymptotic distributional results can be useful in setting pointwise interval estimates, a global measure of goodness of fit of density estimates is often desired. The IMSE introduced in Section 2 has proved popular for this purpose, and also as a basis for choice of bandwidth a. As indicated in Section 2, asymptotic results for the iid case are robust to weak dependence, but not necessarily to strong dependence, as indicated by Robinson (1987).

In this section we report some calculations of IMSE in finite samples from a long-memory process. There has been relatively little study of the finite-sample performance of nonparametric density estimates. Some analytic finite-sample properties of $V\{\widehat{f}(y)\}$, with implications for choice of a, were given by Robinson (1986) in case of a weakly dependent, Gaussian first-order autoregressive y_t. In a similar situation, numerical results of Hart (1984) indicated that IMSE is liable to be substantially affected in finite samples by autocorrelation.

Robinson (1987) studied IMSE under strong dependence. His theoretical results concern both density estimates and derivative-of-density estimates. Consider estimating, for integer r,

$$f^{(r)}(y) = \frac{d^r f(y)}{dy^r},$$

where $f^{(0)}(y) = f(y)$. Put $\widehat{f}^{(0)}(y) = \widehat{f}(y)$ and $\widehat{f}^{(r)}(y) = (d^r/dy^r)\widehat{f}(y)$, $r > 0$, where $\widehat{f}(y)$ is given by (2.1), in which we use the Gaussian kernel

$$K(u) = e^{-\frac{1}{2}u^2}/(2\pi)^{\frac{1}{2}}. \tag{5.1}$$

In the specified situation of Robinson (1987) our $y_t = x_t$, with ρ_j given by

$$\rho_j = \frac{1}{2}\{|j+1|^{2H} - 2|j|^{2H} + |j-1|^{2H}\}, \tag{5.2}$$

for $1/2 < H < 1$. Under (5.2) we may interpret y_t as a fractional Brownian motion, or the increment of a certain self-similar process. The property (3.1) holds for (5.1). From formulae (6.2), (6.3) of Robinson (1987) we obtain

$$\begin{aligned}\mathrm{IMSE}(N,r,a,H) &\triangleq E\int_{-\infty}^{\infty}[\widehat{f}^{(r)}(y) - f^{(r)}(y)]^2 dy \\ &= \frac{(2r)!}{2^{2r+1}r!\sqrt{\pi}N}[a^{-2r-1} - (1+a^2)^{-r-\frac{1}{2}}]\end{aligned}$$

$$+\frac{(2r)!}{2^{2r}r!\sqrt{\pi}(Ha^2)^{r+\frac{3}{8}}N}\sum_{j=1}^{N-1}\Big(1-\frac{j}{N}\Big)\Big[\Big(1-\frac{\rho_j}{1+a^2}\Big)^{-r-\frac{1}{2}}-1\Big]$$

$$+\frac{(2r)!}{2^{2r}r!\sqrt{\pi}}\Big[\frac{1}{2}-\Big(1+\frac{1}{2}a^2\Big)^{-r-\frac{1}{2}}+\frac{1}{2}(1+a^2)^{-r-\frac{1}{2}}-1\Big].$$

It was shown by Robinson (1987) that IMSE(N,r,a,H) possesses the following asymptotic properties:

1. For $H<(2r+3)/(2r+5)$, IMSE(N,r,a,H) exhibits the same asymptotic behaviour as IMSE$(N,r,a,1/2)$, where IMSE$(N,r,a,1/2)$ indicates the IMSE under independent x_t. Thus, modest amounts of long memory dependence leave unchanged the usual IMSE results, thus the usual implications for optimal bandwidth choice also. Moreover as the order of differentiation increases, a greater degree of long memory dependence can be accommodated.
2. For $H=(2r+3)/(2r+5)$, the leading term of IMSE(N,r,a,H) contains an additional component, due to serial dependence, but this does not affect the rate of the optimal bandwidth.
3. For $H>(2r+3)/(2r+5)$, IMSE(N,r,a,H) is asymptotically dominated by the serial dependence, and is insensitive to choice of a.

We emphasize that the above results are all asymptotic, so it would be interesting to investigate what happens in finite samples, especially because convergence can be particularly slow in case of long-memory series, and the asymptotic results give no idea of actual numerical performance. Let

$$R(N,r,a,H)=\frac{\text{IMSE}(N,r,a,H)}{\text{IMSE}(N,r,a,1/2)}.$$

In Tables 1–3 we report numerical values of $R(N,r,a,H)$ and IMSE(N,r,a,H) for $r=0,1,2$, $a=1.06N^{-1/5}\sqrt{\text{Var}(y_t)}$ and $N=50$, 250, 1000. (In the previous version of the paper, results for $r=3$, 4, 5 were also reported, reinforcing the conclusions below.) The values of H used varied somewhat across the tables because of variability with r of the threshold value $H=(2r+3)/(2r+5)$. Each cell in the tables contains $R(N,r,a,H)$ and, in parentheses, IMSE(N,r,a,H).

We draw the following conclusions.

1. As the asymptotic theory predicts, for $H<(2r+3)/(2r+5)$, R is quite close to 1, but for $H>(2r+3)/(2r+5)$ it deteriorates rapidly.
2. The asymptotic theory seems to hold reasonably well for $H<(2r+3)/(2r+5)$, as indicated by the approximate constancy of R against N. For $H>(2r+3)/(2r+5)$, R tends to fall off quite rapidly as N increases. Thus the nonparametric estimates seem less badly affected by high levels of long-memory serial dependence in small samples than in long samples.

Table 1. Value of R (IMSE) for $r = 0$

	$H=0.5$	$H=0.6^*$	$H=0.7$	$H=0.8$	$H=0.9$	$H=0.95$	a
$N=50$	1.000	0.775	0.507	0.275	0.118	0.066	0.457
	(0.009)	(0.016)	(0.018)	(0.033)	(0.076)	(0.137)	
$N=250$	1.000	0.743	0.414	0.168	0.052	0.023	0.332
	(0.003)	(0.004)	(0.007)	(0.017)	(0.056)	(0.123)	
$N=1000$	1.000	0.727	0.345	0.108	0.025	0.010	0.251
	(0.001)	(0.002)	(0.003)	(0.010)	(0.041)	(0.106)	

Table 2. Value of R (IMSE) for $r = 1$

	$H=0.5$	$H=0.6$	$H=0.71^*$	$H=0.8$	$H=0.9$	$H=0.95$	a
$N=50$	1.000	0.892	0.677	0.458	0.211	0.111	0.572
	(0.039)	(0.043)	(0.057)	(0.084)	(0.018)	(0.349)	
$N=250$	1.000	0.896	0.635	0.365	0.123	0.051	0.454
	(0.018)	(0.019)	(0.027)	(0.047)	(0.014)	(0.334)	
$N=1000$	1.000	0.908	0.613	0.303	0.078	0.027	0.373
	(0.008)	(0.009)	(0.013)	(0.027)	(0.105)	(0.304)	

Table 3. Value of R (IMSE) for $r = 2$

	$H=0.5$	$H=0.6$	$H=0.7$	$H=0.78^*$	$H=0.9$	$H=0.95$	a
$N=50$	1.000	0.948	0.837	0.686	0.328	0.168	0.648
	(0.043)	(0.045)	(0.051)	(0.063)	(0.130)	(0.255)	
$N=250$	1.000	0.956	0.836	0.648	0.234	0.092	0.542
	(0.024)	(0.025)	(0.028)	(0.036)	(0.101)	(0.257)	
$N=1000$	1.000	0.966	0.845	0.628	0.178	0.058	0.464
	(0.014)	(0.015)	(0.016)	(0.022)	(0.078)	(0.240)	

$^*H = (2r+3)/(2r+5)$

3. As the asymptotic theory suggests, for comparable values of H, R increases with r.
4. We find that, not surprisingly, IMSE tends to decrease with N, to increase with H, and to increase with r.

6. Density Estimate from Artificial Series

There is little evidence on the actual performance of density estimates with long-memory series. We provide a small illustration using simulated series. Exact simulation of long-memory series is not easy. The approach we employ (cf. Granger (1980)) involved forming

$$y_t = \frac{1}{q}\sum_{j=1}^{q} y_{jt}, \quad t = 1, 2, \ldots, N.$$

The y_{jt} are generated by the random-coefficient autoregressive model for panel data, analyzed by Robinson (1978):

$$y_{jt} = \alpha_j y_{j,t-1} + \epsilon_{jt}, \quad j = 1, \dots, q,$$

where the α_j are independent variates drawn from a $(1/2, 2H-1)$ beta distribution, with $1/2 < H < 1$, and the ϵ_{jt} are iid across j and t, with zero mean and finite variance. As shown by Granger (1980), after averaging with respect to the α_j as well as the ϵ_{jt},

$$\rho_j \sim C|j|^{2(H-1)} \quad \text{as} \quad |j| \to \infty.$$

It seems that for a realization y_t to convincingly exhibit long-memory properties we should choose q large.

We computed $\widehat{f}(y)$ given by (2.1) using kernel (5.1). A variety of values of a was tried. We display results only for $N = 500$, $a = 1.06N^{-1/5}$, the approximately IMSE-optimal bandwidth under independence (see §5). Prior to computing (2.1) we standardized the y_t so as to have sample mean zero and sample variance unity. We computed (2.1) across a fine grid of y values via the fast Fourier transform algorithm. All computations were carried out on LSE's VAX computer.

In Figure 1 we plot $\widehat{f}(y)$ with $q = 250$ and $H = 0.95$ for two different ϵ_{jt} distributions, $N(0,1)$ and uniform $(-1/2, 1/2)$. The density estimate in the normal ϵ case is not quite symmetric though we would expect the unconditional distribution of y_t to be approximately normal. Pronounced asymmetry is detected in the uniform ϵ case.

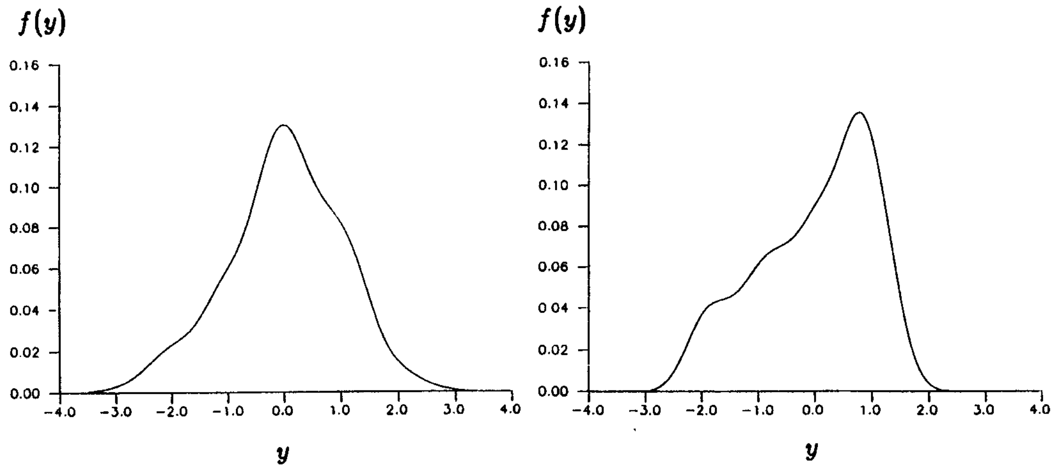

Fig. 1(a) Fig. 1(b)

Figure 1. Density estimates: (a) Normal ϵ, $H = 0.75$, $q = 250$; (b) Uniform ϵ, $H = 0.95$, $q = 250$.

Acknowledgement

This article is based on research funded by the Economic and Social Research Council (ESRC) reference number: R000231441. It is partially based on work presented at the SERC Workshop on Nonlinear Time Series at Edinburgh in July 1989. The research of the first author who is at the Institute of Applied Mathematics, Academia Sinica, Beijing, was carried out while visiting the London School of Economics with the support of the Sino-British Fellowship Trust, administered by the Royal Society. We are grateful to the referees for comments, and to a member of the editorial board for stylistic corrections, which have led to a substantial improvement in presentation.

Appendix 1: Discussion of Condition A1

In this section we attempt to justify and illustrate condition A1, used in Theorem 1. We treat separately the special case where $F(u,v)$ depends on u only, and the more general case where it depends on both u and v.

1. First we assume

$$\widetilde{u} \triangleq \widetilde{F}(u,v) = \widetilde{F}(u).$$

We assume also: (a) $\widetilde{F}$ has continuous first derivative; (b) $\{u : \widetilde{F}'(u) = 0\}$ is a finite point set, where the prime denotes differentiation.

We denote $\{u : \widetilde{F}'(u) = 0\}$ by $\{\tau_1, \dots, \tau_r\}$, with $\tau_1 < \tau_2 < \cdots < \tau_r$. Let $B_1 = (-\infty, \tau_1)$, $B_2 = (\tau_1, \tau_2), \dots, B_r = (\tau_{r-1}, \tau_r)$ and $B_{r+1} = (\tau_r, \infty)$. For brevity put $C_{\mu\nu} = C_{\mu\nu}(y)$. Under A5, (a) and (b),

$$C_{01} = E\Big[K\Big[\frac{y - \widetilde{F}(u_t)}{a}\Big]v_t\Big] = 0,$$

$$C_{10} = E\Big[K\Big[\frac{y - \widetilde{F}(u_t)}{a}\Big]u_t\Big] = \sum_{i=1}^{r+1}\int_{F(B_i)} K\Big[\frac{y-\widetilde{u}}{a}\Big]h_i(\widetilde{u})d\widetilde{u},$$

where φ is the standard normal density and $h_i(\widetilde{u}) = \widetilde{F}_i^{-1}(\widetilde{u})\varphi(\widetilde{F}^{-1}(\widetilde{u}))F_i^{-1}(\widetilde{u})$. Since K has compact support, there is a positive number M such that $\mathrm{SUPP}[K] \subset [-M, M]$. Then

$$C_{10} = -a\sum_{i=1}^{r+1}\int_{(y-\widetilde{F}(B_i))/a\cap[-M,M]} K(u)h_i(y-au)du.$$

However, if $u \in (y - \widetilde{F}(B_i))/a$, then $y - au \in \widetilde{F}(B_i)$. For $y \in \widetilde{F}(B_i)$, by (a), (b), convexity of $\widetilde{F}(B_i)$, and the mean value theorem, we have

$$\int_{F(B_i)} K\Big[\frac{y-\widetilde{u}}{a}\Big]h_i(\widetilde{u})d\widetilde{u} = ah_i(y) + o(a).$$

For $y \notin \widetilde{F}(B_i)$ but either $y = \widetilde{F}(\tau_i)$ or $y = \widetilde{F}(\tau_{i-1})$, there is a difficulty, because $\widetilde{F}'(\tau_i)$ or $\widetilde{F}'(\tau_{i-1})$ are 0. The limit of

$$\int_{F(B_i)} K\left[\frac{y-\widetilde{u}}{a}\right] h_i(\widetilde{u})d\widetilde{u}$$

may not exist as $a \to 0$. For $y \notin \widetilde{F}(B_i)$ and neither $y = \widetilde{F}(\tau_i)$ nor $y = \widetilde{F}(\tau_{i-1})$,

$$\int_{F(B_i)} K\left[\frac{y-\widetilde{u}}{a}\right] h_i(\widetilde{u})d\widetilde{u} \equiv 0,$$

because $\inf\{|\widetilde{u} - y|; \forall \widetilde{u} \in \widetilde{F}(B_i)\} > 0$ and K has compact support, when a is small enough. Overall, we conclude:

(i) When $y \in \widetilde{F}(R^1) = \bigcup_{i=1}^{r+1} \widetilde{F}(B_i)$ and $y \neq \widetilde{F}(\tau_i)$, $i = 1, \ldots, r$, $C_{10}(y) = a\psi_1(y) + o(a)$ where $\psi_1(y) = \sum_{i=1}^{r+1} h_i(y) I_{\widetilde{F}(B_i)}(y)$, and I is the indicator function.

(ii) When $y \notin \widetilde{F}(R^1)$ and $y \neq \widetilde{F}(\tau_i)$, $i = 0, \ldots, r+1$, $C_{10} = o(a)$.

(iii) When $y = \widetilde{F}(\tau_i)$, $i = 0, \ldots, r+1$, the limit $\lim_{a \downarrow 0} a^{-1} C_{10}$ may not exist.

Example 1. $\widetilde{F}(u) = (u+1)^2$, that is, $y_t = (x_{t-p+1} + x_{t-q+1} + 1)^2$. For $y \in (0, \infty)$

$$\begin{aligned} C_{10} &= \int_{-\infty}^{\infty} K\left[\frac{y-(u+1)^2}{a}\right] u\varphi(u)du \\ &= \int_0^{\infty} K\left[\frac{y-u}{a}\right] h_1(u)du + \int_{\infty}^{0} K\left[\frac{y-u}{a}\right] h_2(u)du \end{aligned} \tag{A.1}$$

where $h_1(u) = \frac{1}{2}(1 - u^{-\frac{1}{2}})\varphi(1 - u^{-\frac{1}{2}})$ and $h_2(u) = \frac{1}{2}(1 + u^{-\frac{1}{2}})\varphi(1 + u^{\frac{1}{2}})$. Thus (A.1) is

$$\begin{aligned} &-a\int_{y/a}^{-\infty} K(u)h_1(y-au)du - a\int_{-\infty}^{y/a} K(u)h_2(y-au)du \\ &\sim a[h_1(y) - h_2(y)] \quad \text{as } a \to 0. \end{aligned}$$

Consequently $C_{10} \sim a$ for $y \in (0, \infty)$. For $y \in (-\infty, 0)$ with $y/a \to -\infty$ as $a \to 0$, we get $C_{10} = o(a)$. For $y = 0$, $C_{10} = \int_{-\infty}^{\infty} K(-(u+1)^2/a)u\varphi(u)du$.

2. For general $\widetilde{u} = \widetilde{F}(u, v)$, the situation is much more complicated. But if we assume there exists a function $\widetilde{v} = \widetilde{G}(u, v)$ such that

$$J(u,v) = \frac{\partial \widetilde{u}}{\partial u}\frac{\partial \widetilde{v}}{\partial v} - \frac{\partial \widetilde{u}}{\partial v}\frac{\partial \widetilde{v}}{\partial u} \neq 0,$$

then, by the inverse function theorem, $u = \widetilde{F}^{-1}(\widetilde{u},\widetilde{v})$, $v = \widetilde{G}^{-1}(\widetilde{u},\widetilde{v})$. The joint density function of $(\widetilde{u}_t,\widetilde{v}_t)$ is

$$h(u,v) = \varphi(\widetilde{F}^{-1}(u,v))\varphi(\widetilde{G}^{-1}(u,v))[J(\widetilde{F}^{-1}(u,v),\widetilde{G}^{-1}(u,v))].$$

Hence

$$C_{10} = \iint_{A\times B} K\Big[\frac{y-\widetilde{u}}{a}\Big]\widetilde{F}^{-1}(\widetilde{u},\widetilde{v})h(\widetilde{u},\widetilde{v})d\widetilde{u}d\widetilde{v} \tag{A.2}$$

where $A = \{\widetilde{u};\widetilde{u} = \widetilde{F}(u,v),\forall(u,v)\in R^2\}$, $B = \{\widetilde{v};\widetilde{v} = \widetilde{G}(u,v),\forall(u,v)\in R^2\}$. Thus (A.2) is

$$\int_A K\Big[\frac{y-\widetilde{u}}{a}\Big]\Big[\int_B \widetilde{F}^{-1}(\widetilde{u},\widetilde{v})h(\widetilde{u},\widetilde{v})d\widetilde{v}\Big]d\widetilde{u}.$$

As in the previous discussion, there are three cases:

$$C_{10} \sim \begin{cases} -a\displaystyle\int_B \widetilde{F}^{-1}(y,\widetilde{v})h(y,\widetilde{v})d\widetilde{v} & \text{if } y\in A,\\ \text{the limit may not exist} & \text{if } y\in\partial A,\\ 0 & \text{if } y\notin A\cup\partial A,\end{cases}$$

where ∂A is the boundary of A. Similarly

$$C_{01} \sim \begin{cases} -a\displaystyle\int_A \widetilde{G}^{-1}(\widetilde{u},y)h(\widetilde{u},y)d\widetilde{u} & \text{if } y\in B,\\ \text{the limit may not exist} & \text{if } y\in\partial B,\\ 0 & \text{if } y\notin B\cup\partial B,\end{cases}$$

where ∂B is the boundary of B.

Example 2. $\widetilde{F}(u,v) = uv$, i.e. $y_t = bc(x_{t+p-1}^2 + x_{t+q-1}^2) + (b^2+c^2)x_{t+p-1}x_{t+q-1}$. By some calculations, for each $y\in R^1$, we have $C_{10}(y) = C_{01}(y) \equiv 0$, $C_{11}(y) \sim ayf(y)$ as $a\downarrow 0$, and

$$C_{20}(y) = C_{02}(y) \sim a\int_{-\infty}^{\infty}|u|e^{-\frac{1}{2}(u^2+y^2u^{-2})}du.$$

Appendix 2: Proof of Theorem 1

The basic method of proof is similar to that of Robinson (1987), but the proof is substantially complicated by the nonlinear and non-instantaneous character of F. We first present three preliminary lemmas.

Lemma 1. *Let $f_{t-s}(x)$ be the probability density function of (u_t, v_t, u_s, v_s), where $x = (x_1, x_2, x_3, x_4)$. Then*

$$f_{t-s}(x)=\sum_{r=0}^{\infty}\sum_{m+j+k+\ell=r}\frac{\gamma_{t-s}^{m+b}\delta_{t-s}^{j+k}}{m!j!k!\ell!}H_{m+j}(x_1)H_{k+\ell}(x_2)H_{m+k}(x_3)H_{j+\ell}(x_4)e^{-\frac{1}{2}\Sigma x_j^2}.$$

Proof. $E[\exp\{it_1u_t+it_2v_t+it_3u_s+it_4v_s\}] = \varphi(t)\psi(t)$, where $t = (t_1, t_2, t_3, t_4)'$ and $\varphi(t) = e^{-\frac{1}{2}t't}$, $\psi(t) = \exp\{-[\gamma_{t-s}t_1t_3 + \delta_{t-s}t_1t_4 + \delta_{t-s}t_2t_3 + \gamma_{t-s}t_2t_4]\}$. Now

$$\begin{aligned}\psi(t) &= \sum_{r=0}^{\infty}\frac{(-1)^r}{r!}(\gamma_{t-s}t_1t_3 + \delta_{t-s}t_1t_4 + \delta_{t-s}t_2t_3 + \gamma_{t-s}t_2t_4)^r \\ &= \sum_{r=0}^{\infty}(-1)^r\sum_{m+j+k+\ell=r}\frac{\gamma_{t-s}^{m+\ell}\delta_{t-s}^{j+k}}{m!j!k!\ell!}t_1^{m+j}t_2^{k+\ell}t_3^{m+k}t_4^{j+\ell}.\end{aligned}$$

On the other hand,

$$f_{t-s}(x) = \frac{1}{(2\pi)^4}\int e^{-it'x}\varphi(t)\psi(t)\Pi dt_i.$$

Then apply

$$\int_{-\infty}^{\infty} e^{itx}t^{\nu}e^{-\frac{1}{2}t^2}dt = 2\pi(-i)^{\nu}H_{\nu}(x)e^{-\frac{1}{2}x^2}.$$

Lemma 2. *Let r be a non-negative integer. Then*

$$\sum_{(L)}\frac{1}{m!j!k!\ell!} = \frac{r!}{\mu!\nu!(\mu')!(\nu')!}$$

where $\sum_{(L)}$ sums over $m + j = \mu$, $k + \ell = \nu$, $m + k = \mu'$ and $j + \ell = \nu'$, and where (μ, ν, μ', ν') satisfies $\mu + \nu = r$, $\mu' + \nu' = r$.

Proof. For arbitrary real $t_1, \dots, t_4$,

$$(t_1t_3 + t_1t_4 + t_2t_3 + t_2t_4)^r = \sum_{\substack{\mu+\nu=r\\ \mu'+\nu'=r}}\Big[\sum_{(L)}\frac{r!}{m!j!k!\ell!}\Big]t_1^{\mu}t_2^{\nu}t_3^{\mu'}t_4^{\nu'}.$$

On the other hand,

$$\begin{aligned}(t_1t_3 + t_1t_4 + t_2t_3 + t_2t_4)^r &= (t_1 + t_2)^r(t_3 + t_4)^r \\ &= (r!)^2\sum_{\substack{\mu+\nu=r\\ \mu'+\nu'=r}}\frac{1}{\mu!\nu!(\mu')!(\nu')!}t_1^{\mu}t_2^{\nu}t_3^{\mu'}t_4^{\nu'}.\end{aligned}$$

Then use the arbitrariness of $t_1, \ldots, t_4$.

Lemma 3. *Let G be a measurable function from R^2 to R^1 such that $E(G_1^2) < \infty$, where $G_t = G(u_t, v_t)$, and G has Hermite expansion*

$$G(x_1, x_2) = \sum_{r=0}^{\infty} \sum_{\mu+\nu=r} \frac{\beta_{\mu\nu}}{\mu!\nu!} H_\mu(x_1) H_\nu(x_2),$$

where $\beta_{\mu\nu} = E[G_1 H_\mu(u_1) H_\nu(v_1)]$. If there is a positive integer τ such that $\beta_{\mu\nu} = 0$ for arbitrary μ and ν with $\mu + \nu < \tau$, then there exists a positive integer N_0 such that, for $|t-s| > N_0$,

$$|E(G_t G_s)| \leq C_\tau |\gamma_{t-s}|^\tau E G_1^2,$$

where C_τ is a constant which only depends on τ.

Proof. By Lemma 1,

$$\begin{aligned} E(G_t G_s) &= \int G(x_1, x_2) G(x_3, x_4) f_{t-s}(x) \Pi dx_i \\ &= \sum_{r=\tau}^{\infty} \sum_{\substack{\mu+\nu=r \\ \mu'+\nu'=r}} \Big[\sum_{(L)} \frac{\gamma_{t-s}^{m+\ell} \delta_{t-s}^{k+j}}{m!j!k!\ell!} \Big] \beta_{\mu\nu} \beta_{\mu'\nu'} \\ &= \gamma_{t-s}^\tau \sum_{r=\tau}^{\infty} \sum_{\substack{\mu+\nu=r \\ \mu'+\nu'=r}} \Big[\sum_{(L)} \frac{\gamma_{t-s}^{r-\tau} (\delta_{t-s}/\gamma_{t-s})^{k+j}}{m!j!k!\ell!} \Big] \beta_{\mu\nu} \beta_{\mu'\nu'} \triangleq \gamma_{t-s}^\tau I_\tau. \end{aligned}$$

From (3.1) and (3.3), as $|t-s| \to \infty$,

$$\gamma_{t-s} = (b+c)^2 \rho_{t-s} + o(1), \quad \delta_{t-s} = (b+c)^2 \rho_{t-s} + o(1), \quad \delta_{t-s}/\gamma_{t-s} = 1 + o(1).$$

For fixed $\epsilon > 0$, there exists an integer $N_0 > 0$ such that, for $|t-s| > N_0$, $|\delta_{t-s}/\gamma_{t-s}| < 1 + \epsilon$. For such N_0,

$$\begin{aligned} |I_\tau| &\leq C \sum_{r=\tau}^{\infty} 4^{-r} \sum_{\substack{\mu+\nu=r \\ \mu'+\nu'=r}} \Big[\sum_{(L)} \frac{1}{m!j!k!\ell!} \Big] |\beta_{\mu\nu}|\, |\beta_{\mu'\nu'}| \\ &= C \sum_{r=\tau}^{\infty} 4^{-r} \sum_{\substack{\mu+\nu=r \\ \mu'+\nu'=r}} \frac{r! |\beta_{\mu\nu}|\, |\beta_{\mu'\nu'}|}{\mu!\nu!(\mu')!(\nu')!} \\ &\leq C \sum_{r=\tau}^{\infty} 4^{-r} r! \Big[\sum_{\mu+\nu=r} \frac{\beta_{\mu\nu}^2}{(\mu!)^2(\nu!)^2} \Big], \end{aligned}$$

where the second equality uses Lemma 2 and C is a generic constant. By Hölder's Inequality, $\beta_{\mu\nu}^2 \le \mu!\nu!EG_1^2$. Hence

$$|I_\tau| \le C\sum_{t=\tau}^{\infty} 4^{-r}\Big[\sum_{\mu+\nu=r}\frac{r!}{\mu!\nu!}\Big]EG_t^2 = C\sum_{r=\tau}^{\infty} 2^{-r}EG_1^2 \le CEG_1^2.$$

Corollary. *For* $\{y_t = \widetilde{F}(u_t, v_t)\}$, *if*

$$Ey_1^2 < \infty \quad and \quad E[y_1(u_1+v_1)] \ne 0, \tag{B.1}$$

then

$$E(y_1 - Ey_1)(y_{1+j} - Ey_{1+j}) \sim H(2H-1)j^{2H-2} \quad as \quad j \to \infty.$$

Proof. First expand $\widetilde{F}$ in Hermite polynomials

$$\widetilde{F}(x_1, x_2) = \sum_{r=0}^{\infty}\Big[\sum_{\mu+\nu=r}\frac{\alpha_{\mu\nu}}{\mu!\nu!}H_\mu(x_1)H_\nu(x_2)\Big],$$

where $\alpha_{\mu\nu} = E[y_1 H_\mu(u_1)H_\nu(v_1)]$. By (B.1) $\alpha_{10} + \alpha_{01} = E[y_1(u_1+v_1)] \ne 0$. Now let $G(x_1, x_2) = \widetilde{F}(x_1, x_2) - \alpha_{10}x_1 - \alpha_{01}x_2 - E\widetilde{F}(u_1, v_1)$. From Lemma 1,

$$\left.\begin{aligned} E[H_\mu(u_t)H_\nu(v_t)u_s] &= \begin{cases} \gamma_{t-s}^{\mu}\delta_{t-s}^{\nu}, & \mu+\nu=1,\\ 0, & \text{otherwise,}\end{cases}\\ E[H_\mu(u_t)H_\nu(v_t)v_s] &= \begin{cases} \gamma_{t-s}^{\nu}\delta_{t-s}^{\mu}, & \mu+\nu=1,\\ 0, & \text{otherwise,}\end{cases}\end{aligned}\right\} \tag{B.2}$$

and $E(G_t u_s) = E(G_t v_t) = 0$. Hence $E(y_t - Ey_t)(y_s - Ey_s)$ is

$$\begin{aligned} &E(\alpha_{10}u_t + \alpha_{01}v_t + G_t)(\alpha_{10}u_s + \alpha_{01}v_s + G_s)\\ &\quad = E(\alpha_{10}u_t + \alpha_{01}v_t)(\alpha_{10}u_s + \alpha_{01}v_s) + EG_tG_s.\end{aligned}$$

However

$$\begin{aligned} E(\alpha_{10}u_t + \alpha_{01}v_t)(\alpha_{10}u_s + \alpha_{01}v_s) &= \alpha_{10}^2\gamma_{t-s} + 2\alpha_{10}\alpha_{01}\delta_{t-s} + \alpha_{01}^2\gamma_{t-s}\\ &\sim (\alpha_{10}+\alpha_{01})^2(b+c)^2\rho_{t-s} \quad \text{as} \quad |t-s| \to \infty,\end{aligned}$$

and by Lemma 3, when $|t-s|$ is large enough, $E(G_tG_s) \sim \rho_{t-s}^2$.

We can now proceed to the proof of Theorem 1. As in Robinson (1987), this involves an expansion of the kernel estimate analogous to the Hermite expansion method of Taqqu (1975), but unlike in those references we are dealing here with a two-dimensional function. Let

$$e_y(x_1, x_2) = \sum_{\substack{\mu,\nu=0\\ \mu+\nu\geq 2}}^{\infty} \frac{C_{\mu\nu}}{\mu!\nu!} H_\mu(x_1) H_\nu(x_2),$$

and $e_t = e_y(u_t, v_t)$. Let $d_t = d_y(u_t, v_t)$. Then

$$\hat{f}(y) - E\hat{f}(y) = \frac{1}{Na}\sum_t d_t = \frac{1}{Na}\sum_t (C_{10}u_t + C_{01}v_t) + \frac{1}{Na}\sum_t e_t.$$

We show first that the asymptotic distribution of the left-hand side is governed by that of the first term, which is exactly Gaussian. We have

$$E\Big[\sum_t (C_{10}u_t + C_{01}v_t)\Big]^2 = N(C_{10}^2 + C_{01}^2) + \sum_{\substack{t,s=1\\ t\neq s}}^{N} (C_{10}^2\gamma_{t-s} + 2C_{10}C_{01}\delta_{t-s} + C_{01}^2\gamma_{t-s}).$$

From (3.4), the double sum is

$$(C_{10} + C_{01})^2(b+c)^2 \sum_{\substack{t,s=1\\ t\neq s}}^{N} \rho_{t-s} + o\Big((C_{10} + C_{01})^2 \sum_{\substack{t,s=1\\ t\neq s}}^{N} |\rho_{t-s}|\Big)$$
$$= (\psi_1(y) + \psi_2(y))^2(b+c)^2 N^{2H} a^2 + o(a^2 N^{2H}).$$

Thus $E[\sum_{t=1}^N (C_{10}u_t + C_{01}v_t)]^2 \sim N^{2H}a^2$. We need now to show

$$a^{-2}N^{-2H} E\Big(\sum_t e_t\Big)^2 \to 0, \quad \text{as} \quad N \to \infty. \tag{B.3}$$

By Lemma 3, there exists a fixed positive integer N_0 such that, for $|t-s| > N_0$, $|E(e_t e_s)| \leq C_2\gamma_{t-s}^2 E(e_1^2)$. However

$$E(e_1^2) = \sum_{\substack{\mu,\nu=0\\ \mu+\nu\geq 2}}^{\infty} \frac{C_{\mu\nu}^2}{\mu!\nu!} \leq \sum_{\substack{\mu,\nu=0\\ \mu+\nu\geq 1}}^{\infty} \frac{C_{\mu\nu}^2}{\mu!\nu!} = E(d_1^2)$$
$$= EK^2\Big(\frac{y-y_1}{a}\Big) - \Big[EK\Big(\frac{y-y_1}{a}\Big)\Big]^2 \sim af(y) \;\text{ as }\; a \to 0.$$

For $3/4 < H < 1$, $E(\sum_t e_t)^2 \leq I_1 + I_2$, where

$$I_1 = \sum_{\substack{t,s=1\\ |t-s|\leq N_0}}^{N} |Ee_te_s|, \quad I_2 = \sum_{\substack{t,s=1\\ |t-s|> N_0}}^{N} |Ee_te_s|,$$

and where

$$I_1 \le \sum_{\substack{t,s=1\\|t-s|\le N_0}}^{N} E(e_t^2) \le 2NN_0E(e_1^2) \sim Na,$$

$$I_2 \le C \sum_{\substack{t,s=1\\|t-s|\le N_0}}^{N} \gamma_{t-s}^2 E(e_1^2) \le CE(e_1^2) \sum_{\substack{t,s=1\\t\neq s}}^{N} \gamma_{t-s}^2 = 2CNE(e_1^2)\sum_{s=1}^{N}\gamma_s^2.$$

For $3/4 < H < 1$, (B.3) is true because of condition A6 and

$$\sum_{s=1}^{N}\gamma_s^2 \sim (b+c)^2\sum_{s=1}^{N}\rho_s^2 \sim N^{4H-3}.$$

Now consider $1/2 < H \le 3/4$. For $t \neq s$, $E(e_te_s)$ is

$$\begin{aligned} &E[d_td_s] - Ed_t[C_{10}u_s + C_{01}v_s] - Ed_s(C_{10}u_t + C_{01}v_t)\\ &\qquad\qquad\qquad\qquad + E(C_{10}u_t + C_{01}v_t)(C_{10}u_s + C_{01}v_s)\\ &= J_1 - J_2 - J_3 + J_4. \end{aligned}$$

First

$$J_2 = C_{10}\sum_{\substack{\mu,\nu=0\\\mu+\nu\ge 1}}^{\infty}\frac{C_{\mu\nu}}{\mu!\nu!}E[H_\mu(u_t)H_\nu(v_t)u_s] + C_{01}\sum_{\substack{\mu,\nu=0\\\mu+\nu\ge 1}}^{\infty}\frac{C_{\mu\nu}}{\mu!\nu!}E[H_\mu(u_t)H_\nu(v_t)v_s].$$

From (B.1) and $|\gamma_{t-s}| \le 1$, $|\delta_{t-s}| \le 1$, we have $|J_2| \le (C_{10}+C_{01})^2 \sim a^2$ by A2. Similarly $|J_3| = O(a^2)$. Now

$$J_1 = E(d_td_s) = E\Big[K\Big(\frac{y-y_t}{a}\Big)K\Big(\frac{y-y_s}{a}\Big)\Big] - \Big[EK\Big(\frac{y-y_t}{a}\Big)\Big]^2.$$

The first term is

$$\begin{aligned} &\int\int K\Big(\frac{y-\widetilde{u}}{a}\Big)K\Big(\frac{y-\widetilde{v}}{a}\Big)g_{t-s}(\widetilde{u},\widetilde{v})d\widetilde{u}d\widetilde{v}\\ &\qquad = a^2\int_{-L}^{L}\int_{-L}^{L}K(u)K(v)g_{t-s}(y-au, y-av)dudv \end{aligned}$$

where $L > 0$ satisfies SUPP$[K] \subset [-L, L]$. For $(u,v) \in [-L,L]\times[-L,L]$,

$$|y-au| \le |y| + L, \quad |y-av| \le |y| + L$$

(without loss of generality assume $a < 1$). By A3 there exists a fixed positive integer N_1 such that, for any t and s with $|t-s| > N_1$,

$$\sup_{\substack{|u|\le|y|+L\\|v|\le|y|+L}} |g_{t-s}(u,v)| \le C_{y,L}$$

where $C_{y,L}$ only depends on y and L. Hence for $|t-s| > N_1$,

$$\left|E\left[K\left(\frac{y-y_t}{a}\right)K\left(\frac{y-y_s}{a}\right)\right]\right| \le C_{y,L}a^2\left[\int |K(u)|du\right]^2.$$

Because $EK((y-y_t)/a) \sim af(y)$, we have, for $|t-s| > N_1$, $|J_1| = O(a^2)$. Obviously, $|J_4| \le E(C_{10}u_t + C_{01}v_t)^2 = C_{10}^2 + C_{01}^2 = O(a^2)$. Altogether, for any t and s with $|t-s| > N_1$, $|Ee_te_s| = O(a^2)$. Now let $\overline{N} = \max\{N_0, N_1\}$. Then

$$E\Big(\sum e_t\Big)^2 \le 2N\overline{N}Ee_1^2 + \sum_{\substack{t,s=1\\|t-s|>\overline{N}}}^{N} |Ee_te_s|.$$

Put

$$\Delta_{s-t} = \begin{cases} |Ee_te_s|, & \text{for } |t-s| > \overline{N},\\ 0, & \text{otherwise.}\end{cases}$$

Then, for $1 \le M \le N-2$,

$$\sum_{t,s=1}^{n} \Delta_{s-t} = 2\sum_{s=1}^{M}\sum_{t=1}^{N-M} \Delta_s + 2\sum_{s=M}^{N-1}\sum_{t=1}^{N-s}\Delta_s + 2\sum_{t=N-M+1}^{N-1}\sum_{s=1}^{N-t}\Delta_s$$
$$\le C\Big(NMa^2 + 2NEe_1^2\sum_{s=M}^{N-1}\gamma_s^2\Big) = O\Big(NMa^2 + N\sum_{s=M}^{N-1}\rho_s^2 a\Big).$$

For $1/2 < H < 3/4$, $\sum_M^{N-1}\rho_s^2 = O(M^{4H-3})$. Thus

$$E\Big(\sum_t e_t\Big)^2 = O(Na + NMa^2 + NM^{4H-3}a).$$

Then from the above, for any t, s, $\Delta_{s-t} = O(a^2)$, and by Lemma 3, $\Delta_{s-t} \le O(|\gamma_{t-s}|^2E(e_1^2))$. For any $\epsilon > 0$, take $M \sim \epsilon N^{2H-1}$. Hence (B.3) is true for $1/2 < H \le 3/4$. Therefore by Gaussianity,

$$\frac{N^{1-H}(\hat{f}(y) - E\hat{f}(y))}{(b+c)(\psi_1(y)+\psi_2(y))} \xrightarrow[\text{d}]{} N(0,1).$$

Finally, we must show

$$N^{1-H}[E\hat{f}(y) - f(y)] \to 0. \tag{B.4}$$

By standard arguments

$$E\hat{f}(y) - f(y) \sim (-1)^m \frac{a^m}{m!} \int u^m K(u)du$$

and (B.4) follows from A6.

Appendix 3: Proof of Theorem 2

We first introduce a lemma.

Lemma 4. *For $1 - \frac{1}{2\tau} < H < 1$ and integer $\tau \geq 2$,*

$$\mathrm{Var}\Big[\sum_{t=1}\Big(\sum_{\mu+\nu=\tau} \frac{C_{\mu\nu}}{\mu!\nu!} H_\mu(u_t)H_\nu(v_t)\Big)\Big] \sim \tau!(b+c)^{2\tau}\psi_\tau^2(y)a^2 N^{2-2\tau(1-H)}.$$

Proof. $E\Big\{\sum_{t=1}^{N}\Big(\sum_{\mu+\nu=\tau} \frac{1}{\mu!\nu!} C_{\mu\nu} H_\mu(u_t)H_\nu(v_t)\Big)\Big\}^2 = I_1 + I_2$, where

$$I_1 = \sum_{t=1}^{N} E\Big[\sum_{\mu+\nu=\tau} \frac{C_{\mu\nu}}{\mu!\nu!} H_\mu(u_t)H_\nu(v_t)\Big]^2,$$

$$I_2 = \sum_{\substack{t,s=1\\ t\neq s}}^{N} \Big[\sum_{\substack{\mu+\nu=\tau\\ \mu'+\nu'=\tau}} \frac{C_{\mu\nu}C_{\mu'\nu'}}{\mu!\nu!\mu'!\nu'!} EH_\mu(u_t)H_\nu(v_t)H_{\mu'}(u_s)H_{\nu'}(v_s)\Big].$$

By Lemma 1

$$I_2 = \sum_{\substack{t,s=1\\ t\neq s}}^{N} \sum_{\substack{\mu+\nu=\tau\\ \mu'+\nu'=\tau}} \Big(\sum_{(L)} \frac{\gamma_{t-s}^{m+\ell}\delta_{t-s}^{j+k}}{m!j!k!\ell!}\Big) C_{\mu\nu}C_{\mu'\nu'}.$$

By Lemma 2 this is

$$\tau!(b+c)^{2\tau}\Big[\sum_{\mu+\nu=\tau} \frac{C_{\mu\nu}}{\mu!\nu!}\Big]^2 \sum_{\substack{t,s=1\\ t\neq s}} \rho_{t-s}^{\tau}$$

$$+\tau!\Big[\sum_{\mu+\nu=\tau} \frac{C_{\mu\nu}}{\mu!\nu!}\Big]^2 (b+c)^{2\tau}\rho_{t-s}^{\tau} \sum_{\substack{t,s=1\\ t\neq s}} \Big[\Big(\frac{\gamma_{t-s}}{(b+c)^2\rho_{t-s}}\Big)^\tau - 1\Big]$$

$$+\sum_{\substack{t,s=1\\t\neq s}}\Big[\sum_{\substack{\mu+\nu=\tau\\\mu'+\nu'=\tau}}\Big(\sum\frac{\left(\frac{\delta_{t-s}}{\gamma_{t-s}}\right)^{j+k}-1}{m!j!k!\ell!}\Big)C_{\mu\nu}C_{\mu'\nu'}\Big]\gamma_{t-s}^{\tau}.$$

Since $\gamma_{t-s}/((b+c)^2\rho_{t-s})$ and $\delta_{t-s}/\gamma_{t-s}\to 1$ as $|t-s|\to\infty$,

$$I_2=\tau!(b+c)^{2\tau}\Big[\sum_{\mu+\nu=\tau}\frac{C_{\mu\nu}}{\mu!\nu!}\Big]^2\sum_{\substack{t,s=1\\t\neq s}}\rho_{t-s}^{\tau}+o\Big(a^2\sum_{t,s=1}^{N}\rho_{t-s}^{\tau}\Big)$$
$$\sim\tau!(b+c)^{2\tau}\psi_\tau^2(y)a^2N^{2-2\tau(1-H)},$$

and $I_1\sim a^2N$. But for $1-1/2\tau<H$, we have $1<2-2\tau(1-H)$.

We now prove Theorem 2. As in Robinson (1987)

$$E\Big(\sum e_t\Big)^2=o\big(a^2N^{2-2\tau(1-H)}\big).$$

Here we must divide H into the intervals

$$1-\frac{1}{2(\tau+1)}<H<1,\quad 1-\frac{1}{2\tau}<H<1-\frac{1}{2(\tau+1)}.$$

For the first,

$$E\Big(\sum e_t\Big)^2=O\Big(Na\sum_{1}^{N}\rho_j^{\tau+1}\Big)=O\big(aN^{2-2(\tau+1)(1-H)}\big).$$

For the second,

$$E\Big(\sum_{t=1}^{N}e_t\Big)^2=O\Big(Na+NMa^2+NM^{1-2(\tau+1)(1-H)}a\Big).$$

Thus

$$E\Big(\sum_{t=1}^{N}e_t\Big)^2=o\big(a^2N^{2-2\tau(1-H)}\big).$$

On the other hand,

$$\sum_{\mu+\nu=\tau}\frac{C_{\mu\nu}}{\mu!\nu!}H_\mu(u)H_\nu(v)=\sum_{\mu+\nu=\tau}\frac{C_{\mu\nu}}{\mu!\nu!}H_\mu\Big(\int h_1(x)Z_G(dx)\Big)H_\nu\Big(\int h_2(x)Z_G(dx)\Big).$$

By the formula expressing multiple Wiener-Ito-Dobrushin integrals in terms of

Hermite polynomials (e.g. Ito (1951)), this is

$$\sum_{\mu+\nu=\tau} \frac{C_{\mu\nu}}{\mu!\nu!} \int \cdots \int \prod_{1}^{\mu} h_1(x_i) \prod_{\mu+1}^{\tau} h_2(x_i) \prod_{1}^{\tau} dZ_q(x_i)$$
$$= \frac{1}{\tau!} \int \cdots \int \alpha_0(x) \prod_{1}^{\tau} Z_G(dx_i),$$

where

$$\alpha_0(x) = \sum_{\mu+\nu=\tau} \frac{\tau! C_{\mu\nu}}{\mu!\nu!} \prod_{1}^{\mu} h_1(x_i) \prod_{\mu+1}^{\tau} h_2(x_i).$$

Thus use Lemma 4 and Dobrushin and Major's (1979) Theorem 3 to show that

$$\frac{N^{\tau(1-H)}\{\hat{f}(y) - E\hat{f}(y)\}}{\sqrt{\tau!}(b+c)^{\tau}\psi_{\tau}(y)} \xrightarrow[d]{} \xi,$$

and for $N^{\tau(1-H)}\{E\hat{f}(y) - f(y)\} \to 0$, we straightforwardly employ A3, A5 and A10.

References

Ahmad, I. A. (1982). Integrated mean square properties of density estimation by orthogonal series methods for dependent variables. *Ann. Inst. Statist. Math.* **34**, 339–350.

Dobrushin, R. L. and Major, P. (1979). Non-central limit theorems for non-linear functionals of Gaussian fields. *Z. Wahrsch. verw. Gebiete* **50**, 27–52.

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *J. Econometrics* **14**, 227–238.

Hart, J. D. (1984). Efficiency of a kernel density estimator under an autoregressive dependence model. *J. Amer. Statist. Assoc.* **79**, 110–117.

Ito, K. (1951). Multiple Wiener integral. *J. Math. Soc. Japan* **3**, 157–169.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065–1076.

Prakasa Rao, B. L. S. (1978). Density estimation for Markov processes using delta-sequences. *Ann. Inst. Statist. Math.* **30**, 321–328.

Robinson, P. M. (1978). Statistical inference for a random coefficient autoregressive model. *Scand. J. Statist.* **5**, 163–168.

Robinson, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4**, 185–207.

Robinson, P. M. (1986). On the consistency and finite-sample properties of nonparametric kernel time series regression, autoregression and density estimators. *Ann. Inst. Statist. Math.* **38**, 539–549.

Robinson, P. M. (1987). Nonparametric function estimation for long-memory time series. Forthcoming in *Semiparametric and Nonparametric Methods in Statistics and Econometrics* (Edited by W. Barnett et al.). Cambridge University Press, Cambridge.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832–837.

Rosenblatt, M. (1970). Density estimates and Markov sequences. In *Nonparametric Techniques in Statistical Inference* (Edited by M. L. Puri), 199-210. Cambridge University Press, Cambridge.

Rosenblatt, M. (1987). Scale renormalization and random solutions of the Burgers equation. *J. Appl. Probab.* **24**, 328–338.

Roussas, G. G. (1969). Nonparametric estimation in Markov processes. *Ann. Inst. Statist. Math.* **21**, 73–87.

Taqqu, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. verw. Gebiete* **31**, 287–302.

Taqqu, M. S. (1987). Random processes with long-range dependence and high variability. *J. Geophys. Res.* **92**, 9683–9686.

Institute of Applied Mathematics, Academia Sinica, Beijing.

Department of Economics, London School of Economics and Political Science, University of London, London WC2A 2AE, U.K.

(Received January 1990; accepted January 1991)