MODEL SELECTION FOR BROADBAND SEMIPARAMETRIC ESTIMATION OF LONG MEMORY IN TIME SERIES

By Clifford M. Hurvich

New York University

Abstract. We study the properties of Mallows' C_L criterion for selecting a fractional exponential (FEXP) model for a Gaussian long-memory time series. The aim is to minimize the mean squared error of a corresponding regression estimator $\hat{d}_{FEXP}$ of the memory parameter, d. Under conditions which do not require that the data were actually generated by a FEXP model, it is known that the mean squared error $MSE = E[\hat{d}_{FEXP} - d]^2$ can converge to zero as fast as $(\log n)/n$, where n is the sample size, assuming that the number of parameters grows slowly with n in a deterministic fashion. Here, we suppose that the number of parameters in the FEXP model is chosen so as to minimize a local version of C_L, restricted to frequencies in a neighborhood of zero. We show that, under appropriate conditions, the expected value of the local C_L is asymptotically equivalent to MSE. A combination of theoretical and simulation results give guidance as to the choice of the degree of locality in C_L.

1. INTRODUCTION

Suppose we have an even number of observations $y_0, \ldots, y_{n-1}$ from $\{Y_t\}$, a Gaussian long-memory time series with spectral density

$$f(\omega) = |1 - e^{-i\omega}|^{-2d} f^*(\omega) \qquad \omega \in [-\pi, \pi]$$

where $d \in (-0.5, 0.5)$, and f^* is positive, even, continuous, bounded above, and bounded away from zero on $[-\pi, \pi]$. The long and short memory aspects of $\{Y_t\}$ are controlled by the memory parameter d and the function f^*, respectively. We are interested here in estimating d semiparametrically, i.e., without making fully parametric assumptions on the form of f^*.

Existing semiparametric estimators of d include the GPH estimator (Geweke and Porter-Hudak, 1983; Robinson, 1995a; Hurvich *et al*., 1998), the average periodogram estimator (Robinson, 1994a) and the Gaussian semiparametric estimator (Kunsch, 1987; Robinson, 1995b). These estimators use periodogram values from a vanishingly small neighbourhood of the origin, and therefore require no strong global assumptions on f^*. However, they typically have mean squared error of order no better than $O(n^{-4/5})$, even if f^* is extremely smooth.

A different approach, which we call broadband semiparametric estimation, assumes that f^* is differentiable on $[-\pi, \pi]$ (perhaps excluding zero frequency), but still specifies no parametric form for f^*. This global smoothness

assumption motivates the fitting of a fully parametric (though presumably incorrect) long-memory model to the data, using all $\tilde{n} = n/2 - 1$ available periodogram ordinates. Bhansali and Kokoszka (1999) have studied one such estimator, based on fitting a fractional autoregressive model in the frequency domain. They established that the resulting estimator of d is consistent and asymptotically normal if the autoregressive order approaches infinity at a suitable rate as n increases.

Here, we focus on another broadband semiparametric estimator which we call $\hat{d}_{\text{FEXP}}$ (defined in Section 2), originally due to Janacek (1982) and subsequently discussed by Robinson (1994b). It is based on fitting a fractional exponential, or FEXP model to the log periodogram ordinates at all available Fourier frequencies by linear regression. The FEXP model is a long-memory generalization of the exponential model of Bloomfield (1973), and has been discussed by Beran (1993, 1994). Unfortunately, Janacek (1982) did not provide asymptotic theory for his proposed estimator. This deficiency has been remedied by the two papers of Hurvich and Brodsky (1997) and Moulines and Soulier (1999), which were written independently and at about the same time. Both papers establish the consistency of $\hat{d}_{\text{FEXP}}$ without requiring parametric assumptions on f^*, provided that the order of the exponential model increases at a suitable rate with n. The rate of convergence of the mean squared error of $\hat{d}_{\text{FEXP}}$ improves with the smoothness of $\log(f^*)$, and can be as fast as $(\log n)/n$ if $\log(f^*)$ is infinitely differentiable, as would be the case for example if f^* were the spectral density of a stationary invertible ARMA model. Moulines and Soulier (1999) established asymptotic normality for their version of the estimator, which uses regressor frequencies that are slightly shifted away from the Fourier frequencies. Hurvich and Brodsky (1997) used the Fourier frequencies in the regression, but did not establish asymptotic normality.

A question left unresolved by Hurvich and Brodsky (1997) is the development of a rigorously justifiable method of selecting the number of parameters h to use in the fitted exponential model. Hurvich and Brodsky (1997) proposed to select h so as to minimize

$$C_{\text{L}}^*(h) = \text{RSS}_h + \frac{2h\pi^2}{6}$$

using a residual sum of squares RSS_h taken over all $\tilde{n}$ frequencies used in the regression, but we did not develop any theory to support this choice. If $\{Y_t\}$ were Gaussian white noise, $C_L^*(h)$ would be an exactly unbiased estimator of the mean integrated squared error,

$$\text{MISE}_m = \text{E}\left[\sum_{j=1}^{m}(\log \hat{f}_j - \log f_j)^2\right]$$

(with $m = \tilde{n}$), where $\{\hat{f}_j\}_{j=1}^{\tilde{n}}$ is the spectral estimator corresponding to the fitted FEXP model and $\{f_j\}_{j=1}^{\tilde{n}}$ is the true spectral density at the jth Fourier

frequency $\omega_j = 2\pi j/n$ for $j = 1, \ldots, \tilde{n}$. Moulines and Soulier (2000) have established the optimality with respect to $\text{MISE}_{\tilde{n}}$ of the fitted FEXP model selected by minimizing C_L^*. Brodsky (1997) presented similar though less complete results in an unpublished dissertation. If the primary aim is to obtain a good estimator of d, however, a quantity of more direct interest is the mean squared error of $\hat{d}_{\text{FEXP}}$, i.e.,

$$\text{MSE} = \text{E}[\hat{d}_{\text{FEXP}} - d]^2$$

In this paper, we will assume conditions implying that the $(\log n)/n$ rate of convergence of MSE can be attained but not improved upon.

   We will establish that except for a multiplicative constant, MISE_m and MSE are asymptotically equivalent, under conditions implying that $m = \text{o}(n)$, so that the mean integrated squared error is restricted to a local neighborhood of zero frequency. This equivalence does not hold, however, if $m = \tilde{n}$ as was assumed by Brodsky (1997), Hurvich and Brodsky (1997) and Moulines and Soulier (2000). Under more restrictive conditions on m, we will show further that a local version of C_L, based on using a residual sum of squares restricted to frequencies $\omega_1, \ldots, \omega_m$, is asymptotically unbiased for MISE_m, up to an additive constant. Thus, apart from multiplicative and additive constants which are irrelevant to the choice of h, the local C_L is asymptotically unbiased for MSE. This provides some motivation for selecting the h minimizing the local C_L for use in $\hat{d}_{\text{FEXP}}$.

   In Section 2, we define the local C_L criterion, and give some heuristic motivation for its unbiasedness property. In Section 3, we present and prove our main theoretical results, and set out several Lemmas which are proved in an Appendix. In Section 4, we present the results of a Monte Carlo study exploring the sensitivity of the proposed methodology to the choice of m.

2. THE LOCAL C_L CRITERION

Define the periodogram by

$$I_j = I(\omega_j) = \frac{1}{2\pi n} \left| \sum_{t=0}^{n-1} y_t \exp(-i\omega_j t) \right|^2 \qquad j = 1, \ldots, \tilde{n}$$

Let $y = \log I(\underline{\omega}) + \gamma$, where $\underline{\omega} = (\omega_1, \ldots, \omega_{\tilde{n}})'$ and $\gamma = 0.577216 \ldots$ is Euler's constant. Then we can write $y = \mu + \varepsilon$, where

$$\mu = \log f(\underline{\omega}) = -2d \log \left| 2 \sin\left(\frac{\underline{\omega}}{2}\right) \right| + \log f^*(\underline{\omega})$$

and

$$\varepsilon = \log\left[\frac{I(\underline{\omega})}{f(\underline{\omega})}\right] + \gamma$$

In the case of Gaussian white noise, the entries ε_j ($j = 1, \ldots, \tilde{n}$) of ε are i.i.d., with mean zero and variance $\sigma^2 = \pi^2/6$.

We assume that

$$\log f^*(\omega) = \sum_{k=0}^{\infty} \gamma_k \cos(k\omega) \qquad \omega \in [-\pi, \pi] \tag{1}$$

where the $\{\gamma_k\}$ are unknown constants. Thus,

$$\log f^*(\underline{\omega}) = \sum_{k=0}^{\infty} \gamma_k V_k$$

where $V_k = \cos(k\underline{\omega})$.

For any given positive integer h, define

$$X_h = [V_0, V_1, \ldots, V_h] \qquad \text{and} \qquad X = [X_h, \tilde{V}]$$

where

$$\tilde{V} = \log \left| 2 \sin \left(\frac{\omega}{2} \right) \right|$$

Define

$$\hat{\beta} = (X'X)^{-1}X'y$$

which may be viewed as the ordinary least squares estimator of the parameter $\beta = (\gamma_0, \gamma_1, \ldots, \gamma_h, -2d)'$. By definition, $-2\hat{d}_{\text{FEXP}}$ is the final entry of $\hat{\beta}$. Define the hat matrix,

$$H = X(X'X)^{-1}X'$$

and let $\hat{y} = Hy = X\hat{\beta}$. Let m be a positive integer with $1/m + m/n \to 0$ as $n \to \infty$. Let W be a diagonal matrix with the first m diagonal entries equal to unity, and all other entries equal to zero. For any $\tilde{n}$-vector v, define

$$\|v\|_W^2 = v'Wv = \sum_{j=1}^{m} v_j^2$$

The local version of C_{L} is defined by

$$C_{\text{L}}(h) = \|y - \hat{y}\|_W^2 + 2\sigma^2 Tr(WH) \tag{2}$$

Since $\|y - \hat{y}\|_W^2 = \|(\hat{y} - \mu) - \varepsilon\|_W^2$ and since $\hat{y} - \mu = -(I - H)\mu + H\varepsilon$, we have

$$C_{\text{L}}(h) = \|\hat{y} - \mu\|_W^2 + \|\varepsilon\|_W^2 + 2\mu'(I - H)W\varepsilon + 2[\sigma^2 Tr(WH) - \varepsilon'HW\varepsilon] \tag{3}$$

Note that $\text{E}\|\hat{y} - \mu\|_W^2 = \text{MISE}_m$. If $\{Y_t\}$ were Gaussian white noise, we would then have $\text{E}[C_{\text{L}}(h)] = \text{MISE}_m + m\sigma^2$, so that $C_{\text{L}}(h)$ would be exactly unbiased for MISE_m, except for an additive constant which does not depend on h.

3. THEORETICAL RESULTS

We assume that the γ_k in (1) decay at least exponentially fast, i.e., $|\gamma_k| \leq Ca^k/k^p$ where C, p and a are real constants with $C > 0$, $p \geq 0$ and $0 < a < 1$. For these same values of a and p, we assume the following.

ASSUMPTION 1

$$\frac{h}{2} \sum_{k=h+1}^{\infty} \frac{\gamma_k}{k} \sim \frac{a^h}{h^p} G_h \text{ as } h \to \infty$$

where $\sup_h |G_h| < \infty$ and $\inf_h |G_h| > 0$.

The above sum determines the properties of $Bias(\hat{d}_{\text{FEXP}})$. Since G_h can take on values of both signs, $Bias(\hat{d}_{\text{FEXP}})$ may oscillate between negative and positive values as $h \to \infty$, but $|Bias(\hat{d}_{\text{FEXP}})|$ is bounded both above and below by fixed nonzero multiples of a^h/h^p. Assumption 1 holds with $p = 1$ for f^* corresponding to any AR(1) or MA(1) model, as well as an infinite number of ARMA models of all orders. A detailed explanation of the situation is provided in Appendix 1.

From Theorems 1 and 2 of Hurvich and Brodsky (1997), if

$$\frac{\log \log n}{h} + \frac{h}{n} \log^6 n \to 0$$

then

$$Bias(\hat{d}_{\text{FEXP}}) = \frac{h}{2} \sum_{k=h+1}^{\infty} \frac{\gamma_k}{k} + \text{O}(a^{\tilde{n}}) + \text{O}\left[\frac{h(\log^3 n)}{n}\right]$$

$$= \frac{a^h}{h^p} G_h[1 + \text{o}(1)] + O\left[\frac{h(\log^3 n)}{n}\right]$$

and

$$\text{MSE} \sim \frac{a^{2h}}{h^{2p}} G_h^2 + \sigma^2 h/n$$

Here, we have assumed slightly stronger conditions on h than in the theorems of Hurvich and Brodsky (1997) to guarantee that MSE can be represented as above. By taking h proportional to $\log n$, with a proportionality constant exceeding $-1/(2 \log a)$, we would attain $\text{MSE} \leq C^*(\log n)/n$ where $C^* > 0$. Furthermore, if we take h proportional to $(\log n)^{1+\delta}$ for any fixed $\delta > 0$ and use any positive proportionality constant, then $\text{MSE} \leq C^*(\log n)^{1+\delta}/n$. Unfortunately, it will be difficult in general to obtain an expression for the h which asymptotically minimizes MSE, due to the presence of the potentially oscillatory term G_h^2 in the expression for MSE. Nevertheless, under Assumption 1, MSE cannot converge to

zero at a rate faster than $(\log n)/n$, since $\inf_h |G_h| > 0$. In our theoretical results, we take h to be a deterministic, increasing sequence such that

$$\frac{\log \log n}{h} \to 0 \qquad \text{and} \qquad h = O[(\log n)^K]$$

where K is a fixed positive integer. In certain situations where Assumption 1 does not hold (including some ARMA models for f^*) it may be possible to take advantage of the extremely low bias of $\hat{d}_{\text{FEXP}}$ to attain a faster rate of convergence than $(\log n)/n$, but model selection in such situations would require a far more delicate analysis than we undertake in this paper.

Let β_h and $\hat{\beta}_h$ be $(h+1)$-dimensional column vectors obtained by deleting the final element of β and $\hat{\beta}$, respectively. We have

$$\text{MISE}_m = \text{E}\|\hat{y} - \mu\|_W^2$$

$$= \text{E}\left\|X\hat{\beta} - \left(X\beta + \sum_{k=h+1}^{\infty} \gamma_k V_k\right)\right\|_W^2$$

$$= \text{E}\left\| -2(\hat{d}_{\text{FEXP}} - d)\tilde{V} + X_h(\hat{\beta}_h - \beta_h) - \sum_{k=h+1}^{\infty} \gamma_k V_k\right\|_W^2$$

$$= 4\text{MSE}\|\tilde{V}\|_W^2 + E\|X_h(\hat{\beta}_h - \beta_h)\|_W^2 + \left\|\sum_{k=h+1}^{\infty} \gamma_k V_k\right\|_W^2 + R \qquad (4)$$

where, by the Cauchy–Schwarz inequality, the remainder term R satisfies

$$R = O[\text{MSE}\|\tilde{V}\|_W^2 \cdot \text{E}\|X_h(\hat{\beta}_h - \beta_h)\|_W^2]^{1/2} + O\left(\text{MSE}\|\tilde{V}\|_W^2 \cdot \left\|\sum_{k=h+1}^{\infty} \gamma_k V_k\right\|_W^2\right)^{1/2}$$

$$+ O\left[\text{E}\|X_h(\hat{\beta}_h - \beta_h)\|_W^2 \cdot \left\|\sum_{k=h+1}^{\infty} \gamma_k V_k\right\|_W^2\right]^{1/2} \qquad (5)$$

To obtain our main theoretical results, we will need the following Lemmas, which are proved in Appendix 2. The notation $A = O(B)$, where A and B are matrices, means that the ratio of the j, k entry of A to the j, k entry of B is $O(1)$ for each j, k.

LEMMA 1. *Suppose that*

$$\frac{1}{h} + \left(\frac{h}{n}\right)\log^2 n \to 0$$

Then

$$(X'X)^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

where B_{22} is scalar with $B_{22} \sim 4h/n$. Uniform bounds for the other submatrices are given by

$$B_{11} = \frac{4}{n} I_{h+1} + \mathrm{O} \begin{bmatrix} \dfrac{1}{n} & \dfrac{h \log n}{n^2} \xi_h' \\ \dfrac{h \log n}{n^2} \xi_h & \dfrac{h}{n} \xi_h' \otimes \xi_h \end{bmatrix}$$

$$B_{12} = B_{21}' = \mathrm{O} \begin{bmatrix} \dfrac{h \log n}{n^2} \\ \dfrac{h}{n} \xi_h \end{bmatrix}$$

where I_{h+1} is an $(h+1) \times (h+1)$ identity matrix, and $\xi_h = (1, \frac{1}{2}, \ldots, 1/h)'$.

LEMMA 2. *The entries of H are uniformly $\mathrm{O}[(h/n)\log^2 n]$.*

LEMMA 3. *If $m = \mathrm{o}(n)$, then $\|\tilde{V}\|_W^2 \sim m \log^2(n/m)$.*

LEMMA 4. $\mathrm{E}\|X_h(\hat{\beta}_h - \beta_h)\|_W^2 = \mathrm{O}[m\,\mathrm{MSE} \log^2 h]$.

LEMMA 5. $\left\| \sum_{k=h+1}^{\infty} \gamma_k V_k \right\|_W^2 = \mathrm{O}[m\,\mathrm{MSE}]$

LEMMA 6. *If $m \sim \mathrm{Const} \cdot n^\alpha$ with $\frac{1}{2} < \alpha < 1$, then*

$$\left| \frac{\mathrm{E}[\mu'(I - H)W\varepsilon]}{\mathrm{MISE}_m} \right| \to 0$$

LEMMA 7. *If $m \sim \mathrm{Const} \cdot n^\alpha$ with $\frac{1}{2} < \alpha < 1$, then*

$$|\mathrm{E}[\sigma^2 Tr(WH) - \varepsilon' HW\varepsilon]| = \mathrm{o}(\mathrm{MISE}_m)$$

THEOREM 1. *If $m \sim \mathrm{Const} \cdot n^\alpha$ with $0 < \alpha < 1$, then*

$$\mathrm{MISE}_m = 4\,\mathrm{MSE}\|\tilde{V}\|_W^2[1 + \mathrm{o}(1)]$$

PROOF OF THEOREM 1. We need to show that the final three terms on the righthand side of Equation (4) are $\mathrm{o}(\mathrm{MSE}\|\tilde{V}\|_W^2)$. From Lemma 3 with $m \sim \mathrm{Const} \cdot n^\alpha$ with $0 < \alpha < 1$, we obtain

$$\|\tilde{V}\|_W^2 \sim C_3 \cdot m \log^2 n$$

with $C_3 > 0$. By Lemma 4,

$$\frac{\mathrm{E}\|X_h(\hat{\beta}_h - \beta_h)\|_W^2}{\mathrm{MSE}\|\tilde{V}\|_W^2} = \mathrm{O}\left[\frac{m\,\mathrm{MSE}\log^2 h}{m\,\mathrm{MSE}\log^2 n}\right] = \mathrm{O}\left(\frac{\log^2 h}{\log^2 n}\right) \to 0$$

Next, from Lemma 5, we obtain

$$\frac{\left\|\sum_{k=h+1}^{\infty} \gamma_k V_k\right\|_W^2}{\mathrm{MSE}\|\tilde{V}\|_W^2} = \mathrm{O}\left(\frac{m\,\mathrm{MSE}}{\mathrm{MSE}\|\tilde{V}\|_W^2}\right) = \mathrm{O}\left(\frac{m}{m\log^2 n}\right) \to 0$$

It remains to show that R given in (5) satisfies $R = \mathrm{o}(\mathrm{MSE}\|\tilde{V}\|_W^2)$. By arguments similar to those given above, we have

$$\frac{[\mathrm{MSE}\|\tilde{V}\|_W^2\,\mathrm{E}\|X_h(\hat{\beta}_h - \beta_h)\|_W^2]^{1/2}}{\mathrm{MSE}\|\tilde{V}\|_W^2} = \mathrm{O}\left[\frac{m\,\mathrm{MSE}\log^2 h}{\mathrm{MSE}\|\tilde{V}\|_W^2}\right]^{1/2}$$

$$= \mathrm{O}\left(\frac{\log h}{\log n}\right) \to 0$$

$$\frac{\left(\mathrm{MSE}\|\tilde{V}\|_W^2\left\|\sum_{k=h+1}^{\infty} \gamma_k V_k\right\|_W^2\right)^{1/2}}{\mathrm{MSE}\|\tilde{V}\|_W^2} = \mathrm{O}\left(\frac{m\,\mathrm{MSE}}{\mathrm{MSE}\|\tilde{V}\|_W^2}\right)^{1/2}$$

$$= \mathrm{O}\left(\frac{1}{\log n}\right) \to 0$$

$$\frac{\left[\mathrm{E}\|X_h(\hat{\beta}_h - \beta_h)\|_W^2\left\|\sum_{k=h+1}^{\infty} \gamma_k V_k\right\|_W^2\right]^{1/2}}{\mathrm{MSE}\|\tilde{V}\|_W^2} = \mathrm{O}\left[\frac{(m\,\mathrm{MSE}\log^2 h \cdot m\,\mathrm{MSE})^{1/2}}{\mathrm{MSE} \cdot m\log^2 n}\right]$$

$$= \mathrm{O}\left(\frac{\log h}{\log^2 n}\right) \to 0$$

and the proof is complete.

THEOREM 2. *If $m \sim \mathrm{Const} \cdot n^\alpha$ with $\frac{1}{2} < \alpha < 1$, then*

$$\mathrm{E}[C_\mathrm{L}(h)] = \mathrm{MISE}_m + \mathrm{E}\|\varepsilon\|_W^2 + \mathrm{o}(\mathrm{MISE}_m)$$

PROOF OF THEOREM 2. Taking expectations in Equation (3) yields

$$E[C_L(h)] = \mathrm{MISE}_m + \mathrm{E}\|\varepsilon\|_W^2 + 2\mathrm{E}[\mu'(I - H)W\varepsilon] + 2\mathrm{E}[\sigma^2 Tr(WH) - \varepsilon' HW\varepsilon]$$

The final two terms are both o(MISE$_m$), by Lemmas 6 and 7, respectively. The proof is complete.

4. MONTE CARLO RESULTS

Using the method of Davies and Harte (1987), we simulated ARFIMA(1, d, 0) models with $d = 0.4$. The models are of form

$$(1 - \rho B)(1 - B)^d Y_t = e_t$$

for $t = 0, \ldots, n - 1$, where the $\{e_t\}$ are independent standard normal, and B is the backshift operator, i.e., $BY_t = Y_{t-1}$. We generated 500 simulated realizations for each combination (n, ρ) with $n = 512$, 1024, 4096, and $\rho = 0.2$, 0.8. For clarity, we now introduce the notation $\hat{d}_{\text{FEXP}}^h$, where the superscript denotes the value of h used in computing $\hat{d}_{\text{FEXP}}$. For each realization, we computed $\hat{d}_{\text{FEXP}}^h$, $C_L(h)$ (Equation (2)), and the global version $C_L^*(h)$ for $h = 1, \ldots, 20$. In the calculation of C_L, we tried two different values of $m : m = n^{0.5}$ and $m = n^{0.8}$.

For $n = 4096$, $\rho = 0.8$, $m = n^{0.5}$ (a representative case), Figure 1 plots simulated versions of MISE$_m$, $E[C_L] - m\sigma^2$, $4\,\text{MSE}\|\tilde{V}\|_W^2$ and $E[C_L^*]$ as functions of h, based on averages over the 500 realizations. Since

$$E\|\varepsilon\|_W^2 = m\sigma^2[1 + o(1)]$$

Theorem 2 (if it applies; see comments below) suggests that MISE$_m$ and $E[C_L] - m\sigma^2$ should be similar, and indeed Figure 1 shows very good agreement between these two quantities, with some deterioration as h gets very small. Both curves are minimized at $h = 9$. The quantity $4\,\text{MSE}\|\tilde{V}\|_W^2$ is also minimized at $h = 9$, and has the same overall shape as the plots of MISE$_m$ and $E[C_L] - m\sigma^2$, but (somewhat paradoxically) does not provide a close numerical approximation to these two quantities. A very close approximation of $4\,\text{MSE}\|\tilde{V}\|_W^2$ to MISE$_m$ should not be expected, however, since from the proof of Theorem 1 we have

$$\frac{\text{MISE}_m}{4\,\text{MSE}\|\tilde{V}\|_W^2} = 1 + O\left(\frac{\log h}{\log n}\right)$$

so the remainder term declines only very slowly to zero as n increases.

The close agreement in the numerical values of MISE$_m$ and $E[C_L] - m\sigma^2$ noted above deserves further comment. These two curves were based on $m = n^\alpha$ with $\alpha = \frac{1}{2}$, whereas Theorem 2 requires $\alpha > \frac{1}{2}$. Indeed, the proofs of Lemmas 6 and 7 imply that $E[C_L]/\text{MISE}_m \to \infty$ when $\alpha = \frac{1}{2}$. Our simulation results suggest that the bounds in the proofs of these Lemmas could be improved, and that the conclusion of Theorem 2 holds without the requirement that $\alpha > \frac{1}{2}$.

The curve in Figure 1 for $E[C_L^*]$ has, for large h, a noticeably sharper upward slope than the other curves. Moreover, the $E[C_L^*]$ curve is minimized at

MISE$_m$ ($\cdot$) and E[C_L] $-$ mσ^2 (+)

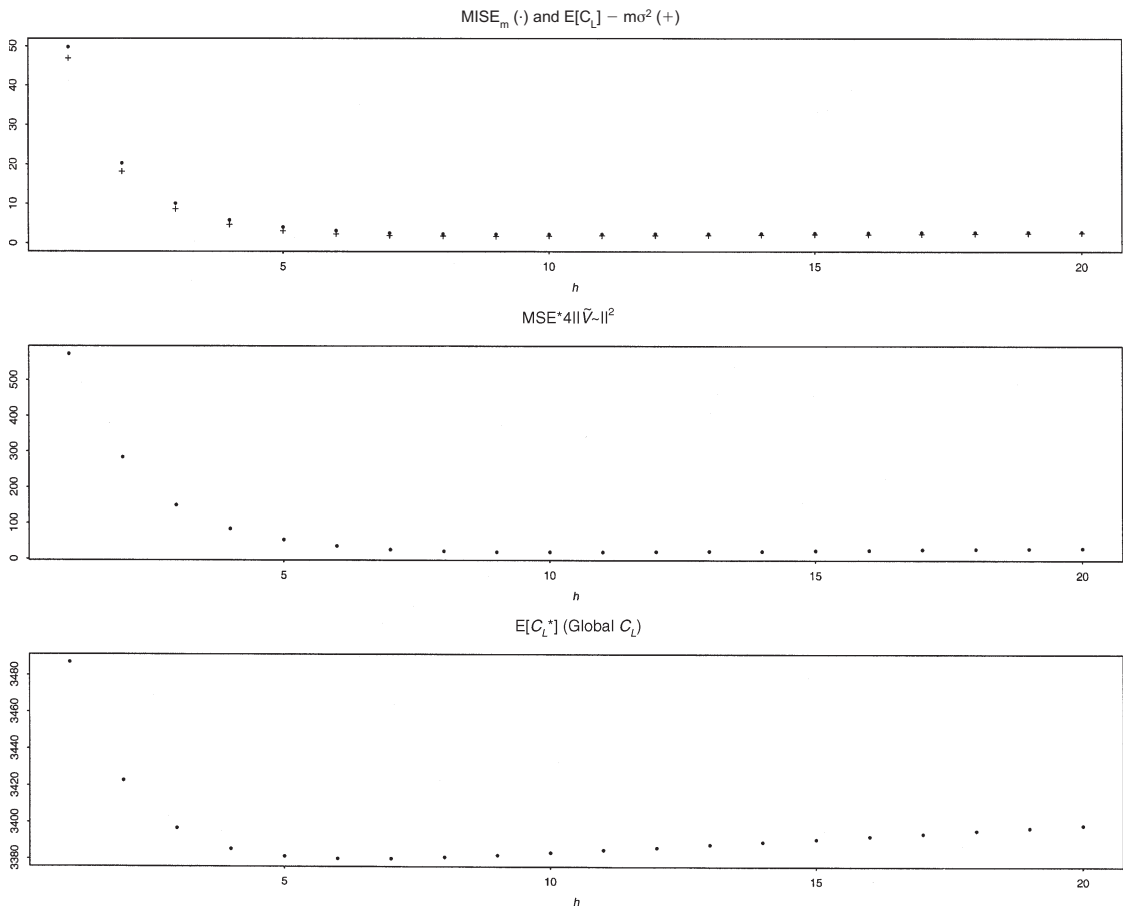MSE*4$\|\tilde{V}\sim\|^2$

E[C_L*] (Global C_L)

FIGURE 1. $n = 4096$, $\rho = 0.8$, $m = n^{0.5}$

$h = 7$, a somewhat smaller value than for the other curves ($h = 9$). Thus, the local criterion C_L provides, on the average, a better approximation than the global C_L^* criterion to the shape of the MSE curve.

Figure 2 ($n = 4096$, $\rho = 0.8$, $m = n^{0.8}$) once again shows that MISE_m and $\text{E}[C_L] - m\sigma^2$ approximate each other well, up to a small additive shift. However these curves have a shape closer to that of $\text{E}[C_L^*]$ than to that of the desired target, MSE. This is not surprising, since in this case $m/\tilde{n} = 0.38$, so the neighbourhood used for computing MISE_m is not sufficiently local to zero. Comparison of Figure 3 ($n = 1024$, $\rho = 0.8$, $m = n^{0.5}$) and Figure 4 ($n = 1024$, $\rho = 0.8$, $m = n^{0.8}$) reveals the same patterns observed above for $n = 4096$.

Next, we present some numerical results, in tabular form, for all situations studied. Let h^{opt}, $h^{\text{E}[C_L^*]}$ and $h^{\text{E}[C_L]}$ denote the minimizers of MSE, $\text{E}[C_L^*]$ and $\text{E}[C_L]$, respectively. (Note that $h^{\text{E}[C_L]}$ depends on m.) Table 1 lists values of n, ρ, h^{opt}, $h^{\text{E}[C_L^*]}$, $\text{MSE}(h^{\text{E}[C_L^*]})/\text{MSE}(h^{\text{opt}})$, $h^{\text{E}[C_L]}$ and $\text{MSE}(h^{\text{E}[C_L]})/\text{MSE}(h^{\text{opt}})$, where $\text{MSE}(h)$ denotes the mean squared error as a function of h. In each of the final two columns, two entries are given, corresponding to $m = n^{0.5}$ and $m = n^{0.8}$, respectively. The fifth and seventh columns are crude measures of the quality of the selections made by C_L^* and C_L, respectively. For a given sample size, the quantities h^{opt}, $h^{\text{E}[C_L^*]}$ and $h^{\text{E}[C_L]}$ all tend to increase with ρ. When $\rho = 0.2$, all of the averaged criteria considered here select the optimal value, $h = 1$. The quantity $h^{\text{E}[C_L]}$ with $m = n^{0.5}$ is typically very close to or identical to h^{opt}, for both values of ρ considered. For $\rho = 0.8$, $h^{\text{E}[C_L]}$ with $m = n^{0.8}$ tends to be smaller than h^{opt}, and $h^{\text{E}[C_L^*]}$ tends to be even smaller. Correspondingly, in terms of the quality measure considered here, C_L with $m = n^{0.5}$ performs virtually as well as the optimal choice h^{opt}, and is much better than the global criterion C_L^*. The local C_L criterion with $m = n^{0.5}$ performs somewhat better than the one with $m = n^{0.8}$, and both are substantially better than C_L^*.

Table 2 presents the means, over the 500 realizations, of the selected values of h. In all cases, the average for the global criterion $\text{E}[h^{C_L^*}]$ is always lowest, followed by $\text{E}[h^{C_L}]$ with $m = n^{0.8}$, followed by $\text{E}[h^{C_L}]$ with $m = n^{0.5}$. When $\rho = 0.2$, $\text{E}[h^{C_L^*}]$ comes closest to $h^{\text{opt}} = 1$. On the other hand, when $\rho = 0.8$, $\text{E}[h^{C_L}]$ with $m = n^{0.5}$ comes closest to h^{opt}. These findings are consistent with the results in Table 1. The largeness of $\text{E}[h^{C_L}]$ when $\rho = 0.2$ may be attributed to excess variability in the local criteria as compared to C_L^*. The closeness of h^{opt} and $\text{E}[h^{C_L}]$ with $m = n^{0.5}$ when $\rho = 0.8$ may be attributed to the near-unbiasedness of C_L in this case.

Table 3 presents the bias and mean squared error of the selected estimators of d, denoted by $\hat{d}_{\text{FEXP}}^{h^{C_L^*}}$ and $\hat{d}_{\text{FEXP}}^{h^{C_L}}$ with $m = n^{0.5}$ and $m = n^{0.8}$. In most cases, $\hat{d}_{\text{FEXP}}^{h^{C_L}}$ with $m = n^{0.5}$ has the least bias, followed by $\hat{d}_{\text{FEXP}}^{h^{C_L}}$ with $m = n^{0.8}$, followed by $\hat{d}_{\text{FEXP}}^{h^{C_L^*}}$. This is consistent with the results of Table 2, since as h increases, the bias of $\hat{d}_{\text{FEXP}}^{h}$ decreases. In terms of mean squared error, for $\rho = 0.2$, $\hat{d}_{\text{FEXP}}^{h^{C_L^*}}$ is clearly best, by a factor of as much as 1.8 compared to its closest competitor. Again, this is consistent with the results of Table 2, since in
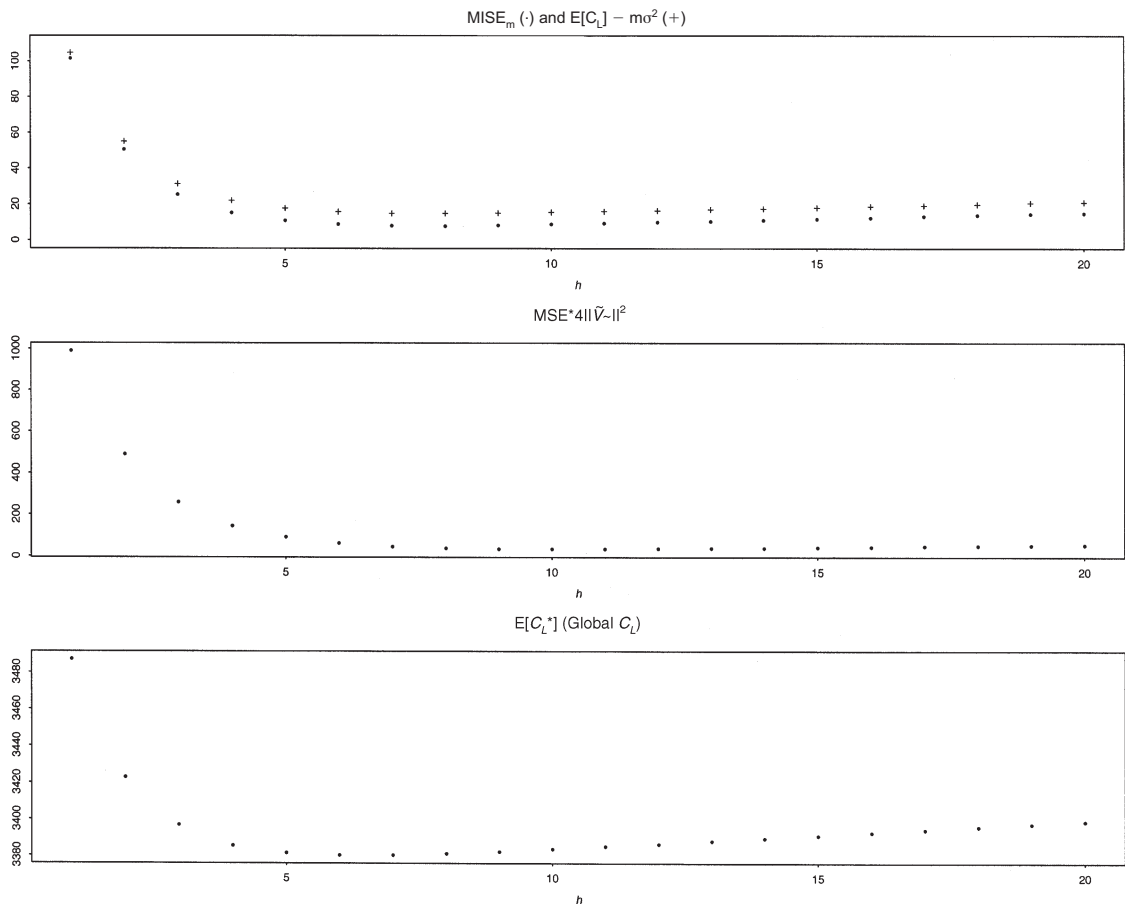
MISE$_m$ ($\cdot$) and E[C_L] $-$ $m\sigma^2$ ($+$)

MSE*4$\|\tilde{V}\|^2$

E[C_L*] (Global C_L)

h

FIGURE 2. $n = 4096$, $\rho = 0.8$, $m = n^{0.8}$

MISE$_m$ ($\cdot$) and E[C_L] − mσ^2 (+)

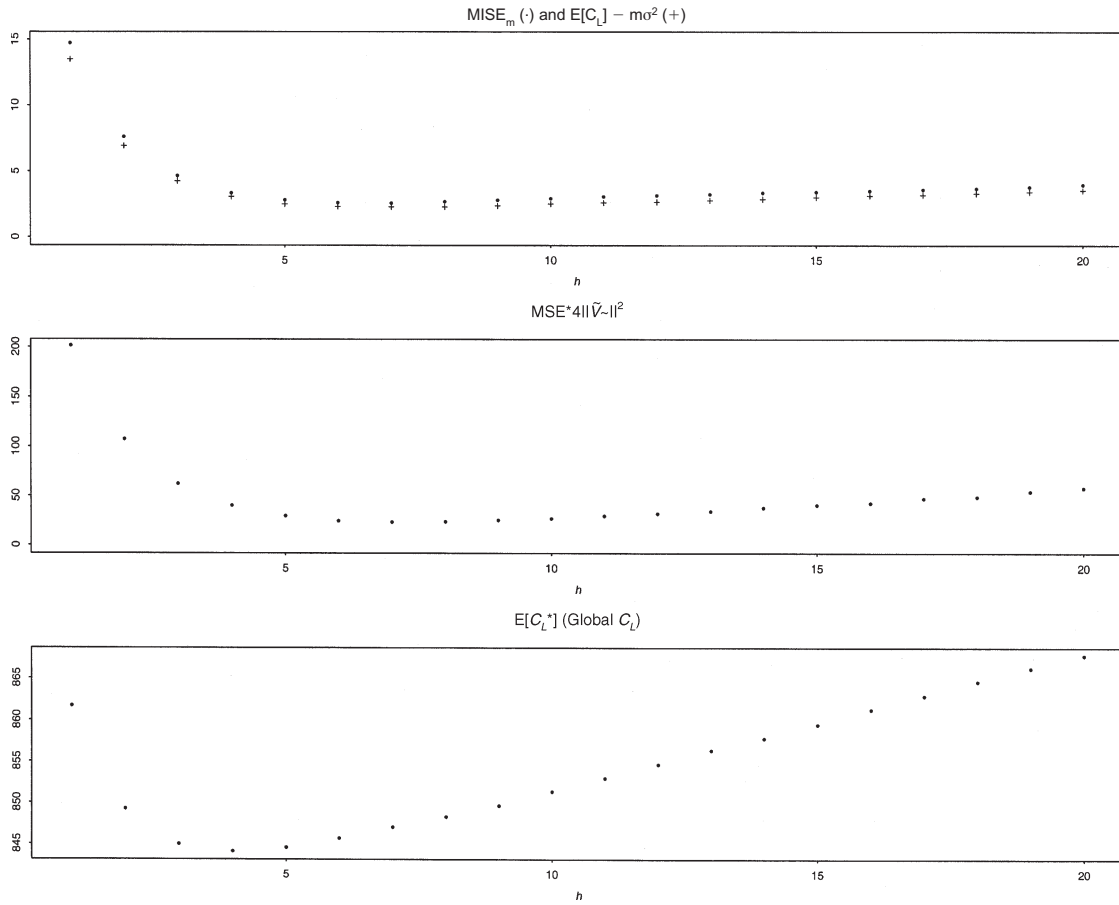MSE*4$\|\tilde{V}\sim\|^2$

E[C_L^*] (Global C_L)

FIGURE 3. $n = 1024$, $\rho = 0.8$, $m = n^{0.5}$
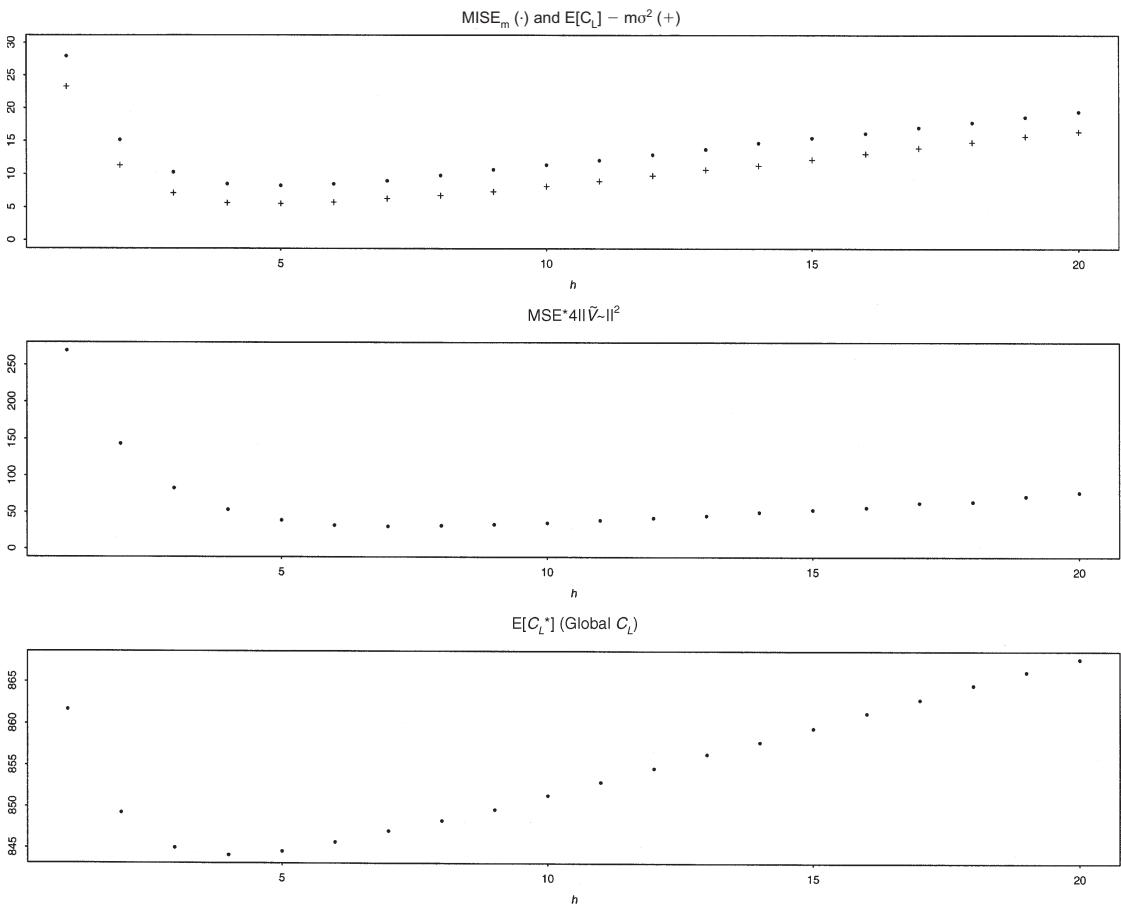
FIGURE 4. $n = 1024$, $\rho = 0.8$, $m = n^{0.8}$

TABLE 1

PERFORMANCE OF AVERAGED SELECTION CRITERIA

n	ρ	h^{opt}	$h^{\text{E}[C_{\text{L}}^*]}$	$\dfrac{\text{MSE}(h^{\text{E}[C_{\text{L}}^*]})}{\text{MSE}(h^{\text{opt}})}$	$h^{\text{E}[C_{\text{L}}]}$	$\dfrac{\text{MSE}(h^{\text{E}[C_{\text{L}}]})}{\text{MSE}(h^{\text{opt}})}$
512	0.2	1	1	1.0	1, 1	1.0, 1.0
512	0.8	7	3	1.78	5, 4	1.02, 1.27
1024	0.2	1	1	1.0	1, 1	1.0, 1.0
1024	0.8	7	4	1.73	7, 5	1.0, 1.27
4096	0.2	1	1	1.0	1, 1	1.0, 1.0
4096	0.8	9	7	1.37	9, 8	1.0, 1.12

TABLE 2

MEANS OF SELECTED VALUES OF h

n	ρ	h^{opt}	$\text{E}[h^{C_{\text{L}}^*}]$	$\text{E}[h^{C_{\text{L}}}], \; m = n^{0.5}$	$\text{E}[h^{C_{\text{L}}}], \; m = n^{0.8}$
512	0.2	1	1.80	3.65	2.48
512	0.8	7	3.64	5.79	4.28
1024	0.2	1	1.98	4.04	2.62
1024	0.8	7	4.71	7.21	5.49
4096	0.2	1	2.04	3.65	3.03
4096	0.8	9	6.78	8.54	8.32

TABLE 3

PERFORMANCE OF THE SELECTED ESTIMATORS OF d

n	ρ	$Bias(\hat{d}^{h^{C_{\text{L}}^*}})$	$\text{MSE}(\hat{d}^{h^{C_{\text{L}}^*}})$	$Bias(\hat{d}^{h^{C_{\text{L}}}})$ $m = n^{0.5}$	$\text{MSE}(\hat{d}^{h^{C_{\text{L}}}})$ $m = n^{0.5}$	$Bias(\hat{d}^{h^{C_{\text{L}}}})$ $m = n^{0.8}$	$\text{MSE}(\hat{d}^{h^{C_{\text{L}}}})$ $m = n^{0.8}$
512	0.2	0.0224	0.0230	0.0104	0.0908	0.0153	0.0424
512	0.8	0.1982	0.1073	0.0987	0.1283	0.1688	0.1121
1024	0.2	0.0179	0.0093	0.0154	0.0301	0.0208	0.0153
1024	0.8	0.1391	0.0479	0.0658	0.0492	0.1137	0.0446
4096	0.2	0.0104	0.0019	0.0126	0.0049	0.0108	0.0030
4096	0.8	0.0657	0.0115	0.0392	0.0098	0.0428	0.0094

this case the best criterion is presumably the one which selects $h = 1$ the most frequently. When $\rho = 0.8$, in terms of mean squared error, the criteria perform more similarly to each other, with $\hat{d}_{\text{FEXP}}^{h^{C_{\text{L}}^*}}$ performing best by 4% for $n = 512$, and $\hat{d}_{\text{FEXP}}^{h^{C_{\text{L}}}}$ with $m = n^{0.8}$ outperforming $\hat{d}_{\text{FEXP}}^{h^{C_{\text{L}}^*}}$ by 7% and 22% for $n = 1024$ and $n = 4096$, respectively. For both values of ρ and for all sample sizes, in terms of mean squared error, $\hat{d}_{\text{FEXP}}^{h^{C_{\text{L}}}}$ with $m = n^{0.8}$ performed better than $\hat{d}_{\text{FEXP}}^{h^{C_{\text{L}}}}$ with $m = n^{0.5}$, although the differences were much more pronounced for $\rho = 0.2$ than for $\rho = 0.8$. Table 3 also reveals that for a given value of ρ, and

for a given selection criterion, the mean squared error of the selected estimator decreases monotonically as n increases.

5. DISCUSSION

We have demonstrated the asymptotic unbiasedness, up to irrelevant additive and multiplicative constants, of a local C_L criterion for the mean squared error (MSE) of a fractional exponential estimator of the memory parameter. This is done by first establishing the asymptotic equivalence of a (local to zero) frequency domain mean integrated squared error and MSE, and then showing that the local C_L is asymptotically unbiased for the corresponding local mean integrated squared error (MISE). The need to work in a local neighbourhood of the origin so that MISE provides a good approximation to MSE arose in a similar context in Hurvich and Beltrao (1994). Our results are analogous in spirit to those of Findley (1985) and Bhansali (1986), who established the asymptotic unbiasedness of the Akaike information criterion (AIC) for the Kullback–Leibler information in autoregressive model selection, assuming that the true model is an infinite order autoregression. As pointed out by Findley (1985), unbiasedness of a model selection criterion is a basic property, worthy of study. We have not, however, established the stronger result that the minimizers of the local C_L and MSE are asymptotically equivalent. Such a result would be more closely analogous to those of Shibata (1980) on the asymptotic efficiency of AIC for autoregressive model selection, and those of Li (1987) on the asymptotic efficiency of C_L for model selection in linear regression. The asymptotic efficiency of a global version of C_L for the global MISE has been established by Moulines and Soulier (2000), assuming a power law decay of the γ_k, but this result does not have a direct bearing on the MSE of the resulting estimator of d.

It seems that there is a trade-off between the quality of model selection methods and the quality of the FEXP estimator of d. If the γ_k decay exponentially fast, as assumed here, then the optimal rate of convergence for the mean squared error of the FEXP estimator is very good, $(\log n)/n$, although it seems difficult to derive strong results on the properties of the estimator selected by C_L. On the other hand, if the γ_k decay as a power law, it may be possible to obtain stronger results on the properties of the estimator of d selected by the local C_L, but then the best possible rate of convergence for the estimator of d is slower than described above. A similar trade-off is known for model selection in linear regression by AIC; see, for example, Hurvich and Tsai (1995).

Another issue not fully resolved is the appropriate choice of the number of frequencies, m, to be used in the local C_L. For MSE and the local MISE to approximate each other well, a reasonably small value of m should be used. Using too small a value for m, however, will result in an excessively variable local C_L criterion. Partly for this reason, the quality of the approximation of the local C_L to the local MISE will degrade if m tends to ∞ at too slow a rate.

Our simulations indicated that the use of $m = n^{0.5}$ was preferable to $m = n^{0.8}$ from the point of view of bias of the local C_L for the shape of the MSE curve, but that the opposite preference holds from the point of view of the mean squared error of the estimator of d. It may seem undesirable to introduce the somewhat arbitrary parameter m in order to obtain a method of selecting the number of parameters to use in an estimator of d. It is noteworthy, however, that similarly arbitrary parameters were needed in treating the analogous problem of bandwidth selection in the Gaussian semiparametric estimator of d (Henry and Robinson, 1996), and in the GPH estimator (Hurvich and Deo, 1999).

APPENDIX 1

Assumption 1 merits further discussion. It would clearly be satisfied if $\gamma_k \sim Ca^k/k^p$, though this specification implies that the γ_k are eventually all of the same sign. If $\{Y_t^*\} = \{(1-B)^d Y_t\}$ is a stationary AR(1) process, $Y_t^* = \alpha_1 Y_{t-1}^* + e_t$ where $\{e_t\}$ is white noise with zero mean, $|\alpha_1| < 1$, $\alpha_1 \neq 0$, it can be shown that $\gamma_k = 2\alpha_1^k/k$ for $k > 0$. Setting $\alpha_1 = ae^{i\theta}$ where $a = |\alpha_1|$ and $\theta = 0$ or π according to whether α_1 is positive or negative, respectively, we obtain

$$\frac{1}{2} \sum_{k=h+1}^{\infty} \frac{\gamma_k}{k} = \sum_{k=h+1}^{\infty} \frac{(ae^{i\theta})^k}{k^2} = \frac{(ae^{i\theta})^{h+1}}{(h+1)^2} \sum_{k=0}^{\infty} \left(\frac{h+1}{h+1+k} \right)^2 (ae^{i\theta})^k \sim \frac{(ae^{i\theta})^{h+1}}{(h+1)^2(1-ae^{i\theta})}$$

where we have used the Dominated Convergence Theorem, together with the absolute summability of a geometric series. Thus,

$$\frac{h}{2} \sum_{k=h+1}^{\infty} \frac{\gamma_k}{k} \sim \frac{a^h}{h} G_h$$

where $G_h = ae^{i(h+1)\theta}/(1-ae^{i\theta})$, so that Assumption 1 with $p = 1$ is satisfied for the AR(1) model. Note that in this case G_h may be expressed as $G_h = \alpha_1/(1-\alpha_1)$ if $\alpha_1 > 0$, or $G_h = (-1)^h \alpha_1/(1-\alpha_1)$ if $\alpha_1 < 0$. A similar argument shows that Assumption 1 holds for the invertible MA(1) model, since the log spectral density of any MA series is the negative of that of a corresponding AR series. Next, we consider the stationary AR(2) model

$$Y_t^* = \alpha_1 Y_{t-1}^* + \alpha_2 Y_{t-2}^* + e_t$$

Denote the roots of the characteristic polynomial $z^2 - \alpha_1 z - \alpha_2$ by r_1, r_2. If these roots are real with $|r_1| > |r_2|$, then for $k > 0$,

$$\gamma_k = \frac{2r_1^k}{k} + \frac{2r_2^k}{k} \sim \frac{2r_1^k}{k}$$

so that Assumption 1 is satisfied with $p = 1$ and with $a = |r_1|$, by the argument given above for the AR(1) case. If the roots are real with $r_1 = r_2$, Assumption 1 holds. If the roots are real with $r_1 = -r_2$ then $\gamma_k = 0$ for $k > 0$, and Assumption 1 does not hold. Then the series is an FEXP model of order 0, and one could attain MSE $\sim C/n$ by holding h constant as $n \to \infty$. If the roots are complex with $r_1 = ae^{i\theta}$, $r_2 = ae^{-i\theta}$ then

$$\gamma_k = \frac{2(ae^{i\theta})^k}{k} + \frac{2(ae^{-i\theta})^k}{k} = \frac{4\cos(k\theta)a^k}{k} \qquad \text{for } k > 0$$

Assuming that

$$\inf_h |\cos(h+1)\theta - a\cos(h\theta)| > 0$$

we obtain

$$\frac{1}{2}\sum_{k=h+1}^{\infty}\frac{\gamma_k}{k} = \sum_{k=h+1}^{\infty}\frac{(ae^{i\theta})^k + (ae^{-i\theta})^k}{k^2} \sim \frac{(ae^{i\theta})^{h+1}}{(h+1)^2(1 - ae^{i\theta})} + \frac{(ae^{-i\theta})^{h+1}}{(h+1)^2(1 - ae^{-i\theta})}$$

$$= \frac{2a^{h+1}}{(h+1)^2}\frac{\cos(h+1)\theta - a\cos(h\theta)}{1 - 2a\cos\theta + a^2}$$

so that Assumption 1 holds with $p = 1$ and

$$G_h = \frac{2a[\cos(h+1)\theta - a\cos(h\theta)]}{1 - 2a\cos\theta + a^2}$$

If θ is a rational multiple of π then $\cos(h\theta)$ will only assume a finite number of values so all but finitely many choices of a will lead to a model which satisfies Assumption 1. Unfortunately, if θ is not a rational multiple of π, there will typically exist no positive lower bound for $|\cos(h+1)\theta - a\cos(h\theta)|$ so that Assumption 1 will fail.

For general stationary and invertible ARMA models, the γ_k are sums of terms of the two forms seen above for real and complex roots, respectively. Assumption 1 holds with $p = 1$ if the root of largest magnitude of either the autoregressive or moving average characteristic polynomial is real, and all other roots are of smaller magnitude. Assumption 1 with $p = 1$ also holds if the two roots of largest magnitude are complex, say $ae^{i\theta}$, $ae^{-i\theta}$, all other roots are of smaller magnitude, and

$$\inf_h |\cos(h+1)\theta - a\cos(h\theta)| > 0$$

APPENDIX 2

PROOF OF LEMMA 1. Index the columns of X by $k = 0, \ldots, h+1$. We can write

$$X'X = \begin{bmatrix} X_h'X_h & c \\ c' & \tilde{V}'\tilde{V} \end{bmatrix} \qquad (X'X)^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

where $c = X_h'\tilde{V} = (c_0, \ldots, c_h)'$, and the entries of $(X'X)^{-1}$ are given by the formula for the inverse of a partitioned matrix. By Lemma 1 of Hurvich and Brodsky (1997), $c_0 = O(\log n)$, and $c_k = -n/(4k) + O(\log n)$ for $k = 1, \ldots, h$. By (10) and Lemma 2 of Hurvich and Brodsky (1997),

$$(X_h'X_h)^{-1} = A^{-1} + M$$

where A is a diagonal matrix with diagonal entries $(n/2, n/4, \ldots, n/4)$, and all entries of M are $\mathrm{O}(1/n^2)$, except for the upper lefthand corner, which is $\mathrm{O}(1/n)$. These properties of M follow from the proof of Lemma 3 of Hurvich and Brodsky (1997), which assumes that $h/n \to 0$. By Lemma 4 of Hurvich and Brodsky (1997), $\tilde{V}'\tilde{V} = n\pi^2/24 + \mathrm{O}(\log^2 n)$.

We focus first on B_{22}, a scalar given by $B_{22} = [\tilde{V}'\tilde{V} - c'(A^{-1} + M)c]^{-1}$, so that

$$B_{22}^{-1} = \frac{n\pi^2}{24} - c'A^{-1}c - c'Mc + \mathrm{O}(\log^2 n)$$

$$= \frac{n\pi^2}{24} - \left(\frac{2}{n}\right)c_0^2 - \left(\frac{4}{n}\right)\sum_{k=1}^{h} c_k^2 - \sum_{j=0}^{h}\sum_{k=0}^{h} c_j M_{jk} c_k + \mathrm{O}(\log^2 n)$$

$$= \frac{n\pi^2}{24} - \frac{n}{4}\sum_{k=1}^{h}\frac{1}{k^2} - c_0^2 M_{00} - 2c_0\sum_{k=1}^{h} c_k M_{0k} - \sum_{j=1}^{h}\sum_{k=1}^{h} c_j M_{jk} c_k + \mathrm{O}(\log^2 n)$$

$$= \frac{n}{4}\left(\frac{\pi^2}{6} - \sum_{k=1}^{h}\frac{1}{k^2}\right) + \mathrm{O}(\log^2 n)$$

$$= \frac{n}{4h} + \mathrm{O}(\log^2 n).$$

Thus, $B_{22} \sim 4h/n$. Next, consider $B_{11} = (A^{-1} + M) + B_{22}(A^{-1} + M)cc'(A^{-1} + M)$, where $(B_{11})_{jk} = (X'X)_{jk}^{-1}$ for $j, k = 0, \ldots, h$. From Lemma 1 of Hurvich and Brodsky (1997),

$$cc' = \mathrm{O}\begin{bmatrix} \log^2 n & n\log n\xi_h' \\ n\log n\xi_h & n^2\xi_h' \otimes \xi_h \end{bmatrix}$$

Similarly, we obtain

$$A^{-1}cc'A^{-1} = \mathrm{O}\left(\frac{cc'}{n^2}\right)$$

$$= \mathrm{O}\begin{bmatrix} \dfrac{\log^2 n}{n^2} & \dfrac{\log n}{n}\xi_h' \\[3mm] \dfrac{\log n}{n}\xi_h & \xi_h' \otimes \xi_h \end{bmatrix}$$

$$A^{-1}cc'M = \mathrm{O}\begin{bmatrix} \dfrac{\log^2 n}{n^2} & \dfrac{\log^2 n}{n^2}1_h' \\[3mm] \dfrac{\log n}{n}\xi_h & \dfrac{\log n}{n}1_h' \otimes \xi_h \end{bmatrix}$$

$$Mcc'A^{-1} = (A^{-1}cc'M)'$$

$$= \mathrm{O}\begin{bmatrix} \dfrac{\log^2 n}{n^2} & \dfrac{\log n}{n}\xi_h' \\[3mm] \dfrac{\log^2 n}{n^2}1_h & \dfrac{\log n}{n}1_h \otimes \xi_h' \end{bmatrix}$$

$$Mcc'M = \mathrm{O}\begin{bmatrix} \left(\begin{array}{cc} \dfrac{1}{n} & \dfrac{1}{n^2}1_h' \\[3mm] \dfrac{1}{n^2}1_h & \dfrac{1}{n^2}1_h' \otimes 1_h \end{array}\right) \left(\begin{array}{cc} \dfrac{\log^2 n}{n} & \dfrac{\log^2 n}{n}1_h' \\[3mm] \log n\,\xi_h & \log n 1_h' \otimes \xi_h \end{array}\right) \end{bmatrix}$$

$$= \mathrm{O}\begin{bmatrix} \dfrac{\log^2 n}{n^2} & \dfrac{\log^2 n}{n^2}1_h' \\[3mm] \dfrac{\log^2 n}{n^2}1_h & \dfrac{\log^2 n}{n^2}1_h' \otimes 1_h \end{bmatrix}$$

Thus,

$$(A^{-1} + M)cc'(A^{-1} + M) = \mathrm{O}\begin{bmatrix} \dfrac{\log^2 n}{n^2} & \dfrac{\log n}{n}\xi_h' \\[3mm] \dfrac{\log n}{n}\xi_h & \xi_h' \otimes \xi_h \end{bmatrix}$$

and

$$B_{11} = \frac{4}{n}I_{h+1} + \mathrm{O}\begin{bmatrix} \dfrac{1}{n} & \dfrac{h\log n}{n^2}\xi_h' \\[3mm] \dfrac{h\log n}{n^2}\xi_h & \dfrac{h}{n}\xi_h' \otimes \xi_h \end{bmatrix}$$

Finally, we have

$$B_{12} = B_{21}' = -B_{22}(A^{-1} + M)c = \mathrm{O}\begin{bmatrix} \dfrac{h\log n}{n^2} \\[3mm] \dfrac{h}{n}\xi_h \end{bmatrix}$$

PROOF OF LEMMA 2. Using the notation established in Lemma 1, we have

$$H = X(X'X)^{-1}X'$$

$$= [X_h, \tilde{V}] \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} X'_h \\ \tilde{V}' \end{bmatrix}$$

$$= X_h B_{11} X'_h + \tilde{V} B_{21} X'_h + X_h B_{12} \tilde{V}' + \tilde{V} B_{22} \tilde{V}'$$

All entries of X_h are O(1). All entries of $\tilde{V}$ are O(log n). By Lemma 1, all entries in the kth column of $X_h B_{11}$ are

$$\mathrm{O}\left[\sum_{j=0}^{h} |(B_{11})_{jk}| \right] = \mathrm{O}\left(\frac{1}{n} + \frac{h \log^2 n}{n^2} \right) \qquad \text{if } k = 0$$

$$= \mathrm{O}\left(\frac{1}{n} + \frac{h \log h}{nk} \right) \qquad \text{if } k > 0$$

Thus, each entry of $X_h B_{11} X'_h$ is

$$\mathrm{O}\left[\frac{1}{n} + \sum_{k=1}^{h} \left(\frac{1}{n} + \frac{h \log h}{nk} \right) \right] = \mathrm{O}\left(\frac{h \log^2 h}{n} \right)$$

Next, we note that $\tilde{V} B_{21}$ is an $\tilde{n} \times (h+1)$ matrix, with

$$(\tilde{V} B_{21})_{jk} = \mathrm{O}\left[\frac{h \log^2 n}{n^2} \right] \qquad \text{if } k = 0$$

$$= \mathrm{O}\left(\frac{h \log n}{nk} \right) \qquad \text{if } k > 0$$

Thus, each entry of $\tilde{V} B_{21} X'_h$ is

$$\mathrm{O}\left(\frac{h \log^2 n}{n^2} + \sum_{k=1}^{h} \frac{h \log n}{nk} \right) = \mathrm{O}\left(\frac{h \log^2 n}{n} \right)$$

The same bound holds for the entries of $X_h B_{12} \tilde{V}'$. Finally, since $B_{22} = \mathrm{O}(h/n)$, we find that each entry of $\tilde{V} B_{22} \tilde{V}'$ is

$$\mathrm{O}\left(\frac{h \log^2 n}{n} \right)$$

The proof is complete.

PROOF OF LEMMA 3. We have

$$\| \tilde{V} \|_W^2 = \sum_{j=1}^{m} \log^2 \left[2 \sin \left(\frac{\omega_j}{2} \right) \right]$$

$$= \sum_{j=1}^{m} \log^2 \omega_j [1 + \mathrm{o}(1)]$$

$$= \sum_{j=1}^{m} \left[\log \left(\frac{2\pi}{n} \right) + \log(j) \right]^2 [1 + \mathrm{o}(1)]$$

Since

$$\log^2 x = \frac{\mathrm{d}}{\mathrm{d}x} [x(\log x - 1)^2 + x]$$

and

$$\log x = \frac{\mathrm{d}}{\mathrm{d}x} [x \log x - x]$$

we conclude that

$$\sum_{j=1}^{m} \log^2 \omega_j = \left[m \log^2 \left(\frac{2\pi}{n} \right) + m(\log m - 1)^2 + m + 2 \log \left(\frac{2\pi}{n} \right) (m \log m - m) \right] [1 + \mathrm{o}(1)]$$

$$= \left\{ m(\log^2 n + \log^2 m - 2 \log n \log m) + \mathrm{o} \left[m \log^2 \left(\frac{n}{m} \right) \right] \right\} \{ 1 + \mathrm{o}(1) \}$$

$$= m \log^2 \left(\frac{n}{m} \right) [1 + \mathrm{o}(1)]$$

PROOF OF LEMMA 4. Define $\beta^* = \mathrm{E}[\hat{\beta}] = (X'X)^{-1} X'[\mu + \mathrm{E}(\varepsilon)]$, and let β_h^* be the $(h+1)$-dimensional column vector obtained by deleting the final element of β^*. We have

$$X_h(\hat{\beta}_h - \beta_h) = X_h(\hat{\beta}_h - \beta_h^*) + X_h(\beta_h^* - \beta_h)$$

It suffices to show that

$$\mathrm{E} \| X_h(\hat{\beta}_h - \beta_h^*) \|_W^2 = \mathrm{O}[m \, \mathrm{MSE} \, \log^2 h] \tag{6}$$

and

$$\| X_h(\beta_h^* - \beta_h) \|_W^2 = \mathrm{O}[m \, \mathrm{MSE} \, \log^2 h] \tag{7}$$

We focus first on establishing (7). We have

$$\beta^* - \beta = (X'X)^{-1} X'[\mu - X\beta + \mathrm{E}(\varepsilon)]$$

$$= \begin{bmatrix} B_{11} X_h' + B_{12} \tilde{V}' \\ B_{21} X_h' + B_{22} \tilde{V}' \end{bmatrix} [\mu - X\beta + E(\varepsilon)]$$

Thus,

$$\beta_h^* - \beta_h = (B_{11} X_h' + B_{12} \tilde{V}')[\mu - X\beta + E(\varepsilon)]$$

Consider first $(B_{11} X_h' + B_{12} \tilde{V}') \mathrm{E}(\varepsilon)$, which is an $(h+1)$-dimensional column vector.

From Lemmas 2 and 5 of Hurvich *et al*. (1998), and from Lemmas 7 and 8 of Hurvich and Brodsky (1997), it follows that

$$\lim_n \sup_{1 \leqslant j \leqslant \log^2 n} |\mathrm{E}(\varepsilon_j)| < \infty$$

and that

$$|\mathrm{E}(\varepsilon_j)| = \mathrm{O}\left[\frac{\log n}{j}\right] \qquad \text{for } j = \log^2 n, \ldots, \tilde{n}$$

Thus, each entry of $X_h' \mathrm{E}(\varepsilon)$ is

$$\mathrm{O}\left[\sum_{j=1}^{\tilde{n}} (1)|\mathrm{E}(\varepsilon_j)|\right] = \mathrm{O}(\log^2 n)$$

and

$$\tilde{V}' E(\varepsilon) = \mathrm{O}\left[\log^2 n \sum_{j=1}^{\tilde{n}} |E(\varepsilon_j)|\right] = \mathrm{O}(\log^4 n)$$

Thus,

$(B_{11}X_h' + B_{12}\tilde{V}')E(\varepsilon)$

$$= \mathrm{O}\begin{bmatrix} \log^2 n\left(\dfrac{1}{n} + \displaystyle\sum_{k=1}^{h} \dfrac{h\log n}{n^2 k}\right) + \log^4 n\left(\dfrac{h\log n}{n^2}\right) \\[2em] \log^2 n\left\{\dfrac{h\log n}{n^2} + \displaystyle\sum_{k=1}^{h}\left(\dfrac{h}{nk} + \dfrac{1}{n}\chi_{\{k=1\}}\right)\right\} + \log^4 n\left(\dfrac{h}{n}\right) \\[2em] \vdots \\[1em] \log^2 n\left\{\dfrac{h\log n}{n^2 j} + \displaystyle\sum_{k=1}^{h}\left(\dfrac{h}{njk} + \dfrac{1}{n}\chi_{\{k=j\}}\right)\right\} + \log^4 n\left(\dfrac{h}{nj}\right) \\[2em] \vdots \end{bmatrix}$$

$$= \mathrm{O}\begin{bmatrix} \dfrac{\log^2 n}{n} \\[1.5em] \dfrac{h\log^4 n}{n}\xi_h + \dfrac{\log^2 n}{n} \end{bmatrix} + \mathrm{O}\begin{bmatrix} \dfrac{\log^2 n}{n} \\[1.5em] \dfrac{h\log^4 n}{n}\xi_h \end{bmatrix}$$

Since the (j, k) entry of $X_h' W X_h$ is $\mathrm{O}(m)$, we conclude that

$\|X_h(B_{11}X_h' + B_{12}\tilde{V}')E(\varepsilon)\|_W^2$

$$= O\left(\frac{m\log^4 n}{n^2} + \frac{m\log^2 n}{n}\sum_{j=1}^{h}\frac{h\log^4 n}{nj} + m\sum_{j=1}^{h}\sum_{k=1}^{h}\frac{h^2\log^8 n}{n^2 jk}\right)$$

$$= O\left[m\left(\frac{h^2}{n^2}\right)\log^{10} n\right] = o[m\,\mathrm{MSE}\log^2 h]$$

Next, we consider $(B_{11}X'_h + B_{12}\tilde{V}')(\mu - X\beta)$. Define

$$g_j = V'_j(\mu - X\beta) \qquad \text{and} \qquad \tilde{g} = \tilde{V}'(\mu - X\beta)$$

Then

$(B_{11}X'_h + B_{12}\tilde{V}')(\mu - X\beta)$

$$= O\begin{bmatrix} \frac{1}{n}|g_0| + \sum_{k=1}^{h}|g_k|\frac{h\log n}{n^2 k} \\[2ex] |g_0|\frac{h\log n}{n^2} + \sum_{k=1}^{h}|g_k|\left(\frac{h}{nk} + \frac{1}{n}\chi_{\{k=1\}}\right) \\[2ex] \vdots \\[2ex] |g_0|\frac{h\log n}{n^2 j} + \sum_{k=1}^{h}|g_k|\left(\frac{h}{njk} + \frac{1}{n}\chi_{\{k=j\}}\right) \\[2ex] \vdots \end{bmatrix} + O\begin{bmatrix} |\tilde{g}|\frac{h\log n}{n^2} \\[2ex] |\tilde{g}|\frac{h}{n} \\[2ex] \vdots \\[2ex] |\tilde{g}|\frac{h}{nj} \\[2ex] \vdots \end{bmatrix} \qquad (8)$$

We now show that, under the assumptions given in Section 3, $|g_k| = O[\mathrm{MSE}^{\frac{1}{2}}]$ uniformly for $0 \leqslant k \leqslant h$, and that $(h/n)|\tilde{g}| = O[\mathrm{MSE}^{\frac{1}{2}}]$. For $0 \leqslant k \leqslant h$, we have

$$V'_k(\mu - X\beta) = V'_k\sum_{j=h+1}^{\infty}\gamma_j V_j = \sum_{j=h+1}^{\infty}\gamma_j V'_k V_j$$

From Equation (10) of Hurvich and Brodsky (1997), it follows that

$$|g_k| = O\left[\sum_{j=h+1}^{\infty}|\gamma_j|(1 + n\chi_{\{j\bmod n=k\}} + n\chi_{\{j\bmod n=n-k\}})\right]$$

$$= O\left[\sum_{j=h+1}^{\infty}|\gamma_j| + n(|\gamma_{n+k}| + |\gamma_{2n+k}| + \cdots) + n(|\gamma_{n-k}| + |\gamma_{2n-k}| + \cdots)\right]$$

$$= O\left(\sum_{L=h+1}^{\infty}|\gamma_L| + n\sum_{L=1}^{\infty}|\gamma_{Ln}| + n\sum_{L=1}^{\infty}|\gamma_{Ln-n/2}|\right)$$

$$= O(a^h/h^p) + O(na^{\tilde{n}})$$

$$= O(\mathrm{MSE}^{\frac{1}{2}})$$

Next, from Lemma 1 and Theorem 1 of Hurvich and Brodsky (1997), we have

$$\frac{h}{n}\tilde{g} = \frac{h}{n}\tilde{V}'(\mu - X\beta)$$

$$= \frac{h}{n}\sum_{j=h+1}^{\infty}\gamma_j\tilde{V}'V_k$$

$$= \frac{h}{n}\sum_{j=h+1}^{\infty}\gamma_j\left[\frac{-n/4}{j\bmod n}\chi_{\{1\leqslant j\bmod n\leqslant\tilde{n}\}} + \mathrm{O}(\log n)\right]$$

$$+ \frac{h}{n}\sum_{j=n/2+1}^{\infty}\gamma_j\left[\frac{-n/4}{n-j\bmod n}\chi_{\{\frac{n}{2}+1\leqslant j\bmod n\leqslant n-h-1\}} + \mathrm{O}(\log n)\right]$$

$$= \mathrm{O}\left(\frac{a^h}{h^p}\right)$$

Thus,

$$\left(\frac{h}{n}\right)^2\frac{|\tilde{g}|^2}{\mathrm{MSE}} = \mathrm{O}\left[\frac{\dfrac{a^{2h}}{h^{2p}}}{\dfrac{a^{2h}}{h^{2p}} + \dfrac{h}{n}}\right] = \mathrm{O}(1)$$

It follows from (8) that

$$(B_{11}X_h' + B_{12}\tilde{V}')(\mu - X\beta) = \mathrm{O}\left[\begin{array}{c}\mathrm{MSE}^{\frac{1}{2}}\dfrac{h\log h}{n}\\[2mm]\mathrm{MSE}^{\frac{1}{2}}\dfrac{h\log h}{n}\xi_h\end{array}\right]$$

$$+ \mathrm{O}\left[\begin{array}{c}\mathrm{MSE}^{\frac{1}{2}}\\[2mm]\mathrm{MSE}^{\frac{1}{2}}\xi_h\end{array}\right]$$

We conclude that

$$\|X_h(B_{11}X_h' + B_{12}\tilde{V}')(\mu - X\beta)\|_W^2 = \mathrm{O}\left[m\,\mathrm{MSE}\left(\frac{h^2\log^4 h}{n^2} + \log^2 h\right)\right]$$

$$= \mathrm{O}(m\,\mathrm{MSE}\log^2 h)$$

thereby establishing (7).

It remains to prove (6). We have

$$E\|X_h(\hat{\beta}_h - \beta_h^*)\|_W^2 = \mathrm{E}[(\hat{\beta}_h - \beta_h^*)'X_h'WX_h(\hat{\beta}_h - \beta_h^*)]$$

$$= \sum_{j=0}^{h}\sum_{k=0}^{h}[\mathrm{Cov}(\hat{\beta})]_{jk}V_j'WV_k$$

where $Cov(\hat{\beta})$ is the covariance matrix of $\hat{\beta}$, given by

$$Cov(\hat{\beta}) = \mathrm{E}[(\hat{\beta} - \beta^*)(\hat{\beta} - \beta^*)'] = (X'X)^{-1}X'Cov(\varepsilon)X(X'X)^{-1}$$

and $Cov(\varepsilon)$ is the covariance matrix of ε. It follows from the proof of Theorem 2 of Moulines and Soulier (1999) that we can write $\varepsilon = \eta + r$ where $\mathrm{E}(\eta_j) = 0$, $\mathrm{Var}(\eta_j) = \sigma^2$,

$|r_j| = \mathrm{O}[(\log j)/j]$ (a non-stochastic bound) uniformly for $1 \leqslant j \leqslant \tilde{n}$, and $Cov(\eta_j, \eta_k) = \mathrm{O}[(\log^2 n)k^{-2|d|}j^{2|d|-2}]$ uniformly for $1 \leqslant k < j \leqslant \tilde{n}$. The conditions required for this theorem are clearly met, in view of the exponential decay of the $\{\gamma_j\}$ sequence. Let $\tilde{r} = r - \mathrm{E}(r)$. We have

$$Cov(\varepsilon) = Cov(\eta) + Cov(r) + 2\mathrm{E}(\tilde{r}\eta')$$

The (j, k)th entry of $Cov(r)$ is $\mathrm{O}((\log j)(\log k)/jk)$ uniformly for $j, k = 1, \ldots, \tilde{n}$. Writing $Cov(\eta) = \sigma^2 I + \sum$, where I is an $(\tilde{n} \times \tilde{n})$ identity matrix, we obtain

$$Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1} + (X'X)^{-1}X'[\Sigma + Cov(r) + 2E(\tilde{r}\eta')]X(X'X)^{-1}$$

so that

$$\mathrm{E}\|X_h(\hat{\beta}_h - \beta_h^*)\|_W^2$$

$$= \sum_{j=0}^{h}\sum_{k=0}^{h} V_j' W V_k \sigma^2 (X'X)_{jk}^{-1} + \sum_{j=0}^{h}\sum_{k=0}^{h} V_j' W V_k \{(X'X)^{-1}X'[\Sigma + Cov(r)]X(X'X)^{-1}\}_{jk}$$

$$+ 2\sum_{j=0}^{h}\sum_{k=0}^{h} V_j' W V_k [(X'X)^{-1}X'E(\tilde{r}\eta')X(X'X)^{-1}]_{jk} \tag{9}$$

The first term on the right-hand side of (9) is of order

$$\mathrm{O}\left[m \sum_{j=0}^{h}\sum_{k=0}^{h} |(X'X)_{jk}^{-1}| \right] = \mathrm{O}\left[m \left(\frac{1}{n} + \sum_{k=1}^{h} \frac{h\log n}{n^2 k} + \sum_{j=1}^{h}\sum_{k=1}^{h} \frac{h}{njk} \right) \right]$$

$$= \mathrm{O}\left[\frac{mh}{n} \log^2 h \right]$$

$$= \mathrm{O}(m\,\mathrm{MSE}\,\log^2 h)$$

Since all entries of X are $\mathrm{O}(\log n)$, it follows that all entries of $X'[\Sigma + Cov(r)]X$ are of order

$$\mathrm{O}\left(\log^2 n \sum_{k=1}^{\tilde{n}}\sum_{j=k+1}^{\tilde{n}} \log^2 n \, k^{-2|d|} j^{2|d|-2} + \log^2 n \sum_{k=1}^{\tilde{n}}\sum_{j=1}^{\tilde{n}} \frac{\log j}{j}\frac{\log k}{k} \right)$$

$$= \mathrm{O}\left(\log^4 n \sum_{k=1}^{\tilde{n}} k^{-2|d|} k^{2|d|-1} + \log^6 n \right)$$

$$= \mathrm{O}(\log^6 n)$$

Thus, the second term on the right-hand side of (9) is

$$\sum_{j=0}^{h}\sum_{k=0}^{h} V_j' W V_k \{(X'X)^{-1}X'[\Sigma + Cov(r)]X(X'X)^{-1}\}_{jk}$$

$$= \mathrm{O}\left(m \log^6 n \sum_{j=0}^{h}\sum_{k=0}^{h}\sum_{L_1=0}^{h+1}\sum_{L_2=0}^{h+1} |(X'X)_{jL_1}^{-1}|\,|(X'X)_{L_2 k}^{-1}| \right) \tag{10}$$

We have

$$\sum_{L_1=0}^{h+1} \sum_{L_2=0}^{h+1} |(X'X)_{jL_1}^{-1}| \, |(X'X)_{L_2 k}^{-1}|$$

$$= \sum_{L_1=0}^{h+1} |(X'X)_{jL_1}^{-1}| \sum_{L_2=0}^{h+1} |(X'X)_{kL_2}^{-1}|$$

$$= O\left[\left(\frac{1}{n}\chi_{\{j=0\}} + \frac{h\log h}{nj}\chi_{\{1\leqslant j \leqslant h\}} \right) \left(\frac{1}{n}\chi_{\{k=0\}} + \frac{h\log h}{nk}\chi_{\{1\leqslant k \leqslant h\}} \right) \right]$$

Using (10) gives

$$\sum_{j=0}^{h} \sum_{k=0}^{h} V_j' W V_k \{(X'X)^{-1} X'[\Sigma + Cov(r)] X (X'X)^{-1}\}_{jk}$$

$$= O\left[m\left(\frac{h^2}{n^2} \right) \log^{10} n \right]$$

$$= O\left[\left(\frac{mh}{n} \right) \log^2 h \right]$$

$$= O(m\,\mathrm{MSE}\,\log^2 h)$$

Let P be the $(h+1) \times \tilde{n}$ matrix obtained by deleting the final row of $(X'X)^{-1}X'$. The absolute value of the last term on the right-hand side of (9) is

$$\left| \mathrm{E} \sum_{j=0}^{h} \sum_{k=0}^{h} V_j' W V_k (P\tilde{r})_j (P\eta)_k \right| = |\mathrm{E}[(X_h P\tilde{r})' W (X_h P\eta)]|$$

$$\leqslant \mathrm{E}(\|X_h P\tilde{r}\|_W \|X_h P\eta\|_W)$$

$$\leqslant (\mathrm{E}\|X_h P\tilde{r}\|_W^2)^{\frac{1}{2}} (\mathrm{E}\|X_h P\eta\|_W^2)^{\frac{1}{2}} \qquad (11)$$

Using similar arguments to those presented above, it can be shown that

$$\mathrm{E}\|X_h P\tilde{r}\|_W^2 = O\left[m\left(\frac{h^2}{n^2} \right) \log^{10} n \right] \qquad \text{and} \qquad \mathrm{E}\|X_h P\eta\|_W^2 = O\left[m\left(\frac{h^2}{n} \right) \log^6 n \right]$$

Thus, the right-hand side of (11) is of order

$$O\left[\frac{m(\log^8 n)h}{n^{3/2}} \right] = O\left[m\left(\frac{h}{n} \right) \log^2 h \right]$$

$$= O[m\,\mathrm{MSE}\,\log^2 h]$$

We conclude that $E\|X_h(\hat{\beta}_h - \beta_h^*)\|_W^2 = O(m\,\mathrm{MSE}\,\log^2 h)$, and the proof is complete.

PROOF OF LEMMA 5. We have

$$\left\| \sum_{k=h+1}^{\infty} \gamma_k V_k \right\|_W^2 = \sum_{k=h+1}^{\infty} \sum_{L=h+1}^{\infty} \gamma_k (V_k' W V_L) \gamma_L$$

$$= O\left[m \left(\sum_{k=h+1}^{\infty} |\gamma_k| \right)^2 \right]$$

$$= O\left(\frac{m a^{2h}}{h^{2p}} \right)$$

$$= O[m\,\text{MSE}]$$

PROOF OF LEMMA 6. We have

$$(I - H)\mu = (I - H)\left[X\beta + \sum_{k=h+1}^{\infty} \gamma_k V_k \right]$$

$$= \sum_{k=h+1}^{\infty} \gamma_k (I - H) V_k$$

By Lemma 2, each entry of $(I - H)V_k$ is of order $h \log^2 n$. Thus,

$$E[\mu'(I - H)W\varepsilon] = \sum_{k=h+1}^{\infty} \gamma_k V_k'(I - H)W E[\varepsilon]$$

$$= O\left[h \log^2 n \sum_{j=1}^{m} |E[\varepsilon_j]| \sum_{k=h+1}^{\infty} |\gamma_k| \right]$$

$$= O(h \log^4 n\,\text{MSE}^{1/2})$$

where we have used the assumptions given at the beginning of Section 3 together with the fact that $E[\varepsilon_j] = O(1)$ for $1 \leqslant j \leqslant \log^2 n$ and $E[\varepsilon_j] = O[(\log n)/j]$ for $\log^2 n \leqslant j \leqslant \tilde{n}$. It follows from Theorem 1 (which will be proved without using Lemma 6) and Lemma 3 that

$$\left| \frac{E[\mu'(I - H)W\varepsilon]}{\text{MISE}_m} \right| = O\left(\frac{h \log^4 n\,\text{MSE}^{\frac{1}{2}}}{m \log^2 n\,\text{MSE}} \right)$$

$$= O\left(\frac{h \log^2 n}{m\,\text{MSE}^{\frac{1}{2}}} \right)$$

$$= O\left[\frac{h \log^2 n}{m(h/n)^{\frac{1}{2}}} \right]$$

$$= O\left[\left(\frac{n^{\frac{1}{2}}}{m} \right) h^{1/2} \log^2 n \right]$$

$$= O[n^{\frac{1}{2}-\alpha} h^{\frac{1}{2}} \log^2 n] \to 0$$

since $\frac{1}{2} - \alpha < 0$, and $h^{\frac{1}{2}} \log^2 n$ is bounded by a power of $\log n$. The Lemma follows.

PROOF OF LEMMA 7. We write $\varepsilon = \eta + r$, where the properties of η and r are described in the proof of Lemma 4. We have

$$\mathrm{E}[\varepsilon' HW\varepsilon - \sigma^2 Tr(WH)] = \mathrm{E}[\eta' HW\eta + r' HWr + r' HW\eta + \eta' HWr - \sigma^2 Tr(HW)]$$

$$= \mathrm{E}\left[\sum_{j=1}^{m}(\eta_j^2 - \sigma^2)H_{jj} + \sum_{j=1}^{\tilde{n}}\sum_{k=1}^{m}\eta_j H_{jk}\eta_k\chi_{\{j\neq k\}} + r' HWr + r' HW\eta + \eta' HWr\right]$$

Note that $\mathrm{E}\sum_{j=1}^{m}(\eta_j^2 - \sigma^2)H_{jj} = 0$. Next, we have

$$\mathrm{E}\sum_{j=1}^{\tilde{n}}\sum_{k=1}^{m}\eta_j H_{jk}\eta_k\chi_{\{j\neq k\}}$$

$$= \mathrm{E}\sum_{k=1}^{m}\sum_{j=1}^{k-1}\eta_j H_{jk}\eta_k + \mathrm{E}\sum_{k=1}^{m}\sum_{j=k+1}^{\tilde{n}}\eta_j H_{jk}\eta_k$$

$$= \mathrm{O}\left[(h/n)(\log^4 n)\sum_{k=1}^{m}\sum_{j=1}^{k-1}k^{2|d|-2}j^{-2|d|} + \left(\frac{h}{n}\right)(\log^4 n)\sum_{k=1}^{m}\sum_{j=k+1}^{\tilde{n}}j^{2|d|-2}k^{-2|d|}\right]$$

$$= \mathrm{O}\left[\left(\frac{h}{n}\right)\log^5 n\right]$$

$$= \mathrm{o}(\mathrm{MISE}_m)$$

Next,

$$\mathrm{E}(r' HWr) = \mathrm{E}\sum_{j=1}^{\tilde{n}}\sum_{k=1}^{m}r_j H_{jk}r_k$$

$$= \mathrm{O}\left[\left(\frac{h}{n}\right)\log^2 n\sum_{j=1}^{\tilde{n}}\frac{\log j}{j}\sum_{k=1}^{m}\frac{\log k}{k}\right]$$

$$= \mathrm{O}\left[\left(\frac{h}{n}\right)\log^6 n\right] = \mathrm{o}(\mathrm{MISE}_m)$$

Next, we have

$$|\mathrm{E}(r' HW\eta)| \leqslant \mathrm{E}|r' HW\eta| = \mathrm{E}\left|\sum_{j=1}^{\tilde{n}}\sum_{k=1}^{m}r_j H_{jk}\eta_k\right|$$

$$\leqslant \mathrm{E}\sum_{j=1}^{\tilde{n}}|r_j|\left|\sum_{k=1}^{m}H_{jk}\eta_k\right| \leqslant Const \sum_{j=1}^{\tilde{n}}\frac{\log j}{j}\mathrm{E}\left|\sum_{k=1}^{m}H_{jk}\eta_k\right|$$

$$\leqslant Const \sum_{j=1}^{\tilde{n}}\frac{\log j}{j}\left[\mathrm{E}\left(\sum_{k=1}^{m}H_{jk}\eta_k\right)^2\right]^{\frac{1}{2}}$$

Now,

$$\mathrm{E}\left(\sum_{k=1}^{m} H_{jk}\eta_k\right)^2 = \mathrm{E}\sum_{k=1}^{m} H_{jk}\eta_k \sum_{L=1}^{m} H_{jL}\eta_L = \mathrm{O}\left\{\left[\left(\frac{h}{n}\right)\log^2 n\right]^2 (m+\log^3 n)\right\}$$

so that

$$|\mathrm{E}(r'HW\eta)| = \mathrm{O}\left[\sum_{j=1}^{\tilde{n}} \frac{\log j}{j} m^{\frac{1}{2}}\left(\frac{h}{n}\right)\log^2 n\right]$$

$$= \mathrm{O}[m^{\frac{1}{2}}(h/n)\log^4 n]$$

$$= \mathrm{o}(\mathrm{MISE}_m)$$

By a similar argument,

$$|\mathrm{E}(\eta'HWr)| = \mathrm{O}\left[n^{\frac{1}{2}}\left(\frac{h}{n}\right)\log^4 n\right] = \mathrm{o}(\mathrm{MISE}_m)$$

since $\alpha > \frac{1}{2}$. The proof is finished.

REFERENCES

BERAN, J. (1993) Fitting long-memory models by generalized linear regression. *Biometrika* 80, 817–22.
—— (1994) *Statistics for Long-memory Processes*. New York: Chapman and Hall.
BHANSALI, R. J. (1986) A derivation of the information criteria for selecting autoregressive models. *Adv. in Appl. Probab.* 18, 360–87.
—— and KOKOSZKA, P. S. (1999) Estimation of the long memory parameter by fitting fractional autoregressive models. Preprint.
BLOOMFIELD, P. (1973) An exponential model for the spectrum of a scalar time series. *Biometrika* 60, 217–26.
BRODSKY, J. (1997) Parameter estimation, model selection and multi-step forecasting for a long memory time series. Doctoral dissertation, New York University Stern School of Business.
DAVIES, R. B. and HARTE, D. S. (1987) Tests for Hurst effect. *Biometrika* 74, 95–101.
FINDLEY, D. F. (1985) On the unbiasedness property of AIC for exact or approximating linear stochastic time series models. *J. Time Ser. Anal.* 6, 229–52.
GEWEKE, J. and PORTER-HUDAK, S. (1983) The estimation and application of long memory time series models. *J. Time Ser. Anal.* 4, 221–37.
HENRY, M. and ROBINSON, P. M. (1996) Bandwidth choice in Gaussian semiparametric estimation of long range dependence. *Athens Conference on Applied Probability and Time Series Analysis, Vol. II*, 19–46, *Lecture Notes in Statistics* 115. New York: Springer.
HURVICH, C. M. and BELTRAO, K. I. (1994) Automatic semiparametric estimation of the memory parameter of a long memory time series. *J. Time Ser. Anal.* 15, 285–302.
—— and BRODSKY, J. (1997) Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential models. Working Paper SOR-97-2, New York University Stern School of Business, Department of Statistics and Operations Research.
—— and DEO, R. S. (1999) Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long-memory time series. *J. Time Ser. Anal.* 20, 331–41.
—— and TSAI, C.-L. (1995) Relative rates of convergence for efficient model selection criteria in linear regression. *Biometrika* 82, 418–25.
——, DEO, R. S. and BRODSKY, J. (1998) The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series. *J. Time Ser. Anal.* 19, 19–46.
JANACEK, G. J. (1982) Determining the degree of differencing for time series via the log spectrum. *J. Time Ser. Anal.* 3, 177–83.
KUNSCH, H. R. (1987) Statistical aspects of self-similar processes. In *Proceedings of the First*

World Congress of the Bernoulli Society (eds Yu. Prohorov and V. V. Sazanov) 1, 67–74. Utrecht: VNU Science Press.

LI, K.-C. (1987) Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation. *Ann. Statist.* 15, 958–75.

MOULINES, E. and SOULIER, P. (1999) Broad band log-periodogram regression of time series with long range dependence. *Ann. Statist.* 27, 1415–39.

—— (2000) Data-driven order selection for projection estimator of the spectral density of time series with long range dependence. *J. Time Ser. Anal.* 21, 193–218.

ROBINSON, P. M. (1994a) Semiparametric analysis of long-memory time series. *Ann. Statist.* 22, 515–39.

—— (1994b) Time series with strong dependence. *Advances in Econometrics, Sixth World Congress, Vol. I*, 47–95. Cambridge University Press.

—— (1995a) Log-periodogram regression of time series with long range dependence. *Ann. Statist.* 23, 1048–72.

—— (1995b) Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* 23, 1630–61.

SHIBATA, R. (1980) Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* 8, 147–64.