Mis-specified Long-Range Dependence

A.I. McLeod, University of Western Ontario, Canada

Department of Statistical and Actuarial Sciences,
University of Western Ontario,
London, Ontario N6A 5B7
email: aim@uwo.ca

August 10, 2005

Abstract
The robustness of AR and ARMA models is investigated in the presence of long-memory autocorrelation. Although previous researchers have already discussed this topic, it was always assumed that the AR or ARMA model order was fixed. The more realistic case which allows the model order to increase as the length of the series increases is presented. It is found that even when the length of the time series is fairly large, it may not be possible to distinguish in terms of goodness-of-fit between the best fitting AR or ARMA model and a suitable long-memory alternative.

Keywords: ARMA modelling and forecasting, Autoregressive approximation, Forecasting, Interval estimates, Intervention analysis, Loess, One-sample problem

1. Introduction

The issue of long-memory dependence and robustness was raised by Hampel et al. (1986, §8.2) and Hampel (1987, 1998). It was suggested, based on an asymptotic argument, that in the presence of long-memory, the use of ARMA and related short memory time series models may yield incorrect results. In other words the most popular methods for time series analysis may not be robust against long-memory autocorrelation alternatives.

The sample mean is well known to be asymptotically efficient (Beran, 1993, §8). Given n consecutive observations, $z_1, \ldots, z_n$ then $\text{var}(\bar{z}) \propto n^{-1}$ in the case of ARMA models and $\text{var}(\bar{z}) \propto n^{-\alpha}$ for hyperbolic decays models. Accordingly, Hampel (1998) suggested that the true standard deviation of the mean will always be underestimated if the actual data are hyperbolic and an ARMA model is used. This is obvious if we hold the ARMA model fixed. But what happens when we allow the order of the model to increase with n? For example, Berk (1974) showed that for n consecutive observations from a general class of short-memory time series, an AR(k) with $k = o(n^{-1/3})$ would provide a consistent autoregressive estimator for the true spectral density. Hampel's claim is based on the asymptotic properties and it may not hold in practice if the model order or complexity is allowed to increase with n.

It is frequently assumed in the practice of statistics that if two statistical models fit the data equally well in terms of model adequacy and model parsimony then it does not matter very much which model is used for practical finite sample inferences. For example, Barnard (1983) has shown that if a sample of data from a Cauchy distribution happens to look approximately normally distributed then standard normal based methods can be used. However, in another case, for inferences on the population total from a simple random sample from a finite population, Rubin (1983) has demonstrated that the although the actual sample may be fit by a variety of plausible models, the inferences from these models can be markedly different. A similar corundrum to Rubin's exists with respect to many inferences in time series analysis and the purpose of this article is to give a rigorous demonstration of this fact.

In actual practice a time series modelling using ARMA models would naturally allow the complexity of the model to increase as the length of the series increases (Hannan, 1970). If the actual data were to follow a long-memory model then the natural question is whether for practical purposes the fitted ARMA model will produce valid inferences. It is the authors experience which is illustrated in the examples below that it is often very difficult to discriminate between the standard ARMA and a long-memory alternative in terms of goodness-of-fit criteria such as the AIC or model diagnostic checks. Often both short-range and long-range models fit actual data equally well. In this situation, one might think that the statistical inferences from such models would be similar. We will show demonstrate that this is not always the case.

In the context of multiple linear regression with long-memory errors, Yajima (1991) showed that the usual least squares estimates were asymptotically efficient and also derived an expression for the asymptotic covariance matrix of the least squares estimates. This might suggest that the problem of mis-specification was solved. But the question of finite-sample properties and the validity or lack thereof of the estimated variances was not discussed.

Beran (1993, §9) has also shown that the usual least squares estimators are asymptotically efficient in many situations but that the estimated precision may be overestimated as was the case for the sample mean.

Kunsch, Beran and Hampel (1993) established that in randomized designs with long-memory errors the standard analysis of contrasts is asymptotically valid and the associated tests and confidence levels are still asymptotically correct. Based on simulation experiments, Kunsch, Beran and Hampel (1993) suggested that these asymptotic properties hold in finite samples. In §2 of this paper we present a non-simulation methodology for investigating the finite sample accuracy of the mis-specified estimates and it is shown that a modest improvement over the standard least squares estimates is sometimes possible using a suitable short-memory approximation.

The main contribution of this paper is to present a finite-sample methodology to compare the fitting of long-memory and short-memory models which allows for the increase in model complexity as the length of the series increases. We also demonstrate that in practice it may be very difficult to distinguish between long-memory and short-memory models. We demonstrate that the finite-sample inference between these competing models may yield quite different conclusions.

2. Long Memory Time Series

2.1 Introduction

Consider a covariance stationary time series, z_t, with theoretical autocovariance function (TACVF), $\gamma_k = \mathrm{Cov}(z_t, z_{t-k})$. We will say a covariance stationary time series exhibits long-memory if the autocorrelation function is not summable and conversely if a stationary time series does not have long-memory, it is described as short-memory. There are many types models that can give rise to autocorrelations long-memory but one of the most widely used can be characterizied by hyperbolic decay,

$$\gamma_k \sim \lambda k^{-\alpha}, \quad (1)$$

where $\lambda > 0$, $\alpha \in (0,\ 1)$. The parameter α is called the persistence parameter. The notation $a_k \sim b_k$ means $a_k / b_k \longrightarrow 1$ as $k \longrightarrow \infty$. The persistence parameter α is related to the famous Hurst coefficient, $\alpha = 2\,(1 - H)$ for $1/2 < H < 1$. In this case the time series exhibits a slowly decaying positive autocorrelation that is plausible for many types of geophysical time series. The more general case with $\alpha \in (-1, 1)$, $\alpha \neq 0$ allows for negative correlations or antipersistence which may occur in over-differenced ARIMA time series.

The fractional ARMA (FARMA) model is an extension of the ARMA model which allows hyperbolic decay,

$$\phi(B)\, \nabla^d\, z_t = \theta(B)\, a_t, \tag{2}$$

$$\phi(B) = 1 - \phi_1 - \ldots - \phi_p\, B^p, \quad \theta(B) = 1 - \theta_1 - \ldots - \theta_q\, B^q,$$

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

where B is the backshift operator on t, $B^k\, z_t = z_{t-k}$. In addition to the usual ARMA parameter requirements for model identifiability and invertibilty, we assume $|\, d\,| < 0.5$. For $d = 0$, the FARMA model is equivalent to the usual ARMA. When $d \in (0,\ 0.5)$, the FARMA model exhibits hyperbolic decay with persistence parameter $\alpha = 1 - 2\, d$ or $H = d + 0.5$. For our robustness study we will concentrate on the special case where $p = q = 0$ which corresponds to fractionally differenced white noise (FDWN). However the results can easily be extended to other FARMA and long-memory models.

2.2 AR Approximation

For a linear process, Shibata (1980) shows that the AIC provides an optimal method of selecting an AR approximation based on observations $z_1, \ \ldots, \ z_n$.

Theorem 1: Let $z_1, \ \ldots, \ z_n$ be generated by a long memory hyperbolic decay time series model and suppose that an AR(k) is fit to $z_1, \ \ldots, \ z_n$ by least squares. The AIC criterion for this model may be written $\text{AIC} = -2 \log L + 2\, k$, where L is the Gaussian AR(k) likelihood function. Then $E\,\{\text{AIC}\} \doteq n \log(\sigma_k^2) + 2\, k$, where σ_k^2 is the variance of the minimum mean square error linear predictor using the last k observations.

The MAICE-AR approximation of order p minimizes $E\,\{\text{AIC}\}$. In other words,

$$p = \underset{\{0 \le k \le n-1\}}{\arg\min}\ n \log \sigma_k^2 + 2\, k,$$

where $\sigma_0^2 = \gamma_0$ and for $k > 0$, $\sigma_k^2 = \sigma_{k-1}^2 (1 - \phi_{k,k}^2)$, where $\phi_{k,k}^2$ is the theoretical partial autocovariance function.

Theorem 2: The order of the MAICE-AR(p) approximation for any fractional ARMA satisfies $p \sim d\, \sqrt{n/2}$.

The proof of these theorems are given in the Appendix.

The BIC often provides a more parsimonious fit than the AIC. This may be useful when it comes to estimating the parameters of the model from actual data but this is not the purpose of the TMAICE-AR approximation. Since $\text{AIC} = -2 \log L + k \log n$, we see that with the BIC the penalty for parameters is larger provided $n > 7$. However if a more parsimonious model were selected than the TMAICE-AR, it would by definition not fit as well in terms of the theoretical forecast error variance. The purpose of the TMAICE-AR approximation is to indicate the approximate upper bound on the best we can hope to achieve in fitting autoregressive approximation to data generated from a hypothetical process. Since the problem of parameter estimation is not taken into account, the TMAICE-AR approximation provides an optimistic upper bound.

Table 1 shows p for various parameters in the case of FDWN.

Table 1. MAICE-AR(p) model orders for FDWN.

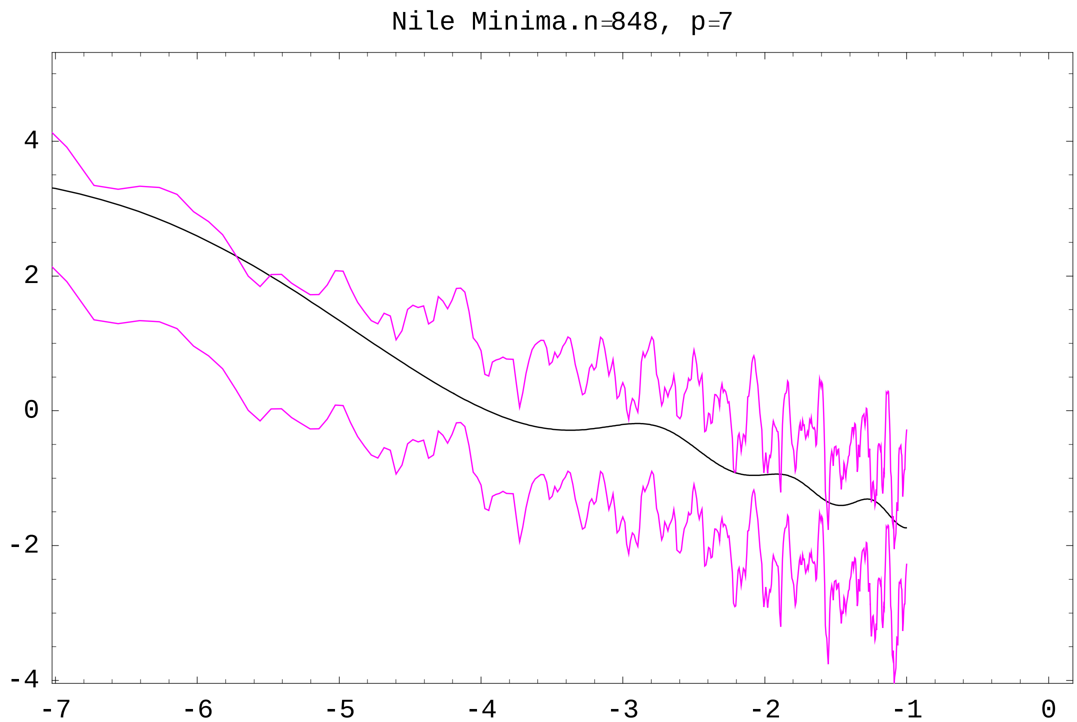
n	d=0.1	d=0.2	d=0.3	d=0.4
25	0	0	1	1
50	0	1	1	2
100	0	1	2	3
200	1	2	3	4
400	1	3	4	6
800	2	4	6	8

The entries in Table 1 were determined by an exact evaluation of AIC_p for the FDWN autocovariance function but it is interesting that Theorem 2 predicts all the entries in Table 1 exactly.

■ 2.3.3 Illustrative Example: Annual Nile Minima

The Nile River Minima, 622-1281, has been widely cited as an example of a long memory time series. The data is available from Beran (1994, §12.2) . To further demonstrate the plausibility of fitting AR models we used the Burg algorithm to compute the MAICE AR model. Then we plotted the spectral density function of the fitted model and 95% nonparametric confidence intervals. The nonparametric confidence intervals were obtained by smoothing the periodogram using a modified Daniell smoother (Bloomfield, 1973) with span selected using the automatic method of Lee (1997) and then using the confidence limits derived by Brillinger (1983). Figure 1 suggests that the MAICE AR(7) model adequately models the low frequencies. It is interesting to note that using exact maximum likelihood we fit a FDWN model and obtained $\hat{d} = 0.38$ and the formula in Theorem 1 gives $p = 7.8$ which agrees well with the actual MAICE model order.

Figure 1. Normalized spectral density from fitted AR(7) model and 95% nonparametric confidence interval. Log-log scale.



We also compared the best fitting ARMA(p, q) and FARMA(p, q) models for this data. Using an exact maximum likelihood algorithm and computing the optimal AIC for each the the 9 possible models with $p, q = 0, 1, 2$ we found the optimal ARMA model was with $p = 2$, $q = 1$ and the optimal FARMA model had $p = q = 0$ and the respective AIC's were 5628.04 and 5626.62. The ARMA and FARMA model fit about equally well according to the AIC criterion. Diagnostic checks including residual autocorrelation analysis and the non-parameteric spectral density confidence limit check illustrated with the AR(7) did not suggest any preference. In this situation, the practical modeler might think that there would be no difference in the practical differences between these models for a estimating confidence interval for the mean. We demonstrate in the next section, that this supposition would be wrong.

3. One-Sample Problem

Given a batch of data $z' = (z_1, \ldots, z_n)$ the problem is to estimate the mean μ and give an estimate of the precision of that estimate. We assume that the data is fractionally differenced Gaussian white noise with parameter $d \in (0, 0.5)$. Then as discussed by Beran (1993, §8) the sample mean is asymptotically efficient and some simple computations show that it has high efficiency even in small samples. We now turn to the effect of misspecification on the confidence intervals for the mean when we fit using a TMAICE-AR approximation. We also compare with the misspecification which assumes the data uncorrelated as in the standard analysis of the one-sample problem.

Our estimate var($\bar{z}$) is computed using the theoretical autocovariances from FDWN with parameter d and the corresponding MAICE-AR approximation,

$$\operatorname{var}(\bar{z}) = \gamma_0 / n + 2\,(1/n)^2 \sum_{k=1}^{n-1} (n-k)\,\gamma_k$$

In most applications it is the width of the confidence interval or equivalently the standard deviation which is relevant. So we look at the percentage underestimation of the error in the standard deviation of the mean, $100\,(\sigma_e - \sigma_x)/\sigma_e$, where σ_e and σ_x denote the exact and approximate standard deviations.

Table 2. Percentage underestimation of the standard deviation of the mean using a: MAICE-AR and i: assuming IID.

d	0.1	0.2	0.3	0.4
n=25:a	26	47	40	43
n=25:i	25	45	60	71
n=50:a	31	39	51	43
n=50:i	30	52	67	78
n=100:a	35	47	52	48
n=100:i	35	58	74	84
n=200:a	32	48	56	55
n=200:i	39	64	79	88
n=400:a	37	51	61	60
n=400:i	43	68	83	91
n=800:a	38	55	64	65
n=800:i	47	72	86	93

Table 2 demonstrates that in practice the confidence limits could be much too narrow if we base on inference on the mean on the TMAICE-AR approximation. For example if $n = 200$ and $d = 0.2$ then the width of any confidence interval will be 48% less than the correct width. The percentage underestimation is seen from Table 2 to increase with n as would be suggested by the asymptotic formula for the sample variance. We can conclude that in the one-sample problem the inferences on the mean may be seriously inadequate if a short memory time series models are used in the presence of hyperbolic decay autocorrelation.

For the Nile River Minima dataset $n = 660$, the 95% confidence intervals for the mean based on the best fitting ARMA ($\phi_1 = 1.2712$, $\phi_2 = -0.2964$, $\theta_1 = 0.8484$, $\sigma = 69.8060$) and FARMA ($d = 0.3931$, $\sigma = 69.8879$) models are respectively (1128, 1169) and (1066, 1230). It is seen that the ARMA interval is 66% too small. This agrees reasonably well with the results in Table 2.

4. Intervention Analysis

In the standard two-sample problem we have two random samples of sizes n and m, denoted by $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$, from normal distributions with means μ_x and μ_y and variance σ^2 and we are interested in testing $\mathbb{H}_0 : \mu_x = \mu_y$ vs the alternative $\mathbb{H}_1 : \mu_x \neq \mu_y$. The standard solution leads to the well known two-sample form of the t-test. Of critical importance to the validity of this approach is the assumption of statistical independence. There are several possible extensions of this problem to the time series setting. We will examine an intervention analysis model.

The basic model we consider is of the form,

$$z_t = \mu + \omega\ P_t(T) + u_t, \quad t = 1, 2, \ldots, n, \tag{3}$$

where n is the number of observations, z_t is the time series, s_t is a step intervention and u_t is the autocorrelated error term. The step intervention is of the form,

$$P_t(T) = \begin{cases} 0 & t \le T \\ 1 & t > T \end{cases}$$

The noise term, u_t, will be assumed to be generated by fractionally differenced white noise. We will examine the effect of misspecifying with a MAICE-AR approximation.

■ 4.1 Efficiency of Misspecified Estimates

Assuming d is known then the optimal estimate of $\beta = (\mu,\ \omega)$, denoted by $\hat{\beta}$, may be computed from generalized least squares theory. We have

$$\hat{\beta} = (X^T \Gamma_n^{-1} X)^{-1} X^T \Gamma_n^{-1} z \tag{4}$$

and

$$\text{var}\{\hat{\beta}\} = (X^T \Gamma_n^{-1} X)^{-1} \tag{5}$$

where X^T, the transpose of X, has 1's in the first row and the second consists of $P_t(T)$, $t = 1, \ \ldots, \ n$ and Γ_n is the covariance matrix of n consecutive error terms, u_t, $t = 1, \ \ldots, \ n$.

If the error term, u_t, is incorrectly specified as an autoregression then the estimate for β can be written

$$\tilde{\beta} = H z$$

where $z = (z_1, \ \ldots, \ z_n)$ and

$$H = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1},$$

where Σ is the covariance matrix of the MAICE-AR approximation. It is easily shown that $\mathcal{E}\{\tilde{\beta}\} = \beta$ and

$$\text{var}\{\tilde{\beta}\} = H \Gamma_n H^T. \tag{6}$$

We will now compare $\text{var}\{\hat{\beta}\}$ and $\text{var}\{\tilde{\beta}\}$ for various d and n. The relative efficiency of $\tilde{\beta}$ vs $\hat{\beta}$ is defined as

$$E = \text{Det}(\text{var}\{\hat{\beta}\}) \big/ \text{Det}(\text{var}\{\tilde{\beta}\})$$

Next we consider the relative efficiency of the ordinary least squares estimate,

$$\vec{\beta} = (X^T X)^{-1} X^T z. \tag{7}$$

Again $\vec{\beta}$ is unbiased. And

$$\text{var}\{\vec{\beta}\} = \mathcal{H} \Gamma_n \mathcal{H}^T. \tag{8}$$

where

$$\mathcal{H} = (X^T X)^{-1} X^T.$$

Table 3. Relative efficiency, in percent, of misspecified estimators, a: MAICE-AR and i: OLS estimator

d	0.1	0.2	0.3	0.4
n=25:a	98.5	94.2	96.9	85.8
n=25:i	98.5	94.2	87.5	78.8
n=50:a	98.1	96.	93.2	87.5
n=50:i	98.1	92.8	84.5	73.5
n=100:a	97.8	93.9	92.5	90.3
n=100:i	97.8	91.8	82.2	69.4
n=200:a	98.	93.5	90.	91.8
n=200:i	97.7	91.1	80.5	66.1

As might be expected from the asymptotic results of Beran (1993, §9) the efficiency of even OLS is quite good. However the MAICE-AR does noticeably better than OLS when $d = 0.3,\ 0.4$.

■ 4.2 Estimate of the Precision of the Estimates

Using the MAICE-AR approximation our estimate of the covariance matrix of $\tilde{\beta}$ would be

$$\text{estvar}\{\tilde{\beta}\} = (X^T \Sigma_n^{-1} X)^{-1} \tag{9}$$

and in the IID case,

$$\text{estvar}\left\{\vec{\beta}\right\} = n^{-1} \gamma_0 (X^T X)^{-1} \tag{10}$$

Table 4. Percentage underestimation of the standard error of the parameter estimates for μ and ω for a: MAICE-AR approximation and i: OLS.

d	0.1	0.2	0.3	0.4
n=25:a μ	21	39	29	35
n=25:a ω	14	25	−1	−18
n=25:i μ	84	87	90	92
n=25:i ω	83	84	85	83
n=50:a μ	26	30	40	34
n=50:a ω	19	14	11	−20
n=50:i μ	89	92	94	96
n=50:i ω	88	90	91	91
n=100:a μ	30	38	41	39
n=100:a ω	24	24	12	−17
n=100:i μ	93	95	97	98
n=100:i ω	92	94	95	95
n=200:a μ	27	40	45	46
n=200:a ω	21	25	17	−11
n=200:i μ	95	97	98	99
n=200:i ω	95	96	97	97

It is well known that the OLS variances are inflated in the presence of autocorrelation so he poor performance of the OLS was expected. The estimate of standard error of $\hat{\omega}$ appears reasonable with the MAICE-AR although the standard error for $\hat{\mu}$ is about the same as in the one-sample case. For example, taking $d = 0.2$ and $n = 200$, the confidence limits for $\hat{\mu}$ are understated by 40% in Table 4 instead of 48% as in Table 2.

4.3 Nile River Intervention Analysis

Hipel and McLeod (1994) fit a simple step intervention model with AR(1) error to the annual flows of the Nile River at Aswan, 1870-1945. The Aswan dam began operation in 1903 which corresponds to about $T = 33$. The model equation can be written,

$$z_t = \mu + \omega S_t^{(T)} + a_t / (1 - \phi B)$$

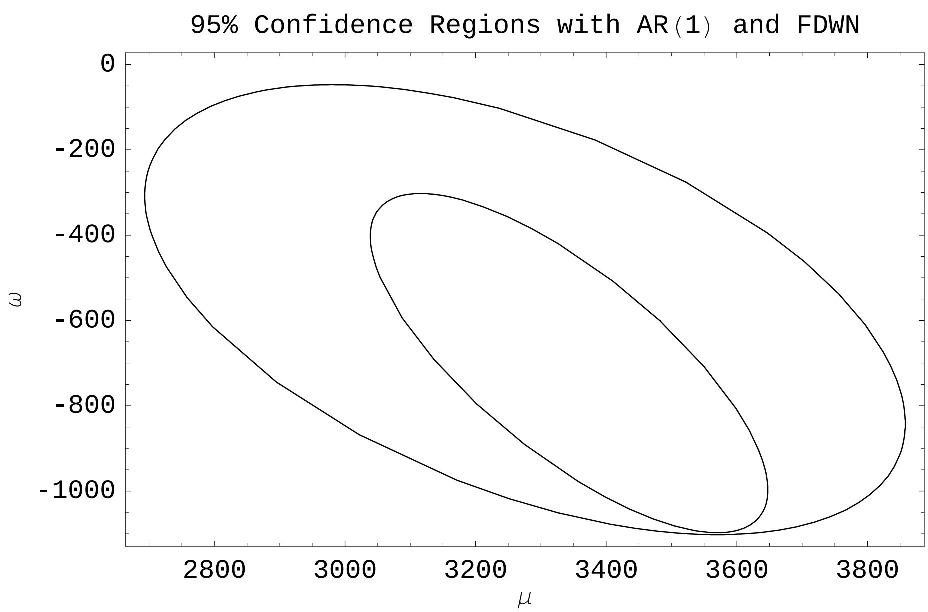
where a_t is assumed NID(0, σ_a^2). Using exact maximum likelihood estimation, $\hat{\phi} = 0.39$ and the observed value of the maximized loglikelihood was -449.8. Using exact maximum likelihood the corresponding model with fractionally differenced white noise was fit, viz.

$$z_t = \mu + \omega S_t^{(T)} + \nabla^{-f} a_t.$$

resulting in $\hat{f} = 0.30$ and a maximized loglikelihood equal to -453.5. In terms of loglikelihood the AR(1) is much better.

The table below compares the exact mle for the parameters μ and ω for these models.

	AR(1)	FDWN
μ	{3343.1, ±124.3}	{3275.6, ±237.9}
ω	{−699.9, ±162.5}	{−574.6, ±215.5}



References

Barnard, G.A. (1983), Pivotal inference and the conditional view of robustness (Why we have managed for so long with normality assumptions?). In *Scientific Inference, Data Analysis and Robustness*, ed. G.E.P. Box, T. Leonard and C.F. Wu, New York: Academic Press.

Berk, K.N. (1974), Consistent autoregressive spectral estimates, *Annals of Statistics* **2**, 489-502.

Beran, J. (1994), *Statistics for Long-Memory Processes*, London: Chapman and Hall.

Bloomfield, P. (1973), *Fourier Analysis of Time Series: An Introduction*, New York: Wiley.

Brillinger, D.R. (1983), *Time Series Analysis: Data Analysis and Theory*, 2nd. ed., San Francisco: Holden-Day.

Brockwell, P.J. & Davis, R.A. (1991), *Time Series: Theory and Methods*, (2nd edn.). New York: Springer-Verlag.

Chambers, J.M. and Hastie, T. (1991), *Statistical Models in S.* Wadsworth & Brooks/Cole.

Cleveland, W.S. (1995), *Visualizing Data*, New Jersey: Summit.

Cleveland, W.S. and Devlin, S.J. (1988). Locally-weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association* **83**, 596-610.

Hampel, F., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986), *Robust Statistics: The Approach Based on Influence Functions.* New York: Wiley.

Hampel, F. (1987). Data analysis and self-similar processes, with discussion. In *Proceedings of the 46th Session of the ISI* **52** (Book 4) 235-264.

Hampel, F. (1998), Is statistics too difficult? *Canadian Journal of Statistics* **26**, 497-513.

Hosking, J.R.M. (1981), Fractional differencing, *Biometrika* **68**, 165-176.

Hosking, J.R.M. (1996), Asymptotic distributions of the sample mean, autocovariances and autocorrelations of long-memory time series, *Journal of Econometrics* **73**, 261-284.

Kunsch, H., Beran, J. and Hampel (1993), Contrasts under long-range correlations, *The Annals of Statistics* **23**, 1025-1030.

Kunsch, H. (1986), Discrimination between monotonic trends and long-range dependence, *Journal of Applied Probability* **23**, 1025-1030.

Lam, H. (1999), *Hyperbolic, Irregular, Spatial and ARCH Time Series.* Ph.D. Thesis, University of Western Ontario.

Ledholter, J. and Abraham, B. (1981), Parsimony and Its Importance in Time Series Forecasting, *Technometrics* 23, 411--414.

Lee, T.C.M. (1997), A simple span selector for periodogram smoothing, *Biometrika* **84**, 965-969.

Li, W.K. & McLeod, A.I. (1986), Fractional time series modelling, *Biometrika* **73**, 217--221.

Matthews, R. (1998), The great health hoax, The Sunday Telegraph, 13 September 1998, http://www.stats.uwo.ca/faculty/aim/misc/TheGreatHealthHoax.htm

McLeod, A.I. (1998), Hyperbolic decay time series, *Journal of Time Series Analysis* **19**, 473-483.

Shibata, R. (1980), Asymptotically efficient selection of the order of the model for estimating the parameters of a linear process, *Annals of Statistics* **8**, 147-164.

Rubin, D. (1983), A case study of the robustness of Bayesian methods of inference. In *Scientific Inference, Data Analysis and Robustness*, ed. G.E.P. Box, T. Leonard and C.F. Wu, New York: Academic Press.

Yajima, Y. (1991), Asymptotic properties of estimates in incorrect ARMA models for long-memory time series. In *New Directions in Time Series Analysis, Part II*, ed. D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt, M. Taqqu. New York: Springer-Verlag.

Yajima, Y. (1991), Asymptotic properties of the LSE in a regression model with long-memory stationary errors, *The Annals of Statistics* 19, 158-177.

Yajima, Y. 1988. On Estimation of a Regression Model with Long-Memory Stationary Errors. *The Annals of Statistics* 16, 791-807.

Appendix: Proof of Theorems 1 and 2

Theorem 1: Let $z_1, \ldots, z_n$ be generated by a long memory hyperbolic decay time series model. If an AR(k) is fit using least-squares or some other Gaussian efficient method then the expected value of the AIC may be written $E\{\text{AIC}\} \doteq n\log(\sigma_k^2) + 2k$, where σ_k^2 is the variance of the minimum mean square error linear predictor using the last k observations.

Proof: The loglikelihood of the AR(k) model may be written (McLeod, 1977; Brockwell and Davis, 1991, §8.7),

$$\log L = n\log\hat{\sigma}^2 + \text{O}_p(1),$$

where $\hat{\sigma}^2$ is the estimate of the residual variance. When the are generated by a long-memory hyperbolic decay time series model, Yajima (1991) showed that

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{L} N(0, v),$$

where v is a constant and $\sigma^2 = \sigma_k^2$. Expanding $\log\dot{\sigma}^2$ in a Taylor series about $\dot{\sigma}^2 = \sigma^2$ and evaluating at $\dot{\sigma}^2 = \hat{\sigma}^2$,

$$\log\hat{\sigma}^2 = \log\sigma^2 + (\hat{\sigma}^2 - \sigma^2)/\sigma^2 + \text{O}_p(1).$$

Hence,

$$E\{-2\log L + 2k\} = n\log\sigma^2 + 2k + \text{O}(1). \quad \blacksquare$$

Theorem 2: For FDWN with $d \in (0, 0.5)$, the order of the MAICE-AR approximation is given by $p \sim d\sqrt{n/2}$.

Proof: From McLeod (1998, eqn. 5) it follows that for FDWN with $d \in (0, 0.5)$,

$$\text{AIC}_k \sim d^2 n/k + 2k. \tag{11}$$

Hence, for large n, p is the value which minimizes AIC_k is given by . Since the high-order autocorrelations are asymptotically equivalent in any fractional ARMA(p, q) with fractional difference parameter d the general theorem follows. ■