

Fitting long-memory models by generalized linear regression

BY JAN BERAN

*Department of Biostatistics, ISPM, University of Zurich, Sumatrastr. 30, 8006 Zürich,
Switzerland*

SUMMARY

There is an increasing awareness of the importance of long-memory models in statistical applications. If long memory is present, it has to be taken into account in order to obtain reliable tests and confidence intervals. One obstacle to using models with long memory in routine statistical analysis has been the lack of easily available and sufficiently versatile statistical software. Here we propose a simple but flexible class of parametric models, which can be used to model such behaviour. We demonstrate that these models can be fitted by generalized linear regression. Standard statistical software packages can be used. A data example is discussed.

Some key words: Fractional ARIMA; Generalized linear models; Long-range dependence; Maximum likelihood estimation.

1. INTRODUCTION

Stationary time series models with long memory have become a very active field of statistical research in the last decade; for a review of recent results see Beran (1992) and Künsch (1986). Their importance in virtually all fields of statistical applications has been demonstrated by numerous examples. References to the relevant literature and to data sets are given by, for example, Mandelbrot (1983), Cox (1984), Hampel (1987), Beran (1992), and by D. B. Percival in Technical Report No. 69, Department of Statistics, University of Washington, Seattle.

In spite of the usefulness of long-memory models, their application in statistical practice has been limited. The main reason is that, apart from some scattered specialized subroutines with limited applicability, no general easily accessible software package for the statistical analysis of such models exists. Moreover, estimation by maximum likelihood or related methods seemed to require excessive CPU time, in particular when many parameters have to be estimated. In this paper we demonstrate how a rich class of long-memory models can be fitted by a Gaussian approximate maximum likelihood method, within the setting of generalized linear models (McCullagh & Nelder, 1983). The statistical analysis, i.e. estimation, tests, confidence intervals and diagnostics, of these models can therefore be done by using standard statistical software for generalized linear regression, e.g. SPLUS, GLIM. Estimation of the model parameters by generalized linear regression is not only fast, but also turns out to be asymptotically efficient for Gaussian processes. Moreover, the parameters are directly interpretable. The model class considered here is an extension of short-memory models considered by Bloomfield (1973); also see Diggle (1990). The use of regression packages for fitting times series models is suggested by Cameron & Turner (1987). They apply generalized linear regression to ARMA-processes, Bloomfield's models

and to a model suggested by Kolmogorov (1941) for the spectrum of turbulence in a fluid. The first two models are short-memory processes, while the third model is a special example of a long-memory process.

The paper is organized as follows. In § 2 we define what we mean by 'long memory' and briefly review results on approximate Gaussian maximum likelihood estimators for parametric models with long memory. In § 3 we introduce the new class of parametric models. In § 4 we show, how for these models, the estimation procedure described in § 2 reduces to generalized linear regression. The practical application of the method is illustrated by a data example in § 5.

2. ESTIMATION OF LONG MEMORY

Let X_t be a linear stationary process with a one-sided moving average representation $X_t = \sum a_k \xi_{t-k}$ such that $\sum a_k^2 < \infty$, where the summation is over the range $k = 0, \dots, \infty$, and ξ_s , $s \in \mathbb{Z}$, are independent identically distributed random variables with $E(\xi_s) = 0$ and $\text{var}(\xi_s) < \infty$. The process X_t is said to have long memory if the spectral density f has a pole at the origin of the form

$$f(x) \sim b|x|^{1-2H} \quad (1)$$

as $x \rightarrow 0$, where $\frac{1}{2} < H < 1$ and $b > 0$. In particular, (1) implies that the sum of all covariances is infinite. More general definitions are possible, but (1) is sufficient in most situations. We consider parametric models $f(x; \theta) = \theta_1 f(x; (1, \gamma))$, where $\gamma = (\theta_2, \dots, \theta_{p+2}) = (H, \theta_3, \dots, \theta_{p+2})$. The first parameter θ_1 is the scale parameter, H determines the long-term behaviour via equation (1), and the additional parameters $\theta_3, \dots, \theta_{p+2}$ allow for flexible modelling of short-term properties.

Central limit theorems for approximate Gaussian maximum likelihood estimators of θ are given in my ETH Zürich Ph.D. thesis, and by Fox & Taqqu (1986), Beran (1992), Dahlhaus (1989) and by H. P. Graf in his ETH Zürich Ph.D. thesis for Gaussian processes, and by Giraitis & Surgailis (1990) for the more general linear processes defined above. In particular, Whittle's approximate maximum likelihood estimator $\hat{\theta}$ (Whittle, 1953) can be obtained by treating the periodogram

$$I(x_{j,n}) = (2\pi n)^{-1} \left| \sum_{t=1}^n X_t e^{itx_{j,n}} \right|^2$$

at Fourier frequencies $x_{j,n} = 2\pi j/n$, for $j = 1, 2, \dots, n^*$, with n^* equal to integer part of $\frac{1}{2}n - \frac{1}{2}$, as independent exponential random variables with means $f(x_{j,n}; \theta)$, and maximizing the corresponding likelihood function. The normalized difference $n^{\frac{1}{2}}(\hat{\theta} - \theta)$ has asymptotically the same normal distribution as the exact Gaussian maximum likelihood estimator. Although other techniques are used to prove this result (see the references above), the same central limit theorem would be obtained by using the simplifying assumption that all periodogram ordinates $I(x_{1,n}), \dots, I(x_{n^*,n})$ are exactly independent exponential random variables with means $f(x_{1,n}), \dots, f(x_{n^*,n})$.

3. DEFINITION OF FEXP-PROCESSES

The best known parametric models, which allow modelling of long memory and short-range properties simultaneously, are fractional ARIMA models (Granger & Joyeux, 1980; Hosking, 1981). Fractional ARIMA models are a natural extension of standard ARIMA

models (Box & Jenkins, 1970). Due to the special structure of these models, computationally fast approximate maximum likelihood methods can be developed; see e.g. Haslett & Raftery (1989). As a simple alternative, we propose another class of models which is an extension of Bloomfield (1973). The advantage of this class is that one can estimate the parameters by using well known methods for generalized linear models. Generalized linear regression is implemented in several standard statistics software packages.

DEFINITION 1. Let $g: [-\pi, \pi] \rightarrow R_+$ be a positive function such that

$$\lim_{x \rightarrow 0} \frac{g(x)}{x} = 1$$

and $g(x) = g(-x)$. Define $h_0 \equiv 1$, and let h_1, h_2, \dots, h_p be functions which are piecewise continuous in the whole interval $[-\pi, \pi]$. Also assume that $h_k(x) = h_k(-x)$ and, for any n , the $n^* \times (p + 1)$ matrix H with column vectors

$$\{h_k(2\pi/n), h_k(2\pi 2/n), h_k(2\pi 3/n), \dots, h_k(2\pi n^*/n)\}^T \quad (k = 0, 1, \dots, p)$$

is nonsingular. Furthermore, let $\theta = (\eta_0, H, \eta_1, \dots, \eta_p)$ be a real vector with $\frac{1}{2} \leq H < 1$. We call X_t a fractional EXP-process (or an FEXP-process) with short-memory components h_1, \dots, h_p and long-memory component g , if its spectral density is given by

$$f(x; \theta) = g(x)^{1-2H} \exp \left\{ \sum_{j=0}^p \eta_j h_j(x) \right\}. \tag{2}$$

Similarly to fractional ARIMA-models, the class of FEXP-processes is very flexible.

Example 1: $g(x) = |1 - e^{ix}|$, $h_k(x) = \cos kx$ ($k = 0, 1, \dots, p$). If $H = \frac{1}{2}$ (short memory), we obtain the model class proposed by Bloomfield (1973). In its generality, this class is comparable to ARMA-models: any piecewise continuous spectral density can be approximated with arbitrary accuracy. If $\theta_j = 0$ for $j \geq 3$, then we obtain a fractional ARIMA $(0, d, 0)$ model with $d = H - \frac{1}{2}$.

Example 2: $g(x) = |1 - e^{ix}|$, $h_k(x) = |x|^k$ ($k = 0, 1, \dots, p$). A data example with $H = \frac{1}{2}$, where a second order polynomial ($p = 2$) makes sense intuitively, is given by Diggle (1990, pp. 125–6). In addition to that, Diggle multiplies the spectrum defined by (2) and $H = \frac{1}{2}$ by an AR(1)-spectrum $1/(1 - 2\alpha \cos x + \alpha^2)$. The same extension can be applied to the models considered here.

4. APPROXIMATE MAXIMUM LIKELIHOOD ESTIMATION FOR FEXP-PROCESSES

For an FEXP-process, estimation of θ by the method described in § 2 becomes very simple. The method followed from the assumption that

$$I(x_{j,n}) = f(x_{j,n}; \theta) Z_j \quad (k = 1, \dots, n^*),$$

with independent exponential random variables Z_j and $E(Z_j) = 1$. For FEXP-models we have

$$\log f(x; \theta) = (1 - 2H) \log g(x) + \sum_{j=0}^p \eta_j h_j(x).$$

Therefore, estimation of θ reduces to the estimation of the regression parameters in a generalized linear model with logarithmic link function, multiplicative exponential error,

intercept $\beta_0 = \eta_0$, slope parameters $(\beta_1, \dots, \beta_{p+1}) = (1 - 2H, \eta_1, \dots, \eta_p)$ and independent x -variables $\log g(x), h_1(x), \dots, h_p(x)$. Estimation, tests, confidence intervals and diagnostics can be obtained by applying standard methods and software for generalized linear models. The introduction of the long-memory parameter does not destroy this convenient property. This is due to the simple modelling of the long-memory component by an exponential function

$$g(x)^{1-2H} = \exp \{(1 - 2H) \log g(x)\}$$

and due to the fact that the limiting properties of the periodogram are essentially, though not exactly, the same as for short-memory processes; see § 2 above.

5. A DATA EXAMPLE

We consider the amount of coded information per frame, measured by the logarithm of the number of so-called ATM cells per frame, for the first 1000 frames of a video signal from a 'low activity' video scene with no camera movement and no change in the background; see Fig. 1. The scene was coded using a variable-bit-rate (VBR) codec, which is designed for high-speed telecommunication networks. About 30 frames per second are processed. More technical details about this series are given by Heeke (1991) and Heyman, Tabatabai & Lakshman (1992). The periodogram in log-log coordinates is displayed in Fig. 2. It was noted by J. Beran, R. Sherman, M. S. Taqqu and W. Willinger in a Bellcore, Morristown technical memorandum that long memory is a typical feature of VBR-data. Moreover, for several of the 20 data sets considered there, the periodogram, in log-log coordinates, turned out to have the same qualitative features as Fig. 2, a negative slope near the origin and otherwise an essentially concave shape. A simple way to model this shape is to use the polynomial FEXP-models of Example 2 in § 3. One might in fact expect that a low degree polynomial can capture the concave shape. This is confirmed by the results in Table 1 and Fig. 2. Figure 2 shows a good agreement between observed and fitted spectrum with $p = 2$. Together with the p -values in Table 1, this suggests that a quadratic polynomial is sufficient.

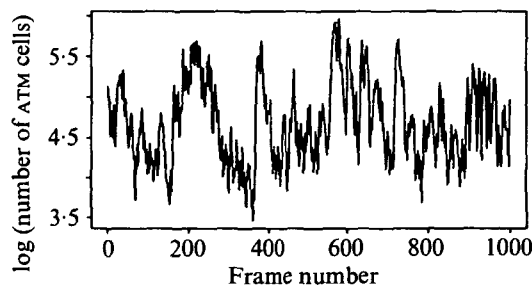


Fig. 1. Variable-bit-rate data by H. Heeke and E. Hundt.

In conclusion, we can say that by embedding the analysis in the framework of generalized linear models, we can obtain a quick first analysis of the data, as well as develop more complex models in an intuitive way. As in most situations where regression is applied, the choice of the functions h_1, \dots, h_p is up to a certain degree arbitrary, and requires some skill and experience. An interesting topic for future research will be to develop procedures which would allow to estimate these functions nonparametrically, for example by using estimation methods for additive models; see e.g. Buja, Hastie & Tibshirani (1989).

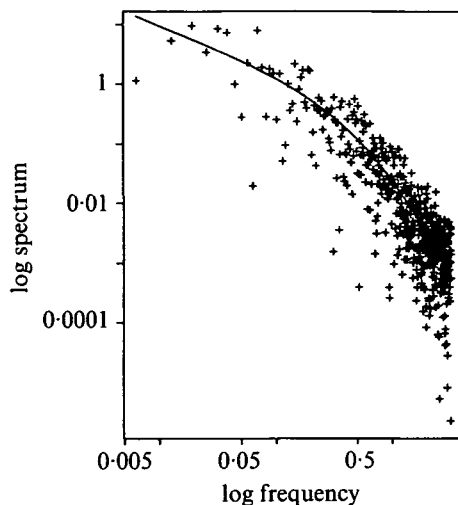


Fig. 2. Periodogram and fitted spectrum of FEXP-model (in log-log coordinates) for VBR data of Fig. 1.

Table 1. Parameter estimates for the FEXP model with spectral density $f(x; \theta) = |1 - e^{ix}|^{1-2H} \exp(\sum \eta_j |x|^j)$, where the summation is over $j = 0, \dots, p$, fitted to the data of Fig. 1, by generalized linear regression

p	Parameter	Estimate	St. dev.	z-statist	p -value
2	β_0	-1.343	0.249	-5.93	0.0000
	β_1	-0.783	0.107	-7.32	0.0000
	β_2	-2.856	0.303	-9.43	0.0000
	β_3	0.428	0.072	5.92	0.0000
3	β_0	-1.209	0.425	-2.84	0.0045
	β_1	-0.741	0.150	-4.94	0.0000
	β_2	-3.124	0.776	-4.03	0.0001
	β_3	0.579	0.422	1.37	0.1698
	β_4	-0.027	0.076	-0.36	0.7218

Notation: $\beta_0 = \eta_0$, $\beta_1 = 1 - 2H$, $\beta_{j+1} = \eta_j$, for $j \geq 1$. The p -values are given for testing $\beta_i = 0$ against the two-sided alternative $\beta_i \neq 0$.

ACKNOWLEDGEMENTS

I would like to thank the referee and the associate editor for their very helpful and constructive comments. Also, I would like to thank H. Heeke and E. Hundt (both at Siemens, Munich) and W. Willinger and A. Tabatabai (both at Bellcore, Morristown) for providing me with the VBR data.

REFERENCES

BERAN, J. (1992). Statistical methods for data with long-range dependence (with discussion). *Statist. Sci.* 7, 404-27.
 BLOOMFIELD, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* 60, 217-26.

- BOX, G. E. P. & JENKINS, G. M. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden Day.
- BUJA, A., HASTIE, T. & TIBSHIRANI, R. (1989). Linear smoothers and additive models. *Ann. Statist.* **17**, 453–510.
- CAMERON, M. A. & TURNER, T. R. (1987). Fitting models to spectra using regression packages. *Appl. Statist.* **36**, 47–57.
- COX, D. R. (1984). Long-range dependence: a review. In *Statistics: An Appraisal. Proceedings 50th Anniversary Conference*, Ed. H. A. David and H. T. David, pp. 55–74. Iowa State University Press.
- DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17**, 1749–66.
- DIGGLE, P. (1990). *Time Series: A Biostatistical Introduction*. Oxford University Press.
- FOX, R. & TAQQU, M. S. (1986). Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14**, 517–32.
- GIRAITIS, L. & SURGAILIS, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to the asymptotic normality of Whittle's estimate. *Prob. Theory Rel. Fields* **86**, 87–104.
- GRANGER, C. W. J. & JOYEUX, R. (1980). An introduction to long-range time series models and fractional differencing. *J. Time Ser. Anal.* **1**, 15–30.
- HAMPEL, F. R. (1987). Data analysis and self-similar processes. In *Proceedings of the 46th Session of ISI, Tokyo*, Book 4, pp. 235–54.
- HASLETT, J. & RAFTERY, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource (invited paper with discussion). *J. Appl. Statist.* **38**, 1–21.
- HEEKE, H. (1991). Statistical multiplexing gain for variable bit rate video codecs in ATM networks. *Int. J. Digit. Analog Commun. Syst.* **4**, 261–8.
- HEYMAN, D., TABATABAI, A. & LAKSHMAN, T. V. (1992). Statistical analysis and simulation of video teleconferencing traffic in ATM networks. *IEEE Trans. Circuits Syst. Video Technol.* **2**, 49–59.
- HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68**, 165–76.
- KOLMOGOROV, A. N. (1941). Local structure of turbulence in fluid for very large Reynolds numbers. Transl. and reprinted (1961) in *Turbulence*, Ed. S. K. Friedlander and L. Topper. New York: Interscience Publishers.
- KÜNSCH, H. R. (1986). Statistical aspects of self-similar processes (invited paper). In *Proc. First World Congress of the Bernoulli Society, Tashkent*, **1**, Ed. Y. Prohorov and V. V. Sazonov, pp. 67–74. Utrecht: VNU Science Press.
- MANDELBROT, B. B. (1983). *The Fractal Geometry of Nature*. New York: Freeman.
- MCCULLAGH, P. & NELDER, J. A. (1983). *Generalized Linear Models*. London: Chapman and Hall.
- WHITTLE, P. (1953). Estimating and information in stationary time series. *Ark. Math.* **2**, 423–34.

[Received August 1992. Revised March 1993]