Nonparametric M-estimation with long-memory errors

Jan Beran
Department of Mathematics and Statistics,
University of Konstanz, Germany

Sucharita Ghosh
Landscape Department,
WSL, Switzerland

Philipp Sibbertsen
Department of Statistics,
University of Dortmund, Germany

July 10, 2000

Abstract

We investigate the behavior of nonparametric kernel M-estimators in the presence of long-memory errors. The optimal bandwidth and a central limit theorem are obtained. It turns out that in the Gaussian case all kernel M-estimators have the same limiting normal distribution. The motivation behind this study is illustrated with an example.

1 Introduction

Consider the regression model

$$y_i = g(t_i) + G(Z_i),\ i = 1, 2, ..., n,\ t_i = \frac{i}{n} \tag{1}$$

where $g(t)$ is an unknown regression function in $C^3[0, 1]$ and the error process $\epsilon_i = G(Z_i)$ is generated by a transformation of a zero mean stationary Gaussian process $Z_i, i \in N$, with long-memory. Here $G : R \to R$ is a measurable function and if Z denotes a standard normal variable, $E[G(Z)] = 0$. This paper deals with the nonparametric M-estimation of the regression function $g(t)$ in this model.

This work was motivated by the recent occurrences of storm events that had disastrous environmental damages in Switzerland and in the neighboring countries. This resulted in increased need for developing suitable models for storms, in particular for wind speed. As part of this research, one basic problem is to estimate time-dependent location parameters of the marginal wind-speed distributions. The estimator should not be affected by outliers, or in this case, unusually large observations (extremely strong wind speed or storms) which in spite of having short duration tend to affects the usual (mean) location estimators. This in turn implies the development of M-estimation procedures for time series, in particular time series which may have long-memory. Figure 1 shows a wind speed series (in 0.1 m/s) measured at a climate station near Zürich (source: SMA). This series consists of hourly maxima during the years 1998-1999. The higher wind speeds during the winters can be seen as well as the storm *Lothar* that occurred in the December of 1999. For the purposes of illustration, the time dependent mean and approximate median functions were estimated by kernel smoothing of the log transformed data. For this, we took a kernel that is proportional to

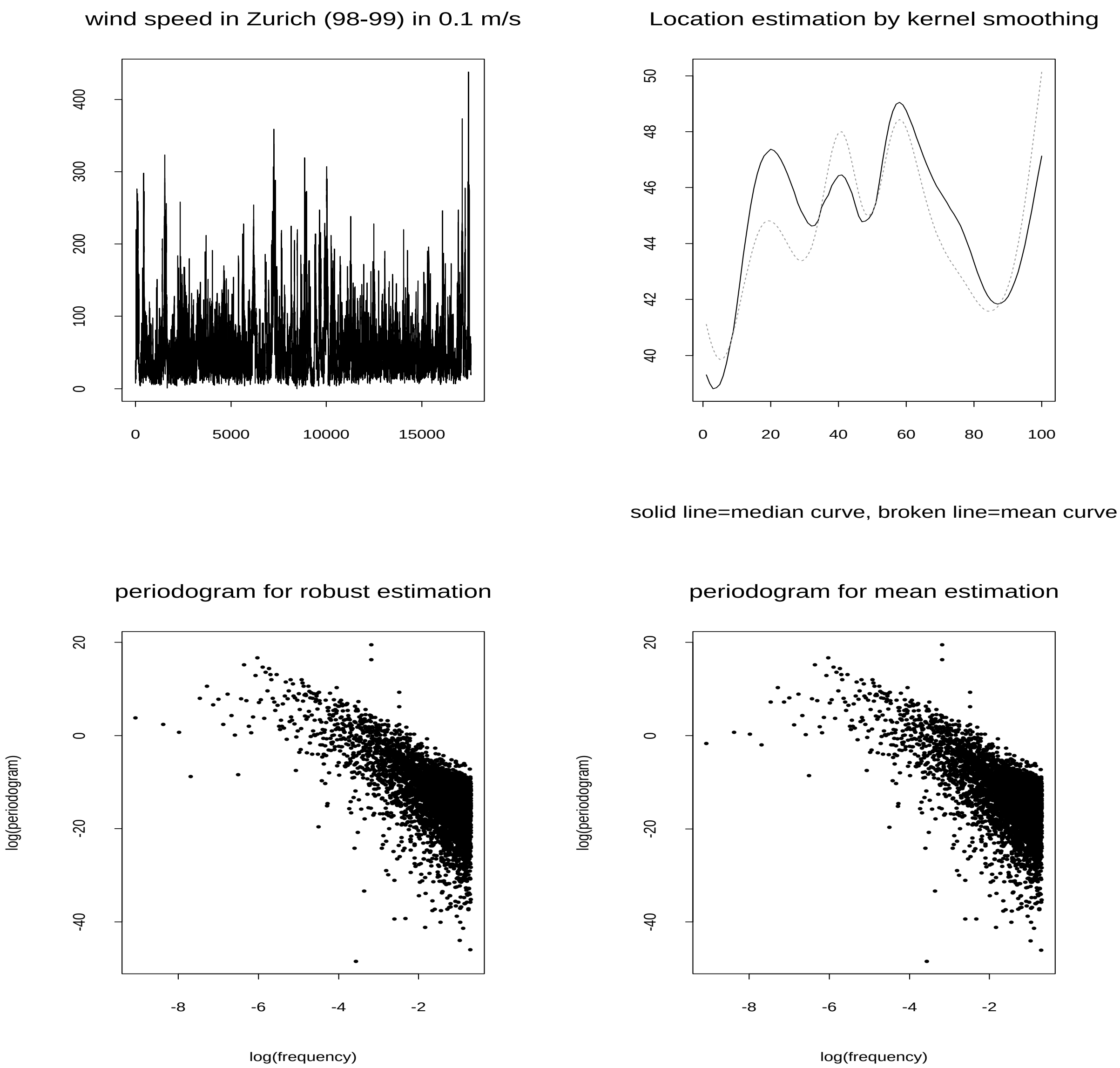

Figure 1: Hourly wind speed maxima in Zurich in 1999, estimated mean (broken line) and median (solid line) curves fitted to the logarithm of the original data, and the peridograms (in log-log-coordinates) of the residuals.

the normal density function (truncated on [-1,1]) for an arbitrary bandwidth equal to 0.05. The estimated mean and the median functions are shown along with the periodograms for the corresponding residual estimates in the log-log coordinates. The presence of long-memory in the residuals is evident as well as the fact that the median estimate is less affected by the fast wind speed due to Lothar. The changes in the probability distribution function over time can be seen in particular by noticing the shifts in the positions of the median curve in relation to the mean. Also, the latter curve is seen to 'explode' towards the end of the series, which is the effect of *Lothar*, whereas the median curve is less affected by these sudden jumps in the observations.

A standard procedure for kernel estimators is to use the Naradaya-Watson estimator $\hat{g}(t)$ of $g(t)$ obtained by minimizing the function

$$H(\theta) = \frac{1}{nb_n} \sum_{i=1}^{n} K(\frac{t_i - t}{b_n})(y_i - \theta)^2,$$

where $b_n > 0$ is a bandwidth and $K(x)$ is a suitably defined kernel. In the presence of long-memory, Naradaya-Watson type estimators were considered among others by Beran (1999), Csörgö & Mielniczuk (1995) and Hall & Hart (1990). Ghosh (2000) considers kernel estimation for replicated time series data including long-memory models. Due to the fact that Naradaya-Watson estimators can be considered as local least squares estimators and the least squares estimators are not robust, the Naradaya-Watson type estimators are also not robust.

In this paper we consider the so-called M-smoothers $\hat{g}$ defined as the solution of the equation:

$$\frac{1}{nb} \sum_{i=1}^{n} K(\frac{t_i - t}{b_n})\psi(y_i - \theta) = 0 \qquad (2)$$

where $\psi : R \to R$ is a measurable function which can be chosen in various ways such as the classical Huber-ψ-function $\psi(x) = \min(c, \max(x, -c))$.

Beran (1991) considers M-estimation for (constant) location parameter for Gaussian Long-memory processes. Robinson(1984) and Härdle(1989) also considered robust kernel M-estimators, but in the case of independent errors. Boente & Fraiman(1989) obtained strong consistency for kernel M-estimators for mixing processes. Sibbertsen (1999) consider robust parametric regression models for long-memory models. The aim of this paper is to study the properties of robust nonparametric M-estimators for long-memory processes. References to long-memory processes include Beran (1992, 1994), Cox (1984), Granger & Joyeux (1980), Hosking (1981) and Mandelbrot & van Ness (1968).

2 Asymptotic results

Define

$$\begin{aligned}
I(g^{"}) &= \int_{\Delta}^{1-\Delta} [g^{''}(t)]^2 dt; , 0 < \Delta < \frac{1}{2} \\
I(K) &= \int_{-1}^{1} x^2 K(x) dx; \\
V_\psi(d) &= \frac{1}{(E\psi')^2} \int_0^1 \int_0^1 K(x)K(y)|x-y|^{2d-1} dx dy.
\end{aligned}$$

Assumptions:

- $\gamma_Z(k) = cov(Z_i, Z_{i+k}) \sim C_Z |k|^{2d-1}$ for $d \in (0, 1/2)$ and $k \to \infty$.
- K is a positive symmetric kernel on $[-1, 1]$ with $\int_{-1}^{1} K(x)dx = 1$;
- $b_n > 0$ is a sequence of bandwidths such that $b_n \to 0$ and $nb_n \to \infty$ as $n \to \infty$.
- ψ is differentiable almost everywhere with respect to the Lebesgue measure;
- $E(\psi(\epsilon_i)) = 0, E(\psi^2(\epsilon_i)) < \infty$;
- $E(\psi'(\epsilon_i)) \neq 0$;
- there exists a measurable function M_2 such that $|\frac{d^2}{dt^2}\psi(t)| < M_2(t),\ E[M_2(t)] < k_2 < \infty$;
- there exists a measurable function M_3 such that $|\frac{d^3}{dt^3}\psi(t)| < M_3(t),\ E[M_3(t)] < k_3 < \infty$;
- $h_\delta(y) = \sup_{x \leq \delta} |\psi(y+x) - \psi(y)| \leq c$ almost everywhere for some $\delta > 0$ and $0 < c < \infty$;
- for δ tending to zero, $h_\delta(y)$ tends to zero almost everywhere;
- $\psi(G(Z))$ has Hermite rank m ($m \geq 1$) and c_m denotes the mth Hermite coefficient.

Theorem 1 *Under the above assumptions and if $d \in (0, 0.5)$, we have*

1. *Uniform consistency: There exists a sequence of solutions $\hat{g}_n$ such that $sup_{t\in[\Delta,1-\Delta]}|\hat{g}_n(t) - g(t)| \longrightarrow 0$ as $n \to \infty$ in probability;*
2. *Bias: $E[\hat{g}(t)-g(t)] = \frac{1}{2}b_n^2 g''(t)I(K)+o(b_n^2)$ uniformly in $\Delta < t < 1-\Delta$;*
3. *Variance: $(nb_n)^{1-2d}var(\hat{g}(t)) = c_m^2 C_Z^m V_\psi(d)$ uniformly in $\Delta < t < 1-\Delta$ if $\frac{1}{2}(1-1/m) < d < \frac{1}{2}$.*
4. *$IMSE$: If $\frac{1}{2}(1-1/m) < d < \frac{1}{2}$, the integrated mean square error in $[\Delta, 1-\Delta]$ is given by*
$$\int_\Delta^{1-\Delta} E\{[\hat{g}(t) - g(t)]^2\}dt = b_n^4 \frac{I(g'')I^2(K)}{4} + (nb_n)^{2d-1}V_\psi(d)+$$
$$o(\max(b_n^4, (nb_n)^{2d-1}));$$
5. *Optimal bandwidth: If $\frac{1}{2}(1-1/m) < d < \frac{1}{2}$, the bandwidth that minimizes the asymptotic $IMSE$ is given by: $b_{opt} = C_{opt}n^{(2d-1)/(5-2d)}$, where $C_{opt} = C_{opt}(\theta) = [\frac{(1-2d)V_\psi(d)}{I(g'')I^2(K)}]^{1/(5-2d)}$.*
6. *Asymptotic distribution: For every fixed t, and $\frac{1}{2}(1-1/m) < d < \frac{1}{2}$, $(nb)^{1/2-d}c_m^{-1}C_Z^{-m/2}V_\psi^{-1/2}[\hat{g}(t) - g(t)]$ converges in distribution to Z_m, where Z_m is a Hermite process (see Rosenblatt 1984 and Taqqu 1975) of order m at time 1.*

If $m = 1$ and $G(x) = x$, then $c_m = E[\psi'(\epsilon)]$ and $V_\psi(d) = \frac{1}{c_m^2}\int_0^1\int_0^1 K(x)K(y)|x-y|^{2d-1}dxdy$ so that we have

Corollary 1 *If $G(x) = x$, then for all ψ functions satisfying the conditions of Theorem 1 and such that $\psi(G(Z))$ has Hermite rank 1, $(nb)^{\frac{1}{2}-d}C_Z^{-1/2}[\hat{g}(t) - g(t)]$ converges to a normal distribution with zero mean and variance equal to $\int_0^1\int_0^1 K(x)K(y)|x-y|^{2d-1}dxdy$.*

Thus the above corollary implies that when $m = 1$ and the error process ϵ_i is Gaussian, then the limiting distribution of $(nb)^{1/2-d}C_m^{-1}C_Z^{-m/2}V_\psi^{-1/2}[\hat{g}(t) - g(t)]$ is the same Gaussian distribution for all M-estimators.

3 Appendix

Proof of Theorem 1: We start with the estimating equation

$$\frac{1}{nb_n}\sum \psi(y_i - \hat{g}(t))K(\frac{t_i - t}{b_n}) = 0.$$

The uniform consistency follows, in a similar way as in Beran (1991), Hall & Hart (1991) and Huber (1967), by Taylor expansion of the estimating equation, the law of large numbers and properties of g and ψ.
As for bias, using Taylor expansion and the assumptions stated above

$$\hat{g}(t) - g(t) = \sum \psi(y_i - g(t))K(\frac{t_i - t}{b_n})[\sum \psi'(y_i - g(t))K(\frac{t_i - t}{b_n})]^{-1} + R_n$$

where $E(R_n) = o(b^2)$. The result now follows by noting that $y_i - g(t) = g(t_i) - g(t) + \epsilon_i$, where the expected value of ϵ_i is zero and

$$g(t_i) - g(t) + \epsilon_i = g'(t)(t_i - t) + g''(t)\frac{(t_i - t)^2}{2} + g'''(t^*)\frac{(t_i - t)^3}{6} + \epsilon_i(t), |t^* - t| \leq |t_i - t|,$$

so that

$$\psi(y_i - g(t)) = \psi(\epsilon_i) + \psi'(\epsilon_i)[g'(t)(t_i - t) + g''(t)\frac{(t_i - t)^2}{2} + g'''(t^*)\frac{(t_i - t)^3}{6}] +$$

$$\frac{\psi''(\xi_i)}{2}[g'(t)(t_i - t) + g''(t)\frac{(t_i - t)^2}{2} + g'''(t^*)\frac{(t_i - t)^3}{6}]^2, |\xi_i - \epsilon_i| \leq |g(t_i) - g(t)|.$$

The final result follows from the properties of ψ and K.
The expression for the variance follows from Taylor expansion of the estimating equation around g and by noting that the ψ function can be expanded via a Hermite expansion of order m. Moreover, $cov(H_p(Z_i), H_q(Z_j)) = 0$ if $p \neq q$, and equal to $[cov(Z_i, Z_j)]^p$ otherwise. Recall that $[cov(Z_i, Z_j)]^p \sim C_Z|i - j|^{2d-1}$ when $|i - j| \to \infty$.
The $IMSE$ and the formula for the optimal bandwidth follows from standard arguments whereas the limiting distribution of $\hat{g}(t)$ follows from Taqqu (1979) and the expansions above.

4 Acknowledgement

Felix Forster (Department of Natural Hazards, WSL, Switzerland) kindly prepared the wind speed data (source: SMA, Zürich) used in this paper. His help is gratefully acknowledged.

5 References

Beran, J. (1991). *M* estimators of location for gaussian and related processes with slowly decaying serial correlations. *Journal of the American Statistical Association* **86**, 704 - 707.

Beran, J. (1992). Statistical methods for data with long-range dependence. *Statistical Science* **7**, 404 - 427.

Beran, J. (1994). *Statistics for long-memory processes.* Chapman & Hall, New York.

Beran, J. (1999) "SEMIFAR Models - A Semiparametric Framework for Modelling Trends, Long Range Dependence and Nonstationarity", Juni 1999 , Discussion paper of the Center of Finance and Econometrics, No. 99/16.

Boente, G., Fraiman, R. (1989). Robust nonparametric regression estimation for dependent observations. *The Annals of Statistics* **17**, 1242 - 1256.

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, Ames: Iowa State University Press, 55 - 74.

Csörgö, S., Mielniczuk, J. (1995) Nonparametric regression under long-range dependent normal errors. Annals of Statistics. 23, 1000-1014.

Ghosh, S. (2000) Nonparametric trend estimation in replicated time series (submitted).

Granger, C., Joyeux, R. (1980). An introduction to long-range time series models and fractional differencing. *J. Time Ser. Anal.* **1**, 15 - 30.

Hall, P. Hart, J.D. (1990) Convergence rates in density estimation for data from infinite-order moving average processes. *Probab. Theory and Related Fields* **87**, 253-274.

Härdle, W. (1989). Asymptotic maximal deviation of M-smoothers. *Journal of Multivariate Analysis* **29**, 163 - 179.

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

Huber, P. (1967) The behaviour of maximum likelihood estimators under nonstandard conditions, in Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability (vol. 1), eds. L. Le Cam and J. Neymann, pp. 221-233.

Mandelbrot, B. B., van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10, No. 4**, 422 - 437.

Robinson, P. M. (1984). Robust nonparametric regression. *Robust and nonlinear time series analysis. Lecture Notes in Statistics* **26**, 247 - 255.

Rosenblatt, M. (1984) Stochastic processes with short-range and long-range dependence. Statistics: An appraisal, Proceedings 50th Anniversary Conference, H. A. David, H.T.David (eds.). The Iowa State University Press, pp. 509-520.

Sibbertsen, P. (1999). *Robuste Parameterschätzung im linearen Regressionsmodell bei Fehlertermen mit langem Gedächtnis.* Verlag für Wissenschaft und Forschung, Berlin (in German).

Taqqu, M. S. (1975). Weak convergence to fractional Brownian Motion and to the Rosenblatt Process. *Probability Theory and Related Fields* **31**, 287 - 302.

Taqqu, M. S. (1979). Convergence of integrated processes of arbitrary Hermite rank. Z. Wahrsch. verw. Geb., 50, 53-83.