On robust local polynomial estimation with long-memory errors

by
Jan Beran[1], Yuanhua Feng[1],
Sucharita Ghosh[2] and Philipp Sibbertsen[3]

Abstract

Prediction in time series models with a trend requires reliable estimation of the trend function at the right end of the observed series. Local polynomial smoothing is a suitable tool because boundary corrections are included implicitly. However, outliers may lead to unreliable estimates, if least squares regression is used. In this paper, local polynomial smoothing based on $M-$estimation is considered for the case where the error process exhibits long-range dependence. In constrast to the iid case, all $M-$estimators are asymptotically equivalent to the least square solution, under the (ideal) Gaussian model. Outliers turn out to have a major effect on nonrobust bandwidth selection, in particular due to the change of the dependence structure.

1 Introduction

1.1 The model

Let $y_1, ..., y_n$ be an observed time series such that

$$y_i = g(t_i) + \epsilon_i, \tag{1}$$

where $t_i = i/n \in [0, 1]$, $g(t)$ is an unknown sufficiently smooth function (e.g. in $C^3[0, 1]$), $\epsilon_i = G(Z_i)$ where Z_i is a zero mean stationary Gaussian process with autocovariances $\gamma_Z(k) = \text{cov}(Z_i, Z_{i+k})$ and spectral density

[1]Department of Mathematics and Statistics, University of Konstanz, D-78457 Konstanz, Germany

[2]Landscape Department, Swiss Federal Research Institute WSL, CH-8903 Birmensdorf, Switzerland

[3]Fachbereich Statistik, Universität Dortmund, D-44221 Dortmund, Germany

$f_Z(\lambda) = (2\pi)^{-1} \sum \gamma_Z(k)\exp(ik\lambda)$ and G is such that $E[G(Z)] = 0$. For defining nonparametric $M-$estimators, a function ψ with the following properties will be needed: ψ is almost everywhere differentiable, $E[\psi(\epsilon)] = 0$, $E[\psi'(\epsilon)] \neq 0$, $E[\psi^2(\epsilon)] < \infty$, and $\psi\{G(.)\}$ has Hermite rank $m \geq 1$. The last property means that, for $c_q = E[\psi\{G(Z)\}H_q(Z)]$, we have $c_q = 0$ $(0 \leq q \leq m-1)$ and $c_m \neq 0$. Here $H_q(x) = \exp(x^2/2)\frac{d^q}{dx^q}\exp(-x^2/2)$ denotes the qth Hermite polynomial. Moreover, it is assumed that the spectral density f_Z is continuous in $[-\pi, 0) \cup (0, \pi]$ and diverges to infinity at the origin at the rate

$$f_Z(\lambda) \sim_{|\lambda| \to 0} c_{f,Z}|\lambda|^{-2d} \tag{2}$$

for some $\frac{1}{2} - \frac{1}{2m} < d < \frac{1}{2}$, $m \in N$, and $0 < c_{f,Z} < \infty$. This condition implies (see e.g. Beran 1992, 1994): 1. $\gamma_Z(k) \sim c_{\gamma,Z}|k|^{2d-1}$ $(0 < c_{\gamma,Z} < \infty)$, as $|k| \to \infty$; 2. the process $\xi_i = \psi\{G(Z_i)\}$ has long memory in the sense that its spectral density f_ξ has a pole at zero of the form $c_{f,\xi}|\lambda|^{-2d_m}$ with $0 < d_m = \frac{1}{2} + m(d - \frac{1}{2}) < \frac{1}{2}$; 3. $\gamma_\xi(k) = \mathrm{cov}(\xi_i, \xi_{i+k}) \sim m! c_m^2 c_{\gamma,Z}^m |k|^{2d_m - 1}$, as $|k| \to \infty$. For additional detailed regularity conditions needed in the context of $M-$smoothing with long-memory errors see e.g. Beran et al. (2000).

1.2 Local polynomial $M-$smoothing and prediction

Suppose, the aim is to predict a future observation y_{n+k} for some $k > 0$. This requires prediction of the stochastic part ϵ_{i+k} and extrapolation of $\hat{g}$. This problem is considered in Beran and Ocker (1999) for so-called $SEMIFAR-$models which include, among others, model (1) with ϵ_i equal to a fractional ARIMA model (Granger and Joyeux 1980, Hosking 1981). Beran and Ocker use optimal linear prediction of the stochastic part and extrapolation of $\hat{g}$ by Taylor expansion. An important problem that has to be solved before extrapolating the trend function is to obtain reliable estimates of $g(t)$ and its derivative(s) at the right boundary (i.e. for t close to 1). Local polynomial smoothers are suitable for this purpose, because of the built-in automatic boundary correction (see e.g. Fan and Gijbels 1996, Beran and Feng 1999). A second problem is that occasional 'outliers' may have an undue influence on local polynomial estimates that are based on least squares regression. (Note that the notion 'outlier' does not necessarily imply that an outlying observation is wrong, but rather that it does not belong to the "majority" of the data or the ideal central model - see Hampel et al. 1986 and Huber 1981). Also, in some cases one may be interested in other location curves, such as the median, instead of the expected value.

This motivates local polynomial smoothing based on $M-$estimation. Two aspects are investigated in this paper: 1. the asymptotic mean squared error and optimal bandwidth; 2. the effect of outliers on automatic bandwidth selection.

1.3 A data example

Figure 1 shows a wind speed series (in 0.1 m/s) measured at a climate station in Disentis, Switzerland (source: SMA) in the years 1997-1999. The series consists of 6-hours-maxima. Thus, for each day, there are four observations corresponding to the maximal wind speed between 0 to 6 am, 6 to 12 am, 12 am to 6 pm and 6 to 12 pm respectively. The dotted line in figure 1 displays the local linear fit based on least squares estimation, whereas the full line is the median (or L_1) local polynomial fit obtained from (3) with $\psi(x) = sign(x)$. (Note that, the derivative of the sign function is zero almost everywhere, but it can be approximated arbitrarily well by $\psi-$functions with $E[\psi'(\epsilon)]\neq 0$.) The bandwidth b was obtained by the iterative plug-in algorithm in Beran and Feng (2000) which is based on least squares estimation.

The following observations can be made. The wind speed series exhibits occasional "outliers" corresponding to sudden high speed winds (storms). There is a seasonal difference between the least squares and the L_1 fit. The mean curve (least squares fit) is clearly above the median curve (L_1 fit) in the middle of winter, whereas the two curves almost coincide in summer. This indicates a seasonal change of the distribution. In summer, high winds are frequent (i.e. correspond to "normal" observations), because of frequent storms. In winter, *occasional* storms lead to rare but extreme wind speeds. Also note that figures 2a and b with the periodogram of the residuals (in log-log coordinates) indicate long-range dependence and an additional strong seasonal component.

In this example, a bandwidth based on the least squares fit was used. This leads to two questions that are discussed in this paper:

1. What is the integrated asymptotic mean square error of and the optimal bandwidth for a local polynomial $M-$estimator, when the random deviations from the deterministic trend function g have long memory?

2. How do "outliers" influence an estimated bandwidth obtained by a least squared based plug-in algorithm?

2 Definition of local polynomial $M-$estimators

Let K be a positive symmetric kernel with support $[-1,1]$ and $\int_{-1}^{1} K(u)du = 1$, let $t \in [0,1]$ and $b \in (0,1)$ a positive bandwidth, and denote by $p \in N$ the degree of the local polynomial. Then a local polynomial $M-$estimator of $g(t)$ is defined by $\hat{g}_{\psi}(t) = x'(t)\hat{\beta}(t)$ where $x(t) = (1, t, t^2, ..., t^p)' \in R^{p+1}$, x' denotes the transposed of x and $\hat{\beta}(t) \in R^{p+1}$ solves the system of $p+1$ equations

$$\frac{1}{nb}\sum_{i=1}^{n} K(\frac{t_i - t}{b})\psi\{y_i - x'\hat{\beta}(t)\}x_j(t) = 0 \ (j = 0, 1, ..., p). \tag{3}$$

Note that for $\psi(u) \equiv u$, equation (3) defines the local polynomial fit based on least squares regression. In the following the least squares estimator of g will be denoted by $\hat{g}_{LSE}$. Robust estimates are obtained by using bounded functions ψ (see e.g. Huber 1981, Hampel et al. 1986). A standard example is the Huber function $\psi(x) = \min(c, \max(x, -c))$ with $0 < c < \infty$. Different $\psi-$functions can also be used to obtain estimates of location curves other than the mean. For instance, the Huber function with $c = 0$ (or c close to zero) estimates the median function. The following results hold for arbitrary $\psi-$functions and location curves.

3 Asymptotic mean squared error

3.1 Variance

To simplify presentation, the rectangular kernel $K(u) = \frac{1}{2}1\{-1 \leq u \leq 1\}$ is used here. The extension to general kernels is straightforward. The asymptotic variance and bias of least squares local polynomial estimates for long-memory processes is considered in Beran and Feng (1999). For local polynomial $M-$estimators, the following result holds:

Theorem 1 *Let $\hat{\beta}$ be the solution of (3). Define the following $(p+1)\times(p+1)$ matrices: $M_n = (m_{ij})_{i,j=1,...,p+1}$ with $m_{ij} = \text{cov}\{\hat{\beta}_{i-1}(t), \hat{\beta}_{j-1}(t)\}$, $P = (p_{ij})_{i,j=1,...,p+1}$ with $p_{ij} = 0$ for $i+j$ odd and $p_{ij} = \sqrt{(2j-1)(2l-1)}/(j+l-1)$ for $i+j$ even, $\kappa_{ij}(d_m) = \sqrt{(2i-1)(2l-1)}\Gamma(1-2d_m)/[\Gamma(d_m)\Gamma(1-d_m)]$, $Q = (q_{ij})_{i,j=1,...,p+1}$ with*

$$q_{ij} = \kappa_{ij}(d_m)\int_{-1}^{1}\int_{-1}^{1} x^{i-1}y^{j-1}|x-y|^{2d_m-1}dxdy,$$

$D_n = (d_{ij}(n))_{i,j=1,\ldots,p+1}$ with $d_{ij} = 0$ $(i \neq j)$ and $d_{jj} = 2(nb)^j/(2j-1)$. Then, as $n \to \infty$, $b \to 0$ and $nb \to \infty$,

$$(2nb)^{-2d_m} D_n M_n D_n \longrightarrow \frac{2\pi c_{f,Z}^m c_m^2}{m! E^2[\psi'(\epsilon)]} P^{-1} Q P^{-1}. \tag{4}$$

In particular, for $m = 1$ and $\epsilon_i = Z_i$, the first Hermite coefficient c_1 is equal to $E[\psi'(\epsilon)]$ so that the asymptotic variance of $\hat{\beta}(t)$ does not depend on the $\psi-$function. We thus have

Corollary 1 *Let $\hat{g}_\psi(t)$ be defined by (3), then*

$$(nb)^{1-2d_m} \mathrm{var}\{\hat{g}_\psi(t)\} \to v(t) \tag{5}$$

where $0 < v(t) < \infty$ does not depend on ψ.

An explicit expression for $v(t)$ is given in Beran and Feng (1999) (also see Ghosh 2000 for the case of repeated time series). A stronger version of Corollary 1 can also be proved, stating that $(nb)^{\frac{1}{2}-d_m}\{\hat{g}_\psi(t) - \hat{g}_{LSE}(t)\}$ converges to zero in probability. This is analogous to location estimation (Beran 1991), parametric regression (Giraitis et al. 1996) and kernel $M-$estimation (Beran et al 2000).

3.2 Bias

The bias of $\hat{g}_{LSE}$ for long-memory processes is considered in Beran and Feng (1999). Taylor expansion implies that the same asymptotic formula holds for all $M-$estimators. Thus, define $I(g^{(p+1)}) = \int_0^1 [g^{(p+1)}(t)]^2 dt$ and $I(K) = \int_{-1}^1 x^{(p+1)} K^*_{(0,p)}(x) dx$ where $K^*_{(0,p)}$ is the so-called equivalent kernel (see Beran and Feng 1999). Let $0 < \Delta < \frac{1}{2}$ be a small positive number. Then

$$E[\hat{g}_\psi(t) - g(t)] = b^{p+1} \frac{g^{(p+1)}(t) I(K)}{k!} + o(b^{p+1}) \tag{6}$$

uniformly in $\Delta < t < 1 - \Delta$. For boundary points, the order of the bias is the same, when p is odd, with $K^*_{(0,p)}$ replaced by an equivalent boundary kernel. Note that this result is the same as for $d = 0$ (see e.g. Fan and Gijbels, 1996) and also holds for $-0.5 < d < 0$.

3.3 IMSE and optimal bandwidth

The asymptotic integrated mean squared error (IMSE) follows from the results above:

$$\int_0^1 E\{[\hat{g}(t)-g(t)]^2\}dt \approx b^{2(p+1)}\frac{[g^{(p+1)}(t)]^2 I^2(K)}{[(p+1)!]^2}+(nb)^{2d_m-1}\int_0^1 v(t)dt \quad (7)$$

The bandwidth that minimizes the asymptotic IMSE is thus given by

$$b_{opt} = C_{opt} n^{(2d_m-1)/(2p+3-2d_m)} \quad (8)$$

where

$$C_{opt} = \{\frac{(1-2d)[(p+1)!]^2 \int_0^1 v(t)dt}{2(p+1)I(g^{(p+1)})I^2(K)}\}^{1/(2p+3-2d_m)}. \quad (9)$$

Similar results for robust local polynomial fits with independent errors may be found in Fan and Gijbels (1996) (p. 201, see also pp. 63ff). Note that the formula for the asymptotic MISE is given on the interval $[0,1]$, since a local polynomial estimator adapts automatically at the boundary.

3.4 Asymptotics for other values of d

The asymptotic formula (6) for the bias is correct in the whole range $-\frac{1}{2} < d < \frac{1}{2}$. This is not the case for the variance. If $d < \frac{1}{2} - \frac{1}{2m}$, then the process $\xi_i = \psi(\epsilon_i)$ is no longer long-range dependent. Therefore, the asymptotic equivalence between all $M-$smoothers no longer holds, and the variance depends on the function ψ. A general formula for the variance with $\psi(x) \equiv x$ (least squares local polynomial estimator) that is valid in the whole range $-\frac{1}{2} < d < \frac{1}{2}$ is given in Beran and Feng (1999). For general $M-$estimators, the situation is more complicated. Consider, for instance, the case of a Gaussian error process with $m = 1$ (and hence $d_m = d$). Then the following holds, in analogy to location estimation (Beran 1991): a) For $d > 0$, all $M-$estimators have the same asymptotic variance (see above); b) for $d = 0$, nonlinear $M-$estimators lose efficiency compared to $\hat{g}_{LSE}$, the efficiency is however still positive; c) for $d < 0$, all nonlinear $M-$estimators have asymptotic efficiency zero compared $\hat{g}_{LSE}$. Detailed formulas are omitted here, since the focus is on long memory ($d > \frac{1}{2} - \frac{1}{2m}$).

4 Effect of outliers on bandwidth selection

4.1 Motivation of the simulation study

Consider the case of Gaussian errors ϵ_i and $m = 1$. Then the results in the previous section imply that all local polynomial fits have the same asymptotic mean squared error and the same asymptotically optimal bandwidth. To choose an optimal bandwidth one might thus be tempted to always use automatic bandwidth selection based on least squares regression (see e.g. Ray and Tsay 1997, Beran 1999 and Beran and Feng 2000 for algorithms in the long-memory context), independently of the $\psi-$function used in the estimation. However, the difference between least squares and robust $M-$estimation comes into play under departures from the ideal Gaussian process. The following simulation study illustrates how outliers can affect the least squared based bandwith selection method in Beran and Feng (2000).

4.2 Theoretical considerations and design of the simulation study

The method in Beran and Feng (2000) is an iterative plug-in method that yields an estimated optimal bandwidth $\hat{b}$. Starting with an initial bandwidth, an initial value of $\hat{g}_{LSE}$ and preliminary residuals are calculated. In the next iteration, a new bandwidth is obtained from the residuals and so on. Outliers can influence the solution $\hat{b}$ by: 1. changing the estimated values of $\hat{g}_{LSE}(t)$; 2. changing the variance of the residuals; 3. changing the spectral density f_ϵ at the origin. The influence of outliers on the dependence structure is often stronger than that on the estimation of the mean function. The reason is that estimation of the long-memory parameter (and other dependence parameters) of the "uncontaminated" process becomes more difficult. The asymptotic formula for b_{opt} (equations 8 and 9) that is used in the plug-in method not only depends on g and the marginal variance of ϵ_i, but also on the behaviour of f_ϵ at zero.

In the simulation study, the following outlier model is used: The error process ϵ_i is a standardized mixture of a zero mean Gaussian FARIMA$(0, d, 0)$ model X_i with $\mathrm{var}(X_i) = 1$, and a "mild outlier process" consisting of observations $W_i = \frac{2}{\sqrt{3}} U_i$ with U_i iid t_3-distributed variables independent of the process X_i. Thus, $\epsilon_i = \{(1 - I_i)X_i + I_i W_i\}/\sqrt{(1 - p) + 4p}$ where I_i are iid Bernoulli variables with $P(I_i = 1) = 1 - P(I_i = 0) = p$. Note that $\mathrm{var}(W_i) = 4$ so that $\mathrm{var}(\epsilon_i) = 1$ for all values of p. Since ϵ_i is standardized, the change of $b_{opt} = b_{opt}(p)$ as a function of p (when calculated for the contaminated process ϵ_i) is due to the change in the dependence structure only. The effect of W_i on the dependence structure is that, compared to the uncontaminated process X_i, dependence is weaker (ϵ_i is closer to independence than X_i). More exactly, $\rho_\epsilon(k) = (1-p)/(1+3p)\rho_X(k)$ $(k \neq 0)$ where $\rho_X(k)$ are the autocorrelations of the FARIMA process X_i. This means that

1. For $d < 0$, $0 < |\sum \rho_\epsilon(k)| < \infty$ whereas $\sum \rho_X(k) = 0$. Thus ϵ_i loses the property of antipersistence. As a result, $b_{opt}(p)$ *increases* with increasing p and is of the order $O(n^{-1/5})$ which is the same as in the independent case.

2. For $d = 0$, the dependence structure does not change, whatever the value of p is. Thus, $b_{opt}(p)$ does not change.

3. For $d > 0$, adding iid observations does not change the long-memory parameter d. However, the constant $c_{f,\epsilon}$ is smaller than the corresponding constant $c_{f,X}$ for X_i, since the spectral density is now a weighted average of the spectral densities of X and W_i. Thus, with increasing p, the asymptotic bandwidth $b_{opt}(p)$ *decreases*.

The simulations were carried out for the two trends $g_1(t) = 2\tan\{5(t-0.5)\}$ and $g_2(t) = 4\sin^2\{(t-0.5)\pi\}$. Moreover, $d = -0.3, 0, 0.3$, $p = 0, 0.05, 0.1, 0.2, 0.5$ and the sample sizes $n = 500$ and 1000 were used. For each case, four hundred simulations were carried out. The estimated optimal bandwidth $\hat{b}$ for a local (unweighted) linear fit was selected for each replication by the data-driven SEMIFAR algorithm in Beran and Feng (2000).

4.3 Results of the simulation study

Table 1 gives the asymptotically optimal bandwidth $b_{opt} = b_{opt}(0)$ for the least squares local polynomial fit and $p = 0$, i.e. for the "ideal" distribution with no outliers. Tables 2 and 3 give the simulated mean, standard deviation, minimum, median and maximum of $\hat{b}$ for $n = 500$ and 1000 respectively. Box-plots of $\hat{b}$ as a function of p are shown in figures 3 and 4.

The simulation results confirm the theoretical considerations. In the case of an antipersistent error process X_i, iid outliers lead to an increase of $\hat{b}$. The estimated bandwidth therefore tends to increase with increasing p. The opposite is the case for $d > 0$, whereas practically no change can be observed for $d = 0$.

5 Final remarks

The simulations in this paper illustrate that bandwidth choice may be strongly influenced by "outliers", when the algorithm is based on least squares regression. The aim of robust $M-$estimation is to reduce the influence of outliers and/or to estimate other quantities than the mean function. In both cases, least squared based algorithms for the choice of an optimal bandwidth need to be modified. How to obtain general, computationally feasible algorithms will need to be considered in future research.

6 Appendix

Outlined proof of theorem 1: Consistency of $\hat{\beta}$ follows by standard arguments from the law of large numbers and properties of ψ (for technical details see e.g. Beran et al. 2000). For $j = 0, ..., p$, let

$$A_{n,j} = \sum_{i=[n(t-b)]}^{[n(t+b)]} \psi(\epsilon_i)(i - nt)^j$$

and

$$B_{n,j} = \sum_{l=0}^{p} \sum_{i=[n(t-b)]}^{[n(t+b)]} \psi'(\epsilon_i)(i - nt)^j (i - nt)^l n^{-l} (\hat{\beta}_l - \beta_l).$$

Here $[x]$ denotes the integer part of x. By Taylor expansion (3), implies $A_{n,j} - B_{n,j} \approx 0$ $(j = 0, ..., p)$. Hermite expansion yields

$$A_{n,j} \approx c_m/(m!)(nb) \sum_{i=[n(t-b)]}^{[n(t+b)]} H_m(Z_i)(i - nt)^j = A_{n,j}^{(m)}.$$

Moreover, $(B_{n,0}, ..., B_{n,p})' = C_n(\hat{\beta} - \beta)$ where C_n can be approximated by $(nb)E\{\psi'(\epsilon_i)\}D^{-1}(n)(X'X)$ with $X = (x_{ij})_{i,j=1,...,p+1}$ and $x_{ij} = (i - nt)^{j-1}$. Let $a_m = c_m/\{m!E[\psi'(\epsilon)]\}$ and $\zeta_m = (A_{n,0}^{(m)}, ..., A_{n,p}^{(m)})'$. Then $\hat{\beta} - \beta \approx a_m D(n)(X'X)^{-1}\zeta_m$. Now the right hand expression is equal to $a_m D(n)$ times $\tilde{\beta} - \beta$ where $\tilde{\beta}$ is the least squres estimator of β based on observations in the linear regression equation $y_i = X\beta + \eta_i$, with $\eta_i = H_m(Z_i)$ and $-[n(t-b)] \le i \le [n(t+b)]$. Moreover, $\mathrm{cov}(H_m(Z_i), H_m(Z_{i+k})) = m!\gamma_Z^m(k)$, so that for the spectral density f_η of η_i we have $f_\eta(\lambda) \sim m!c_{\gamma,Z}^m|\lambda|^{-2d_m}$ as $\lambda \to 0$. By analogous arguments as in Yajima (1989) it follows that

$$D(n)\mathrm{var}(\tilde{\beta} - \beta)D(n) \to \frac{2\pi c_m^2 c_{f,Z}^m}{m!E^2[\psi'(\epsilon)]} P^{-1}QP^{-1}.$$

This completes the proof.

7 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.

8 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.

Beran, J. and Feng, Y. (1999) Local polynomial fitting with long-memory errors. Preprint.

Beran, J. and Feng, Y. (2000) Data-driven estimation of semiparametric fractional autoregressive models. June 2000, Discussion paper of the Center of Finance and Econometrics.

Beran, J., Ghosh, S. and Sibbertsen, P. (2000) Nonparametric kernel M-estimation with long-memory errors. Submitted.

Beran, J. and Ocker, D. (1999) SEMIFAR forecasts, with applications to foreign exchange rates. *Journal of Statistical Planning and Inference*, **80**, 137-153.

Fan, J. and Gijbels, I. (1996). Local Polynomial Modeling and its Applications. Chapman & Hall, London.

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

Giraitis, L., Koul, H.L. and Surgailis, D. (1996) Asymptotic normality of regression estimators with long memory errors. *Statist. Probab. Lett.* **29**, 317-335.

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

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986) Robust statistics - the approach based on influence functions. Wiley, New York.

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

Huber, P. (1981) Robust Statistics. Wiley, New York.

Ray, B.K. and Tsay, R.S. (1997). Bandwidth selection for kernel regression with long-range dependence. *Biometrika* **84** 791–802.

Table 1: Asymptotically optimal bandwidths for $p = 0$ (no outliers), multiplied by 10.

Trend	g_1						g_2					
n	500			1000			500			1000		
d	-0.3	0	0.3	-0.3	0	0.3	-0.3	0	0.3	-0.3	0	0.3
$b_{\text{opt}}(0) \times 10$	0.77	1.06	1.59	0.63	0.92	1.50	0.56	0.75	1.07	0.46	0.65	1.00

Table 2: Simulated mean, standard deviation, minimum, median and maximum of $\hat{b}$ for $n = 500$, based on 400 simulations. All values are multiplied by 10.

		g_1					g_2				
d	p_a	Mean	SD	Min	Med	Max	Mean	SD	Min	Med	Max
	0	0.69	0.064	0.57	0.68	0.99	0.58	0.024	0.52	0.58	0.66
	0.05	0.71	0.079	0.54	0.70	1.06	0.61	0.055	0.51	0.60	0.98
-0.3	0.1	0.75	0.115	0.56	0.73	1.93	0.63	0.064	0.51	0.62	1.17
	0.2	0.80	0.119	0.52	0.79	1.40	0.67	0.071	0.52	0.66	0.98
	0.5	0.89	0.141	0.54	0.89	1.53	0.73	0.081	0.55	0.73	1.06
	0	0.92	0.139	0.61	0.91	1.33	0.75	0.079	0.57	0.75	1.06
	0.05	0.92	0.146	0.62	0.91	1.33	0.75	0.079	0.55	0.75	1.04
0	0.1	0.93	0.142	0.59	0.92	1.44	0.76	0.084	0.56	0.74	1.07
	0.2	0.93	0.154	0.54	0.93	1.51	0.76	0.091	0.56	0.75	1.29
	0.5	0.93	0.165	0.55	0.93	1.55	0.76	0.101	0.53	0.76	1.45
	0	1.60	0.793	0.70	1.40	5.00	1.09	0.172	0.59	1.09	1.76
	0.05	1.41	0.638	0.47	1.27	5.00	1.01	0.254	0.61	1.00	5.00
0.3	0.1	1.28	0.447	0.64	1.22	5.00	0.96	0.241	0.60	0.95	5.00
	0.2	1.14	0.370	0.57	1.10	5.00	0.89	0.235	0.60	0.87	5.00
	0.5	1.01	0.187	0.54	1.00	1.76	0.79	0.106	0.52	0.78	1.23

Table 3: Simulated mean, standard deviation, minimum, median and maximum of $\hat{b}$ for $n = 1000$, based on 400 simulations. All values are multiplied by 10.

		g_1					g_2				
d	p_a	Mean	SD	Min	Med	Max	Mean	SD	Min	Med	Max
	0	0.58	0.032	0.49	0.58	0.79	0.49	0.012	0.47	0.49	0.54
	0.05	0.62	0.052	0.52	0.61	0.95	0.52	0.035	0.47	0.52	0.92
-0.3	0.1	0.66	0.067	0.52	0.65	0.97	0.55	0.049	0.47	0.54	1.06
	0.2	0.71	0.083	0.52	0.71	1.19	0.59	0.052	0.49	0.58	0.86
	0.5	0.79	0.112	0.53	0.78	1.37	0.65	0.062	0.50	0.64	0.92
	0	0.80	0.106	0.54	0.80	1.11	0.66	0.062	0.53	0.66	0.88
	0.05	0.81	0.112	0.55	0.80	1.25	0.66	0.058	0.52	0.65	0.95
0	0.1	0.81	0.110	0.47	0.81	1.24	0.66	0.060	0.52	0.66	0.87
	0.2	0.82	0.112	0.52	0.82	1.23	0.66	0.055	0.51	0.66	0.83
	0.5	0.83	0.109	0.52	0.84	1.23	0.66	0.064	0.52	0.66	0.94
	0	1.38	0.465	0.71	1.29	5.00	1.05	0.134	0.67	1.04	1.37
	0.05	1.20	0.302	0.63	1.15	3.15	0.95	0.129	0.49	0.95	1.98
0.3	0.1	1.11	0.308	0.59	1.07	5.00	0.88	0.109	0.55	0.88	1.37
	0.2	0.99	0.169	0.48	0.98	1.59	0.80	0.095	0.52	0.80	1.20
	0.5	0.86	0.127	0.48	0.87	1.38	0.70	0.071	0.53	0.70	1.20

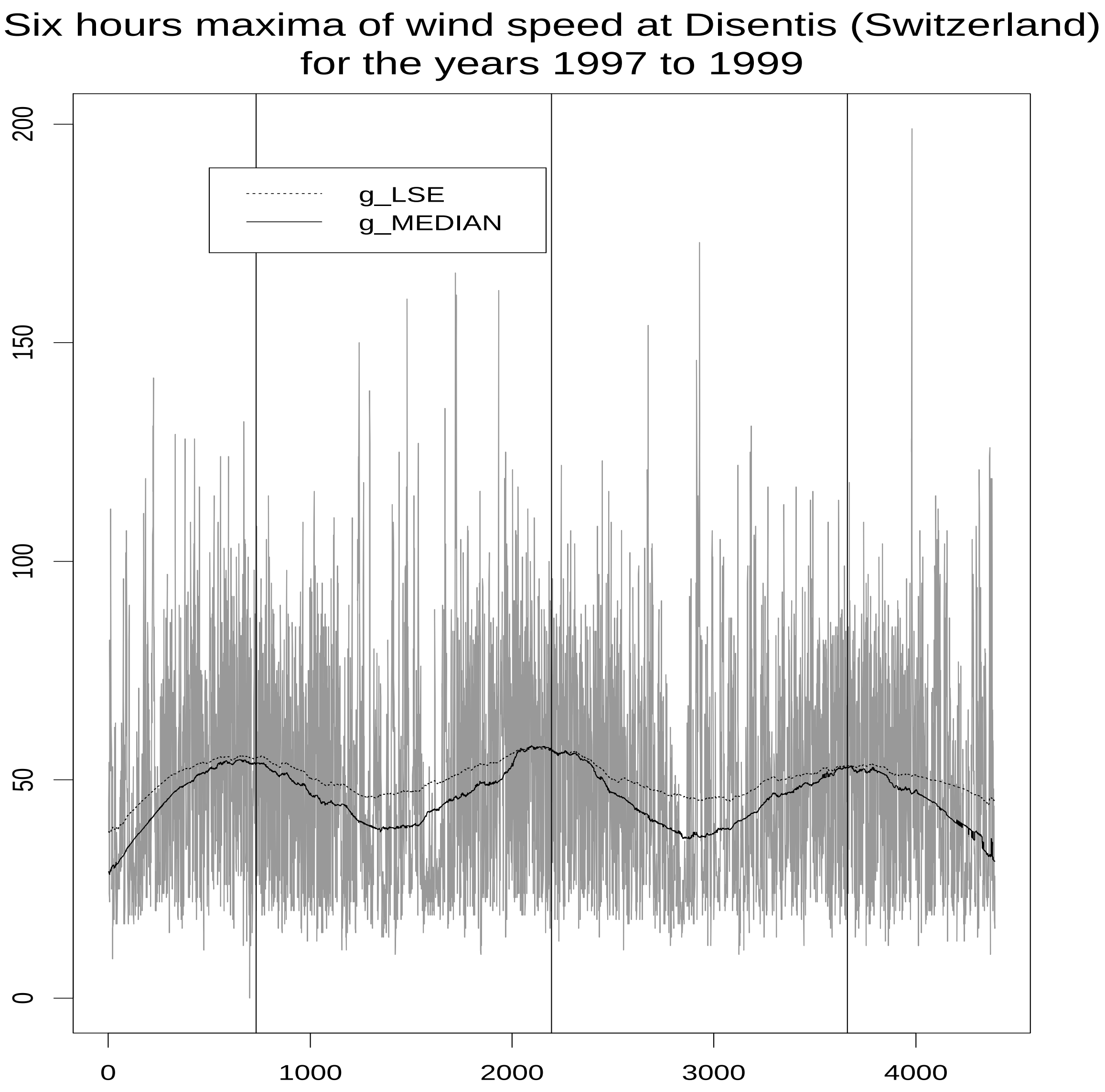

Figure 1: Wind speed (6-hours-maxima) at Disentis (Switzerland) for the years 1997 to 1999, fitted least squares (dotted line) and median local linear trends (full line).

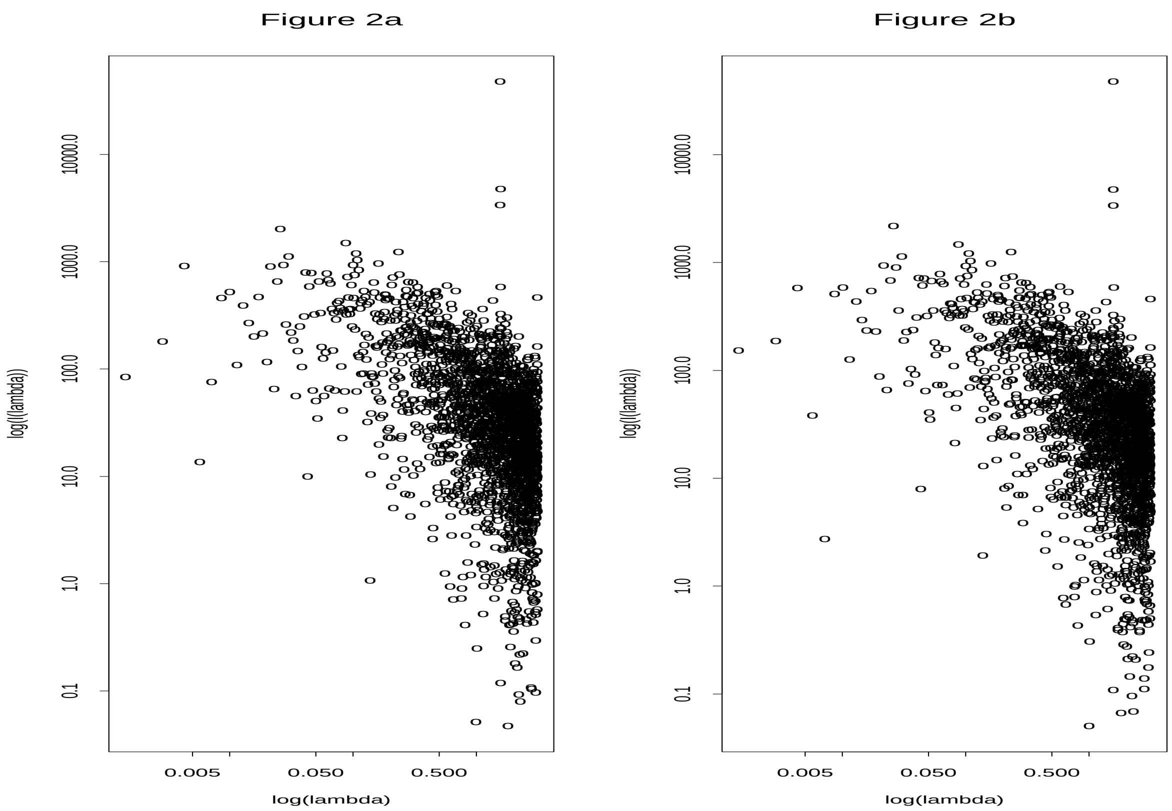

Figure 2: Periodogram (in log-log-coordinates) of estimated residuals $\hat{\epsilon}_i = y_i - \hat{g}$ for the least squares (figure 2a) and the median (figure 2b) local linear fit.

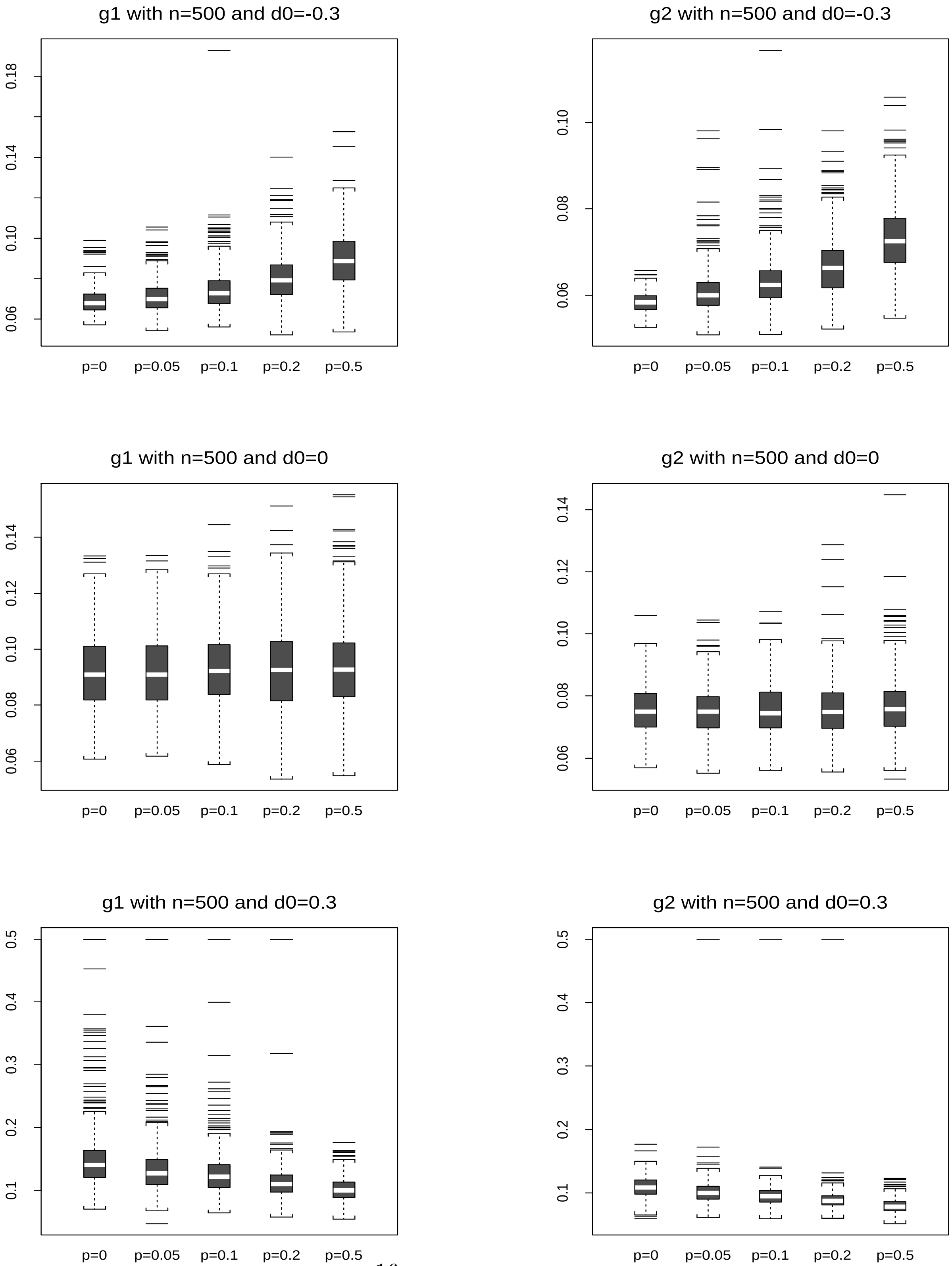

Figure 3: Boxplots of $\hat{b}$, each based on 400 simulations with $n = 500$, as a function of p and different values of d. The trend functions were g_1 and g_2 respectively.

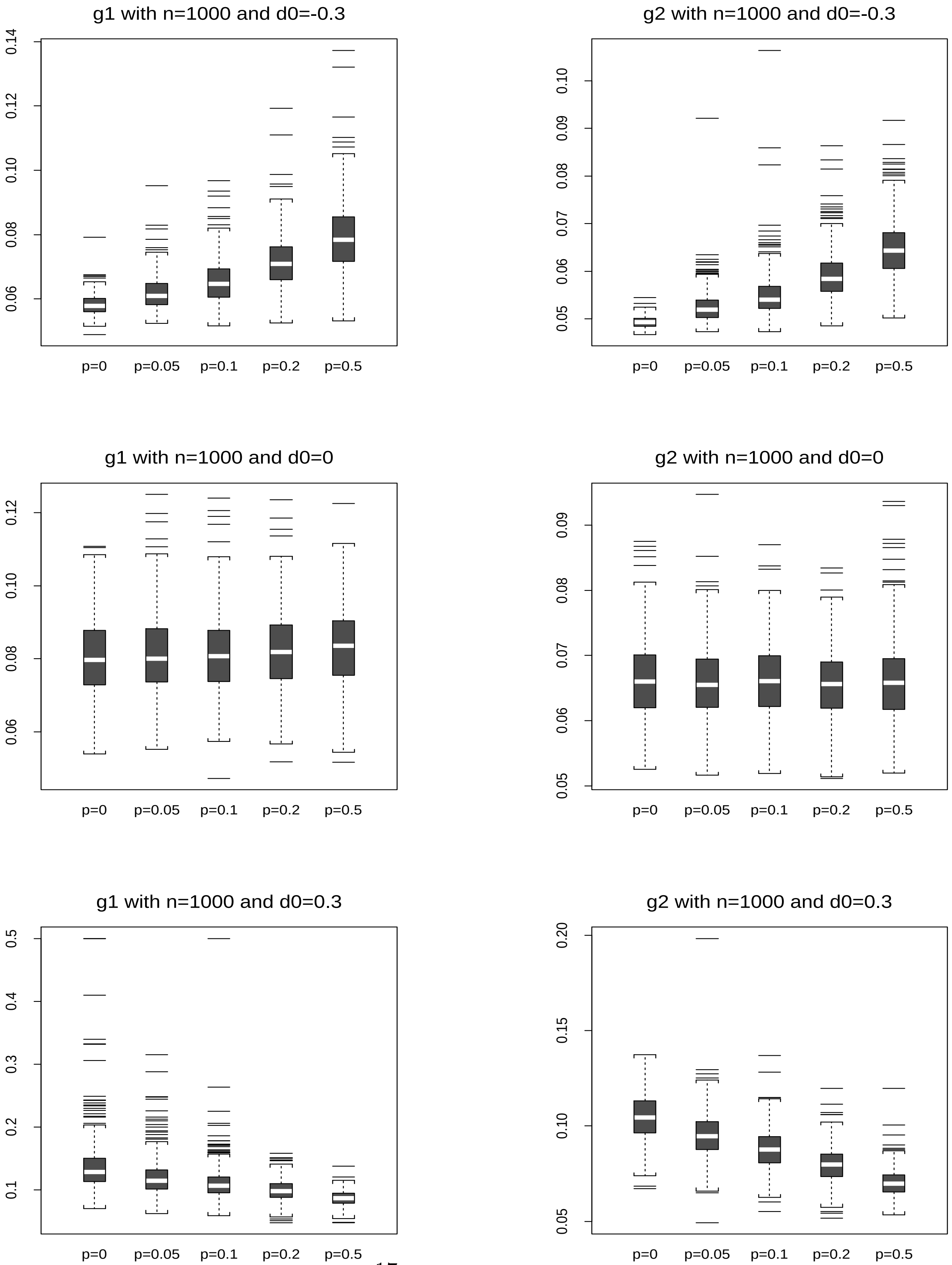

Figure 4: Boxplots of $\hat{b}$, each based on 400 simulations with $n = 1000$, as a function of p and different values of d. The trend functions were g_1 and g_2 respectively.