



INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386



Einbeck, Diva Saldiva de André, Singer:
Local Smoothing with Robustness against Outlying
Predictors

Sonderforschungsbereich 386, Paper 290 (2002)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



Local Smoothing with Robustness against Outlying Predictors

Jochen Einbeck

Ludwig Maximilians University, Institut für Statistik
Akademiestr. 1, 80799 München, Germany

Carmen D. S. André

Julio M. Singer

Departamento de Estatística, Universidade de São Paulo
Caixa Postal 66281, São Paulo, SP, 05311-970, Brazil

28th of August 2002

Abstract

Outlying pollutant concentration data are frequently observed in time series studies conducted to investigate the effects of atmospheric pollution and mortality/morbidity. These outliers may severely affect the estimation procedures and even generate unexpected results like a protective effect of pollution. Although robust methods have been proposed to downweight the effect of outliers in the response variable distribution, little has been done to handle outlying explanatory variable values. We consider a robust local polynomial smoothing technique which may be useful for such purposes. It is based on downweighting points with a small design density and may also be used as a diagnostic tool to identify outliers. Using data from a study conducted in São Paulo, Brazil, we show how an unexpected form of the relative risk curve of mortality attributable to pollution by SO_2 obtained via nonrobust methods may be completely reversed when the proposed technique is employed.

Key Words: Atmospheric pollution, nonparametric curve fitting, outliers, robust methods.

1 Introduction

Smoothing methods are widely used to eliminate random noise in regression problems involving time series of explanatory and response variables. Typical examples are studies of the association between daily measures of pollutant concentrations in the atmosphere and mortality as considered in Schwartz (1994), Braga et al. (2001) and Singer et al. (2002), among others. In such ecological studies, the frequent presence of outlying observations, requires robust smoothing techniques and for such purposes, the LOWESS (LOcally WEighted Scatter plot Smoothing) technique has been successfully employed to downweight the effect of outliers in the response variable (Cleveland, 1979). The idea of the LOWESS technique is to carry out a series of iteratively reweighted local polynomial fits, where, in each step, the points with the largest residuals in the previous step are downweighted. Alternatively, one of the several recently published robust nonparametric methods may also be considered. For example, a common approach to robustification is to replace the quadratic loss function $l(z) = z^2$ by functions which are less sensitive to outliers, e.g. the L_1 norm $l(z) = |z|$ as proposed by Wang & Scott (1994).

More specifically, in the context of local constant fitting, the estimate of a function $m(\cdot)$ at point x , given the data $(X_i, Y_i), i = 1, \dots, n$, is

$$\hat{m}(x) = \operatorname{argmin}_a \sum_{i=1}^n w_i(x) l(Y_i - a)$$

with weights $w_i(x) = K[(X_i - x)/h]$, where K is a kernel function and h is the bandwidth. These local M -estimators are discussed in Härdle & Gasser (1984), Truong (1989) and Hall & Jones (1990). An improvement on such estimators involve a local linear instead of constant fit, as discussed in Tsybakov (1986), Fan, Hu & Truong (1994) and Yu & Jones (1998). Honda (2000) enriched the concept by accounting for correlated errors. These papers, however, deal with robustness against outlying responses. The task of how to treat outliers in the predictors remains unexamined, a fact which was already noted by Hastie & Tibshirani (1990).

To illustrate the importance of the development of such techniques we consider

the data set analyzed by Conceição et al. (2001) and Singer et al. (2002) to evaluate the association between mortality attributed to respiratory causes of children under five and the concentration of PM_{10} , SO_2 , O_3 and CO in the city of São Paulo, Brazil, from 1994 to 1997 (the data is available in www.ime.usp.br/~jmsinger). The number of daily respiratory deaths as a function of the SO_2 concentration is depicted in Figure 1. Days with high pollutant concentrations (as compared to the majority of the data) are clearly identified. The effect of such observations is to "pull" the fitted curve downward (dotted line), suggesting that the effect of the pollutant on children mortality decreases for concentrations beyond $50 (\mu g/m^3)$, a fact that has no biological plausibility. To better understand the effect, observe the two data points at the lower right side of the picture. Although they do not seem to correspond to vertical outliers, they definitely disturb the local linear fit. A possible reason for this is that high concentrations of the pollutant are not coupled with a large number of deaths, contrary to what is expected; this is probably due to the sparse design for large concentrations. It seems clear that some robust method must be employed to bypass this inconsistency. In the light of this example, we consider a robust local polynomial smoothing technique which downweights the effect of outliers both in the response and in the explanatory variables (application of this method on the given data yields the solid line in Figure 1). It may also be used as a diagnostic tool to identify outliers in the explanatory variables.

In Section 2 we give a short review over some existing related concepts. In Section 3 we introduce the new outlier robust smoother and in Section 4 we return to the example addressed above. We finish with a discussion in Section 5.

2 An overview of related concepts

2.1 Ridging

Seifert & Gassser (1996, 2000) show that the conditional variance of a local linear fit can be unbounded in situations where the design is clustered or sparse. In many cases, the presence of sparse data is a problem highly related to outlying

predictors. As a solution, they propose to use data adaptive ridging, i.e. they replace the local linear fit by a weighted sum of a local linear and a local constant fit. An appropriate choice of a data adaptive selected ridge parameter achieves a balance between these two estimators and is successful in robustifying the procedure against unbounded variance. However, this form of robustification is not exactly what we desire here. Note that outlying predictors correspond to regions with sparse design and in those situations the ridge estimator performs a local constant fit. Thus, the estimator will more or less reproduce the response value associated to the outlying observations, which is the opposite of what we expect of an outlier robust method.

2.2 Variable bandwidth

Fan & Gijbels (1992) discuss a local linear estimator based on a global variable bandwidth, i.e., a bandwidth which depends on the predictors. In particular, let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a random sample from a population (X, Y) . Assume that $m(x) = E(Y|X = x)$ is the mean regression function of Y given X and let $f(\cdot)$ denote the (design) density of X . Then

$$\sum_{j=1}^n (Y_j - a(x) - b(x)(x - X_j))^2 \alpha(X_j) K \left[\frac{x - X_j}{h_n} \alpha(X_j) \right] \quad (1)$$

is minimized in terms of $a(x)$ and $b(x)$, applying a variable bandwidth $h(X_j) = h_n/\alpha(X_j)$, where $\alpha(\cdot)$ is some nonnegative function. This leads to the local estimator $\hat{m}(x) = \hat{a}(x)$. Fan & Gijbels (1992) show that the optimal variable bandwidth, i.e., the bandwidth minimizing the asymptotic MSE, is achieved by setting $\alpha(x)$ proportional to $(f(x)[m''(x)]^2/\sigma^2(x))^{1/5}$, where m'' denotes the second derivative of m and $\sigma^2(x)$ denotes the conditional variance of Y . For $\alpha(x) = f(x)$, the estimator $\hat{m}(x)$ obtained by minimizing (1) is asymptotically equivalent to a nearest-neighbour estimator.

Although not explicitly stated by the authors, this kind of estimator may be considered as a first step towards robustification against outlying predictors. Let us assume that $\alpha(\cdot)$ is any monotone increasing function of $f(\cdot)$. Then the factor $\alpha(X_j)$ in the minimization problem (1) downweights all points with

a small design density, which is what we expect for the outlying covariates. There is, however, a serious drawback with this approach to robustification. The function $\alpha(\cdot)$ appears again in the argument of the kernel K and covariates lying in sparse regions, (e.g. outliers) become associated to huge bandwidths, $h(X_j)$; thus they will have a large influence on the estimation at remote (i.e., all other!) data points. This effect is also contrary to the desired one. To overcome this problem, one could either replace the function $\alpha(\cdot)$ in K by a more suitable function $\beta(\cdot)$ or simply leave it out. For simplicity and transparency of the concept we will focus on the last alternative.

3 Robustness against outlying predictors

3.1 Soft robustification

In the light of the above discussion, we consider the estimator

$$\hat{m}(x, \alpha) = \hat{a}(x),$$

obtained by minimizing

$$\sum_{j=1}^n (Y_j - a(x) - b(x)(x - X_j))^2 \alpha(X_j) K\left(\frac{x - X_j}{h_n}\right) \quad (2)$$

where $\alpha(\cdot)$ is any monotone increasing function of $f(\cdot)$.

In order to avoid singularities due to sparse designs, we propose to use kernels with unbounded support in the presence of outlying predictors. In this paper we use Gaussian kernels in all examples.

Note that asymptotically, i.e., for $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$, the estimator (2) is equivalent to a local linear estimator. In fact, the factor $\alpha(\cdot)$ is vanishing in the leading terms of asymptotic bias as well as asymptotic variance expressions as may be deduced from the results presented in Fan & Gijbels (1996). This however is not surprising, since the asymptotic bias and variance of local linear estimators do not depend on the design density $f(\cdot)$. The weight function $\alpha(\cdot)$ essentially modifies the influence of the design density, regardless if $\alpha(\cdot)$ depends on $f(\cdot)$ or not. However, asymptotics for horizontal outliers seem not

to make much sense, since for $n \rightarrow \infty$ the data will be arbitrarily dense at any location x for which $f(x) > 0$. Consequently we will not focus on asymptotic considerations in this paper.

Normally the function $\alpha(\cdot)$ is unknown and we obtain the estimator $\hat{m}(x, \hat{\alpha})$ by minimizing (2) with $\alpha(\cdot)$ replaced by a consistent estimator, $\hat{\alpha}(\cdot)$.

A near-at-hand idea is to use $\alpha(\cdot) = f(\cdot)$. We estimate the density by

$$\hat{f}(x) = \frac{1}{ng_n} \sum_{i=1}^n K\left(\frac{X_i - x}{g_n}\right).$$

To select the bandwidth g_n , we choose the modified normal reference bandwidth selector proposed by Silverman (1986), namely

$$g_n = 0.9An^{-1/5},$$

where

$$A = \min(\text{standard deviation}, \text{interquartile range}/1.34). \quad (3)$$

We defer the task of how to select the bandwidth h_n to Subsection 3.5.

As a further improvement, one could imagine to use not the density $f(\cdot)$, but a power $f^k(\cdot)$ with $k > 1$ as the weight function $\alpha(\cdot)$. As we shall see, the larger the exponent, the better is the robustification. However, the exponent cannot increase arbitrarily, since then estimation becomes unstable.

In the sequel we will refer to the method introduced above as *soft robustification*. Under this method, outliers are downweighted but not eliminated. When one is convinced that the outliers do not contain useful information, it might be desirable to eliminate them from the estimation procedure. This approach, called *hard robustification*, will be introduced in the following subsection.

3.2 Hard robustification

Under soft robustification procedures, outliers still influence estimated values associated to predictors lying in their neighbourhood. To avoid this, one could

consider automatically cutting off points associated to estimated density values which fall beyond a certain barrier. This barrier can be calculated data-adaptively by applying an idea similar to that of the normal reference bandwidth selector. In a (very) rough approximation, one can assume

$$\hat{f}(\cdot) \approx \phi_{\mu, A^2}(\cdot),$$

where ϕ_{μ, A^2} denotes the density function of a normal distribution with mean $\mu = \text{median}(X_1, \dots, X_n)$ and standard deviation A given in (3). Let p denote the proportion of expected outliers (typically, $p = 0.05$ or $p = 0.01$). Then the required barrier is given by

$$\delta = \phi_{\mu, A^2}(x_{p/2}) = \phi_{\mu, A^2}(\mu + A \cdot z_{p/2}) = \frac{1}{A} \phi(z_{p/2}),$$

where $x_{p/2}$ and $z_{p/2}$ are the $p/2$ quantiles of the distribution of X and of the $N(0, 1)$ distribution, respectively. The estimator $\hat{m}(x)$ is now obtained by minimizing

$$\sum_{j=1}^n (Y_j - a(x) - b(x)(x - X_j))^2 \alpha(X_j) 1_{\{f(X_j) > \delta\}} K\left(\frac{x - X_j}{h_n}\right). \quad (4)$$

Surely the question arises whether one can rely on estimation results in areas where the data were downweighted or even cut off. This, however, is a question inherent to any robust method. In particular, when applying soft robustification techniques, we must face the question of whether it is correct to downweight the data on the one hand, i.e., to pretend not to trust the data, but to believe in the estimation results in the same region, on the other hand. Some decision has to be made and we suggest to base it on *areas of confidence*, which can be selected by means of density estimation. Within the areas of confidence, i.e., for all x with $\hat{f}(x) > \delta$, the estimation is considered to be reliable. Outside these areas, the reliability of the estimation procedures is questionable and interpretation of the estimated curve must be taken cautiously. This will become clearer in the following example.

3.3 Example

We now apply the proposed methods to a simulated data set, generated by contaminating the underlying function $m(x) = (x - 3)^2$ with Gaussian noise ($\sigma = 0.75$). The predictor values are assumed to be uniformly distributed in the interval $[0,6]$.

In Figure 2 we illustrate how the soft robust fit is affected by successively adding outlying predictors. We start without outliers and finish with a cluster of seven outliers. In each case we take the estimated density and the third power of the estimated density as weights.

A similar investigation was carried out under the hard robustification method; the results are depicted in Figure 3. The development of the kernel density and the cut off barrier are shown in Figure 4.

An analysis of these figures leads us to conclude that

- The proposed soft method succeeds to robustify the local estimator in areas where the data is associated with a non-negligible design density, i.e., within the areas of confidence. However, the fit is still affected in a neighbourhood of the outlier.
- Weighting with the third power of the density yields a better robustification effect than weighting with the density itself.
- When only one or two outliers are present, the hard robustification method produces a fit which is completely unaffected by the outliers (see top of Figure 3). Like desired, only the outliers are removed from the data set.
- For bigger clusters of outliers soft and hard robustification methods yield the same results since in this case the density does not get cut off. Note that the cut off barrier is decreasing with an increasing number of outliers (see Figure 4).
- The bigger the cluster, the smaller is the effect of robustification as expected, since a big cluster is probably not just a group of outliers but rather contains genuine information.

- From Figure 4 we may observe that values at the boundary are down-weighted in general, even if they are not outlying, due to the smoothing effect of the kernel density estimator. This is a desirable property, since boundary points are known to provide possibly spurious information.

In all estimations in this example we used the same global bandwidth $h_2 = 0.6$, motivated by the result of one-sided cross-validation (OSCV) discussed in Subsection 3.5.

Note that the outliers considered in this example could be regarded as outlying predictors as well as outlying responses. Thus methods which robustify against outlying responses should work here as well. In fact, when one outlier is present, the S-Plus function loess yields the same effect of a hard robustification procedure after 2 iterations. We chose this example only to demonstrate the effect with a particularly difficult kind of horizontal outliers. In Subsection 3.4 and Section 4 we will provide examples where vertical robustification methods fail.

3.4 Simultaneous robustness for predictor and response variables

Now we show that robust methods for outlying predictors and responses can be combined successfully. We chose the robust LOWESS method of Cleveland (1979), which is one of the most widely used robustification methods. It is implemented in S-Plus.

The data shown in Figure 5 were generated by contaminating the underlying function $m(x) = 6\sqrt{x}$ with Gaussian noise ($\sigma = 3$) and the predictors are uniformly distributed in the interval $[0, 6]$. One vertical outlier at point $(2; 25)$ and two horizontal outliers at $(7.5; 11)$ and $(8; 10.5)$ were intentionally added by hand, yielding a total of $n = 51$ data points. Note that the observations with outlying predictors cannot be regarded as outlying responses, since, when compared to the other data points, they are not located at a considerable distance from the function m .

Figure 5 (top) shows the result of a simple local linear fit and a LOWESS fit after 4 iterations. The LOWESS fit succeeds to eliminate the influence of the

vertical outlier, but fails to handle the horizontal outliers.

An illustration that the local linear as well as the LOWESS fit can be robustified against outlying predictors via the soft robustification method (with the estimated density as weight function) is given in the bottom part of Figure 5.

All the estimation procedures were carried out by means of the S-Plus function *loess* with smoothing parameter equal to 0.45.

3.5 Some notes about bandwidth selection

In principle, any arbitrary local linear (constant or variable) bandwidth selection routine can be applied to select the bandwidth h_n . Possible methods to select constant bandwidths are, among others, cross-validation (CV), one-sided cross-validation (Hart & Yi, 1998), plug-in methods (Ruppert, Sheather & Wand, 1995), methods based on the AIC (Hurvich, Simonoff & Tsai, 1998) or the RSC criteria (Fan & Gijbels, 1995). For local variable bandwidths, i.e., bandwidths of the form $h(x)$, we may also refer to Fan & Gijbels (1995), and further to Fan, Gijbels, Hu & Huang (1996) or Doksum, Petersen & Samarov (2000).

However, we should point out that the results of bandwidth selection routines can be seriously affected by horizontal outliers. As an example, we demonstrate the effect of outliers on cross-validation and one-sided cross-validation techniques for the simulated data examined in Subsection 3.3. The results of the selection of a bandwidth for a local linear estimator under increasing numbers of outliers is shown in Figure 6 for each method (CV or OSCV). As may be deduced from the plot, the bandwidth obtained under CV is considerably affected by a single outlier, while that obtained under OSCV seems to be somewhat more robust. The corresponding bandwidths are bigger under OSCV than under CV as the number of outliers increases. A similar analysis was conducted for the example of Subsection 3.4; the selected bandwidths under CV and OSCV for none, one and two horizontal outliers are summarized in the following table.

Selection method	Horizontal outliers		
	0	1	2
<i>CV</i>	1.60	1.97	1.36
<i>OSCV</i>	1.25	1.28	1.08

Here again, *OSCV* yields more stable bandwidth values than *CV*. The seemingly better robustness of *OSCV* to outlying predictors is in conformity to other robustness properties of this methodology (Hart & Lee, 2002).

We finally remark that the above results do not change significantly when using a soft robustified estimator instead of a local linear estimator under the *CV* (*OSCV*) routines. The problem is intrinsic to the bandwidth selector and not to the smoothing method.

4 Relative Risk curves for respiratory deaths

We now return to the example addressed in the introduction. Following a standard analysis strategy for this type of data as in Schwartz (1994), a generalized additive "core" model including terms to control for trend, days of the week, seasonality, temperature, humidity and non-respiratory deaths was initially fitted.

A scatter plot of the deviance residuals from this "core" model versus the SO_2 concentrations is presented in Figure 7 (top) along with a LOWESS smoother (dotted line). The resulting curve is falling for high concentrations, whereas the soft robustified fit is slightly rising.

Regarding the plot, this effect seems to be not so serious - however the misleading effect of the horizontal outliers becomes much more dramatic when regarding relative risk curves similar to those presented in Singer et al. (2002). The generalized additive model considered by those authors is typical for count data like the ones investigated here. For our purposes, it suffices to know that the model may be generally expressed as

$$\ln[E(\text{respiratory death})] = \alpha + \sum_{i=1}^{p-1} f_i(X_i) + f(SO_2)$$

where X_i , $i = 1, \dots, p-1$ denote variables like temperature, humidity, etc. The relative risk of death at a concentration $SO_2(j)$ of the pollutant SO_2 relative to the risk of death at the minimum concentration $SO_2(\text{min})$ is given by

$$RR(j) = \frac{E(\text{respiratory death}|SO_2(j))}{E(\text{respiratory death}|SO_2(\text{min}))} = \exp[f(SO_2(j)) - f(SO_2(\text{min}))].$$

In the center portion of Figure 7, we show the relative risk curve (\cdot) and its soft robustified counterpart ($+$). The plot shows a tremendous influence of the horizontal outliers. The relative risk curve obtained by a simple local fit suggests that the risk of respiratory death decreases with increasing pollutant concentration, what is obviously unacceptable. The relative risk curve obtained by a soft robustified local linear fit (weighted with the density) behaves as desired.

The form of the kernel density estimation for these data is presented in the bottom portion of Figure 7 and suggests that there is no need for a hard robustified version of the local linear fit in this case.

5 Discussion and Outlook

We showed that local linear and LOWESS smoothers can be robustified against outlying predictors. The main idea is to plug in the estimated density into the minimization problem. Such an idea is not restricted to these estimators and can certainly be applied to local polynomial estimators in general, but also simulations with smoothing splines led to the desired results. In fact, we believe that it is not an exaggeration to claim that any smoothing method which is based on minimization (maximization) of any loss (likelihood) function can be robustified against outlying responses by applying the concept introduced in this paper.

We further verified that robustification against outlying responses can be achieved for multivariate smoothing and for derivative estimation, results which we do not present here, since they would burst the framework of the paper.

We feel that there is still plenty of room for further research in this area. Beyond the topics mentioned above, a challenging task seems to be the problem of bandwidth selection; in particular, we mention the problem of robustness of common bandwidth selection routines to horizontal outliers. Up to now, this task is only rudimentarily treated even for vertical outliers. In the context of local L_1 regression, Wang & Scott (1994) introduced a version of CV with robustness against outlying responses.

Acknowledgements

This research was conducted while the first author was visiting the Department of Statistics, University of São Paulo, Brazil and was partially supported by grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), FINEP (PRONEX), Brazil; and the SFB 386 (DFG), Germany.

References

- Braga, A. L. F., Saldiva, P. H. N., Pereira, L. A. A., Menezes, J. J. C., Conceição, G. M. S., Lin, C. A., Zanobetti, A., Schwartz, J., and Dockery, D. W. (2001). Health effects of air pollution exposure on children and adolescents in São Paulo, Brazil. *Pediatr Pulmonol* **31**, 106–113.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74**, 829–836.
- Conceição, G. M. S., Miraglia, S. G. E. K., Kishi, H. S., Saldiva, P. H. N., and Singer, J. M. (2001). Air pollution and children mortality: a time series study in São Paulo, Brazil. *Environ Health Perspect* **109**, 347–350.
- Doksum, K., Petersen, D., and Samarov, A. (2000). On variable bandwidth selection in local polynomial regression. *Journal of the Royal Statistical Society, Series B* **62**, 431–448.

- Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Ann. Statist.* **20**, 2008–2036.
- Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaption. *Journal of the Royal Statistical Society, Series B* **57**, 371–395.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London: Chapman and Hall.
- Fan, J., Gijbels, I., Hu, T.-C., and Huang, L.-S. (1996). A study of variable bandwidth selection for local polynomial regression. *Statistica Sinica* **6**, 113–127.
- Fan, J., Hu, T.-C., and Truong, Y. K. (1994). Robust nonparametric function estimation. *Scandinavian Journal of Statistics* **21**, 433–446.
- Hall, P. and Jones, M. (1990). Adaptive M-estimation in nonparametric regression. *Ann. Statist.* **18**, 1712–1728.
- Härdle, W. and Gasser, T. (1984). Robust nonparametric function fitting. *Journal of the Royal Statistical Society, Series B* **46**, 42–51.
- Hart, J. D. and Lee, C.-L. (2002). Robustness of one-sided cross-validation to autocorrelation. *Submitted to: Journal of Multivariate Analysis*.
- Hart, J. D. and Yi, S. (1998). One-sided cross-validation. *J. Amer. Statist. Assoc.* **93**, 620–631.
- Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. London: Chapman and Hall.
- Honda, T. (2000). Nonparametric estimation of a conditional quantile for α -mixing processes. *Ann. Inst. Statist. Math.* **52**, 459–470.
- Hurvich, C., Simonoff, J., and Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society, Series B* **60**, 271–293.
- Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* **90**, 1257–1270.
- Schwartz, J. (1994). Nonparametric smoothing in the analysis of air pollution and respiratory illness. *Canadian Journal of Statistics* **22**, 471–487.
- Seifert, B. and Gasser, T. (1996). Finite-sample analysis of local polynomials:

- Analysis and solutions. *J. Amer. Statist. Assoc.* **91**, 267–275.
- Seifert, B. and Gasser, T. (2000). Data adaptive ridging in local polynomial regression. *Journal of Comput. and Graph. Statistics* **9**, 338–360.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.
- Singer, J. M., André, C. D. S., Lima, P. L., and Conceição, G. M. S. (2002). Association between atmospheric pollution and mortality in São Paulo, Brazil: Regression models and analysis strategy. In Y. Dodge (Ed.), *Statistical Data Analysis based on the L_1 norm and related methods*, pp. 439–450. Birkhäuser, Berlin.
- Truong, Y. K. (1989). Asymptotic properties of kernel estimators based on local medians. *Ann. Statist.* **17**, 606–617.
- Tsybakov, A. B. (1986). Robust reconstruction of functions by the local approximation approach. *Problems of Information Transmission* **22**, 133–146.
- Wang, F. T. and Scott, D. W. (1994). The L_1 method for robust nonparametric regression. *J. Amer. Statist. Assoc.* **89**, 65–76.
- Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93**, 228–237.

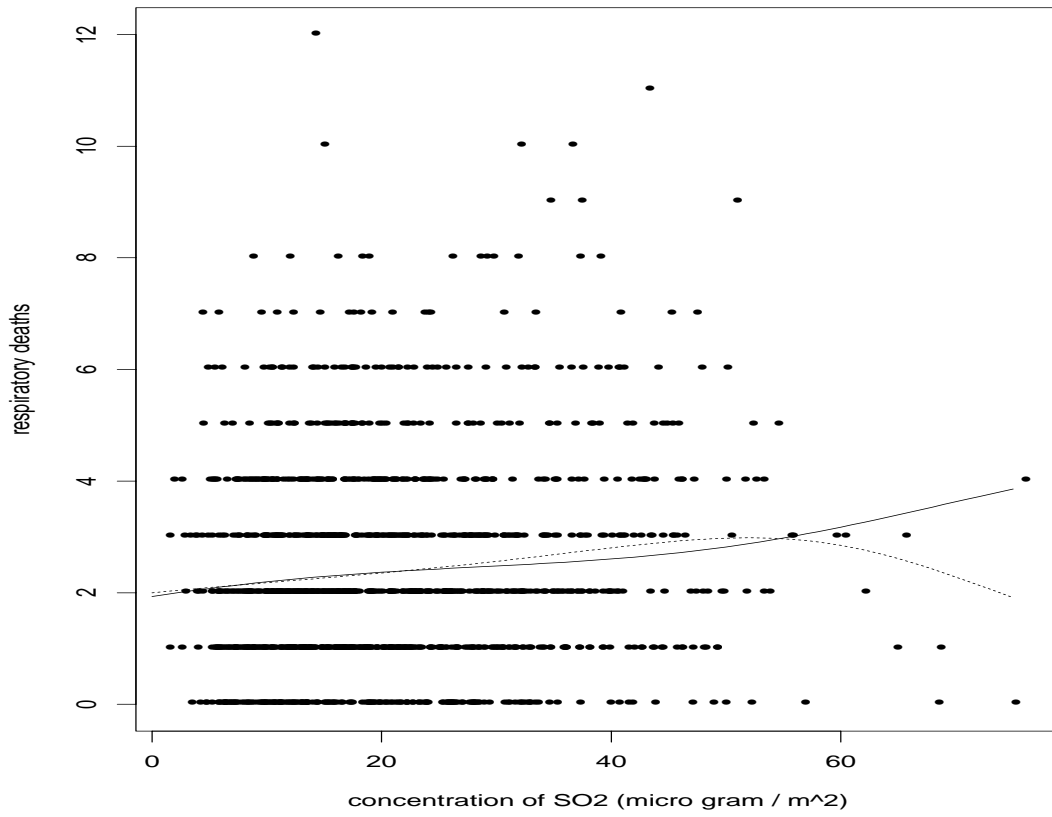


Figure 1: Respiratory deaths versus SO_2 concentration, local linear fit (dotted) and fit with robustness to horizontal outliers (solid).

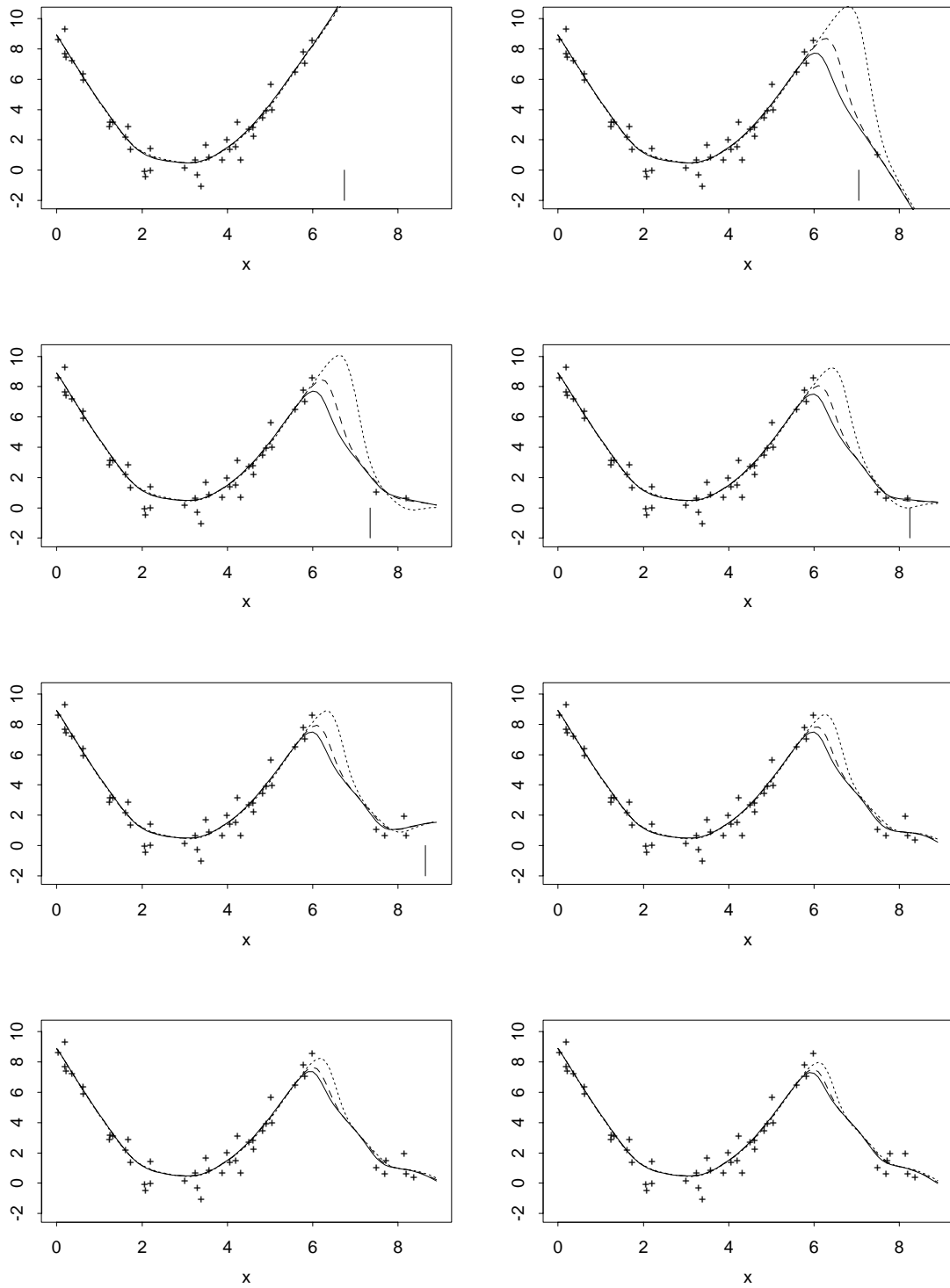


Figure 2: Data (+), local linear fit (solid line) and soft robustification by weighting with the density (dashed line) and third power of the density (dotted line) for varying numbers of outliers. Vertical lines indicate the end of the confidence area.

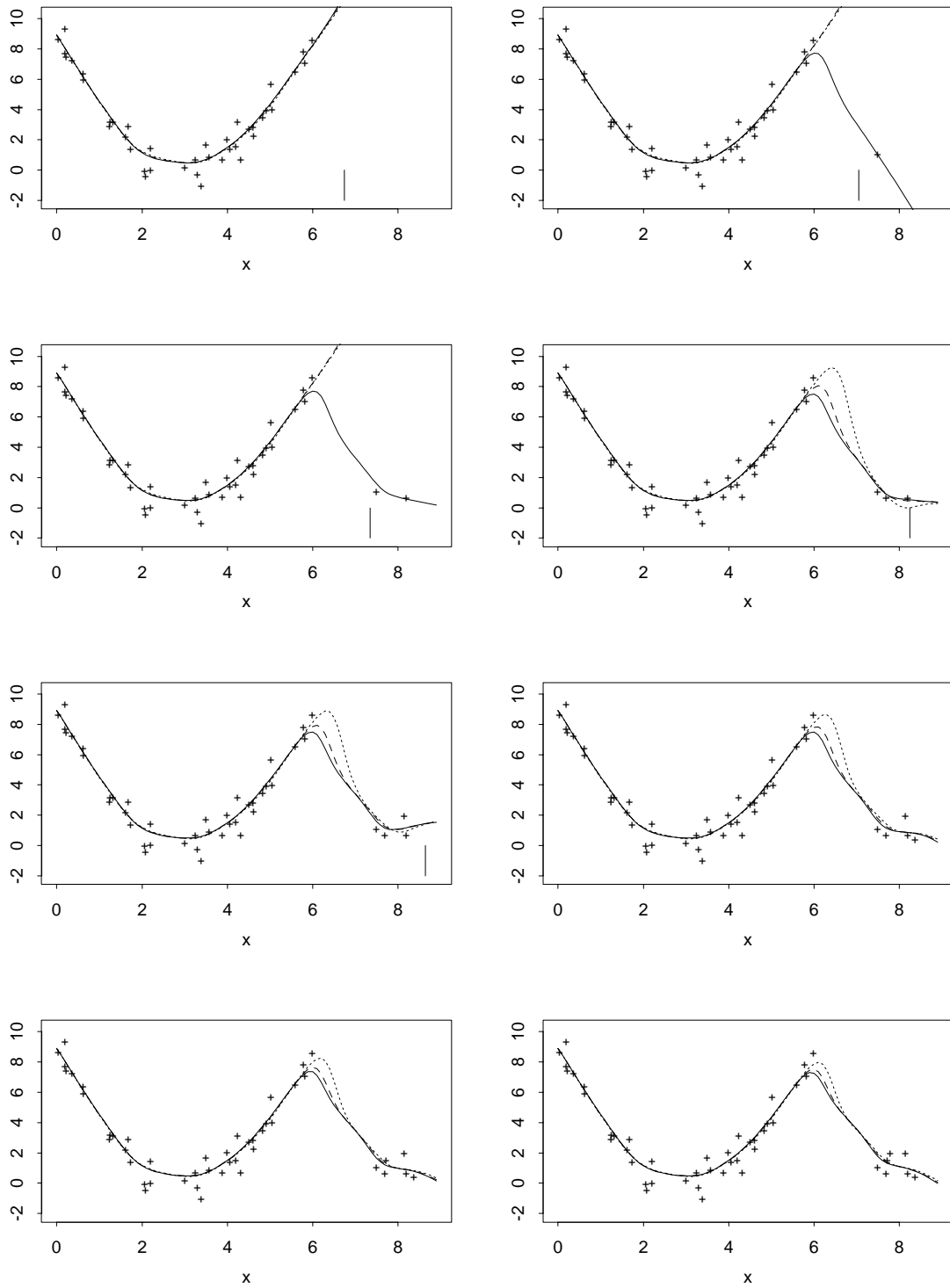


Figure 3: Data (+), local linear fit (solid line) and hard robustification by weighting with the density (dashed line) and third power of the density (dotted line) for varying numbers of outliers. Vertical lines indicate the end of the confidence area.

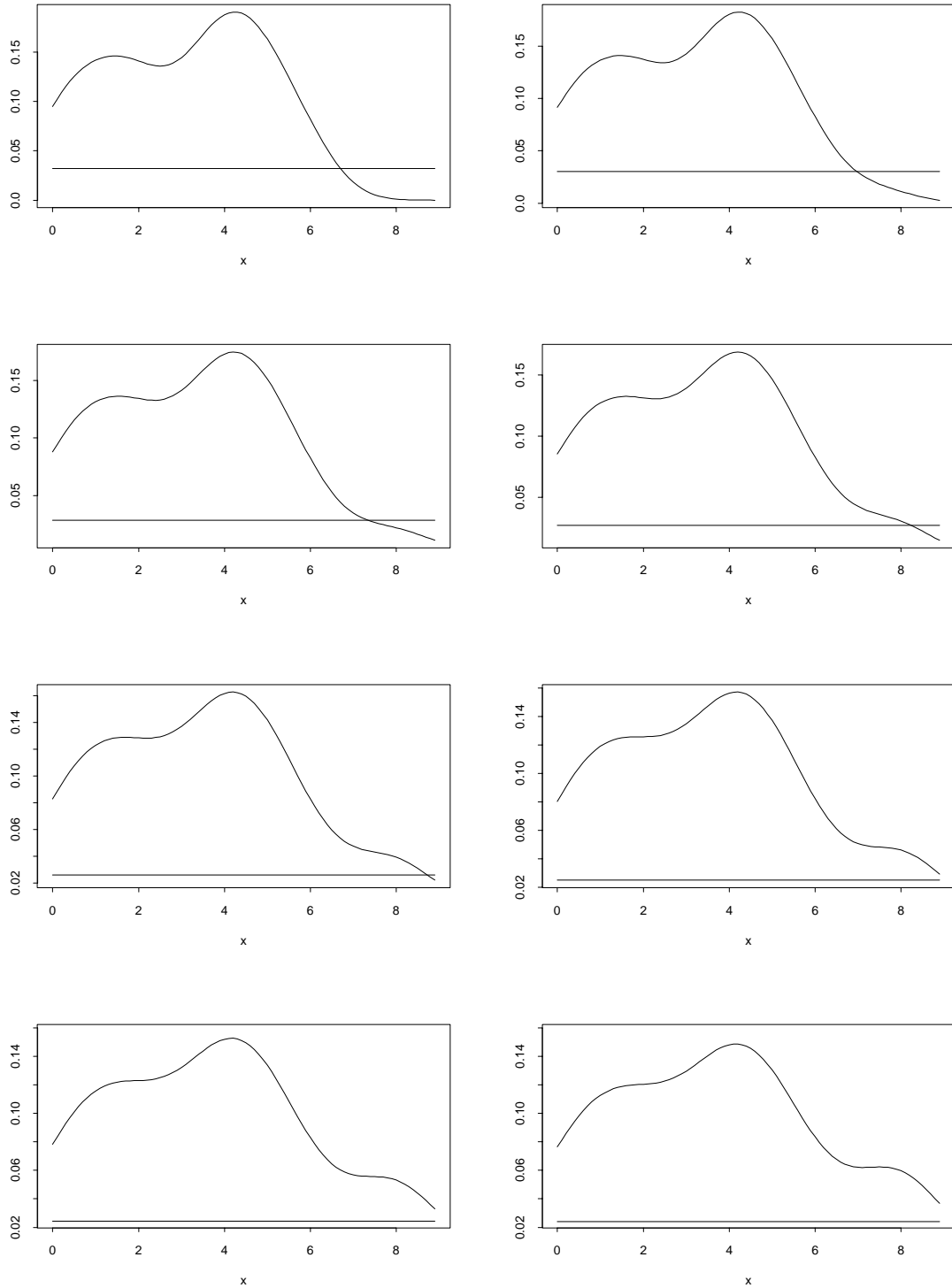


Figure 4: Density estimations and cut off barriers of the data set used in Example 3.3.

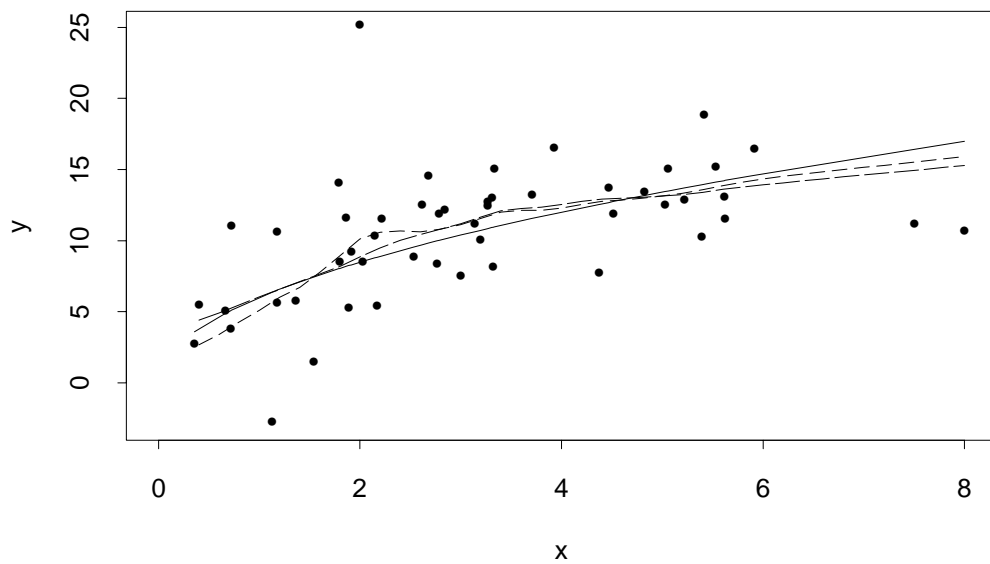
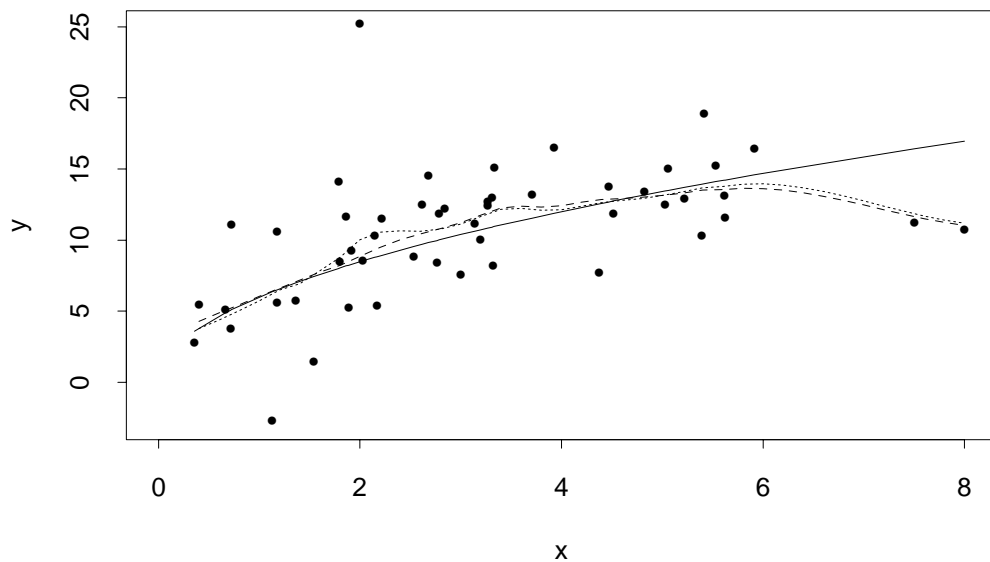


Figure 5: Top: Simulated data, underlying function (solid line), local linear (dashed line) and loess fit (dotted line); bottom: soft robustified local linear (short-dashed) and loess fit (long-dashed).

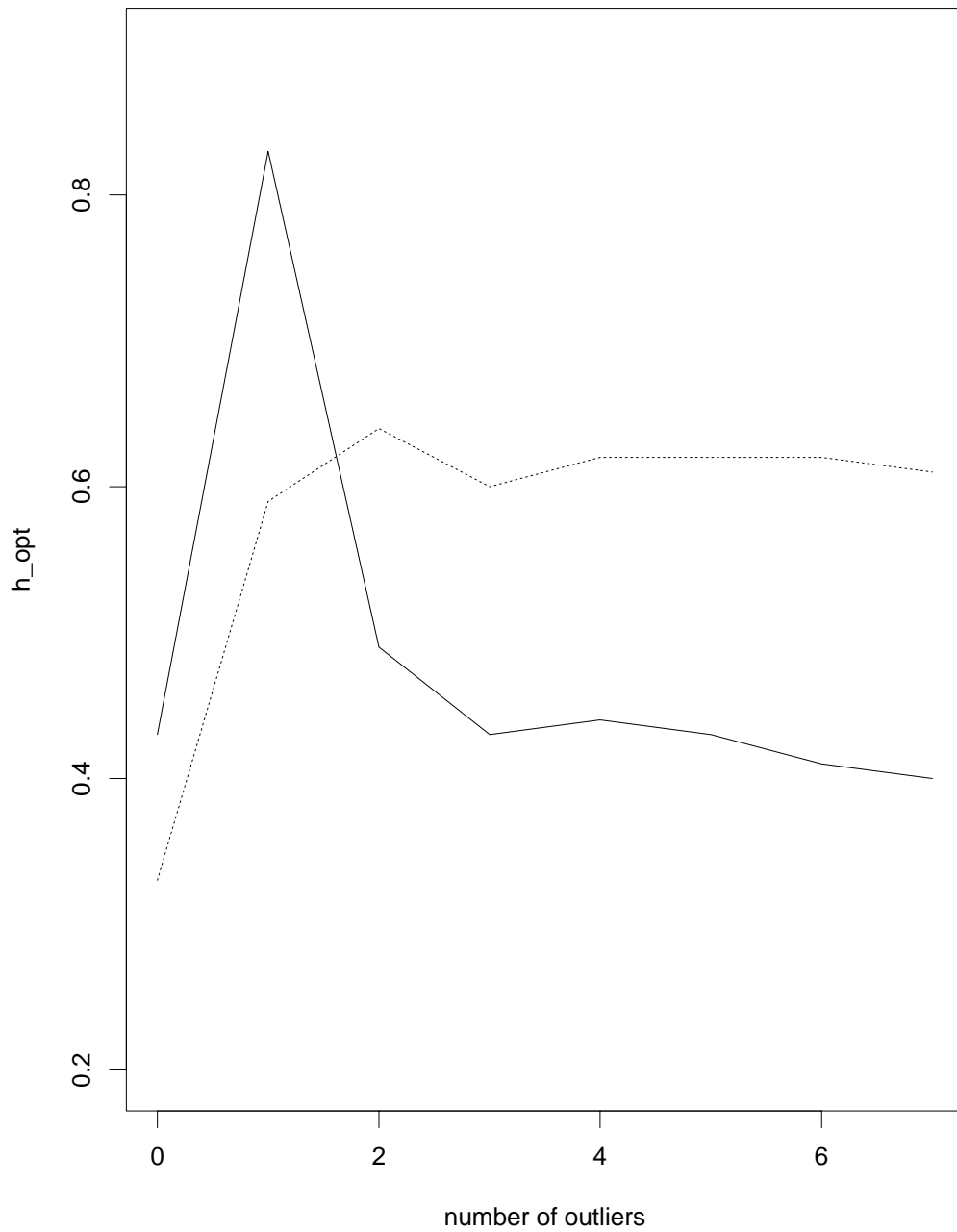


Figure 6: Bandwidth selected by CV (solid) and OSCV (dotted) for increasing numbers of outliers.

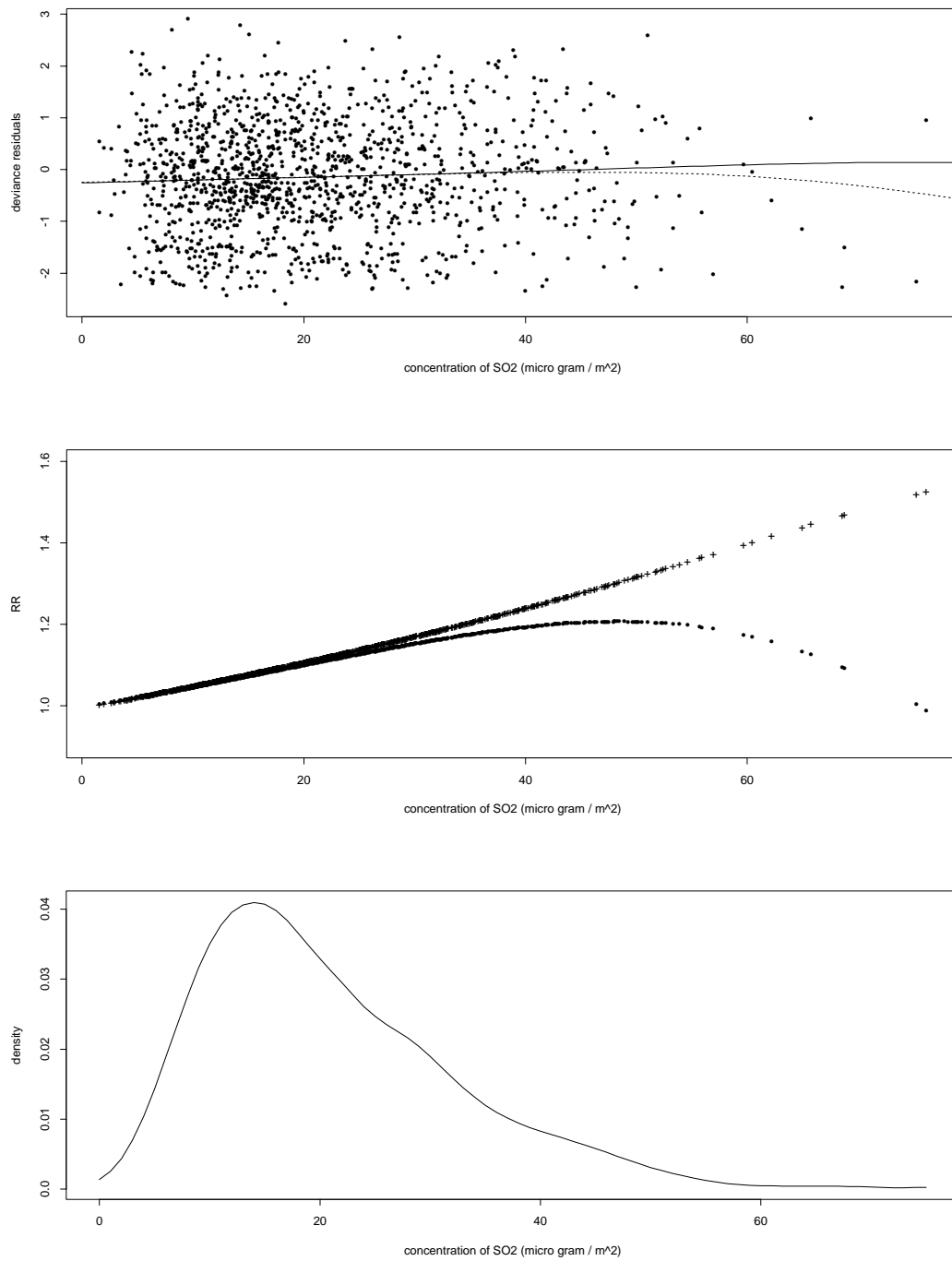


Figure 7: Top: Deviance residuals versus SO_2 concentration with a loess fit (dotted line) and a soft robustified fit (solid line); Middle: Relative risk curves versus SO_2 concentration resulting from a local linear (\cdot) and a soft robustified fit ($+$), each evaluated at all measured values of SO_2 concentration; bottom: Kernel density estimation of the SO_2 concentration.