Resistant Selection of the Smoothing Parameter for Smoothing Splines

Cantoni, Eva; Ronchetti, Elvezio

How to cite

CANTONI, Eva, RONCHETTI, Elvezio. Resistant Selection of the Smoothing Parameter for Smoothing Splines. In: Statistics and computing, 2001, vol. 11, n° 2, p. 141–146. doi: 10.1023/A:1008975231866

This publication URL: https://archive-ouverte.unige.ch/unige:22900

Publication DOI: 10.1023/A:1008975231866

Resistant Selection of the Smoothing Parameter for Smoothing Splines

Eva Cantoni and **Elvezio Ronchetti**

Department of Econometrics

University of Geneva

CH - 1211 Geneva 4, Switzerland

August 1998

Revised December 1999

Abstract

Robust automatic selection techniques for the smoothing parameter of a smoothing spline are introduced. They are based on a robust predictive error criterion and can be viewed as robust versions of C_p and cross-validation. They lead to smoothing splines which are stable and reliable in terms of mean squared error over a large spectrum of model distributions.

Key words: nonparametric models; M-type smoothing splines; robust C_p; robust cross-validation; tail length.

1 Introduction

Smoothing splines are flexible techniques for data modeling and basic building blocks in nonparametric models like Generalized Additive Models. Consider the model

$$Y_i = f(x_i) + u_i, \quad i = 1, \dots, n, \tag{1}$$

where $x_1, \dots, x_n$ are the design points and u_i are independent random variables with expectation $E(u_i) = 0$ and variance $V(u_i) = \sigma^2$. We suppose, without loss of generality, that $a < x_1 < \dots < x_n < b$. Then a (natural cubic) spline minimizes the penalized criterion

$$\sum_{i=1}^{n} \left(\frac{y_i - f(x_i)}{\sigma} \right)^2 + \frac{1}{2}\lambda \int_a^b \{f''(t)\}^2 \, dt, \tag{2}$$

where λ is a positive constant which controls the amount of smoothness. Typically the smoothing parameter λ is selected automatically to minimize the average predictive squared error by means of cross-validation or the C_p statistic. General references include de Boor (1978), Wahba (1990), Härdle (1990), Hastie and Tibshirani (1990), and Green and Silverman (1994).

Although smoothing splines are local fits in nature, they can still suffer from potential robustness problems due to a few outlying points. To avoid these problems, Huber(1979) introduced M-type smoothing splines by replacing the classical criterion (2) by

$$\sum_{i=1}^{n} \sigma\rho \left(\frac{y_i - f(x_i)}{\sigma} \right) + \frac{1}{2}\lambda \int_a^b \{f''(t)\}^2 \, dt, \tag{3}$$

where $\lambda > 0$ and $\rho(t)$ is a convex function. Although an appropriate choice of $\rho(.)$ ensures resistance to outlying points, from a robustness point of view the selection of the smoothing parameter λ is crucial and must be based on some robust criterion. This point seems to have been neglected in the literature.

[Figure 1 about here.]

Figure 1 shows some data with a few outlying points in the upper left corner. Fives curves are fitted: the true curve which generated the data, a classical spline, a M-type spline based on criterion (3) with a *classical criterion for the selection of* λ, a M-type spline with a selection of λ based on robust cross-validation (cf. Section 3), and a classical spline with λ chosen by robust cross-validation. It is clear that the classical fit misses some of the features of the data. The fit based on the robust criterion (3) but with λ chosen by an unmodified cross-validation criterion presents the same behavior as the classical spline. Both curves are over-smoothed. This shows that the robust selection of the smoothing parameter is crucial. Indeed, the best fit is obtained with a M-type spline with the smoothing parameter selected by the robust criterion. The value of the smoothing parameter obtained automatically by classical cross-validation is about 200 times larger than that obtained by the robust version of cross-validation. This affects the fit *everywhere* and not only locally where the outlying points appear. Finally, note that a classical spline with the robust selection of the smoothing parameter leads to an unsatisfactory fit which lies somewhere in between.

The aim of this paper is to introduce robust selection techniques for the smoothing parameter by means of robust versions of cross-validation and C_p. They are based on a robust predictive error criterion which takes into account the predictive performance *for the majority* of the data. Similar ideas have been developed for robust model selection in regression; cf. Ronchetti and Staudte (1994) and Ronchetti, Field, and Blanchard (1997).

The paper is organized as follows. In Section 2 we review M-type smoothing splines. In Section 3 we motivate and develop robust versions of cross-validation and C_p for the automatic selection of the smoothing parameter. Section 4 presents the results of a small simulation study which shows the stability and reliability of the new techniques for a large spectrum of error distributions in model (1). Section 5 presents some conclusions.

2 M-type Smoothing Splines

The nonparametric estimation of the regression function f in model (1) by M-type smoothing splines goes back to Huber (1979) and consists of minimizing the penalized criterion (3) with respect to f over $S_2[a,b]$[1]. Suppose, for the moment, that σ is known.

[1] $S_2[a,b]$ is the space of functions that are differentiable on $[a,b]$ and have absolutely continuous first derivative.

The solution of this problem is a compromise between closeness to the data and smoothness. Whereas in the ordinary spline case closeness to the data is measured by the sum of squared residuals, with M-type smoothing splines goodness-of-fit is evaluated through a loss function ρ applied to residuals. The parameter λ controls this compromise, which is in fact a trade-off between bias and variance. We will discuss the choice of this parameter in the next section.

If $\rho(.)$ is convex, it can be shown that the solution of problem (3) is a cubic spline. We can then write its finite dimensional form as follows

$$\sum_{i=1}^{n} \sigma\rho(\epsilon_i) + \frac{1}{2}\lambda\, \mathbf{f}^T K \mathbf{f}, \tag{4}$$

where $\epsilon_i = \frac{y_i - f(x_i)}{\sigma}$, $\mathbf{f} = (f(x_1), \dots, f(x_n))$, $K = N^{-T}\Omega N^{-1}$, $\Omega_{ij} = \int N_i''(x) N_j''(x) dx$ and N is a natural-spline basis matrix. By differentiating (4) with respect to $\mathbf{f}$, we obtain the set of estimating equations

$$-\psi(\boldsymbol{r}) + \lambda K \hat{\mathbf{f}} = 0, \tag{5}$$

where $\psi(t) = \frac{\partial}{\partial t}\rho(t)$ and $\psi(\boldsymbol{r})$ is the vector whose components are the function ψ applied to each component of $\mathbf{r} = (r_1, \dots, r_n)$, and $r_i = \frac{y_i - \hat{f}_i}{\sigma}$. The choice of a bounded function ψ will ensure robustness with respect to outliers in the residuals; $\psi(t) = t$ leads to ordinary splines.

A first order Taylor expansion $\psi(\boldsymbol{r}) \simeq \psi(\boldsymbol{\epsilon}) - \frac{1}{\sigma} M(\hat{\mathbf{f}} - \mathbf{f})$, where $M = \mathrm{diag}(\psi'(r_1), \dots, \psi'(r_n))$, leads to the one-step representation of the solution of (5)

$$\hat{\mathbf{f}} \simeq \left(I + \frac{\lambda\sigma}{E\psi'} K\right)^{-1} \left(\mathbf{f} + \frac{\sigma}{E\psi'}\psi(\boldsymbol{\epsilon})\right), \tag{6}$$

where the matrix M is replaced by its expectation. This representation can be viewed as the result of applying an ordinary spline to the unobservable pseudodata $\tilde{\mathbf{y}} = \mathbf{f} + \frac{\sigma}{E\psi'}\psi(\boldsymbol{\epsilon})$. This shows that the resistance of the M-type smoothing spline is achieved by down-weighting the standardized errors ϵ_i by the nonlinear (bounded) transformation $\frac{\sigma}{E\psi'}\psi(\epsilon_i)$. Moreover, it allows one to derive limit theorems and rates of convergence for the estimator $\hat{\mathbf{f}}$, cf. Cox (1983).

While in classical smoothing splines σ is concentrated out in the calculation of the smoothing parameter, in our proposal σ needs to be estimated. A robust estimator of σ is crucial in order to guarantee the global robustness of the procedure. Based on our experience, we recommend to use Huber's Proposal 2 (see Huber, 1981 and Hampel, Ronchetti, Rousseeuw, and Stahel, 1986) which consists of solving (5) and

$$\sum_{i=1}^{n} \chi(r_i) = 0 \tag{7}$$

simultaneously, where $\chi(t) = t\psi(t) - \rho(t) - \beta$, and β is a constant which ensures Fisher consistency for the estimation of σ. More details of the computation of robust splines can be found in Utreras (1981).

In this development, care has only been taken of robustness with respect to residuals. If we suspect leverage points in the x's (design points), resistance can be achieved by introducing weights. In this case (5) is replaced by $-W\psi(\boldsymbol{r}) + \lambda K\hat{\mathbf{f}} = 0$, where W is a diagonal matrix of weights depending on the x's. The solution of this new problem becomes a weighted spline.

3 Robust Selection of Smoothing Parameter

Assume model (1) with symmetric errors and a M-type smoothing spline defined by (5), where ψ is an odd function. Although the parameter λ can be chosen by eye, it is more reasonable to have an automatic selection procedure. To construct this rule we can proceed by defining a robust criterion for prediction. Minimizing an estimate of this criterion will give us the optimal value of λ. For example, in the case of classical procedures, cross-validation and Mallows's C_p are estimators of the predictive squared error. From a robustness point of view, a robust criterion for prediction should not penalize values of λ which lead to good fits for the majority of the data except perhaps at a few points.

Therefore, as in Ronchetti and Staudte (1994) we define the rescaled weighted predictive squared error by

$$WPSE(\lambda) = \frac{1}{\sigma^2} E\left[\sum_{i=1}^{n} \widehat{w}_i^2 \left(\hat{f}(x_i) - f(x_i)\right)^2\right], \quad (8)$$

where $\widehat{w}_i = \psi(\frac{y_i - \hat{f}(x_i)}{\sigma})/(\frac{y_i - \hat{f}(x_i)}{\sigma}) = \psi(r_i)/r_i$. (8) has the attractive form of a *weighted mean squared error* but other general loss functions could be used. The weights $\hat{w}_i$ in (8) have the effect of reducing possible large contributions of $(\hat{f}(x_i) - f(x_i))^2$ at a few outlying points. This in turn does not penalize robust fits which perform well for the majority of the data except at a few outlying points.

If we define the weighted sum of squared residuals by

$$WSR(\lambda) = \sum_{i=1}^{n} \widehat{w}_i^2 r_i^2 = \sum_{i=1}^{n} \psi^2(r_i), \quad (9)$$

and let $\delta_i = \frac{\hat{f}(x_i) - f(x_i)}{\sigma}$, it is easy to see that

$$WPSE(\lambda) = E(WSR(\lambda)) - \sum_{i=1}^{n} E(\widehat{w}_i^2 \epsilon_i^2) + 2 \sum_{i=1}^{n} E(\widehat{w}_i^2 \epsilon_i \delta_i). \tag{10}$$

This suggests the following robust version of Mallows's C_p

$$RC_p(\lambda) = WSR(\lambda) - \sum_{i=1}^{n} E(\widehat{w}_i^2 \epsilon_i^2) + 2 \sum_{i=1}^{n} E(\widehat{w}_i^2 \epsilon_i \delta_i). \tag{11}$$

The term $-\sum_{i=1}^{n} E(\widehat{w}_i^2 \epsilon_i^2) + 2\sum_{i=1}^{n} E(\widehat{w}_i^2 \epsilon_i \delta_i)$ is a correction term in order to make $WSR(\lambda)$ unbiased for $WPSE(\lambda)$.

For the computation of criterion (11) we need an expression for the correction term $-\sum_{i=1}^{n} E(\widehat{w}_i^2 \epsilon_i^2) + 2\sum_{i=1}^{n} E(\widehat{w}_i^2 \epsilon_i \delta_i)$. Its value is derived in Appendix A and allows us to write the final formula of $RC_p(\lambda)$

$$\begin{aligned} RC_p(\lambda) = \sum_{i=1}^{n} \psi^2(r_i) - nE\psi^2 + 2\frac{E(\psi^2\psi')}{E\psi'} Tr(S) - \\ \frac{1}{\sigma^2}\left[E\psi'^2 - E\left(\frac{\psi^2}{\varepsilon^2}\right)\right] Tr\left((S-I)ff^T(S-I)^T\right) - \\ \frac{1}{E^2\psi'}\left[E(\psi^2\psi'^2) - E\left(\frac{\psi^4}{\varepsilon^2}\right)\right] Tr(SS^T). \end{aligned} \tag{12}$$

One can check that putting $\psi(t) = t$ in the above expression yields the usual definition of Mallows's C_p.

Equation (12) depends on σ, which is in general unknown. As suggested in the literature about classical splines (see Hastie and Tibshirani, 1990), it is important to take an external estimation that does relatively little smoothing. We use a robust estimation of σ obtained as a solution of

$$\sum_{i=1}^{n} \chi\left(\frac{\tilde{r}_i}{\sigma\sqrt{a_i^2 + b_i^2 + 1}}\right) = 0,$$

where $\tilde{r}_i = (x_{i+1} - x_i)/(x_{i+1} - x_{i-1})y_{i-1} + (x_i - x_{i-1})/(x_{i+1} - x_{i-1})y_{i+1} - y_i = a_i y_{i-1} + b_i y_{i+1} - y_i$, $i = 2, \ldots, n-1$, and $\chi(.)$ is defined as in (7). The residuals $\tilde{r}_i$ are obtained by taking design points x_{i-1}, x_i, x_{i+1}, joining the two outer observations by a straight line and then computing the difference between the straight line and the middle observation. The estimator proposed is a robust version of the estimator of Gasser, Sroka, and Jennen-Steinmetz (1986). A different robust version of this estimator is discussed by Cunningham, Eubank, and Hsing (1991).

Another robust criterion for choosing the optimal value of λ can be obtained by exploiting the pseudodata structure we discussed in Section 2. We outline here a heuristic argument for its derivation. By applying ordinary cross-validation to the pseudodata, we have

$$\begin{aligned} RCV(\lambda) &= \frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - \hat{\tilde{y}}_i^{-i})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\tilde{y}_i - \hat{\tilde{y}}_i}{1 - S_{ii}}\right)^2, \end{aligned} \tag{13}$$

where $\tilde{y}_i = f(x_i) + \frac{\sigma}{E\psi'}\psi(\epsilon_i)$, $\hat{\tilde{y}}_i^{-i}$ is the fit obtained leaving out the ith data point, and S_{ii} is the ith diagonal element of the matrix $S = (I + \frac{\lambda\sigma}{E\psi'}K)^{-1}$. But $\hat{\tilde{y}}_i = \hat{f}(x_i)$, and by estimating $f(x_i)$ by $\hat{f}(x_i)$, we finally obtain

$$RCV(\lambda) = \frac{1}{n}\frac{\sigma^2}{(E\psi')^2}\sum_{i=1}^{n}\frac{\psi^2(r_i)}{(1 - S_{ii})^2}. \tag{14}$$

As in the case of $RC_p(\lambda)$, putting $\psi(t) = t$ in (14) recovers the ordinary cross-validation. A similar proposal was suggested by Leung, Marriott, and Wu (1993) for robust kernel M-smoothers.

4 Simulation Study

For our simulation study we consider the model

$$Y_i = \sin(2\pi(1-x_i)^2) + 0.5 * \epsilon_i, \; i = 1, \dots, n. \tag{15}$$

The function ψ is the Huber function defined by

$$\psi(t) = \psi_c(t) = \begin{cases} t & \mid t \mid \leq c \\ c \text{ sgn(t)} & \mid t \mid > c. \end{cases} \tag{16}$$

We perform a comparison of the classical procedure with the robust procedure proposed in this paper. We choose $c = 1.345$ in (16) which ensures 95% efficiency with respect to the normal model in a location problem. The scale parameter in the robust procedure is estimated by means of Huber's Proposal 2 given by (7), where $\chi(t) = t\psi_{\tilde{c}}(t) - \rho_{\tilde{c}}(t) - \beta$, $\frac{\partial}{\partial t}\rho_{\tilde{c}} = \psi_{\tilde{c}}$, and $\tilde{c} = 1.95$. This last value corresponds to 80% efficiency for the scale parameter with respect to the normal model.

Errors are generated from several symmetric distributions and from an asymmetric one. These distributions are described in Hoaglin, Mosteller, and Tukey (1983), Chapter 10, and classified with respect to their tail index defined by

$$\tau(F) = \frac{F^{-1}(0.99) - F^{-1}(0.5)}{F^{-1}(0.75) - F^{-1}(0.5)} \Big/ \frac{\Phi^{-1}(0.99) - \Phi^{-1}(0.5)}{\Phi^{-1}(0.75) - \Phi^{-1}(0.5)}, \tag{17}$$

where Φ is the cumulative distribution function of a standard normal distribution. The asymmetric distribution is $0.9\mathcal{N}(0,1) + 0.1\mathcal{N}(30,1)$. The spread parameters of the distributions considered are fixed by quartile matching with respect to the quartiles of the standard normal distribution. The expectations appearing in the definition of $RC_p(\lambda)$ (formula (12)) are calibrated at the normal model.

[Table 1 about here.]

A list of the distributions used is given in Table 1, together with their corresponding tail index. This design covers a large spectrum of possible error distributions from very short-tailed to very long-tailed distributions. For each distribution, we simulated 100 replications of a sample of size $n = 100$.

For each sample and for each technique (classical and robust), we compute the performance criterion defined by

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}(x_i) - f(x_i)\right)^2, \tag{18}$$

where the smoothing parameter is obtained by means of classical and robust cross-validation and classical and robust C_p.

All our simulations were performed with the software S-PLUS, MathSoft, Seattle. The implementation is simple and reasonably fast when exploiting the built-in functions for the treatment of ordinary splines.

[Figure 2 about here.]

Let us now look at the results. Figure 2 shows the boxplots of the $\log_{10} MSE$ ratio between the classical procedure and the robust one when the errors are generated from each distribution considered in Table 1. The smoothing parameter is chosen by cross-validation. It appears that the robust procedure is essentially equivalent to the classical one for short-tailed and normal-like distributions, but it has a smaller MSE by a factor of 5 under distributions with moderate tails and by a factor of 10 under more extreme cases. The higher variability of the ratio in the extreme cases is due to the MSE of the classical procedure (cf. Figure 3) and to the high variability of its smoothing parameter (not shown).

The same simulations were carried out using C_p and RC_p for determining the best value of λ.

[Figure 3 about here.]

Figure 3 shows that in the robust case, the MSE is stable across the underlying distributions and shows a moderate variability even when the error distribution is heavy-tailed. On the other hand, the MSE of the classical spline increases with $\tau(F)$. In fact, when the underlying distribution has a large tail index, $C_p(\lambda)$ is an increasing function and the minimum is attained at $\lambda = 0$. This particular behavior reflects once more the non-resistance of the procedure.

5 Conclusions

We have shown that the selection of the smoothing parameter of a smoothing spline must be based on a robust prediction criterion if we want to obtain a stable quality of the fit over a large spectrum of error distributions. This seems to be even more important in nonparametric regression where one does not want to specify a model for the errors. As is the case for its classical counterpart, the robust technique proposed in this paper will depend on the assumption of an underlying smooth target signal. Based on our experience, the robust procedure performs better than its classical counterpart in the presence of mild non-smoothness of the underlying signal. However, it cannot cope with situations where the signal shows clear discontinuities. In these situations other techniques like wavelets would probably be more appropriate.

Open research directions include the generalizations of these techniques in more complex models like the Generalized Additive Models where smoothing splines play an important role as building blocks of the estimation procedure. Some related work in this area includes Gu (1992a) and Gu (1992b).

Acknowledgments

The authors would like to thank the Editor, the Associate Editor, and two Referees for their constructive comments.

A Appendix

A.1 Derivation of the Correction Term in the Formula of $RC_p(\lambda)$

We derive the value of the constant $C = -\sum_{i=1}^n E(\widehat{w}_i^2 \epsilon_i^2) + 2\sum_{i=1}^n E(\widehat{w}_i^2 \epsilon_i \delta_i)$. Consider first the Taylor expansion of the weights $\hat{w}_i = \psi(\epsilon_i - \delta_i)/(\epsilon_i - \delta_i)$ at $\delta_i = 0$:

$$\widehat{w}_i = w_i - w_i'\delta_i + \frac{w_i''}{2}\delta_i^2 - \frac{1}{6}(w_i''')^* \delta_i^3, \tag{19}$$

where $w_i = \psi(\epsilon_i)/\epsilon_i$ and $(w_i''')^* = w_i'''(\epsilon^*)$ with $\epsilon^* \in [\epsilon, \hat{\epsilon} = r]$. Squaring (19), taking into account terms up to order δ_i^2 and grouping by order of δ_i, we have

$$\begin{aligned} C = -nE(w_i^2\epsilon_i^2) + 2\sum_{i=1}^n E\left[\left(w_i^2\epsilon_i + w_i w_i'\epsilon_i^2\right)\delta_i\right] + \\ -\sum_{i=1}^n E\left[\left(w_i'^2\epsilon_i^2 + w_i w_i''\epsilon_i^2 + 4w_i w_i'\epsilon_i\right)\delta_i^2\right]. \end{aligned} \tag{20}$$

Moreover, by $E(w_i^2\epsilon_i^2) = E\psi^2$ and (6) we have

$$\sum_{i=1}^n E\left[\left(w_i^2\epsilon_i + w_i w_i'\epsilon_i^2\right)\delta_i\right] = \frac{E(\psi^2\psi')}{E\psi'} Tr(S), \tag{21}$$

and

$$\begin{aligned} \sum_{i=1}^n E\left[\left(w_i'^2\epsilon_i^2 + w_i w_i''\epsilon_i^2 + 4w_i w_i'\epsilon_i\right)\delta_i^2\right] = \\ \frac{1}{\sigma^2}\left[E\psi'^2 - E\left(\frac{\psi^2}{\epsilon^2}\right)\right] Tr\left((S-I)ff^T(S-I)^T\right) + \\ \frac{1}{E^2\psi'}\left[E(\psi^2\psi'^2) - E\left(\frac{\psi^4}{\epsilon^2}\right)\right] Tr(SS^T). \end{aligned} \tag{22}$$

Putting all this together we find the final formula (12) for $RC_p(\lambda)$.

References

Cox, D. D. (1983). Asymptotics for M-type smoothing splines. *The Annals of Statistics*, **11**, 530–551.

Cunningham, J. K., Eubank, R. L., and Hsing, T. (1991). M-type smoothing splines with auxiliary scale estimation. *Computational Statistics and Data Analysis*, **11**, 43–51.

de Boor, C. (1978). *A Practical Guide to Splines.* Berlin/New York: Springer-Verlag.

Gasser, T., Sroka, L., and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, **73**, 625–633.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* London: Chapman & Hall.

Gu, C. (1992a). Cross-validating non-Gaussian data. *Journal of Computational and Graphical Statistics*, **1**, 169–179.

Gu, C. (1992b). Diagnostics for nonparametric regression models with additive terms. *Journal of the American Statistical Association*, **87**, 1051–1058.

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

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge: Cambridge University Press.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models.* London: Chapman & Hall.

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (Eds.) (1983). *Understanding Robust and Exploratory Data Analysis.* New York: Wiley.

Huber, P. J. (1979). Robust smoothing. In R. L. Launer and G. N. Wilkinson (Eds.), *Robustness in Statistics*, pp. 33–48. Ne York; London: Academic Press.

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

Leung, D. H. Y., Marriott, F. H. C., and Wu, E. K. H. (1993). Bandwidth selection in robust smoothing. *Journal of Nonparametric Statistics*, **2**, 333–339.

Ronchetti, E., Field, C., and Blanchard, W. (1997). A robust linear model selection by cross-validation. *Journal of the American Statistical Association*, **92**, 1017–1023.

Ronchetti, E. and Staudte, R. G. (1994). A robust version of Mallows' C_p. *Journal of the American Statistical Association*, **89**, 550–559.

Utreras, F. I. (1981). On computing robust splines and applications. *SIAM Journal on Scientific and Statistical Computing*, **2**, 153–163.

Wahba, G. (1990). *Spline Models for Observational Data.* Philadelphia: SIAM.

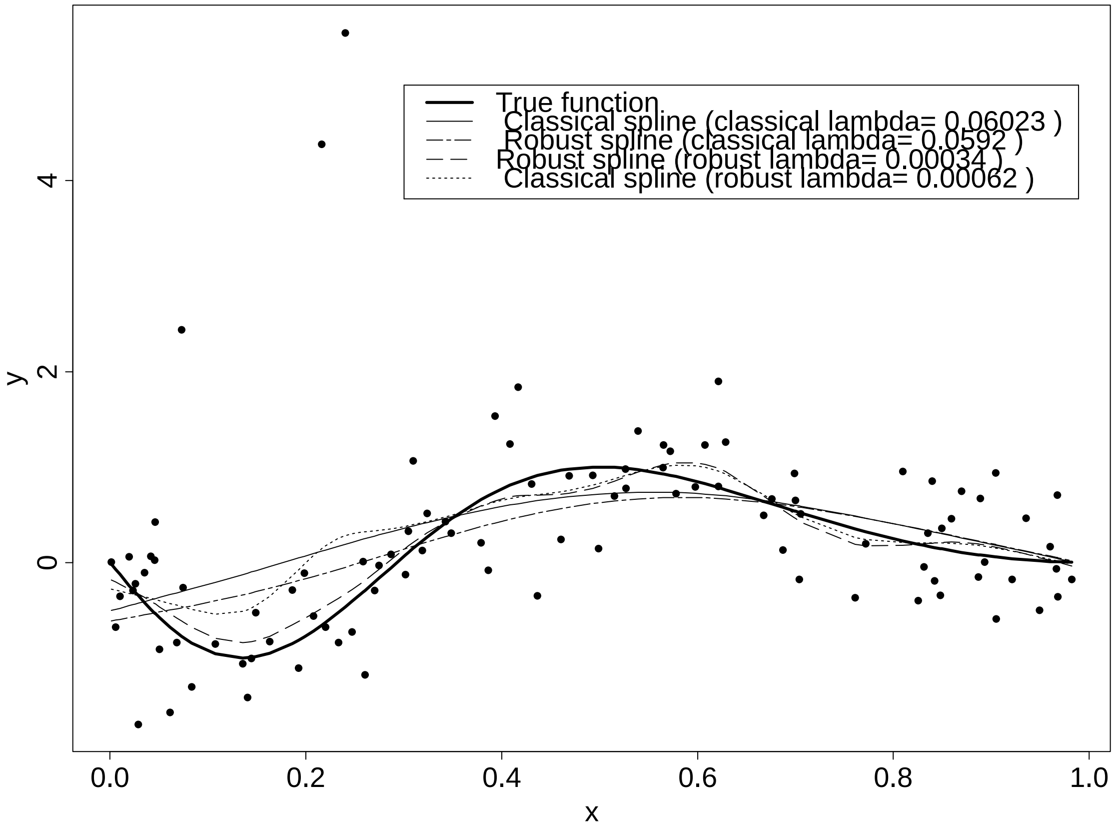


Figure 1: Example from model (1).

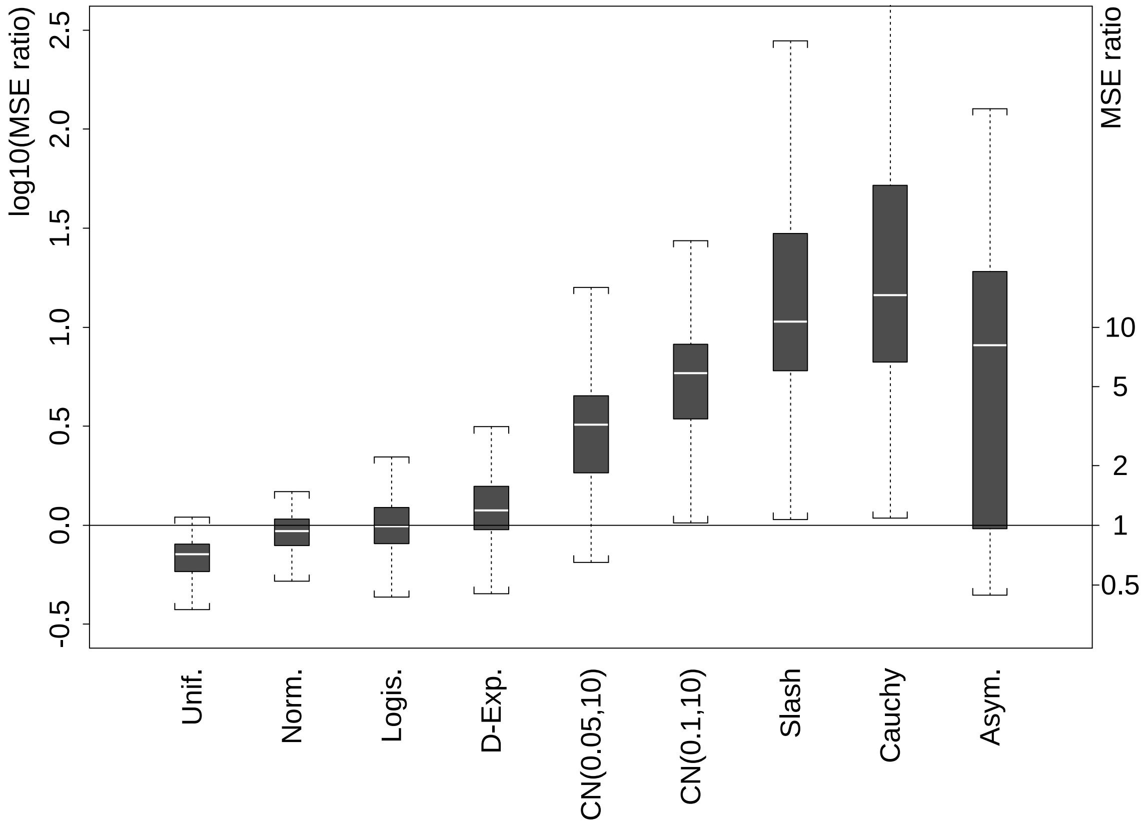
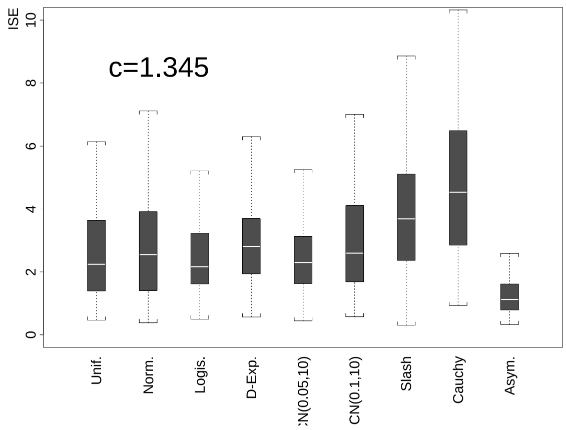


Figure 2: $\log_{10}[MSE(classical)/MSE(robust)]$. The smoothing parameter is chosen by classical and robust cross-validation respectively.



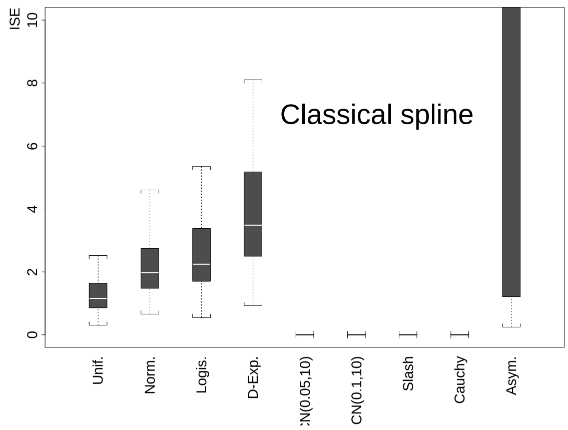


Figure 3: MSE corresponding to the solution minimizing $RC_p(\lambda)$ (robust spline with $c = 1.345$) and $C_p(\lambda)$ (classical spline).

Distribution	$\tau(F)$
Uniform	0.57
Normal	1
Logistic	1.21
Double Exponential	1.63
Contaminated Normal (0.05,10)	3.42
Contaminated Normal (0.10,10)	4.94
Slash	7.85
Cauchy	9.22

Table 1: Distributions used in the simulation study and their corresponding tail index.