

Comparing Local Fitting to Other Automatic Smoothers

Maderbacher, Michael; Müller, Werner

Published: 01/01/1995

Document Version

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Maderbacher, M., & Müller, W. (1995). *Comparing Local Fitting to Other Automatic Smoothers*. (March 1995 ed.) (Forschungsberichte / Institut für Statistik; No. 45). Department of Statistics and Mathematics, WU Vienna University of Economics and Business.

Comparing Local Fitting to Other Automatic Smoothers



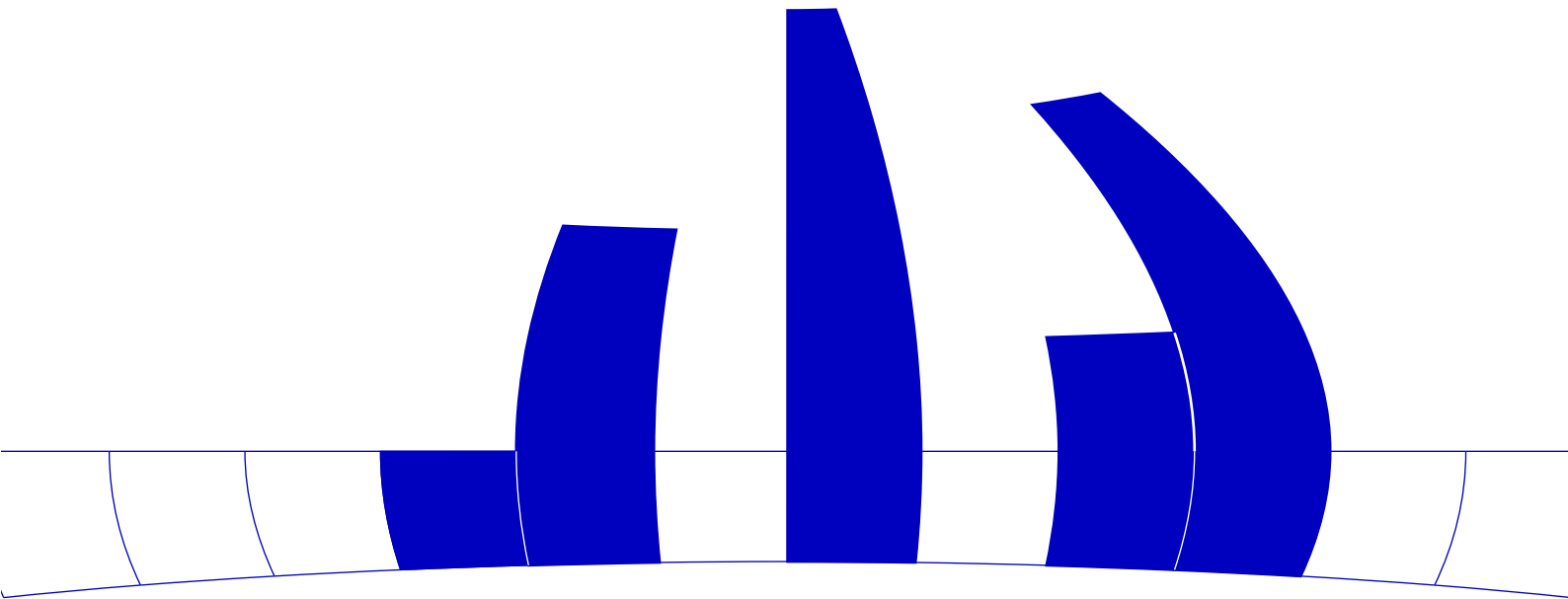
Michael Maderbacher, Werner Müller

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 45
March 1995

<http://statmath.wu-wien.ac.at/>



Comparing Local Fitting to Other Automatic Smoothers

(with full tables)

Michael Maderbacher, Werner G. Müller

No. 45, March 1995

1 Introduction

In a public service enterprise by Breiman and Peters (1991) various automatic smoothers, such as the supersmoothen (SSMU), cross-validated smoothing splines (BART), delete-knot regression splines (DKS) and the cross-validated kernel smooth (KERNEL) were compared by simulation on a variety of sample sizes, noise levels and functions. The intention was to give practitioners guidelines when to use which type of smoother. The given work completes those simulations by including the increasingly popular local fitting approach, that was introduced to the statistical literature by Cleveland (1979). Fedorov *et al.* (1993) have modified the technique in order to take account possible misspecification bias, termed ‘optimized moving local regression’, and here we use an automated version (by cross-validation) of it as given in Fedorov *et al.* (1994).

At each point ξ in the regressor space the underlying local model

$$y = \theta + \delta\varphi(|\xi - x|) + \varepsilon, \quad (1.1)$$

is assumed, where $\varphi(\cdot)$ is a polynomial term. Weighted least squares are employed to estimate θ and hence provide a smoother for y . It turns out (see Fedorov *et al.* (1993)) that optimal weights have the general form

$$\lambda^* = a + b\varphi(\cdot). \quad (1.2)$$

2 Simulation

In order to make the results of Breiman and Peters (1991) and our simulation study comparable, it was necessary to duplicate their data generation mechanism

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, \dots, n$$

for three different sample sizes n , namely 25, 75 and 225. First the independent variable x is generated using three different designs: equispaced, normally distributed and gap distributed (at each x the distance to the neighbouring point is exponentially distributed with parameter 1 with probability p and with parameter μ with probability $(1 - p)$), all on the interval $[0, 1]$. For gap distribution the values for p and μ were:

$$\begin{array}{lll} N=25 & p=0.75 & \mu=8 \\ N=75 & p=0.60 & \mu=6 \\ N=225 & p=0.75 & \mu=20 \end{array}$$

To allow for a non-zero intercept a column of ones was included as second regressor.

In generating the dependent variable y seven different increasingly complex underlying functions $f(x)$ were employed, see Figure 1.

- $f(x) = 2x$
- ‘broken line’: $f(x) = \begin{cases} -1 + 3x & 0 \leq x \leq 2/3 \\ 3 - 3x & 2/3 < x \leq 1 \end{cases}$
- $f(x) = \sin \pi x$
- $f(x) = \sin 2\pi x$
- ‘spike’: $f(x) = \begin{cases} -8 + 20x & 0.4 \leq x \leq 0.5 \\ 12 - 20x & 0.5 < x \leq 0.6 \\ 0 & \text{else} \end{cases}$

- Friedman's function: $f(x) = 3 \sin(2\pi(1-x)^2)$

- 'sawtooth': $f(x) = \begin{cases} 2x & 0 \leq x \leq 0.5 \\ -2 + 2x & 0.5 < x \leq 1 \end{cases}$

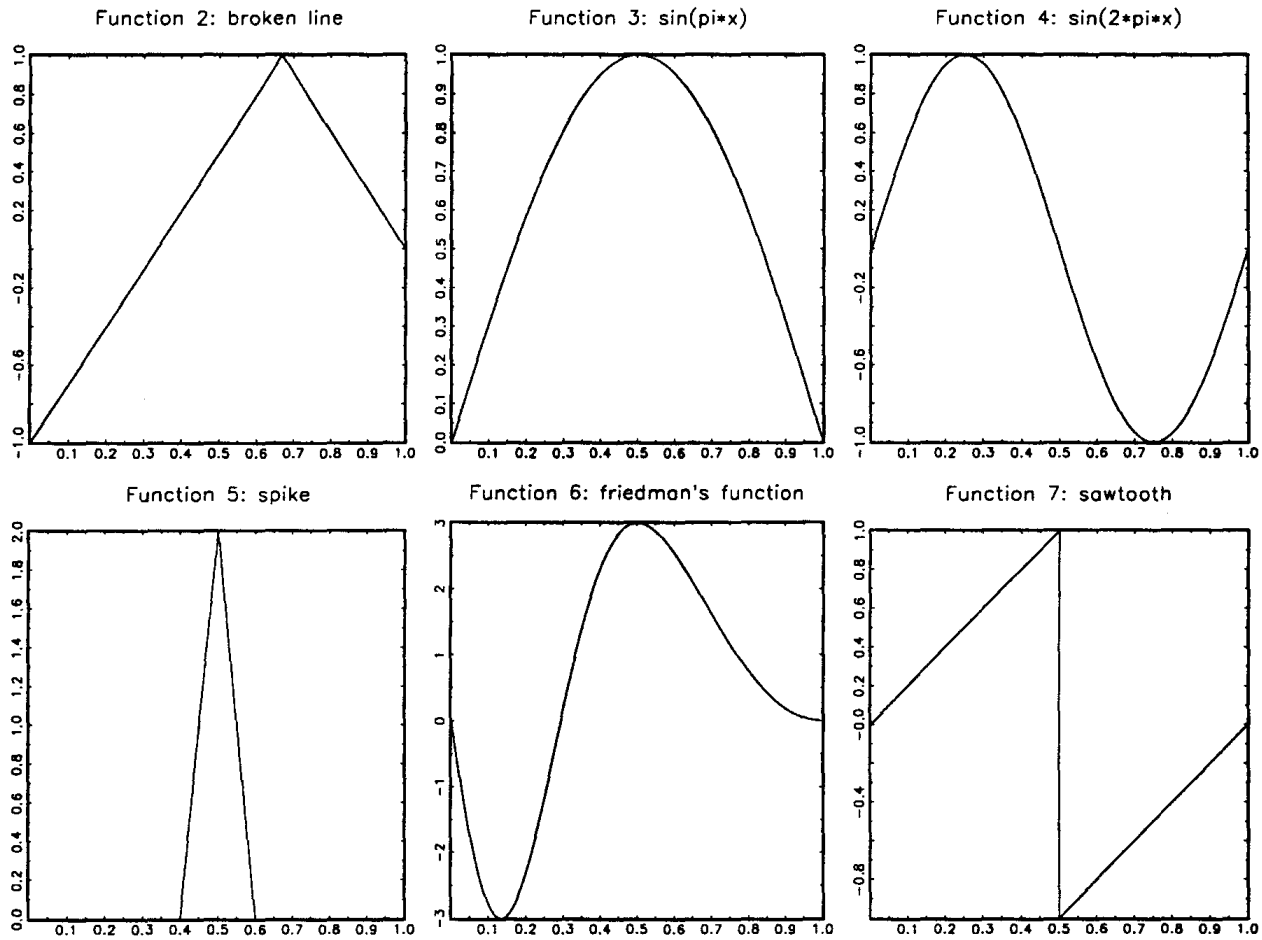


Figure 1: Functions $f(x)$ used for the simulation.

At each simulation run identically, independently and normally distributed noise ϵ is added to the underlying function, using one of the three different levels of variation ($\sigma = 0.5, 1, 2$). To automatize the smoothing procedure the optimal smoothing parameter δ was determined by cross-validation.

For each setting $J = 500$ runs were simulated and six different performance criteria were calculated:

- root mean squared error

$$rmse = \sqrt{\frac{1}{J} \sum_{j=1}^J mse_j} \quad (2.1)$$

with

$$mse = \frac{1}{n} \sum_i (f(x_i) - \hat{f}(x_i))^2 \quad (2.2)$$

- root mean squared bias

$$rmsb = \sqrt{\frac{1}{n} \sum_i (f(x_i) - \bar{f}(x_i))^2} \quad (2.3)$$

with

$$\bar{f}(x_i) = \frac{1}{J} \sum_{j=1}^J \hat{f}(x_i) \quad (2.4)$$

- maximum deviation

$$mxdev = \frac{1}{J} \sum_{j=1}^J \left(\max_i |(f(x_i) - \hat{f}(x_i))| \right) \quad (2.5)$$

- smoothness

$$smth = \frac{1}{J} \sum_{j=1}^J \left(\frac{\sum_{i=1}^{n-1} (|\hat{f}(x_{i+1}) - \hat{f}(x_i)|)}{\sum_{i=1}^{n-1} (|f(x_{i+1}) - f(x_i)|)} \right) \quad (2.6)$$

- average width of the 2-standard-deviation-band

$$band = \frac{2}{n} \sum_i sd(\hat{f}(x_i)) \quad (2.7)$$

with $sd(\hat{f}(x_i))$ being the standard deviation of $\hat{f}(x_i)$ in 500 runs.

- and a proportional root mean squared error

$$prmse = \frac{rmse}{sd(f)}. \quad (2.8)$$

Optimal moving local regression in our simulation is applied in two variants referring to different assumptions about φ . The first one (model *L*), which assumes a true local linear behaviour, applies a linear weight-function in the form of a straight line with falling weights as distance increases. The second (model *Q* assumes a locally quadratic scheme) results in a quadratic form of the weight-function. The parameter δ controls the width of the window and hence defines the amount of locality. Since the two models guarantee a fixed shaped function the computational efficiency in searching for the optimal weight-function can be enhanced. One only has to determine the window size with minimal mean squared error using cross-validation, i.e. finding

$$a^*, b^* = \arg \min_{a,b} \sum_i (y_i - \sum_{j \neq i} \lambda_j(a,b) y_j)^2$$

given (1.2), such that $\lambda_i \geq 0$ for all i and $\sum_i \lambda = 1$. Windows with 10, 20, ..., 100 percent of available datapoints are examined. Within those windows weight-functions are applied in the forms given above, always making sure that the weights sum to one.

In addition to the results given by Breiman and Peters (1991) we computed the average over the degrees of freedom (see above) and of the selected window size over the 500 runs to show that model *Q* on average uses smaller windows than model *L*.

All simulations were undertaken with the help of the program system GAUSS 3.15 (see Apteck(1993)). Further details and the corresponding code can be found in Maderbacher (1994).

3 Results

As in Breiman and Peters (1991) results are shown in detail only for the medium noise level ($\sigma = 1$). For comparative reasons all the following tables were produced in the same or a similar form like those in Breiman and Peters (1991).

sample size: 25						
	f(x)=2x			f(x)=broken line		
	equi	normally	gapped	equi	normally	gapped
L	0,44	0,40	0,38	0,52	0,46	0,46
Q	0,44	0,38	0,36	0,52	0,48	0,47
BART	0,36	0,38	0,84	0,46	0,47	0,53
DKS	0,32	0,33	0,30	0,48	0,49	0,45
SSMU	0,44	0,45	0,46	0,46	0,47	0,47
	f(x)=sin(pi*x)			f(x)=sin(2*pi*x)		
	equi	normally	gapped	equi	normally	gapped
L	0,55	0,45	0,44	0,56	0,52	0,52
Q	0,52	0,47	0,49	0,55	0,50	0,54
BART	0,44	0,43	0,51	0,47	0,48	0,52
DKS	0,44	0,46	0,46	0,54	0,56	0,49
SSMU	0,44	0,46	0,49	0,46	0,50	0,49
sample size 75						
	f(x)=2x			f(x)=broken line		
	equi	normally	gapped	equi	normally	gapped
L	0,22	0,21	0,22	0,29	0,32	0,31
Q	0,22	0,24	0,24	0,30	0,32	0,29
BART	0,22	0,27	0,28	0,27	0,29	0,29
DKS	0,19	0,18	0,19	0,29	0,34	0,27
SSMU	0,33	0,34	0,34	0,33	0,38	0,33
KERNEL	0,46	0,49	0,59	0,34	0,38	0,40
	f(x)=spike			f(x)=sin(2*pi*x)		
	equi	normally	gapped	equi	normally	gapped
L	0,44	0,50	0,48	0,30	0,31	0,32
Q	0,44	0,38	0,41	0,30	0,38	0,32
BART	0,42	0,41	0,43	0,27	0,28	0,29
DKS	0,44	0,38	0,55	0,32	0,35	0,32
SSMU	0,42	0,39	0,55	0,34	0,35	0,35
KERNEL	0,42	0,47	0,48	0,28	0,31	0,34
	f(x)=Friedman					
	equi	normally	gapped			
L	0,35	0,40	0,39			
Q	0,35	0,43	0,38			
BART	0,32	0,32	0,33			
DKS	0,33	0,36	0,35			
SSMU	0,34	0,38	0,37			
KERNEL	0,31	0,35	0,37			

Table 1: PRMSE

3.1 Analysis of PRMSE

Table (3.1), giving the *prmse* for the medium noise level, shows the advantage of *Q* over *L* in the case of sharp curvature. For both settings neither advantages nor drawbacks in comparison to alternative methods are distinguishable.

Table (3.1) gives the average of *prmse* over all considered functions. It seems like there is no significant difference in performance between *L* and *Q*. On average over functions and *x*-distributions they produce middle-class results with the best values at large sample sizes and in every case beating at least one of their competitors. Local fitting seems to be the safe option even with rather carelessly selected weight functions but with the opportunity to do extremely well if the optimal weight function is used.

3.2 Analysis of SMTH

Table (3.2) shows the average of SMTH over functions for medium noise level. Unlike the others *L* and *Q* seem to get worse with increasing sample size steadily but on the other hand at very moderate speed. The only one to cope with local fitting in this respect is DKS.

3.3 Computational Efficiency

The average time per run for the whole procedure of smoothing i.e. finding optimal smoothing parameter and carrying out the estimation on a significantly slower computer (486DX-33, 4MB) than Breiman and Peters (1991) (SUN 3/160 with FPA) were 3, 20 and 187 seconds for the sample sizes 25, 75, 225 respectively.

4 Conclusions

The values of all the criterion variables are given for the two models L and Q and for the smoothers used by Breiman and Peters (1991) in Table 4. These figures show that L and Q do considerably better on 'simple' functions, on larger sample sizes (where they do extremely well compared to the alternative smoothers) and on more unregular designs like 'normally' and 'gapped'. Comparing L and Q , the first one dominates the other in the case of medium samplesizes and on functions that are linear or show sharp but local changes (broken line, spike, sawtooth) or with low and steady curvature ($\sin(x\pi)$), whereas model Q dominates in the case of functions with relatively high and steady curvature ($\sin(2x\pi)$, Friedman's function).

Our simulations produced values of the criterion *band* of about two times the magnitude that the ones of Breiman and Peters (1991) but with no evidence from graphical output that the variability of our smoothers is that much larger. Therefore we assume the formula of *band* given by Breiman and Peters (1991) was used in their simulation without the factor '2'. Our results are corrected to reflect this discrepancy.

In addition to the simulation results the graphical output of Breiman and Peters (1991) was duplicated as well, showing the average of $\hat{f}(x)$ over the runs and the curves given by the average of $\hat{f}(x) \pm sd(\hat{f}(x))$ superimposing the underlying function (all of them linearly interpolated between the neighbouring x -values). They also give the variability of the estimates over the 500 runs, all for sample size 75 and gap-distribution. The graphic for model Q and Friedman's function shows that even with rather bad design (due to large gaps) local fitting can give good smoothing results.

Summarizing, local fitting procedures such as 'optimized moving local regression' can, besides their theoretical advantages, also be regarded as serious competitors to more traditional smoothers from the point of view of practical performance.

References

- Aptech Systems, Inc., Maple Valley, (1993). *GAUSS Volume 2, Command Reference*.
- Breiman, L. and Peters, S. (1991). Comparing automatic smoothers (a public service enterprise). *International Statistical Review*, 60(3):271–290.
- Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836.
- Fedorov, V.V., P.Hackl, and Müller, W.G. (1993). Moving local regression: The weight function. *Journal of Nonparametric Statistics*, 2(4):355–368.
- Fedorov, V.V., P.Hackl, and Müller, W.G. (1994). Optimized moving local regression: Computational methods for the moving average case. In Dutter, R. and Grossmann, W., editors, *COMPSTAT: Proceedings in Computational Statistics*, pages 476–481. Physica.
- Maderbacher, M. (1994). Vergleich automatischer Glättungsverfahren und Implementierung der gleitenden lokalen Regression in GAUSS. Master's thesis, Wirtschaftsuniversität Wien.

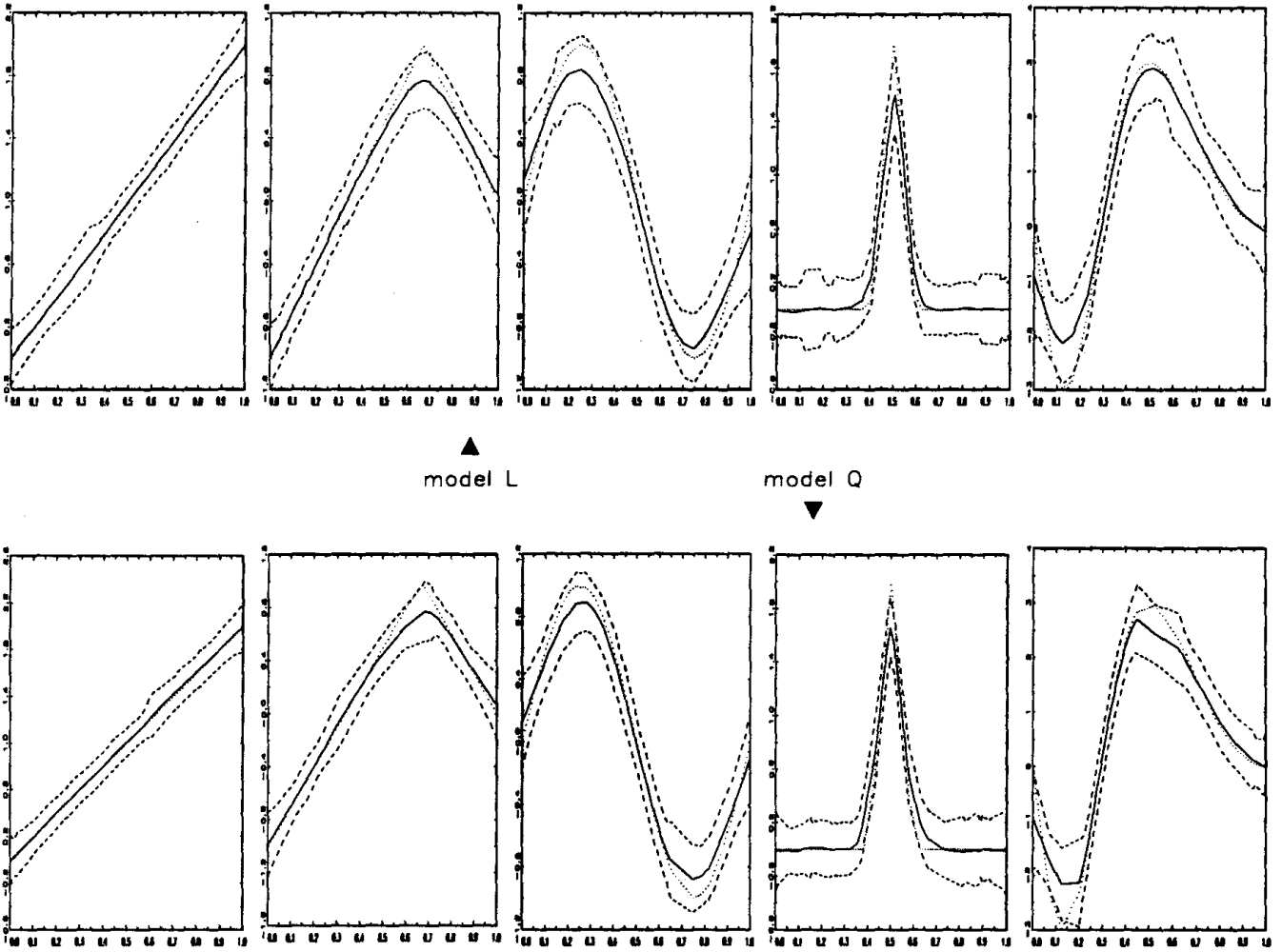


Figure 2: Averages of $\hat{f}(x)$ and confidence curves.

	sample size 25				sample size 75				sample size 225			
	equi	normally	gapped	average	equi	normally	gapped	average	equi	normally	gapped	average
L	0,52	0,46	0,45	0,48	0,32	0,35	0,34	0,34	0,22	0,23	0,22	0,22
Q	0,51	0,46	0,47	0,48	0,32	0,35	0,33	0,33	0,22	0,22	0,23	0,22
BART	0,43	0,44	0,60	0,49	0,30	0,31	0,32	0,31	0,22	0,22	0,23	0,22
DKS	0,45	0,46	0,43	0,45	0,31	0,32	0,34	0,32	0,22	0,22	0,23	0,22
SSMU	0,45	0,47	0,48	0,47	0,35	0,37	0,39	0,37	0,22	0,21	0,23	0,22
KERNEL					0,32	0,40	0,44	0,40				

Table 2: averaged PRMSE

<i>SMTH</i>	ss25	ss75	ss225
L	1.11	1.11	1.12
Q	1.09	1.10	1.11
BART	1.32	1.22	1.26
DKS	1.09	1.08	1.07
SSMU	1.63	2.02	1.06
KERNEL		1.50	

Table 3: SMTH

$f(x)=2x$	equi-spaced								normally spaced								gap spaced							
	rmse	rmsb	mxdv	smth	band	dof	q	rmse	rmsb	mxdv	smth	band	dof	q	rmse	rmsb	mxdv	smth	band	dof	q			
0.07	0.00	0.13	1.05	0.08	221.24	0.87	0.04	0.00	0.12	1.01	0.04	221.41	0.91	0.07	0.00	0.13	1.05	0.07	221.20	0.86				
0.07	0.00	0.13	1.03	0.07	221.77	0.82	0.05	0.00	0.11	1.02	0.05	221.68	0.87	0.07	0.00	0.13	1.03	0.07	221.72	0.83				
0.08	0.00	0.20	1.14	0.09			0.06	0.00	0.27	1.11	0.06			0.13	0.01	0.36	1.55	0.13						
0.07	0.00	0.14	1.07	0.08			0.04	0.00	0.12	1.04	0.05			0.07	0.00	0.14	1.08	0.08						
0.09	0.00	0.25	1.11	0.09			0.05	0.00	0.20	1.07	0.05			0.09	0.00	0.26	1.26	0.09						
	equi-spaced								normally spaced								gap spaced							
	rmse	rmsb	mxdv	smth	band	dof	q	rmse	rmsb	mxdv	smth	band	dof	q	rmse	rmsb	mxdv	smth	band	dof	q			
0.10	0.04	0.24	0.93	0.09	217.64	0.35	0.08	0.03	0.40	0.94	0.08	214.89	0.31	0.11	0.04	0.26	0.92	0.10	217.38	0.33				
0.10	0.04	0.25	0.94	0.09	218.50	0.31	0.08	0.03	0.32	0.92	0.08	217.96	0.34	0.10	0.04	0.24	0.96	0.10	218.50	0.31				
0.09	0.03	0.24	0.96	0.09			0.08	0.02	0.37	1.04	0.08			0.12	0.01	0.33	1.19	0.12						
0.10	0.02	0.26	1.00	0.11			0.09	0.03	0.35	0.97	0.09			0.10	0.02	0.26	1.04	0.10						
0.10	0.03	0.28	0.97	0.09			0.08	0.03	0.34	0.96	0.08			0.09	0.02	0.25	1.12	0.09						
	equi-spaced								normally spaced								gap spaced							
	rmse	rmsb	mxdv	smth	band	dof	q	rmse	rmsb	mxdv	smth	band	dof	q	rmse	rmsb	mxdv	smth	band	dof	q			
0.13	0.05	0.39	1.46	0.12	204.01	0.10	0.14	0.06	0.45	1.24	0.13	211.66	0.19	0.13	0.05	0.37	1.52	0.12	203.91	0.10				
0.13	0.05	0.40	1.37	0.12	208.85	0.10	0.15	0.07	0.53	1.24	0.13	214.57	0.20	0.15	0.07	0.46	1.35	0.13	210.94	0.14				
0.13	0.05	0.39	1.37	0.12			0.16	0.06	0.63	1.55	0.14			0.09	0.04	0.39	1.27	0.09						
0.13	0.06	0.39	1.21	0.12			0.15	0.03	0.47	1.22	0.15			0.12	0.08	0.56	0.97	0.08						
0.12	0.07	0.39	1.10	0.10			0.15	0.07	0.49	1.21	0.13			0.11	0.09	0.84	0.89	0.07						
	equi-spaced								normally spaced								gap spaced							
	rmse	rmsb	mxdv	smth	band	dof	q	rmse	rmsb	mxdv	smth	band	dof	q	rmse	rmsb	mxdv	smth	band	dof	q			
0.13	0.05	0.31	0.96	0.13	216.43	0.28	0.13	0.05	0.82	0.95	0.11	214.01	0.29	0.13	0.05	0.34	0.93	0.12	216.61	0.28				
0.13	0.05	0.31	0.95	0.12	217.77	0.26	0.13	0.06	0.88	0.92	0.11	216.20	0.26	0.13	0.05	0.30	1.00	0.12	217.62	0.26				
0.12	0.03	0.30	0.99	0.12			0.11	0.01	0.50	1.05	0.11			0.14	0.01	0.39	1.18	0.14						
0.15	0.03	0.37	1.06	0.15			0.15	0.05	0.63	0.90	0.14			0.13	0.02	0.35	1.04	0.14						
0.14	0.05	0.40	1.00	0.13			0.13	0.07	0.62	0.85	0.11			0.14	0.06	0.37	1.05	0.13						
	equi-spaced								normally spaced								gap spaced							
	rmse	rmsb	mxdv	smth	band	dof	q	rmse	rmsb	mxdv	smth	band	dof	q	rmse	rmsb	mxdv	smth	band	dof	q			
0.40	0.15	1.05	0.99	0.37	213.85	0.21	0.40	0.21	1.94	0.94	0.33	212.86	0.23	0.42	0.16	1.06	1.00	0.38	214.85	0.23				
0.40	0.14	1.04	1.01	0.37	215.59	0.19	0.40	0.21	2.63	0.95	0.32	216.08	0.26	0.39	0.15	1.00	1.01	0.35	215.50	0.20				
0.36	0.12	1.03	1.00	0.35			0.34	0.10	1.62	1.05	0.32			0.36	0.03	1.01	1.17	0.36						
0.39	0.07	1.06	0.98	0.40			0.40	0.12	1.55	1.03	0.38			0.36	0.05	1.06	1.07	0.36						
0.39	0.17	1.19	0.98	0.35			0.38	0.17	1.50	1.03	0.33			0.40	0.17	1.34	1.12	0.35						
	equi-spaced								normally spaced								gap spaced							
	rmse	rmsb	mxdv	smth	band	dof	q	rmse	rmsb	mxdv	smth	band	dof	q	rmse	rmsb	mxdv	smth	band	dof	q			
0.21	0.15	1.02	1.37	0.15	205.78	0.12	0.25	0.17	1.03	1.56	0.18	208.20	0.16	0.20	0.13	0.90	1.32	0.15	208.90	0.15				
0.21	0.16	1.04	1.30	0.14	209.78	0.11	0.24	0.15	1.06	1.58	0.17	210.99	0.15	0.23	0.18	1.06	1.39	0.14	208.79	0.11				
0.21	0.14	1.03	1.57	0.16			0.28	0.18	1.21	2.07	0.20			0.14	0.06	0.63	1.33	0.13						
0.20	0.16	1.04	1.12	0.13			0.24	0.17	1.24	1.24	0.16			0.15	0.09	0.77	1.09	0.11						
0.19	0.15	1.03	0.98	0.11			0.22	0.16	1.26	1.22	0.14			0.14	0.10	0.80	1.24	0.10						

Table 4: Full results of the simulation.