Regensburger
DISKUSSIONSBEITRÄGE
zur Wirtschaftswissenschaft

University of Regensburg Working Papers in Business,
Economics and Management Information Systems

A hat matrix for monotonicity constrained B-spline and P-spline regression

Kathrin Kagerer*

March, 05 2015

Nr. 484

JEL Classification: C14, C52

Key Words: Spline, monotonicity, penalty, hat matrix, regression, Monte Carlo simulation.

* Kathrin Kagerer was research assistant at the Chair of Econometrics, Faculty of Business, Economics and Management Information Systems at the University of Regensburg, 93040 Regensburg, Germany.

A hat matrix for monotonicity constrained B-spline and P-spline regression

Kathrin Kagerer*

March 5, 2015

Summary. Splines constitute an interesting way to flexibly estimate a nonlinear relationship between several covariates and a response variable using linear regression techniques. The popularity of splines is due to their easy application and hence the low computational costs since their basis functions can be added to the regression model like usual covariates. As long as no inequality constraints and penalties are imposed on the estimation, the degrees of freedom of the model estimation can be determined straightforwardly as the number of estimated parameters. This paper derives a formula for computing the hat matrix of a penalized and inequality constrained splines estimator. Its trace gives the degrees of freedom of the model estimation which are necessary for the calculation of several information criteria that can be used e.g. for specifying the parameters for the spline or for model selection.

Key words. Spline, monotonicity, penalty, hat matrix, regression, Monte Carlo simulation.

*Most of the work was done as research assistant at the Chair of Econometrics, Faculty of Business, Economics and Management Information Systems, University of Regensburg, 93040 Regensburg, Germany.

1 Introduction

Compared to classical parametric models, nonparametric models have the advantage of a flexible functional form that is determined by the available data. With increasing computing power the opportunities for nonparametric methods have improved and hence these methods are applied more and more. Though estimating models with a spline component basically corresponds to estimating a linear model, optimization methods are required when inequality constraints such as monotonicity are incorporated. Further, for estimating spline models with monotonicity constraints, the degrees of freedom of the estimated model are not equal to the number of estimated parameters like it is for non-constrained parametric models. The degrees of freedom of a model estimation are required for example when information criteria like the Akaike or Schwarz criterion are applied for model selection. In this paper, a formula is presented to calculate the hat matrix of the estimated spline regression with monotonicity constraints and with or without additional penalty terms. The trace of this hat matrix then gives the degrees of freedom of the estimation like in Ruppert et al. (2003).

The remainder of this paper is organized as follows: Section 2 summarizes the splines specifications used in this paper and derives for each the hat matrix / smoothing matrix. In a Monte Carlo study, the hat matrix is computed for estimations for data from different DGPs in Section 3 and an empirical example using the LIDAR data set is presented in Section 4. Section 5 concludes.

2 Splines and their hat matrix

2.1 Splines with penalties and monotonicity constraints

A spline can be used to approximate other functions (e.g. Ruppert et al., 2003, de Boor, 2001, or Dierckx, 1993 as general references for splines). It consists of piecewise polynomial functions which are connected at knots and satisfy certain continuity conditions at these knots. The order of the piecewise polynomial functions is determined by the order k of the spline. The knot sequence $\boldsymbol{\kappa} = (\kappa_{-(k-1)}, \ldots, \kappa_{m+k})$ consists of $m + 2k$ non-decreasing knot positions κ_j, $j = -(k-1), \ldots, m+k$, where κ_0 and κ_{m+1} are the boundary knots that usually coincide with the bounds of the interval of interest, i.e. in the case of a scalar covariate x these are $\kappa_0 = \min_i(x_i)$ and $\kappa_{m+1} = \max_i(x_i)$. A spline s can be constructed

as a linear combination of basis functions. Using B-spline basis functions $B_j^{\boldsymbol{\kappa},k}$, the spline is

$$s(\cdot) = \sum_{j=-(k-1)}^{m} \alpha_j B_j^{\boldsymbol{\kappa},k}(\cdot). \tag{1}$$

For a definition of the B-spline basis functions see e.g. de Boor (2001, Chapter IX) or Schumaker (1981, Chapter 3). Each of the basis functions $B_j^{\boldsymbol{\kappa},k}$ is positive on the interval (κ_j, κ_{j+k}) and zero outside and the unweighted sum of the B-spline basis functions (i.e. $\alpha_j = 1$ for all j) is 1 on $[\kappa_0, \kappa_{m+1}]$.

For regression purposes, splines can be used to estimate an unknown sufficiently smooth regression curve. Consider the bivariate functional relationship $y = f(x) + u$ with

$$\mathrm{E}(y|x) = f(x), \tag{2}$$

where the regression function f is to be estimated using spline regression, i.e. minimizing

$$\sum_{i=1}^{n} \left(y_i - \tilde{f}(x_i)\right)^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=-(k-1)}^{m} \tilde{\alpha}_j B_j^{\boldsymbol{\kappa},k}(x_i)\right)^2 \tag{3}$$

with respect to the $m + k$ parameters $\tilde{\alpha}_j$ for a given sample $i = 1, \ldots, n$.

Restricting the estimated parameters $\hat{\alpha}_j$ such that they are in non-decreasing order,

$$\hat{\alpha}_j \leq \hat{\alpha}_{j+1}, \quad j = -(k-1), \ldots, m-1, \tag{4}$$

ensures a monotone increasing estimated spline function and analogously, a decreasing function results for non-increasing parameters (e.g. Dierckx, 1993, Section 7.1). For $k \geq 4$, this restriction is not necessary to obtain a monotone function, but it is a sufficient condition and is easy to implement in the used software.

Cubic splines ($k = 4$) are commonly used in practice (e.g. Bollaerts et al., 2006, Eilers & Marx, 1996). They are easy to handle, exhibit a good fit and can be subject to several constraints as for example monotonicity or convexity (cf. Dierckx, 1993, Sections 3.2, 7.1). Hence, cubic splines are also applied here.

Together with the knot sequence, the order of the spline fully determines the functions $B_j^{\boldsymbol{\kappa},k}$ of the B-spline basis. Eilers & Marx (1996) and Ruppert et al. (2003, Section 3.4) state some studies with an automatic choice of the knot sequence (i.e. the number and

location of the knots) which, however, is computationally expensive. But if the knot sequence is restricted to be equidistant, only the number of knots has to be chosen. Using many knots can result in a rough fit, while using only few knots may not reflect the conditional relationship (2) well. Hence, Eilers & Marx (1996) propose the use of quite many equidistant knots while penalizing a rough fit (P-splines). This is achieved for example by avoiding large second-order differences of the estimated parameters $\hat{\alpha}_j$, i.e. by penalizing large $\Delta^2\tilde{\alpha}_j = \tilde{\alpha}_j - 2\tilde{\alpha}_{j-1} + \tilde{\alpha}_{j-2}$. The objective function of the resulting minimization problem then is

$$\sum_{i=1}^{n}\left(y_i - \sum_{j=-(k-1)}^{m}\tilde{\alpha}_j B_j^{\boldsymbol{\kappa},k}(x_i)\right)^2 + \lambda \sum_{j=-(k-1)+2}^{m}(\Delta^2\tilde{\alpha}_j)^2, \tag{5}$$

where λ is the smoothing parameter which controls the amount of smoothing and has to be chosen by the researcher (see below). Note that for $\lambda = 0$ the unpenalized fit as in Equation (3) results and for $\lambda \to \infty$ the fit is given by a straight line for $k = 4$ (cf. Eilers & Marx, 1996, with cubic splines and a penalty on the second-order differences of the estimated parameters). Still the number of knots has to be specified, though this is not that influential (see Ruppert et al., 2003, Sections 5.1, 5.5). For example, Ruppert (2002) proposes to use roughly

$$\min(n/4, 35) \tag{6}$$

inner knots as a rule of thumb.

Now only the smoothing parameter λ is left to be specified. It can be chosen for example by the (generalized) cross validation criterion (CV, GCV) or the Akaike information criterion (AIC) (e.g. summarized in Ruppert et al., 2003, Section 5.3). Several of these criteria are based on the elements of the diagonal of the hat matrix of the estimation. Ruppert et al. (2003, Section 3.13) use the trace of the hat matrix of a penalized spline estimation as the equivalent to the degrees of freedom in linear models. For linear models, the degrees of freedom (df) are given by the number of parameters which in this case equals the trace of the hat matrix. Section 2.2 explains how to obtain the hat matrix for the spline estimators considered in this section.

For more details on spline estimation using the same notation see Kagerer (2013). That work also covers the general case with one or more than one covariate which is considered in the next section.

2.2 A hat matrix for monotonicity constrained P-splines

The hat matrix $\mathbf{H}$ of the minimization problem $\min_{\tilde{f}} \sum_{i=1}^{n}(y_i - \tilde{f}(\mathbf{x}_i))^2$ with $q \times 1$ covariate vector $\mathbf{x}_i$ is defined to be the matrix for which $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$. In case of a linear regression, i.e. $\min_{\tilde{\boldsymbol{\alpha}}} \sum_{i=1}^{n}(y_i - \mathbf{x}_i\tilde{\boldsymbol{\alpha}})^2$, with $\mathbf{X} = \begin{pmatrix}\mathbf{x}_1^T & \cdots & \mathbf{x}_i^T & \cdots & \mathbf{x}_n^T\end{pmatrix}^T$, the hat matrix is given by

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T. \tag{7}$$

For penalized estimations with a general symmetric penalty matrix $\mathbf{D}$ (for an example see Equation (11) below) where $\sum_{i=1}^{n}(y_i - \mathbf{x}_i\tilde{\boldsymbol{\alpha}})^2 + \lambda\tilde{\boldsymbol{\alpha}}^T\mathbf{D}\tilde{\boldsymbol{\alpha}}$ is minimized with respect to $\tilde{\boldsymbol{\alpha}}$, the hat matrix can be determined as

$$\mathbf{H}_\lambda = \mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{D})^{-1}\mathbf{X}^T \tag{8}$$

(e.g. Ruppert et al., 2003, Section 3.10).

For regression problems with general inequality constraints but without penalty, i.e. $\min_{\tilde{\boldsymbol{\alpha}}} \sum_{i=1}^{n}(y_i - \mathbf{x}_i\tilde{\boldsymbol{\alpha}})^2$ subject to $\mathbf{C}\tilde{\boldsymbol{\alpha}} \geq \mathbf{0}$ (for an example see Equation (12) below), the hat matrix can be derived from the work of Paula (1999) and is given by

$$\mathbf{H}_{\text{constr}} = \mathbf{X}\left(\mathbf{I} - (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}_R^T(\mathbf{C}_R(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}_R^T)^{-1}\mathbf{C}_R\right)(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T, \tag{9}$$

where the $q \times r$ matrix $\mathbf{C}_R$ contains the r rows of $\mathbf{C}$ satisfying $\mathbf{C}\hat{\boldsymbol{\alpha}} = \mathbf{0}$ (cf. Paula, 1993, 1999). Since $\hat{\boldsymbol{\alpha}}$ is a random variable, it is also random which rows of $\mathbf{C}$ satisfy $\mathbf{C}\hat{\boldsymbol{\alpha}} = \mathbf{0}$, hence $\mathbf{C}_R$ and with it $\mathbf{H}_{\text{constr}}$ and its trace are random. This can also be observed in the simulation in Section 3 (e.g. Figure 5) and is also discussed in the empirical Section 4.

Penalized estimations with inequality constraints are obtained by minimizing $\sum_{i=1}^{n}(y_i - \mathbf{x}_i\tilde{\boldsymbol{\alpha}})^2 + \lambda\tilde{\boldsymbol{\alpha}}^T\mathbf{D}\tilde{\boldsymbol{\alpha}}$ subject to $\mathbf{C}\tilde{\boldsymbol{\alpha}} \geq \mathbf{0}$ with respect to $\tilde{\boldsymbol{\alpha}}$. Since the penalized estimation without constraints can be interpreted as ordinary least-squares problem with $\mathbf{X}^* = \begin{pmatrix}\mathbf{X}^T & \sqrt{\lambda}(\mathbf{D}^{1/2})^T\end{pmatrix}^T$ and $\mathbf{y}^* = \begin{pmatrix}\mathbf{y}^T & \mathbf{0}^T\end{pmatrix}^T$ (e.g. Eilers & Marx, 1996), these two hat matrices can be combined, resulting in the hat matrix for inequality constrained penalized estimations:

$$\mathbf{H}_{\lambda,\text{ constr}} = \mathbf{X}\left(\mathbf{I} - (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{D})^{-1}\mathbf{C}_R^T(\mathbf{C}_R(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{D})^{-1}\mathbf{C}_R^T)^{-1}\mathbf{C}_R\right)(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{D})^{-1}\mathbf{X}^T. \tag{10}$$

Note that for the hat matrices from the constrained estimations (9) and (10), $\mathbf{C}_R$ is empty if the inequality constraint on $\mathbf{C}\hat{\boldsymbol{\alpha}}$ is not necessary (i.e. none of the entries of $\mathbf{C}\hat{\boldsymbol{\alpha}}$ is zero) and hence, the hat matrices reduce to the unconstrained case in (7) and (8).

Applying the properties of the trace operator, the trace of the hat matrix reduces to q (i.e. the number of estimated parameters) in case of the unconstrained unpenalized estimation (hat matrix (7)) and to $q-r$ in case of the constrained but unpenalized estimation (hat matrix (9)). Hence for $\min_{\tilde{\boldsymbol{\alpha}}} \sum_{i=1}^{n}(y_i - \mathbf{x}_i\tilde{\boldsymbol{\alpha}})^2$ subject to $\mathbf{C}\tilde{\boldsymbol{\alpha}} \geq \mathbf{0}$, the degrees of freedom of the estimation equal the number of estimated parameters minus the number of rows of the constraint matrix $\mathbf{C}$ for which $\mathbf{C}\hat{\boldsymbol{\alpha}} = \mathbf{0}$ holds. For the case of penalized estimations, the trace of the hat matrices (8) and (10) can not be simplified in an analog way and have to be determined after having estimated the model with the given data.

For penalized monotonicity constrained spline estimations ((5) with constraint (4)) with equidistant knots for a scalar covariate x, the (i,j)-th entry of the $n\times(m+k)$ matrix $\mathbf{X}$ is $B_{j-k}^{\boldsymbol{\kappa},k}(x_i)$. The penalty matrix $\mathbf{D}$ is the matrix for which $\sum_{j=-(k-1)+2}^{m}(\Delta^2\tilde{\alpha}_j)^2 = \tilde{\boldsymbol{\alpha}}^T\mathbf{D}\tilde{\boldsymbol{\alpha}}$ holds and hence is given by

$$\mathbf{D} = \begin{pmatrix} 1 & -2 & 1 & & & & & \\ -2 & 5 & -4 & 1 & & & & \\ 1 & -4 & 6 & -4 & \cdot & & & \\ & 1 & -4 & 6 & \cdot & \cdot & & \\ & & 1 & -4 & \cdot & \cdot & 1 & \\ & & & 1 & \cdot & \cdot & -4 & 1 \\ & & & & \cdot & \cdot & 6 & -4 & 1 \\ & & & & & \cdot & -4 & 5 & -2 \\ & & & & & & 1 & -2 & 1 \end{pmatrix}. \tag{11}$$

Note that the penalty matrix can also be obtained for non-equidistant knot sequences (cf. Kagerer, 2013).

The constraint matrix $\mathbf{C}$ required to satisfy the constraint (4) which results in a monotonically increasing fit is

$$\mathbf{C} = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & -1 & 1 & \\ & & & \ddots & \ddots \end{pmatrix} \tag{12}$$

and for a monotone decreasing fit it has to be multiplied by -1.

3 Monte Carlo simulation

3.1 Data generating processes

In this section only bivariate data generating processes (DGPs) are considered, i.e.

$$y = f(x) + u.$$

For all DGPs, the scalar covariate x and the errors u are assumed to be distributed as

$$x \sim U(0,1), \qquad u|x \sim N(0,\sigma^2).$$

To be able to use the same knot sequence for all replications, x is rescaled such that $\min_i(x_i) = 0$ and $\max_i(x_i) = 1$ for each sample.

Six different regression functions f are studied:

$$
\begin{aligned}
f_1(x) &= x^2, \\
f_2(x) &= 4\,(x - 0.5)^3 + 0.5, \\
f_3(x) &= 34.1\,x^5 - 85.3\,x^4 + 78.23\,x^3 - 32\,x^2 + 6\,x, \\
f_4(x) &= \sum_{j=-3}^{5} \alpha_j\, B_j^{\boldsymbol{\kappa},4}(x) - 0.2 \quad \text{for } \boldsymbol{\alpha}^T = \big(\,2\;2\;7\;7\;8\;9\;16\;16\;20\,\big)/\,14\,, \\
f_5(x) &= x - \sin(5\pi\,x)/16, \\
f_6(x) &= \frac{\exp(10(x - 0.5))}{1 + \exp(10(x - 0.5))}.
\end{aligned}
$$

The functions f_1, f_2 and f_3 are polynomial functions of different degrees, f_4 is a cubic spline, f_5 is the sine function with higher periodicity and a trend and f_6 is the CDF of the logistic distribution with parameters $a = 0.5$ and $b = 0.1$. All functions are chosen such that they are monotonically increasing on the interval $[0, 1]$. Further, they are (approximately) scaled on the interval $[0, 1]$, i.e. $f(x) \in [0, 1]$ for $x \in [0, 1]$, hence the same error variance σ^2 is appropriate for all DGPs and is chosen to equal $\sigma^2 = 0.09$. Figure 1 presents the respective functions and Figure 2 shows one simulated sample for each DGP.

For all estimations the open source software `R` (R Core Team, 2014, version 3.1.2, 32 bit) is used. The spline regressions are based on the base package `splines` and the constraint is implemented using the function `pcls` from the `mgcv` package from Wood (2014).

3.2 Simulation results

For each replication $r = 1, \ldots, R$, $R = 1000$, of the Monte Carlo simulation, a sample of size $n = 500$ is drawn for x and u and the corresponding y for the regression functions f_1 to f_6 are calculated. According to Equation (6), the knot sequence for the spline basis $\boldsymbol{\kappa}$ contains $m = 35$ equidistant inner knots and is subject to the constraint (4) for all functions f_1 to f_6. For each function the smoothing parameter λ has to be chosen. In the simulation the true function f is known. Hence, λ can be selected for each of the six functions f from Section 3.1 by minimizing the mean integrated squared error ($MISE$).

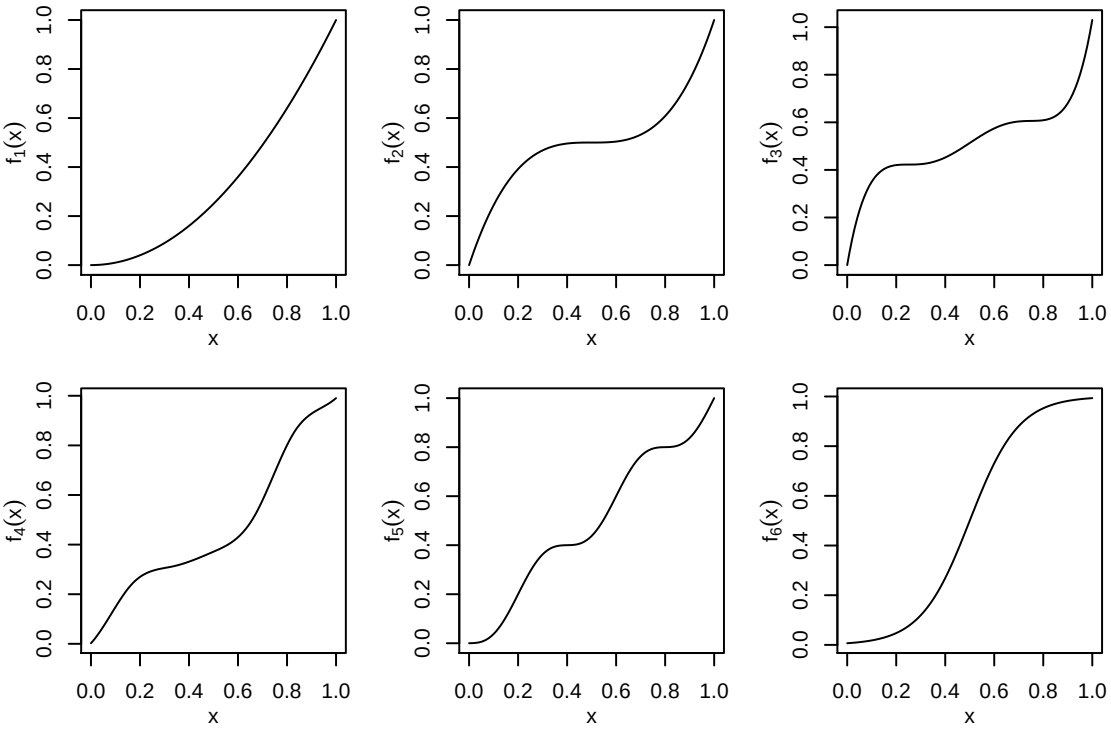

Figure 1: Plot of the regression functions of each DGP used in the simulation.

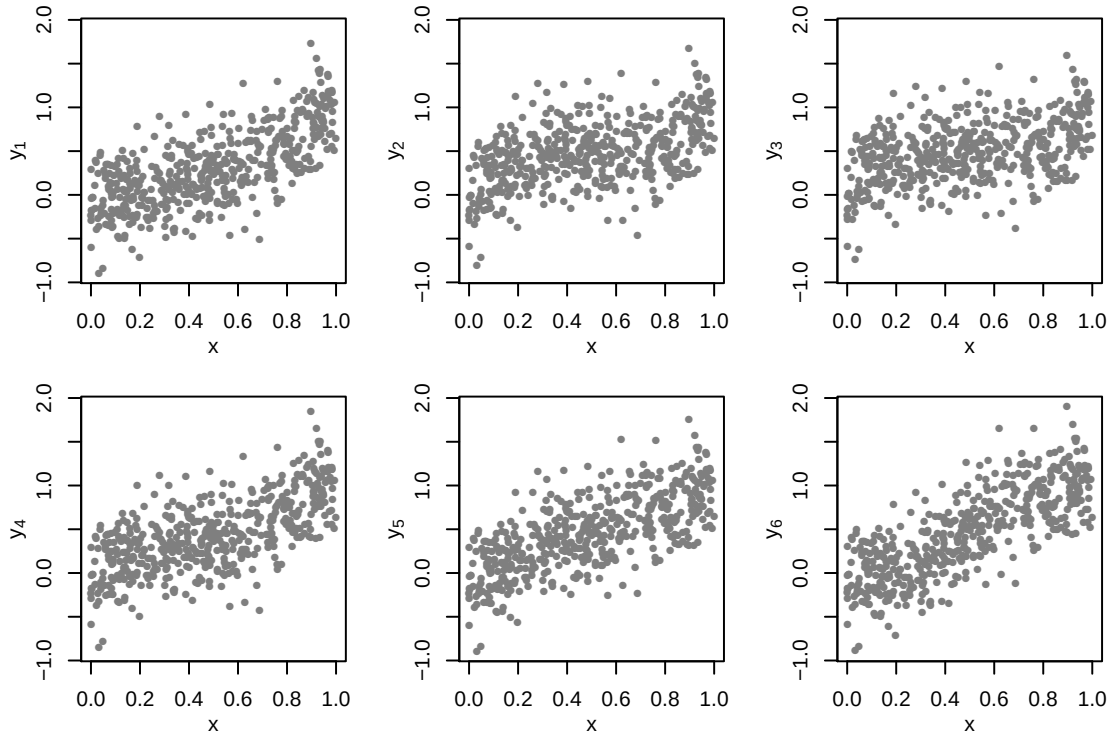

Figure 2: Plot of one exemplary sample for each DGP.

Then the smoothing parameter is chosen as

$$\lambda = \arg\min_{\tilde{\lambda}} \frac{1}{R} \sum_{r=1}^{R} \left(\frac{1}{n} \sum_{i=1}^{n} \left(f(x_{i,r}) - \hat{f}_{\tilde{\lambda},r}(x_{i,r}) \right)^2 \right),$$

where $x_{i,r}$ is the ith observation in the rth replication and $\hat{f}_{\tilde{\lambda},r}$ is the estimate of f for the rth replication and a given value $\tilde{\lambda}$ for the smoothing parameter. The results for λ chosen by minimizing the $MISE$ can be found in Figure 3 and are comparable to the smoothing parameter analogously obtained by the mean GCV or the mean AIC criterion (cf. Table 1). Selection criteria like GCV or AIC are feasible for real data problems when f is unknown and hence are applied for the empirical example in Section 4. The advantage of using the $MISE$ for the simulation is that it does not contain the trace of the hat matrix of the estimations and hence can be used to compare the results to those for e.g. GCV or AIC. Note that for choosing the optimal λ in the simulation, only 100 of the $R = 1000$ samples are included to reduce the computational effort.

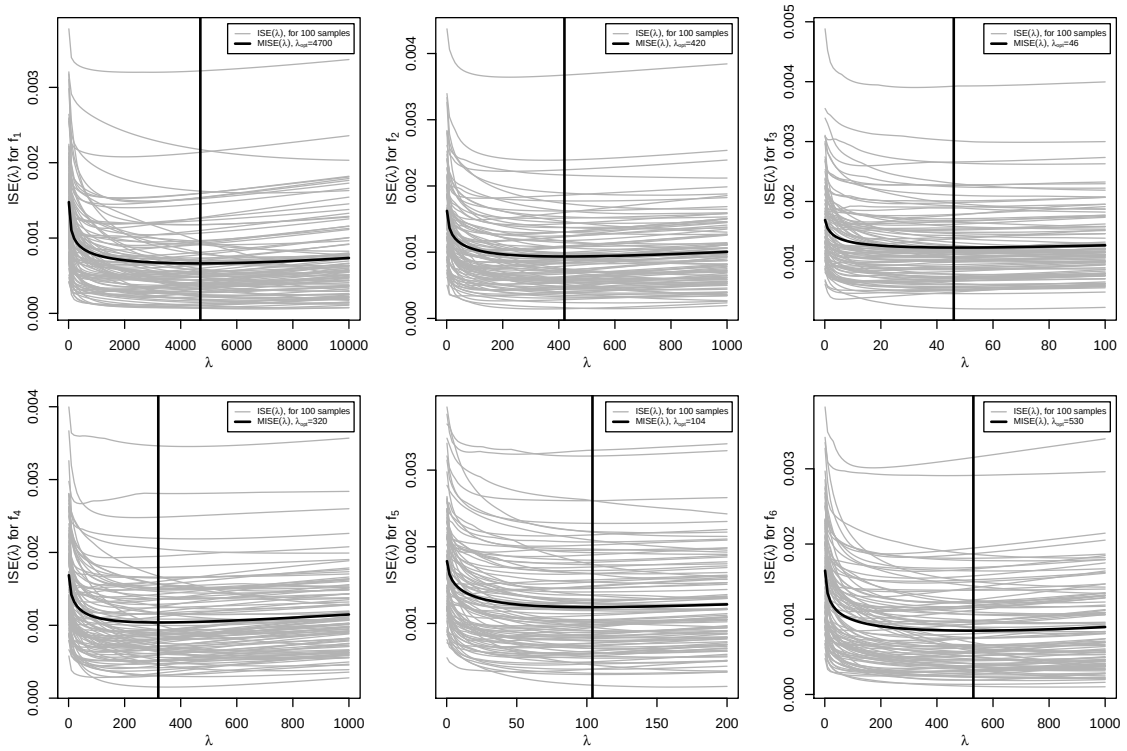

Figure 3: Integrated squared error from 100 of the $R = 1000$ samples (ISE, grey) and mean integrated squared error ($MISE$, black) depending on λ.
The vertical line is the optimal λ with respect to $MISE$.

function	f_1	f_2	f_3	f_4	f_5	f_6
λ_{opt} from $MISE$	4700	420	46	320	104	530
λ_{opt} from $\overline{GCV}$	5100	620	43	360	120	490
λ_{opt} from $\overline{AIC}$	4400	620	30	360	120	490

Table 1: Optimal λ for the different regression functions chosen with respect to $MISE$, mean GCV and mean AIC from 100 of the $R = 1000$ samples.

For each of the $R = 1000$ samples, the six functions are estimated using cubic splines with $m = 35$ inner knots, the optimal smoothing parameter λ with respect to $MISE$ and monotonicity constraint (4). The corresponding estimated regression curves can be regarded in Figure 4 where it can be observed that they fit the DGP functions very well. Therefore, the smoothing parameters appear to be well-chosen.

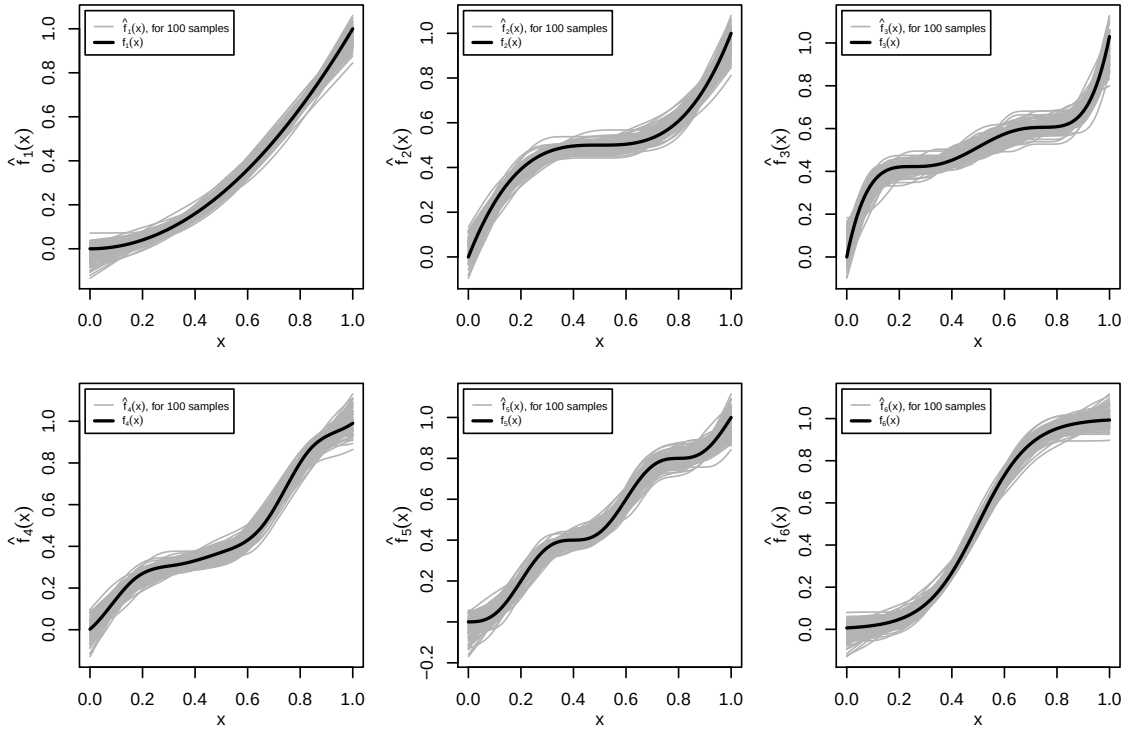

Figure 4: Fitted regression curves from 100 of the $R = 1000$ samples and corresponding DGP function. For each estimation the same optimal λ w.r.t. $MISE$ (see Table 2) is used.

Using the estimation results, the hat matrix (10) and its trace are calculated each time. The results are aggregated in the density plots in Figure 5. It can be seen that

for each DGP the obtained degrees of freedom do not vary by much (compared to the maximal degrees of freedom of $m + k = 39$) and the estimated standard errors lie between 0.08 and 0.70 for the six chosen functions (cf. also Table 2). For the polynomial functions f_1, f_2 and f_3 of degree 2, 3 and 5 (i.e. at most 3, 4, 6 parameters to estimate) the degrees of freedom are about 4, 5-6 and 7-9, hence only slightly overestimated. Note that the extent of over-/underestimation of the degree of the functions also depends on the chosen error variance. For example a larger error variance of $\sigma_u^2 = 0.25$ leads to traces of the hat matrix of 3-4, 4-5 and 6-7 for these three functions, which is also quite close to the degrees of the three polynomial functions.

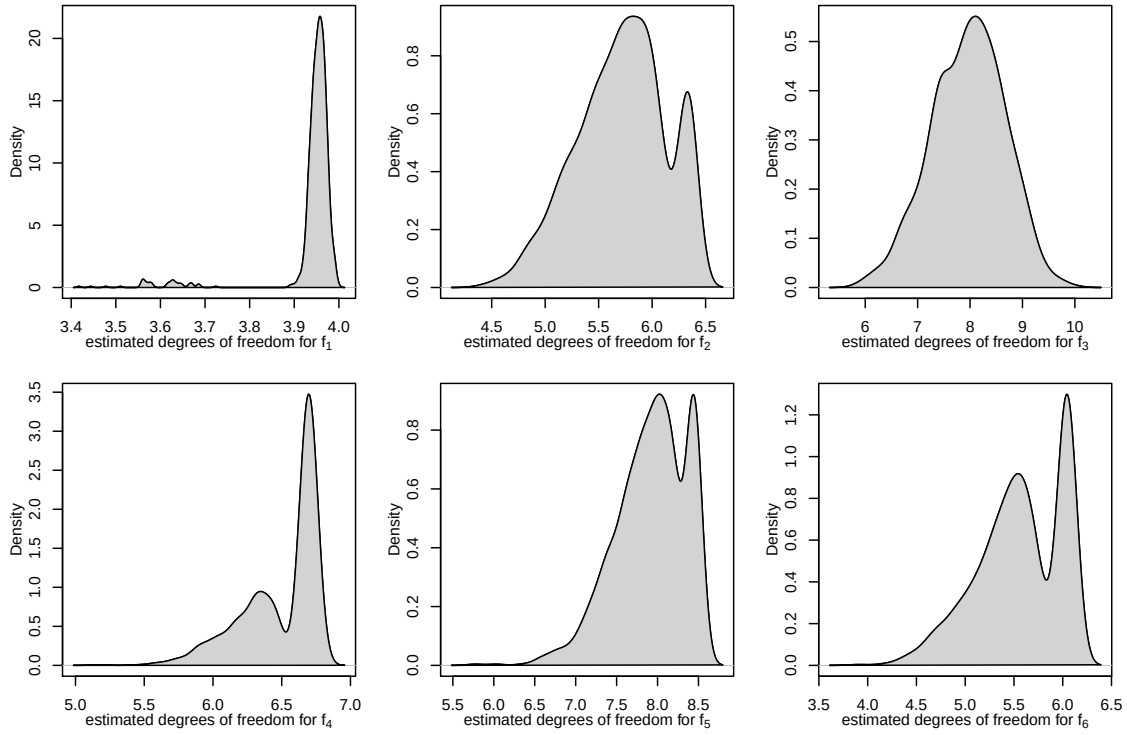

Figure 5: Density plot of the degrees of freedom of the estimations for the $R = 1000$ simulated samples. For each estimation the same optimal λ w.r.t. $MISE$ (see Table 2) is used.

The selection of the optimal smoothing parameter λ is mostly guided by selection criteria, but these often depend on the hat matrix or functions of its elements (e.g. AIC). This was not the case in the simulation since λ was selected by minimizing the $MISE$. Figure 6 shows the dependence of the trace of the hat matrix of the estimated model from the selected smoothing parameter. The vertical line corresponds to the optimal λ

function	f_1	f_2	f_3	f_4	f_5	f_6
min	3.4	4.4	5.8	5.2	5.8	3.9
max	4.0	6.4	10.0	6.8	8.5	6.1
mean	3.9	5.7	8.0	6.5	7.9	5.6
sd	0.08	0.42	0.70	0.27	0.43	0.43

Table 2: Sample minimum, maximum, mean and standard deviation of the traces of the hat matrices for the different DGPs from the $R = 1000$ samples.

according to $MISE$.

For increasing λ the degrees of freedom of the estimation decreases as was to be expected since the fitted curve converges to a straight line (cf. Eilers & Marx, 1996). Note that if the plot was drawn further to $\lambda \to \infty$, the degrees of freedom would converge to 2 corresponding to a straight line with a non-zero slope (cf. Section 2.1).

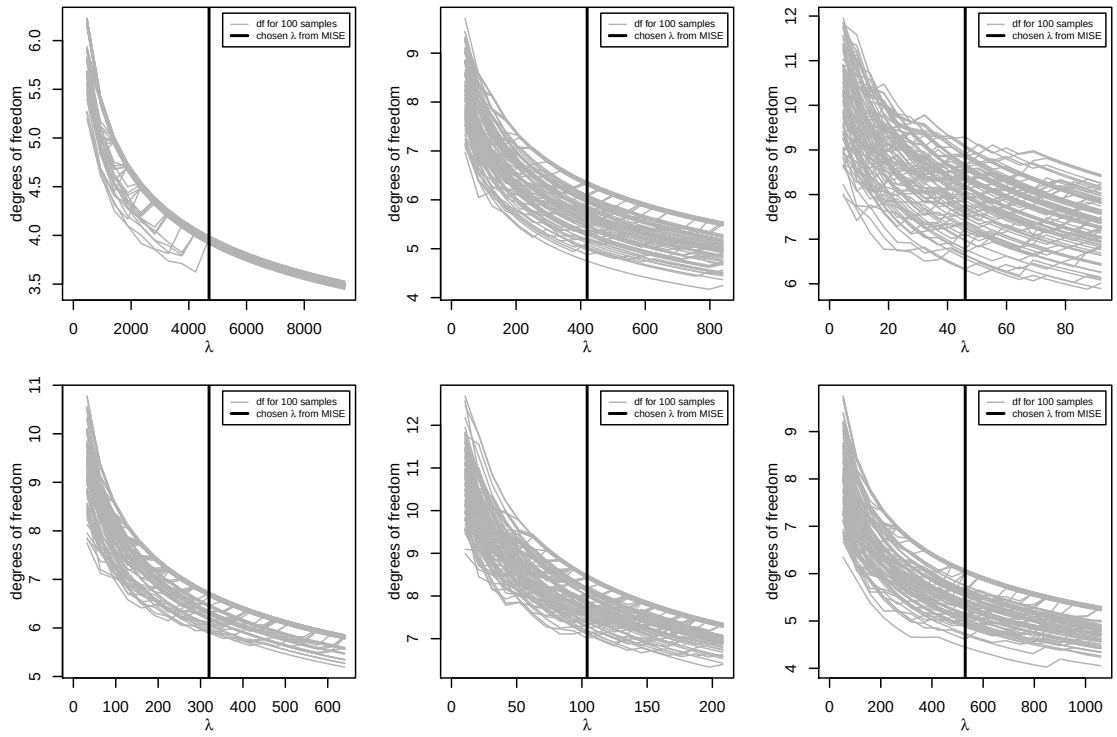

Figure 6: Degrees of freedom of the estimations depending on the smoothing parameter λ for 100 of the $R = 1000$ samples.

The vertical line is the optimal λ with respect to $MISE$.

4 Empirical example

In the following, the well-known data set containing LIDAR data is analyzed.

The LIDAR (light detection and ranging) data set which has been examined especially with nonparametric techniques (e.g. Ruppert et al., 2003, Ruppert & Carroll, 2000) can be found on the homepage of the book of Ruppert et al. (2003, `http://stat.tamu.edu/~carroll/semiregbook/`). The data set includes information on $n = 221$ observations from a LIDAR experiment. The dependent variable y is *logratio*, the logarithm of the ratio of received light from two laser sources, which is explained by the covariate $x = range$, the distance the light traveled before being reflected back to its source.

Since a larger distance before the reflection of the light is assumed to lead to less received light, the relationship between *range* and *logratio* is modeled as a monotone decreasing spline function. Figure 7 shows the scatter plot for the data set and also contains the estimated regression curve for the chosen λ.

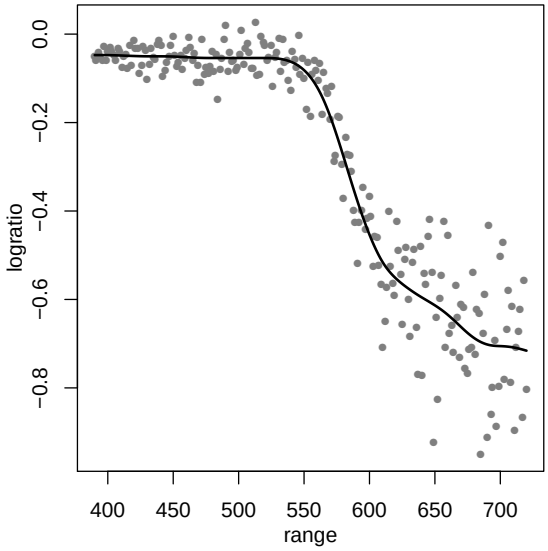

Figure 7: Scatter plot for the LIDAR data set with covariate *range* and dependent variable *logratio* and fitted regression curve from the monotonicity restricted P-spline estimation with optimal λ chosen with respect to GCV.

For this example, cubic splines are applied and the fitted regression curve is restricted to be monotone decreasing what is implemented via the constraint $\alpha_j \geq \alpha_{j+1}$ (cf. the analog Equation (4)). The number of inner knots for the spline specification is

$m = 35$ according to Equation (6). The smoothing parameter λ is chosen with respect to $GCV(\lambda) = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{(1-\frac{1}{n}\text{tr}(\mathbf{H}))^2}$ (e.g. Ruppert et al., 2003, Section 5.3) with $\mathbf{H} = \mathbf{H}_{\lambda,\text{ constr}}$ being the respective hat matrix as in Equation (10).

Figure 8 shows the GCV and $\text{tr}(\mathbf{H}_{\lambda,\text{ constr}})$ depending on λ. For $\lambda = 7.9$, GCV is minimized and the corresponding $\text{tr}(\mathbf{H}_{\lambda,\text{ constr}})$ equals 9.4. Hence, for the given example the degrees of freedom of the estimation are 9.4.

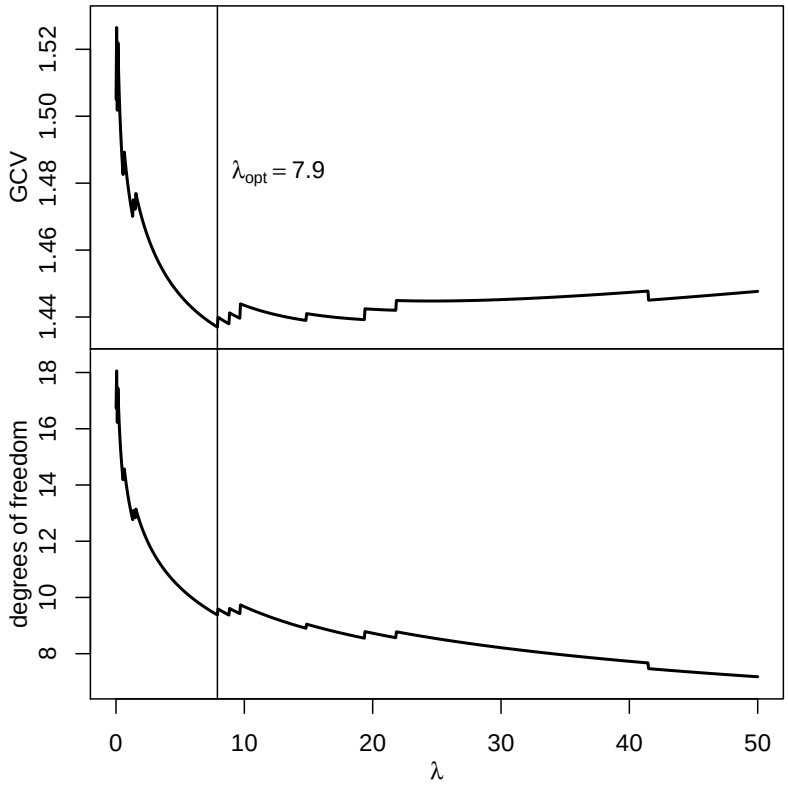

Figure 8: Smoothing parameter λ vs. GCV (top) and λ vs. degrees of freedom (bottom). Optimal λ with respect to GCV is 7.9 and corresponding degrees of freedom are 9.4.

At some values of λ, small jumps in the trace of the estimated hat matrix and hence also in GCV occur. This is due to the matrix $\mathbf{C}_R$ in the hat matrix formula. Remember, the matrix $\mathbf{C}_R$ contains those rows of the constraint matrix $\mathbf{C}$ for which $\mathbf{C}\,\hat{\boldsymbol{\alpha}} = \mathbf{0}$ holds. With increasing λ, the penalty term forces the estimated regression curve to a straight line and the monotonicity constraint becomes less important, hence the number of rows in $\mathbf{C}_R$ decreases. Whenever the number of rows of $\mathbf{C}_R$ changes, a little jump in Figure 8 occurs.

5 Conclusion

A formula for calculating the hat matrix for an estimated regression model using a monotonicity constrained and penalized spline is derived. The trace of this hat matrix can be interpreted as the equivalent to the degrees of freedom of the estimated model. It can be used for example for model selection when criteria such as the Akaike information criterion or generalized cross validation are applied. In the context of penalized spline estimation, it can be applied for the selection of the optimal smoothing parameter according to one of those criteria. For non-penalized as well as for penalized estimations, the order of the spline, the number of knots and/or the position of the knots can, if not fixed in advance, be chosen analogously using the same selection criteria.

In an extensive Monte Carlo study, the hat matrices for six different DGPs and $R = 1000$ samples are obtained. The results suggest that the degrees of freedom from the estimations fit the DGP functions appropriately. In an empirical example, the LIDAR data set is analyzed and the degrees of freedom are found to be 9.4.

The presented hat matrix also works for the general case of an inequality constrained estimation with restriction $\mathbf{C}\boldsymbol{\beta} \geq \mathbf{0}$. This general case includes semiparametric models using monotonicity constrained splines for the nonparametric part of the model. In this case, the matrices $\mathbf{D}$ and $\mathbf{C}$ have to be filled with zeros up to the appropriate dimension and the remaining parts are just like in the case with only a single spline component.

Overall, this work helps practitioners to calculate the hat matrix of an estimated monotonicity constrained spline model and hence the degrees of freedom of an estimation, what is an important task for example for model selection purposes including the search for an optimal smoothing parameter in the penalized case.

References

Bollaerts, K., Eilers, P. H. C., & Aerts, M. (2006). Quantile regression with monotonicity restrictions using P-splines and the L_1-norm. *Statistical Modelling*, 6(3), 189–207.

de Boor, C. (2001). *A Practical Guide to Splines*, volume 27 of *Applied Mathematical Sciences*. Berlin: Springer, revised edition.

Dierckx, P. (1993). *Curve and Surface Fitting with Splines*. Numerical Mathematics and Scientific Computation. Oxford: Oxford University Press.

Eilers, P. H. C. & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2), 89–121.

Kagerer, K. (2013). *A short introduction to splines in least squares regression analysis*. University of Regensburg Working Papers in Business, Economics and Management Information Systems 472, University of Regensburg, Department of Economics.

Paula, G. A. (1993). Assessing local influence in restricted regression-models. *Computational Statistics & Data Analysis*, 16(1), 63–79.

Paula, G. A. (1999). Leverage in inequality-constrained regression models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 48(4), 529–538.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, 11(4), 735–757.

Ruppert, D. & Carroll, R. J. (2000). Spatially-adaptive penalties for spline fitting. *Australian & New Zealand Journal of Statistics*, 42(2), 205–223.

Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression*, volume 12 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press.

Schumaker, L. L. (1981). *Spline Functions Basic Theory*. Pure and applied mathematics. John Wiley & Sons.

Wood, S. N. (2014). *mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation*. R package version 1.7-29.