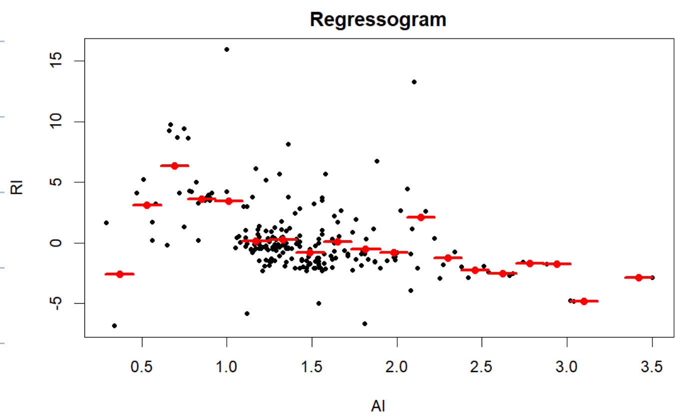
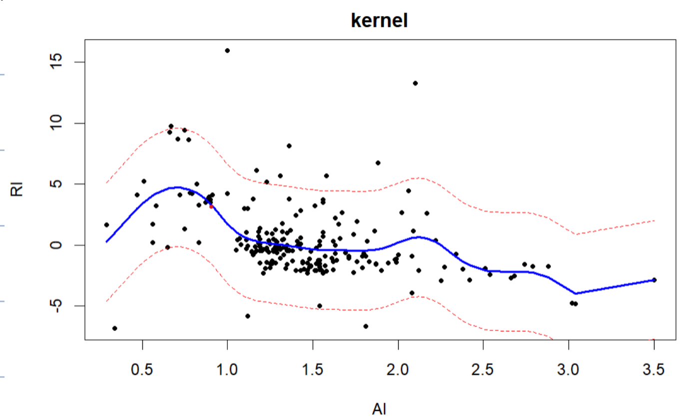
3. Get the data on fragments of glass collected in forensic work from the book website. Let Y be refractive index and let x be aluminium content (the fourth variable). Perform a nonparametric regression to fit the model $Y = r(x) + \epsilon$. Use the following estimators: (i) regressogram, (ii) kernel, (iii) local linear, (iv) spline. In each case, use cross-validation to choose the amount of smoothing. Estimate the variance. Construct 95 percent confidence bands for your estimates. Pick a few values of x and, for each value, plot the effective kernel for each smoothing method. Visually compare the effective kernels.

(i) Regresogram. h chosen by cross-validation

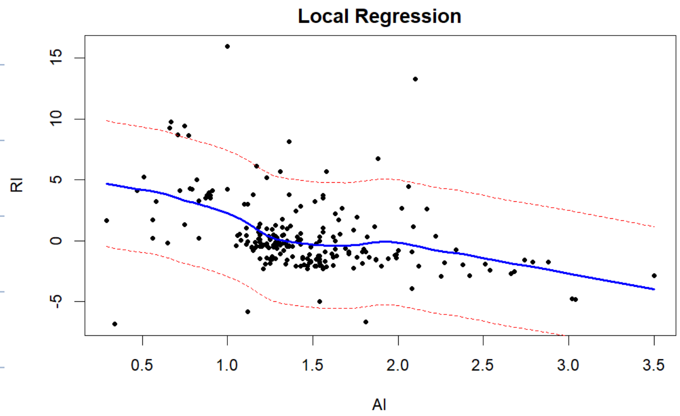


(ii) Kernel:



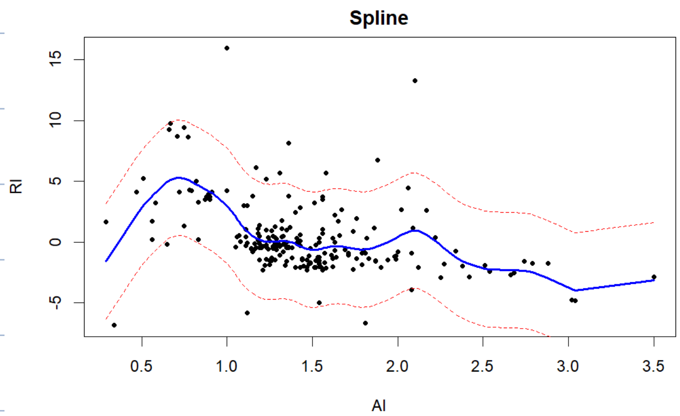
$\widehat{Var} = 6.15$

(iii) Local Linear Regression



$\widehat{Var} = 6.93$

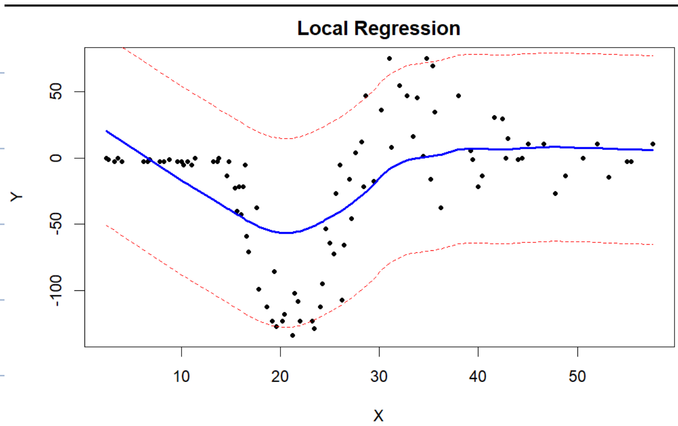
(iv) Spline



$\widehat{Var} = 5.87$.

All the regressions selected by CV

4. Get the motorcycle data from the book website. The covariate is time (in milliseconds) and the response is acceleration at time of impact. Use cross-validation to fit a smooth curve using local linear regression.



selected by cross-validation

$\widehat{Var} = 1319$.

7. Prove Theorem 5.60.

Local Linear Smoothing

5.60 Theorem. *When $p = 1$, $\widehat{r}_n(x) = \sum_{i=1}^n \ell_i(x) Y_i$ where*

$$\ell_i(x) = \frac{b_i(x)}{\sum_{j=1}^n b_j(x)},$$

$$b_i(x) = K\left(\frac{x_i - x}{h}\right)(S_{n,2}(x) - (x_i - x)S_{n,1}(x)) \qquad (5.61)$$

and

$$S_{n,j}(x) = \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)(x_i - x)^j, \quad j = 1, 2.$$

$$X = \begin{bmatrix} 1 & x_1 - x \\ 1 & x_2 - x \\ \vdots & \vdots \\ 1 & x_n - x \end{bmatrix}$$

$$W = \operatorname{diag}[w_1(x), w_2(x) \cdots w_n(x)]$$

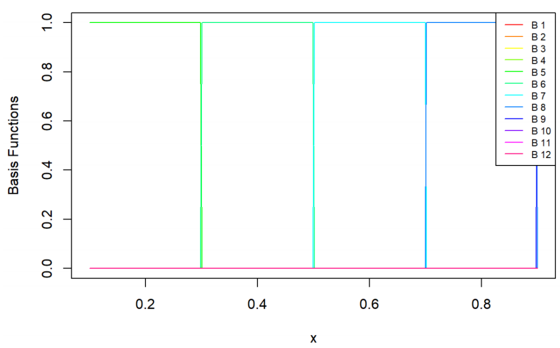
$$w_i = K\left(\frac{x_i - x}{h}\right), \quad i = 1, 2, \cdots n.$$

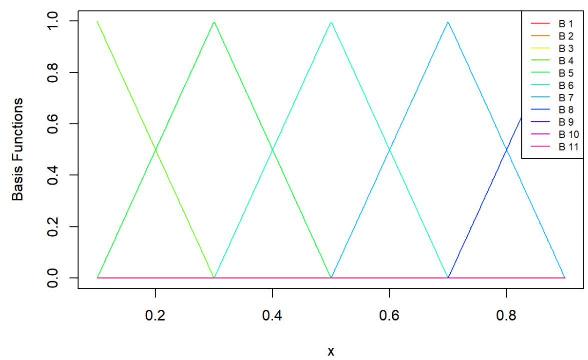
$$\Rightarrow \hat{a} = \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \end{bmatrix} = \frac{1}{S_{n,0}(x) S_{n,2}(x) - S_{n,1}^2(x)} \begin{bmatrix} S_{n,2}(x) & -S_{n,1}(x) \\ -S_{n,1}(x) & S_{n,0}(x) \end{bmatrix} \begin{bmatrix} \sum w_i Y_i \\ \sum w_i (x_i - x) Y_i \end{bmatrix}$$

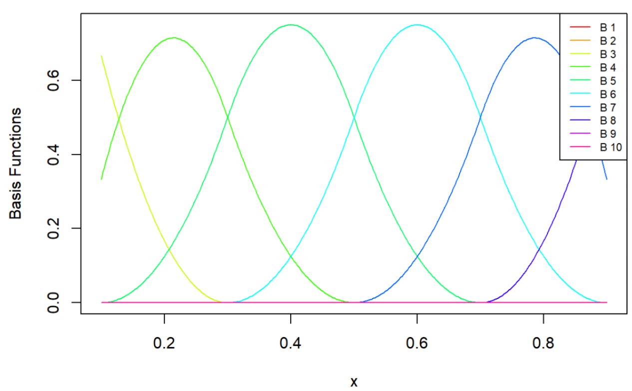
$$\hat{r}_n(x) = \frac{1}{S_{n,0}(x) S_{n,2}(x) - S_{n,1}^2(x)} \sum_{i=1}^n (S_{n,2}(x) - (x_i - x) S_{n,1}(x)) w_i Y_i$$

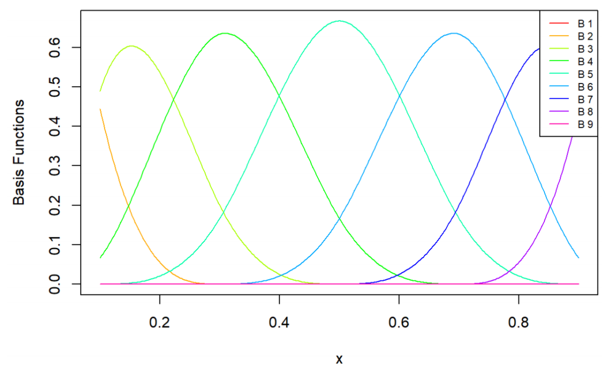
$$= \sum_{i=1}^n \frac{b_i(x)}{\sum_{j=1}^n b_j(x)} Y_i = \sum_{i=1}^n \ell_i(x) Y_i$$

10. Using five equally spaced knots on (0,1), construct a B-spline basis of order M for $M = 1, \ldots, 5$. Plot the basis functions.



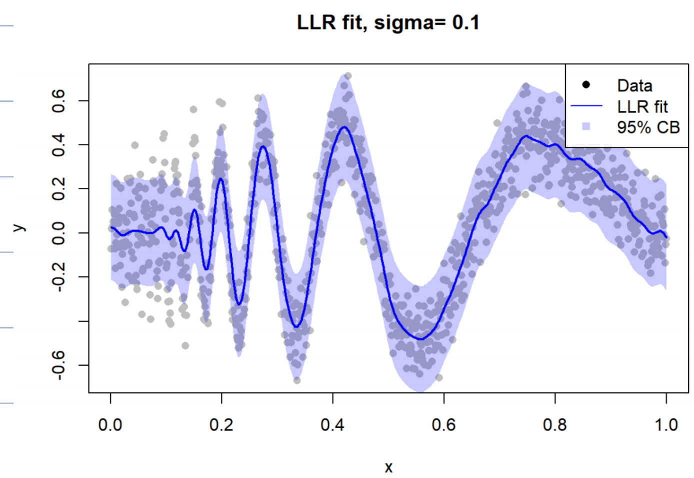


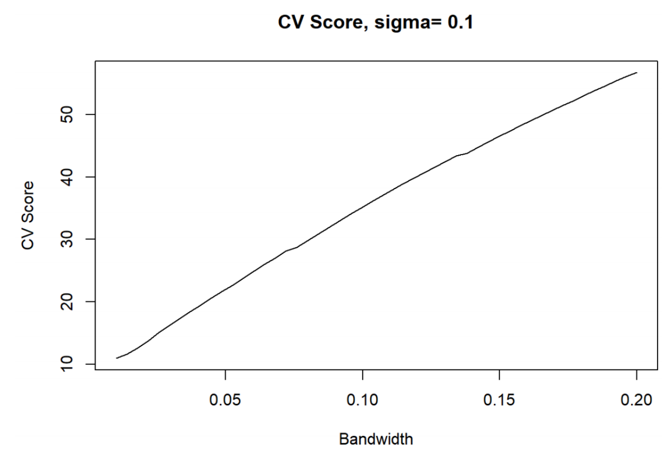




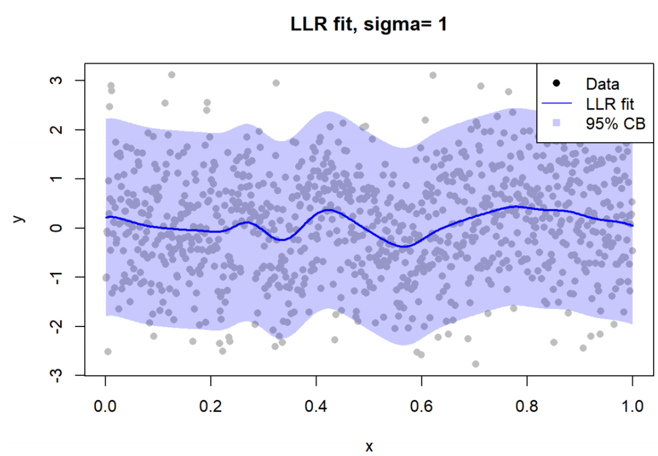
12. Recall the Doppler function defined in Example 5.63. Generate 1000 observations from the model $Y_i = r(x_i) + \sigma\epsilon_i$ where $x_i = i/n$ and $\epsilon_i \sim N(0,1)$. Make three data sets corresponding to $\sigma = .1$, $\sigma = 1$ and $\sigma = 3$. Plot the data. Estimate the function using local linear regression. Plot the cross-validation score versus the bandwidth. Plot the fitted function. Find and plot a 95 percent confidence band.

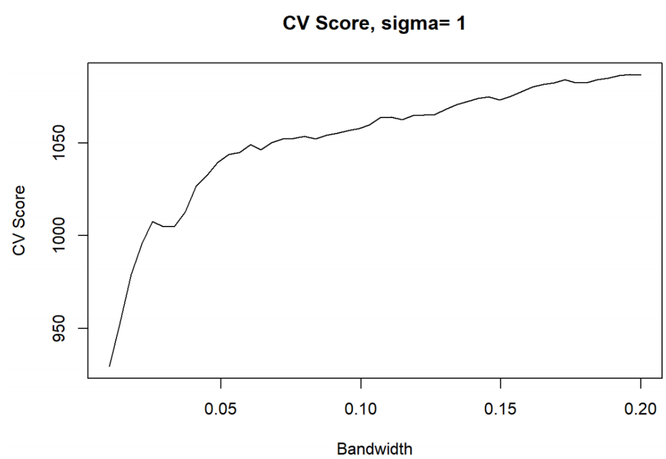
$\sigma = 0.1$



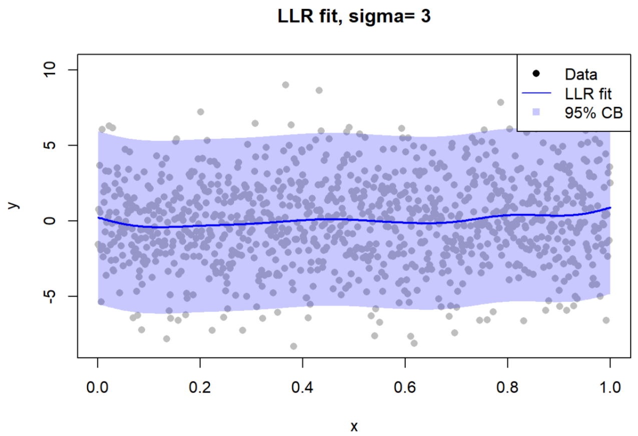


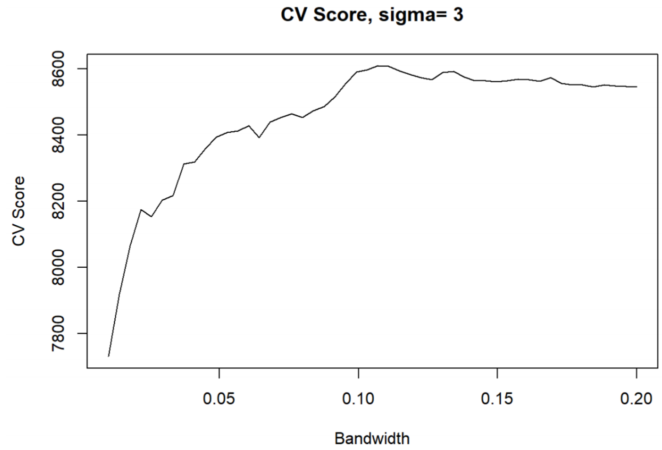
$\sigma = 1$





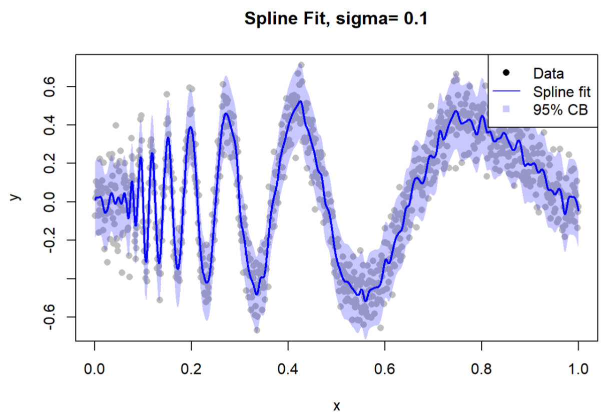
$\sigma = 3$

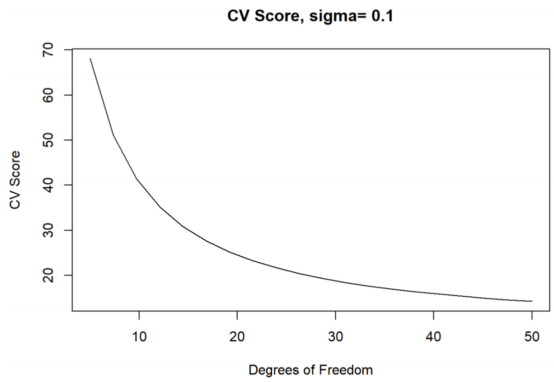




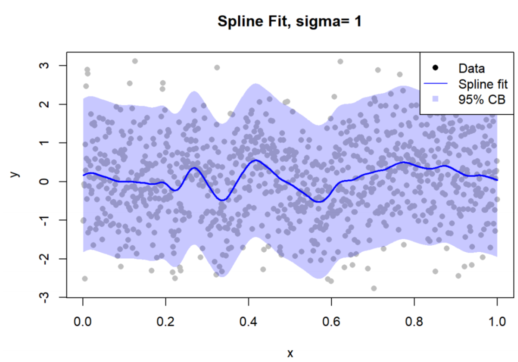
13. Repeat the previous question but use smoothing splines.

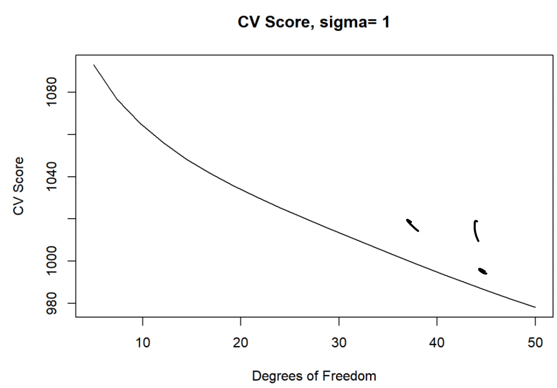
$\sigma = 0.1$



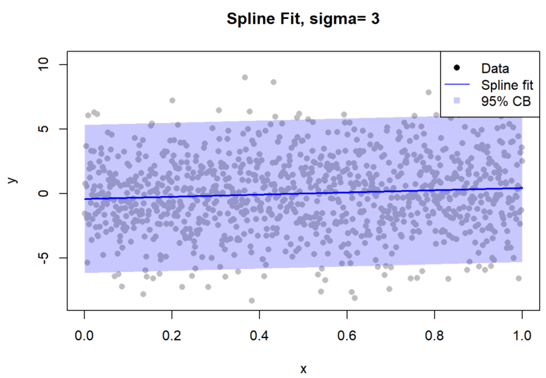


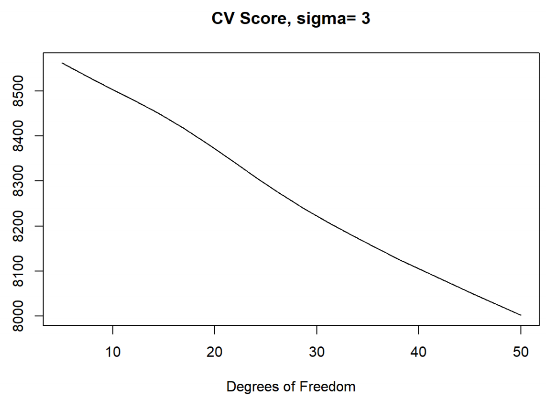
$\sigma = 1$





$\sigma = 3$


Spline Fit, sigma= 3
Data
Spline fit
95% CB
y
x
0.0
0.2
0.4
0.6
0.8
1.0
-5
0
5
10


CV Score, sigma= 3
8000
8100
8200
8300
8400
8500
10
20
30
40
50
Degrees of Freedom