ASM Practice

Local Linear Regression

Maria Gkotsopoulou & Ricard Monge Calvo & Amalia Vradi

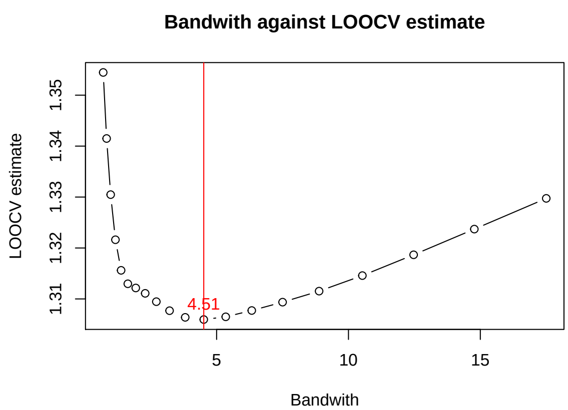
20/11/2019

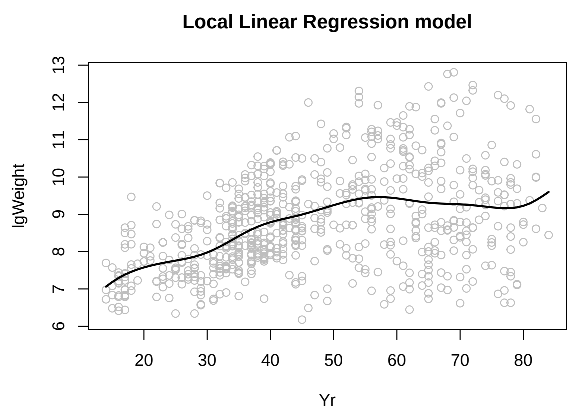
The aim of this project is to compute the conditional variance $\sigma^2(x)$ of the variable $lgWeight = log(Weight)$ of the *aircraft* dataset (in *sm* package) given the year, *Yr*, variable.

Estimation using *locpolreg* function

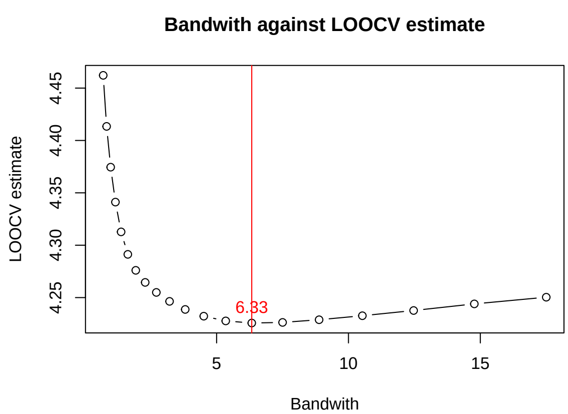
We load our local function *locpolreg* along with the *bandwith_selection* script which contains different functions for bandwidth selection. We will choose the bandwidth hyper-parameter by LOOCV. We use the appropriate function in the *bandiwth_selection* script to get the LOOCV and GCV estimates (which uses the *locpolreg* function). Regarding the Kernel choice, we decide to use the *normal* kernel.

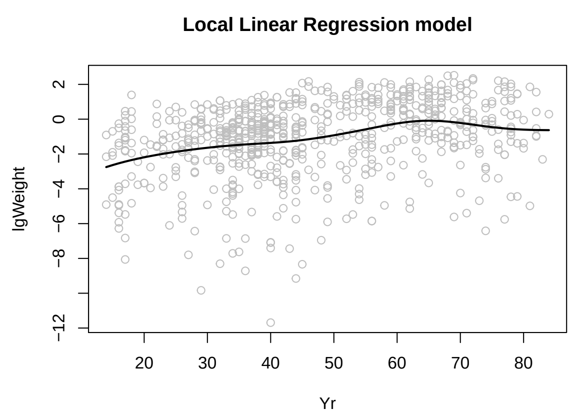
The first thing we need to do is to compute the local linear regression for the predicted variable *lgWeight* depending on *Yr*. We choose the bandwidth as the one that minimizes the LOOCV estimate. After choosing the bandwidth, we build the local linear regression model using the *locpolreg* function and compute the residual values.





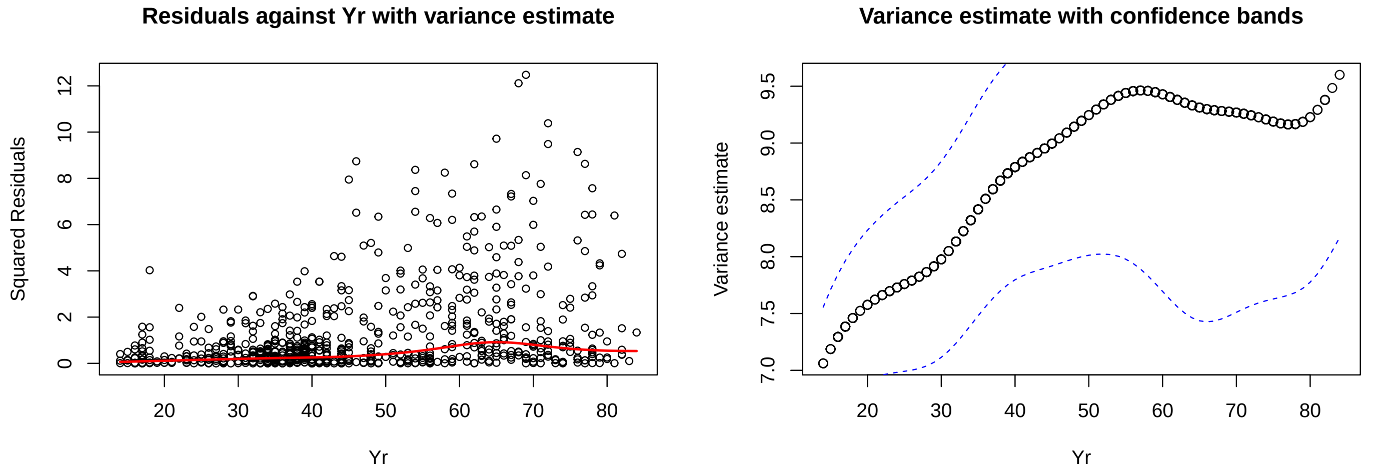
Having obtained the residual values $\hat{\epsilon}_i = y_i - \hat{m}(x_i)$ we compute their logarithm $z_i = \log \hat{\epsilon}_i^2$. We need to build a new model for z_i against x_i.





Finally, the conditional variance $\hat{\sigma}^2(x) = \exp \hat{q}(x)$ where $\hat{q}(x)$ is the estimate of the previous model.

To sum up the results, we plot the value of ϵ_i^2 against x_i superimposing the values of $\hat{\sigma}^2(x)$, and also the values of $\hat{m}(x)$ with the bands $\hat{m}(x) \pm 1.96\hat{\sigma}(x)$.



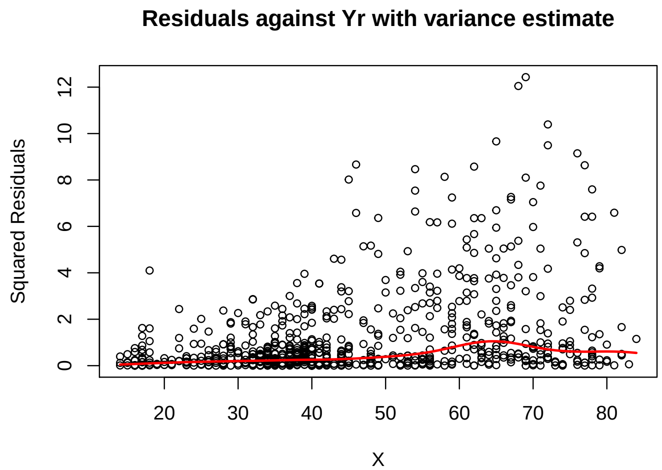
Estimation using *sm.regression* function

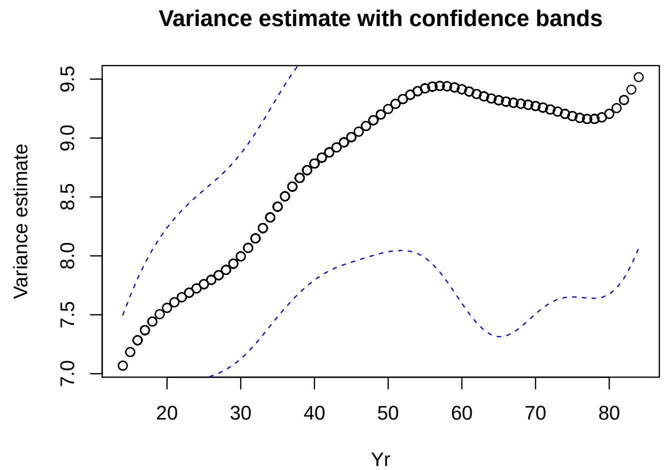
We use the function *dpill* in the *KernSmooth* package in order to compute the *Plug-in* bandwidth parameter. Afterwards, we use this bandwidth with the *sm.regression* function in the *sm* package to compute the local linear regression models, both $\hat{m}(x)$ and $\hat{q}(x)$. Regarding the Kernel choice, we decide to use the *normal* kernel.

After choosing the bandwidth, we build the local linear regression model using the *sm.regression* function and compute the residual values $\hat{\epsilon}_i = y_i - \hat{m}(x_i)$ and their logarithm $z_i = \log \hat{\epsilon}_i^2$. We then build a new model for z_i against x_i.

Finally, the conditional variance $\hat{\sigma}^2(x) = \exp \hat{q}(x)$ where $\hat{q}(x)$ is the estimate of the previous model.

To sum up the results, we plot the value of ϵ_i^2 against x_i superimposing the values of $\hat{\sigma}^2(x)$, and also the values of $\hat{m}(x)$ with the bands $\hat{m}(x) \pm 1.96\hat{\sigma}(x)$.



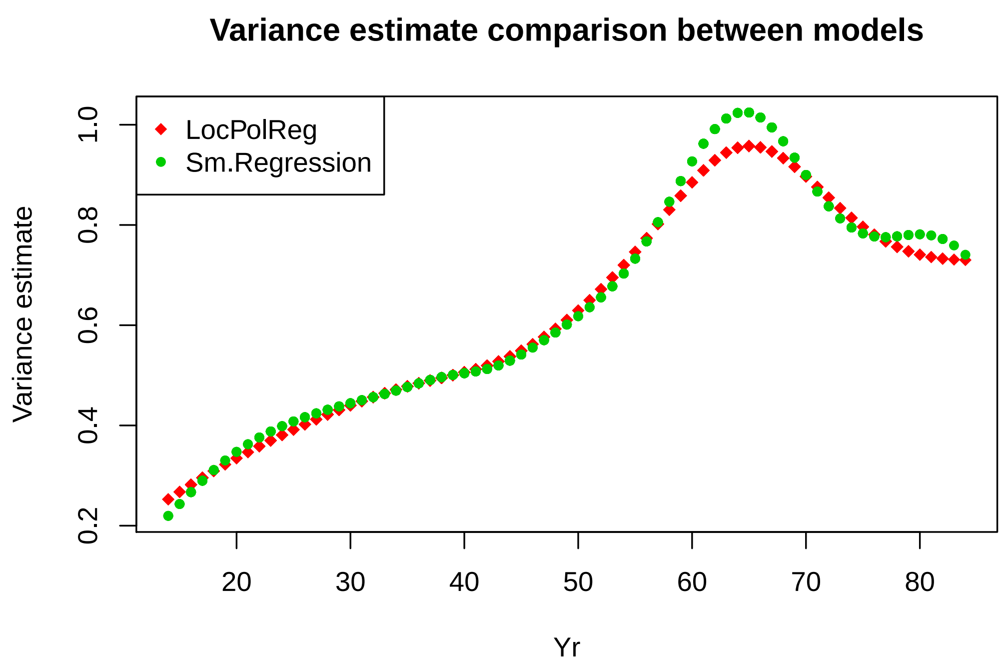


Method comparison

To compare the results of the methods, we present a table with the chosen bandwiths for both models:

Table 1: Model Bandwith comparison

	Bandwith_Model_1	Bandwith_Model_2
Locpolreg	4.513	6.333
Sm.Regression	5.021	4.288



The bandwith values do not differ greatly. In addition, by looking at the comparison plot, we confirm the previous idea. However, we observe a slightly more extreme behaviour for the *Sm.Regression* estimate, due to its lower bandiwth (for model 2).