

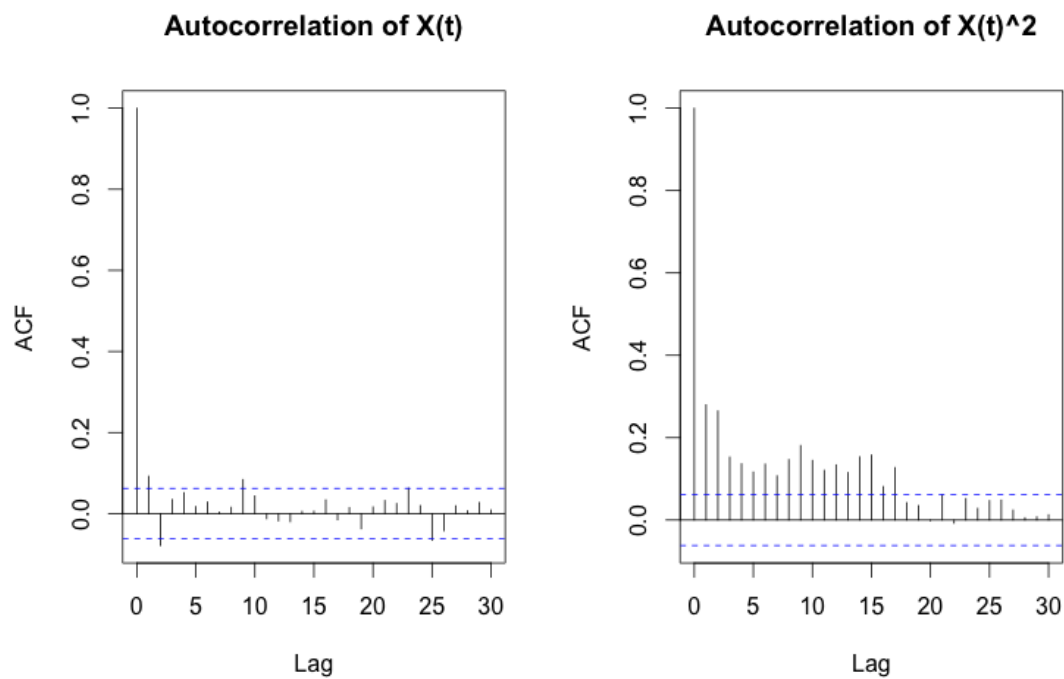
Computational Statistics: Homework 4

Michal Porvaznik
Nemesis key = 8676cc2

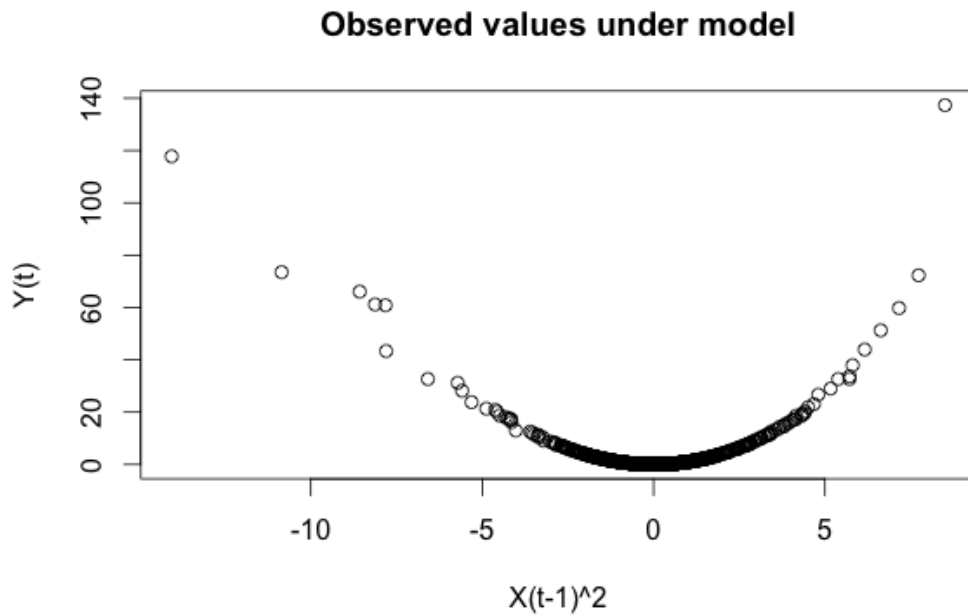
March 24, 2015

Exercise 1

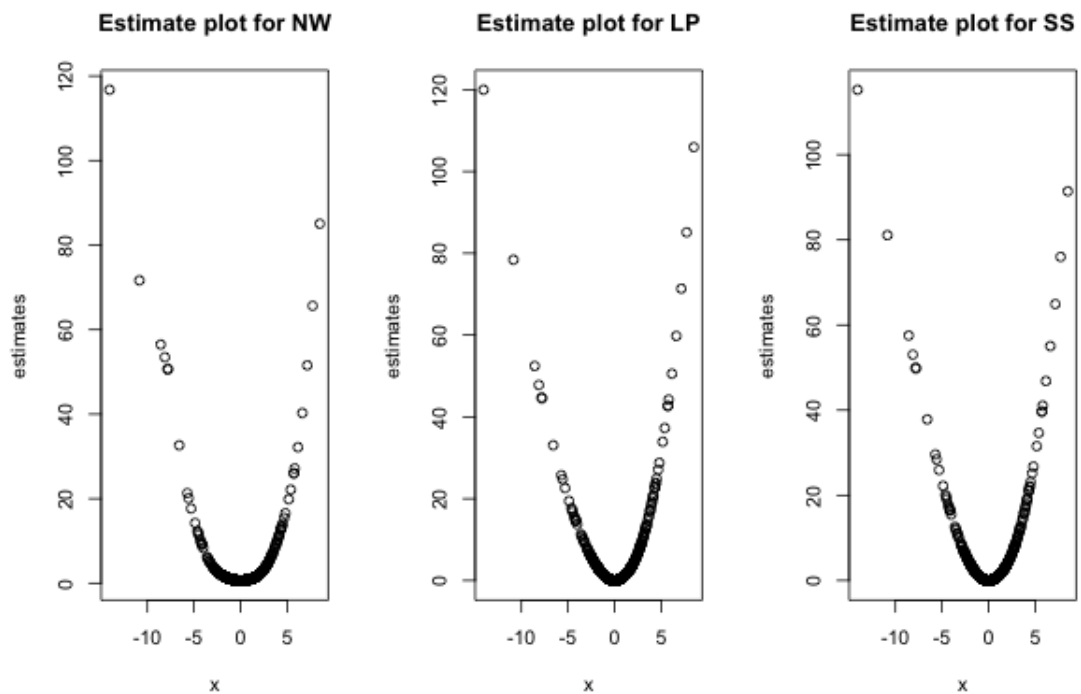
a) Bellow are the autocorrelation plots of the **bmw** dataset as well as of the squared values. For most lags the autocorrelation of the original data is within the 5% acceptance region of for testing the null hypothesis of uncorrelated data. This is however not true for the transformed values.

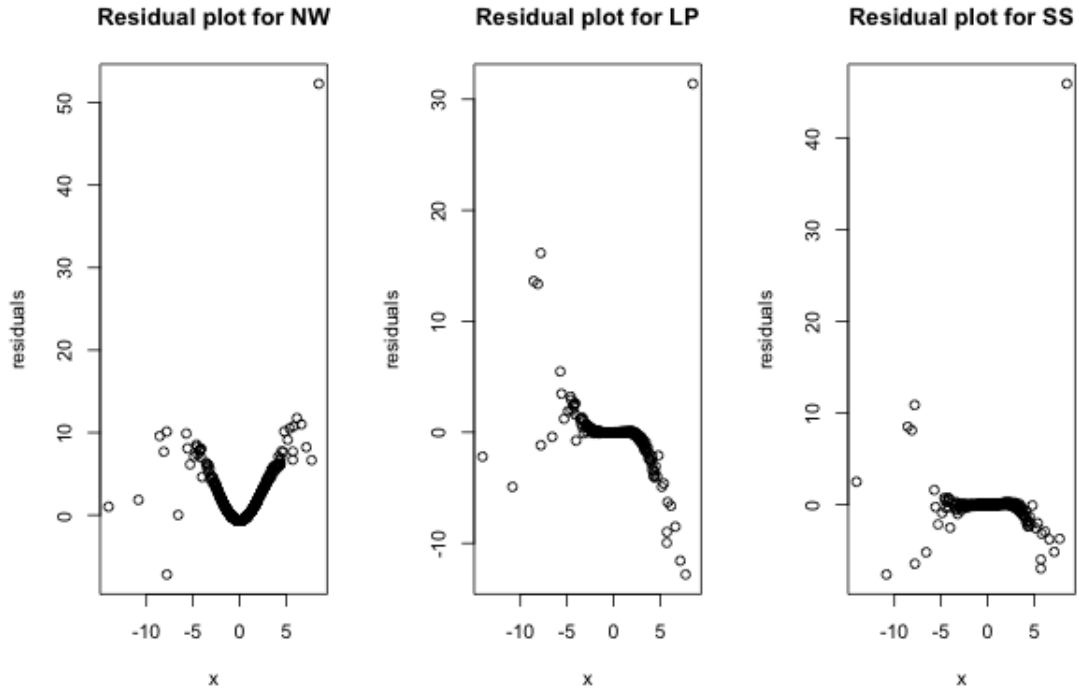


b) Below we see the plot of $Y_t = X_t^2 = v(X_{t-1}) + \eta_t$ for the observed values.

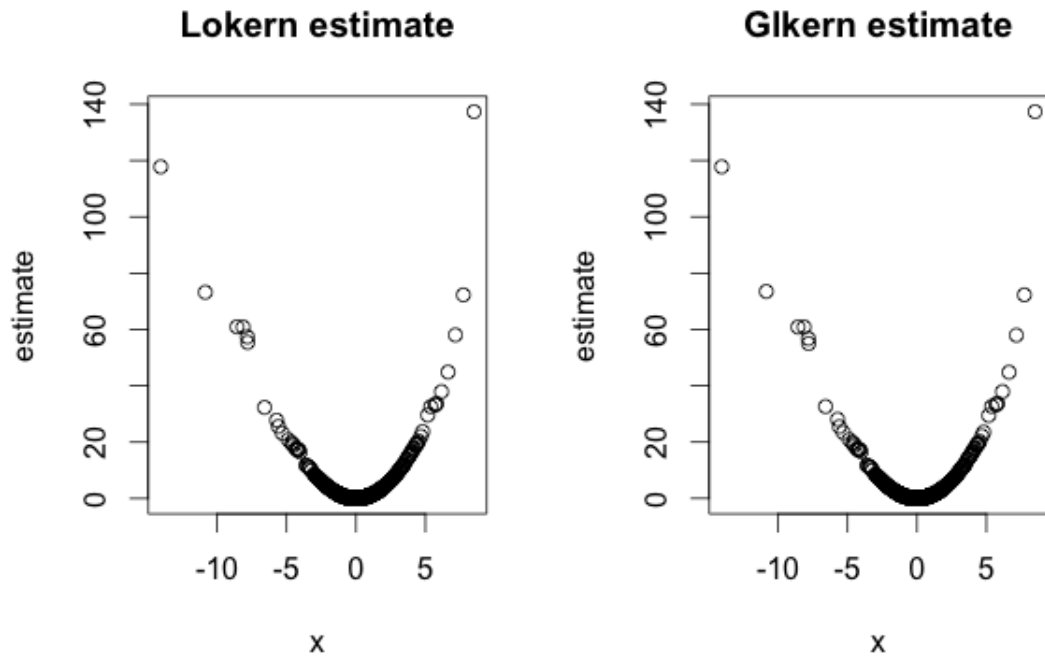


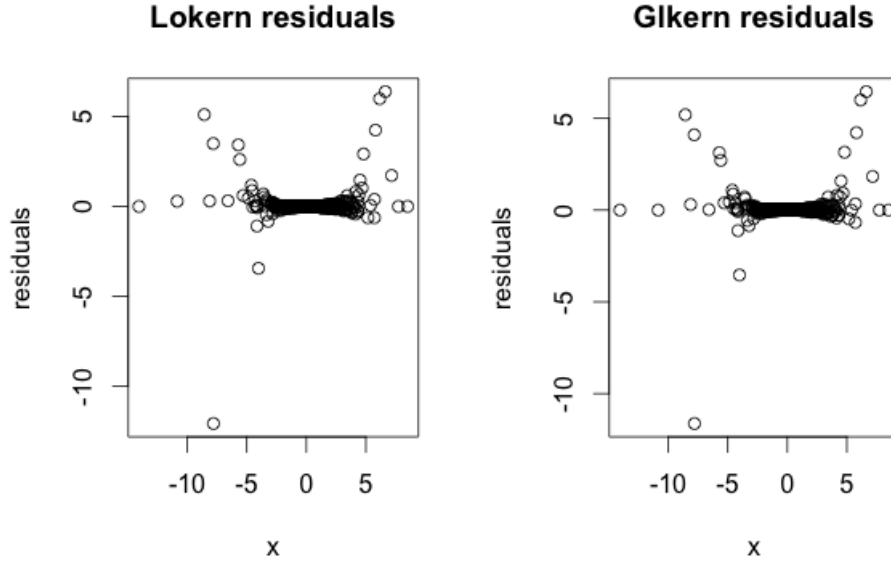
Upon fitting the three models to the data I obtained the estimate and residual plots bellow. These show that the expected noise is not zero and there the errors are heavily correlated. However, I am not sure what to deduce from this. Does this mean that we should not model the data set using regression?





c) Plots below show estimates and residuals of **lokern** and **glkern** fits. In both cases the estimate is good around the center, where we have a lot of observations, but fluctuates closer to boundaries where the observations are sparse. These estimates are so similar that the difference is not even visible on the plots.





Here we see the log of local and global bandwidth plots. As expected, the local bandwidth is much smaller where data is dense, but as we can see from the **rug(x)** the bandwidth is not decreasing strictly with the density. Here again I do not understand why the local bandwidth peaks locally at 0 or why does it even vary. The optimal bandwidth that **lokern** implements is

$$h_{opt}(x) = n^{-1/5} \left(\frac{\sigma_\epsilon^2 \int K^2(z) dz}{(m''(x) \int z^2 K(z) dz)^2} \right) \quad (1)$$

and I don't see which term of this expression influences the value of the bandwidth, since $m''(x)$ is constant in this case.

