

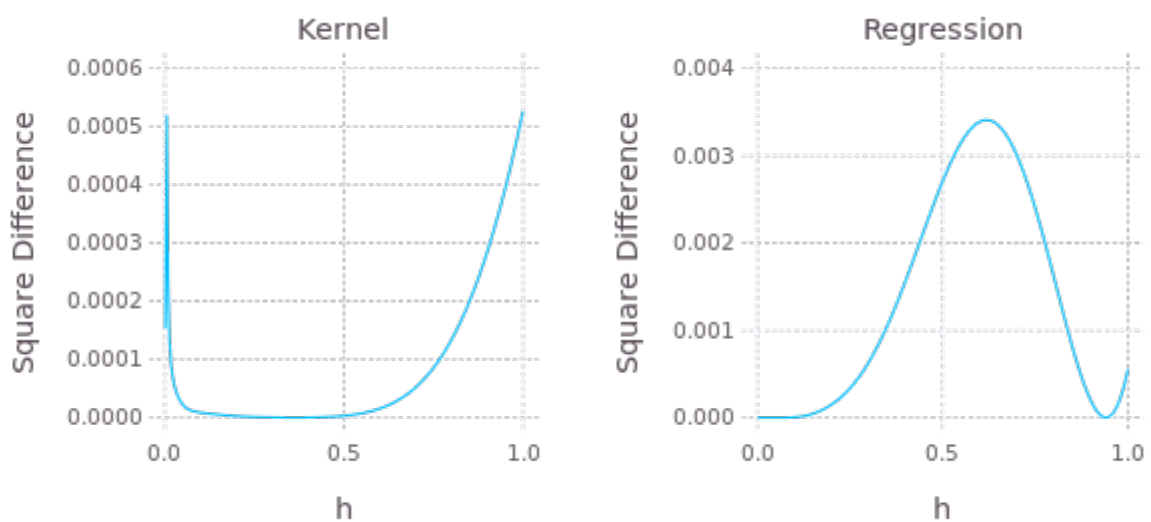
# Econometrics: Homework

Caio Figueiredo

October 27, 2019

## 1 Question 1

In order to save the bootstrap for the next question I decided to estimate this only once, using a high  $n = 10000$ , the results is as follows. The data was created using an Gumbel Distribution, for the regression estimator I used a simple exponential function.

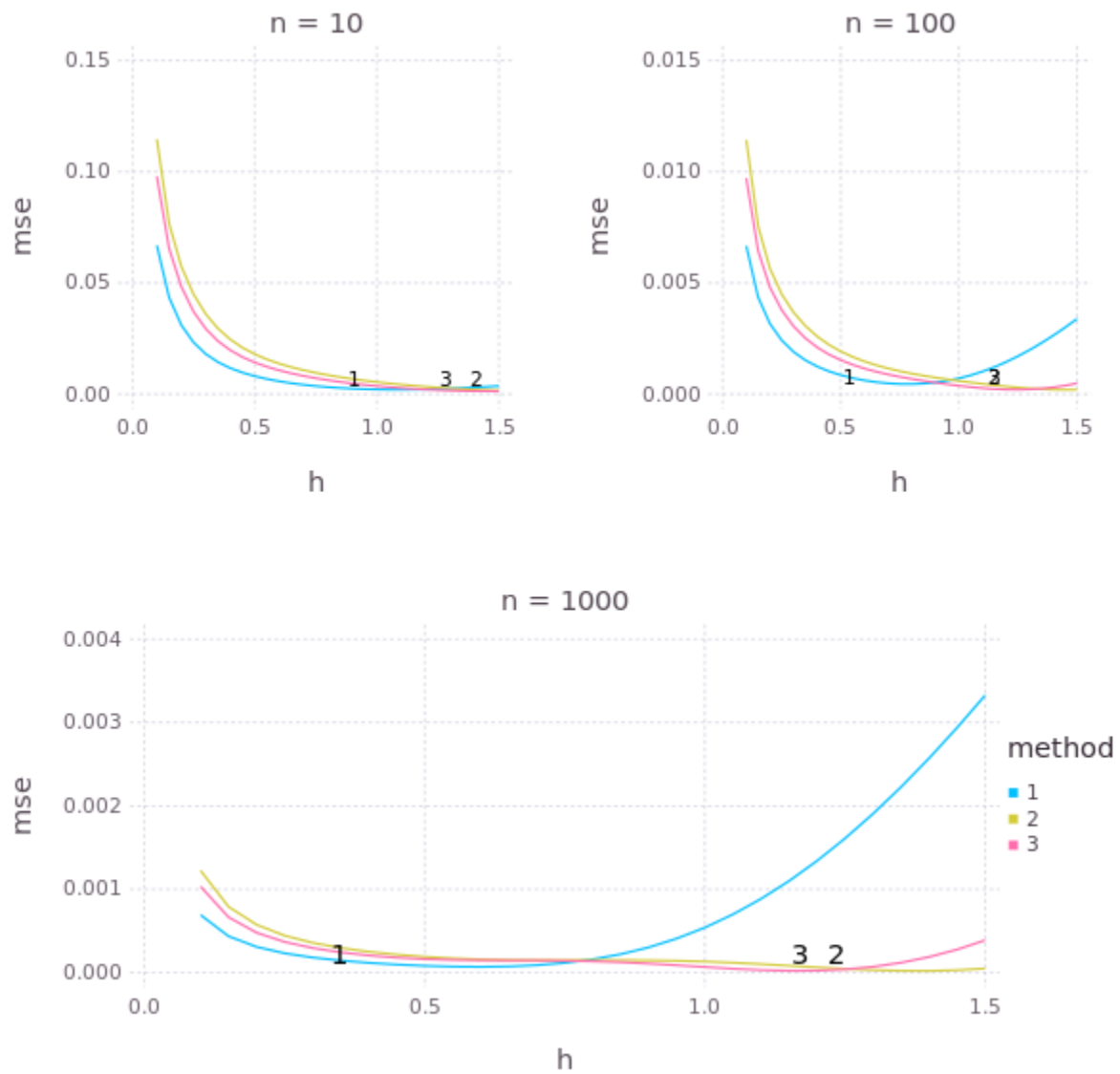


We can see nice that the error of the kernel estimator is very small for low values of  $h$ , but as  $h \rightarrow 0$  the estimator goes awry, as  $h \rightarrow \infty$  the error increases smoothly. The same seems to apply for the regression estimator but we can observe a second valley near 1.

## 2 Question 2

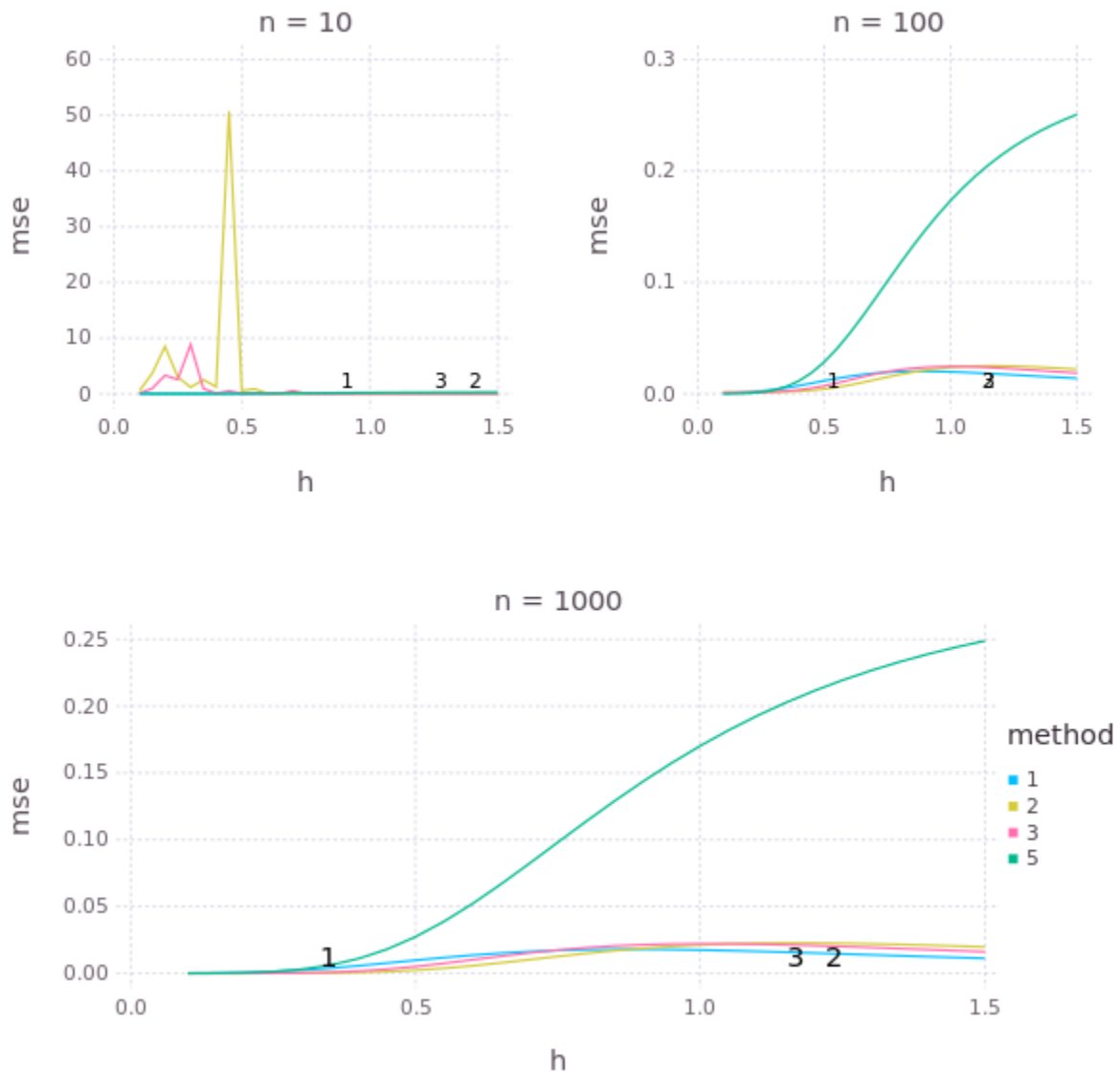
### 2.1 a

In the following graphs the kernel number 4 will be ignore, since it's MSE is much worse than the others that it undermines readability.



The points 1, 2 and 3 are the suggested value for the bandwidth using the cross validation method for kernel number 1, 2 and 3 respectively. The method does not match the value of  $h$  that minimizes the  $MSE$  but overall behave really well and stays in a decently close neighborhood of that point.

## 2.2 b



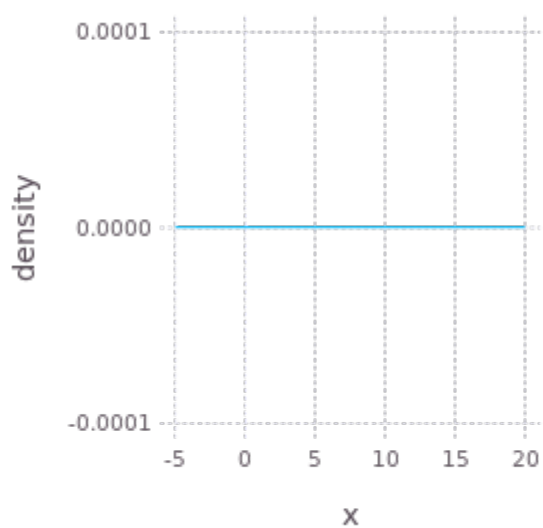
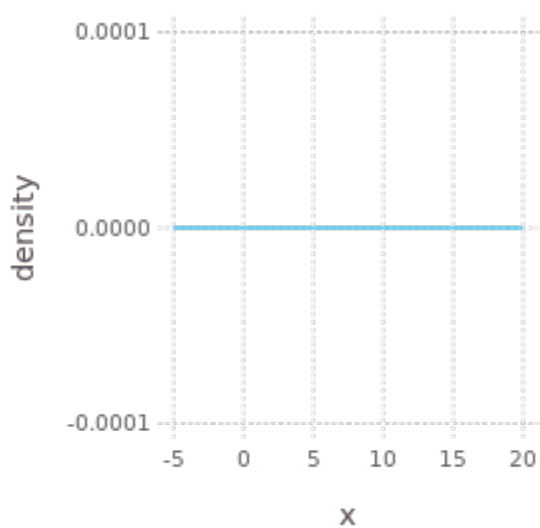
Here we have a different story, ignoring the case for  $n = 10$  where the our estimator does not behave really well the  $MSE$  seems to be a increasing function of the bandwidth, that is the lower the  $MSE$  the better, which is not at all that intuitive.

Of course because of this the values suggested by the cross validation method does not work well if the regression estimator.

This graph also have an additional line (#5) which reference the local linear estimator, the  $MSE$  of this line does not really compare to the other lines since we are estimating different objects but the behavior along  $h$  seems to be the same.

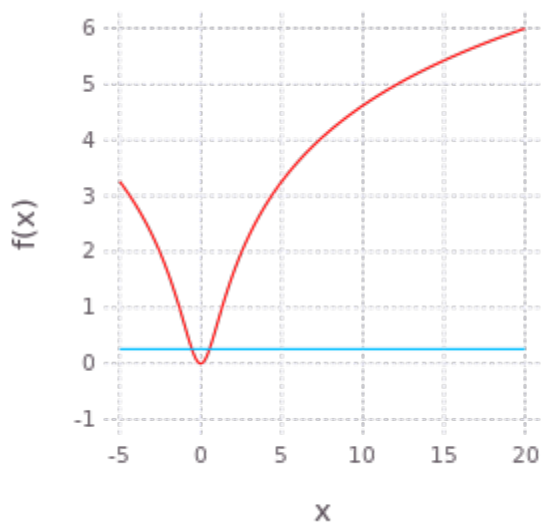
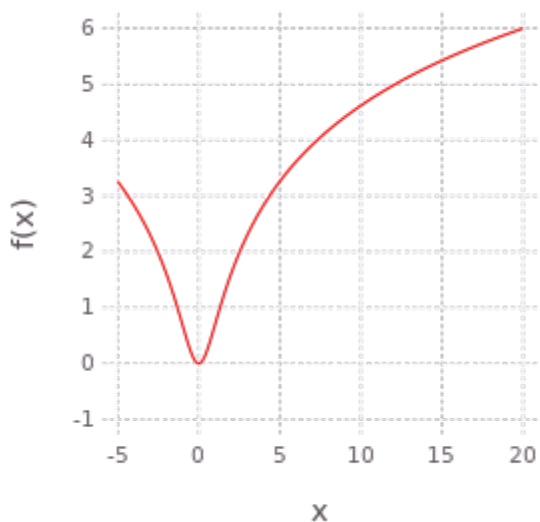
### 3 Question 3

#### 3.1 a

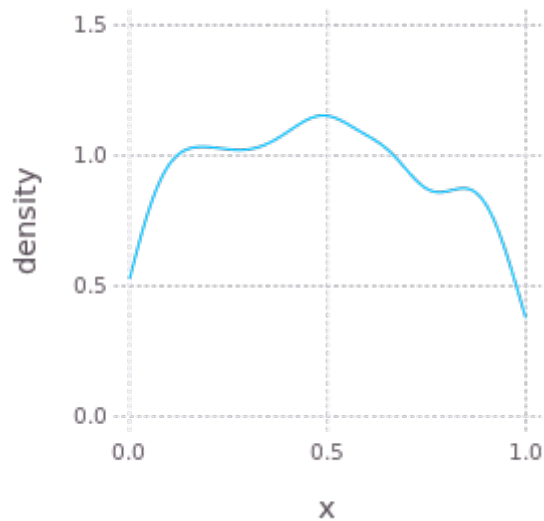


For both the high and the low bandwidth we see that the kernel collapse to zero, this probably happens because  $10^{-6}$  is so low that most intervals have zero measure, and for the  $10^6$  the bandwidth covers all observations and it is therefore impossible to distinguish between them.

#### 3.2 b

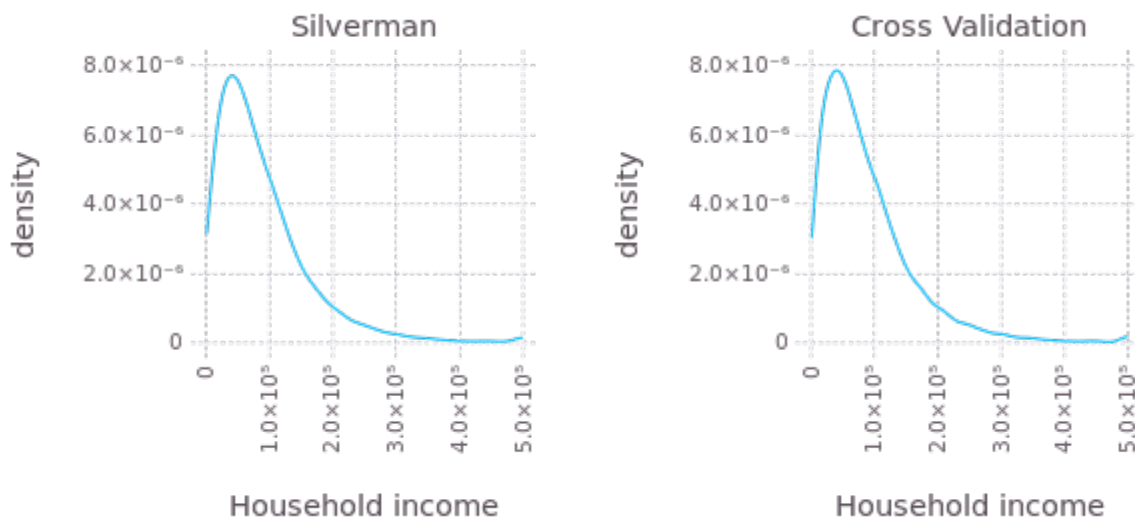


For the regression estimator things don't get much better the very low bandwidth causes most estimators to be *NaN* and therefore the plot is empty, for the very high bandwidth we estimation is no better than a simple average

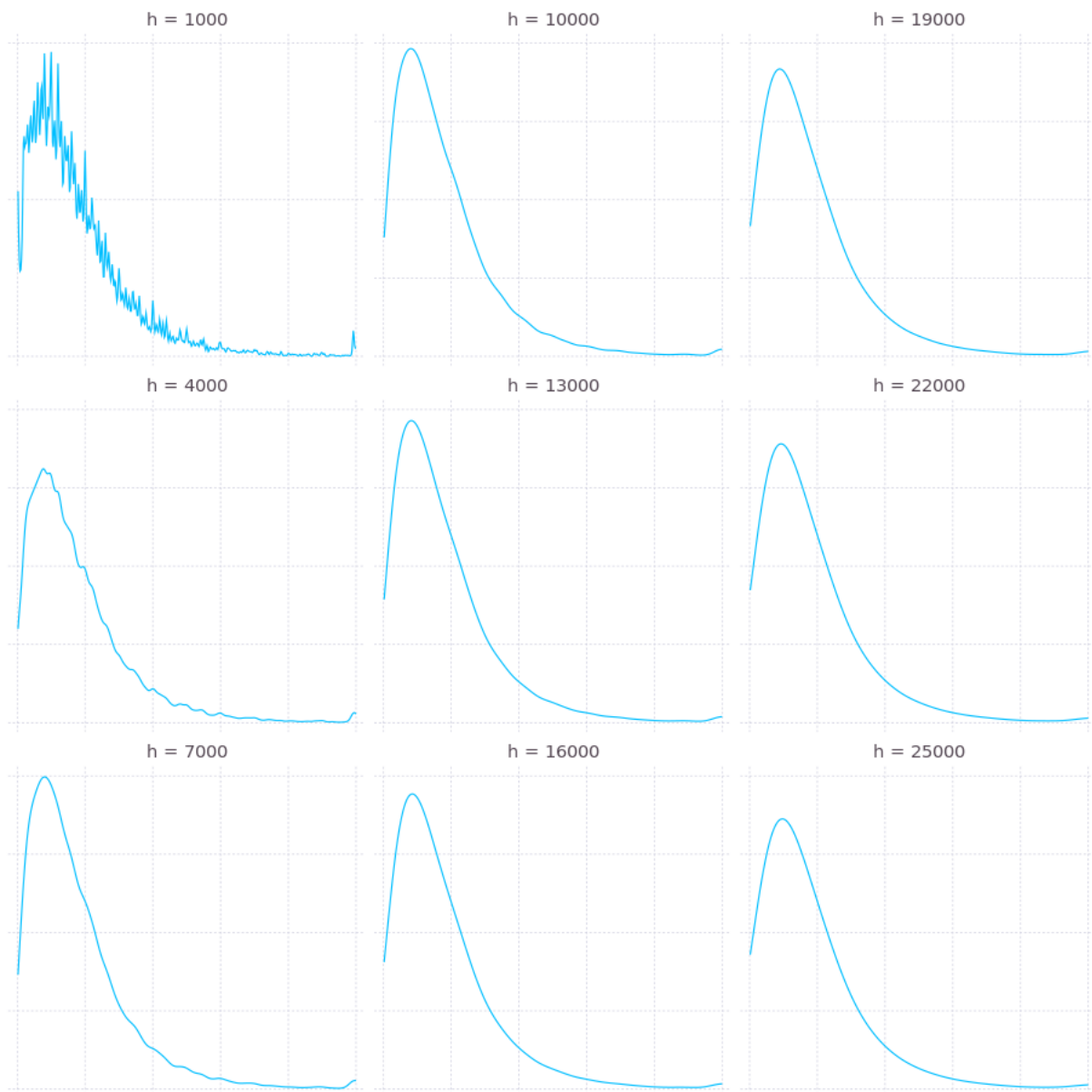


We can observe that the we have a hard time fitting the curvy behavior of the normal kernel to the flat line that is the uniform. This is not however a bandwidth problem, testing with higher are lower bandwidths showed similar results. Moreover we can observe that the problem get worse at the boundaries.

#### 4 Question 4



Visual inspection:



Silverman Rule of thumb suggested a bandwidth around 13,000, Cross Validation one around 9,000, visually both results seem appealing.