

ECON 7020
Philip Shaw
Problem Set 4
Due date: May 2, 2024

Problem 1. Using the NP package in R, complete the following tasks:

a. Using the data set `wagesub.csv`¹, compute the optimal bandwidths for the conditional pdf of $f(lwage|educ)$. Use both maximum likelihood and least squares cross-validation.² Report the optimal bandwidths for each method. Also, time each method by utilizing the code below:

```
start.time <- Sys.time()
...Relevantcodes...
end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

How much faster is the bandwidth selection using maximum likelihood versus least squares cross validation?

b. Now using the command `npcdens.R`, estimate and plot the conditional pdf using each choice of bandwidth above. Is there a significant difference in the estimates?

¹Be sure to only use a subset of the data. Using `set.seed(1)` and `ind=sample(1:5000,500,replace=T)` to construct subsamples `lwagesub=lwage[ind]` and `Educsub=Educ[ind]`.

²The `npcdensbw.R` script file will do this for you. Be sure to classify each variable as continuous or discrete (ordered or unordered).