Advanced Econometrics: Exercise #8

September 22, 2025

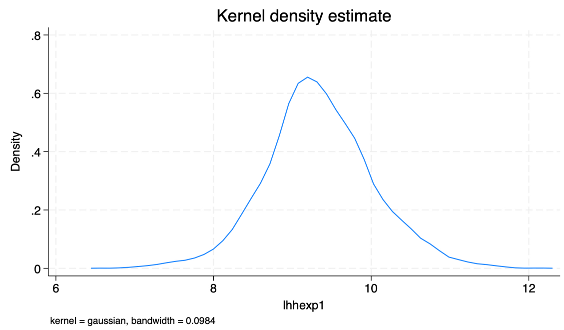
Professor Paulo Parente

Amy Qian

Problem 1

Obtain a non-parametric estimate of the density of lhhexp1. Compare the bandwith used by STATA with the one suggested by Silverman's rule of thumb.

STATA
kdensity lhhexp1, kernel(gaussian)



Bandwith used by STATA: 0.0984

Silverman's Rule of Thumb
sum lhhexp1, d

This gives us

- Standard deviation: 0.6877458
- n: 5,999
- IQR: 9.759566 - 8.919547 = 0.840019

To calculate Silverman's bandwith, we use the formula: $h = \min(\sigma, \frac{IQR}{1.34})\frac{0.9}{n^{-1/5}}$
Plugging in our values, we get:

- display ((.6877458*0.9)/5999^(0.2)) = .10865624
- display (((9.759566 - 8.919547)/1.34*0.9)/5999^(0.2)) = .09904009

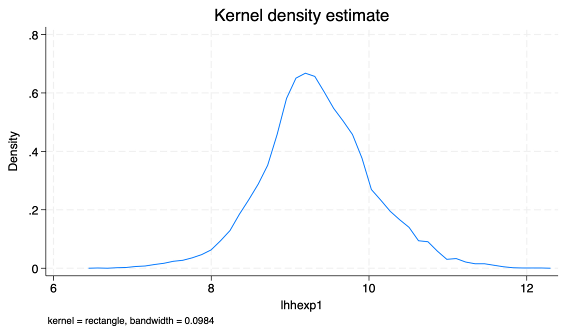
So the bandwidth suggested by Silverman's rule of thumb is 0.09904009, which is very close to the bandwidth used by STATA.

Problem 2

Repeat the previous exercise using different kernels.

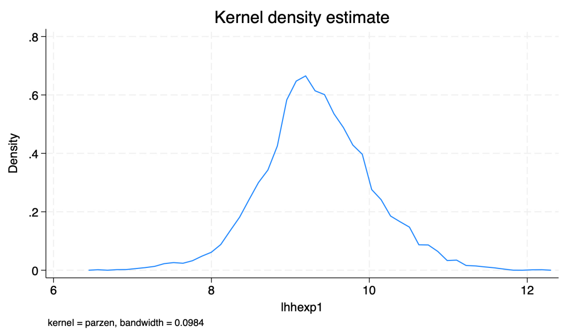
Rectangular Kernel

Problem 2 continued on next page...



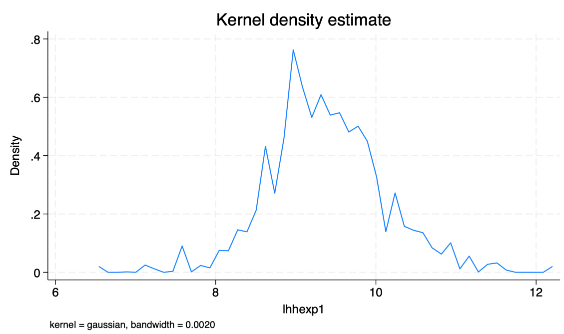
kdensity lhhexp1, kernel(rec)
This is less smooth than the Gaussian kernel.

Parzan Kernel

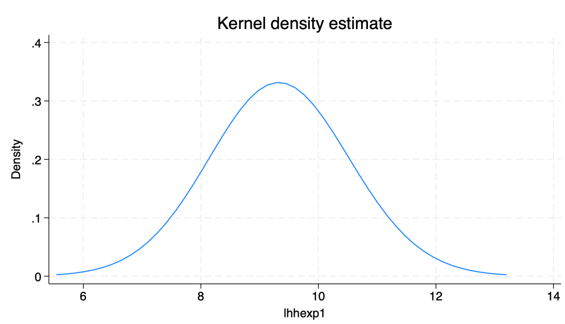


kdensity lhhexp1, kernel(parz)
This is also less smooth than the Gaussian kernel.

Gaussian with different bandwidth



kdensity lhhexp1, kernel(gaussian) bwidth(0.002)



kdensity lhhexp1, kernel(gaussian) bwidth(1)

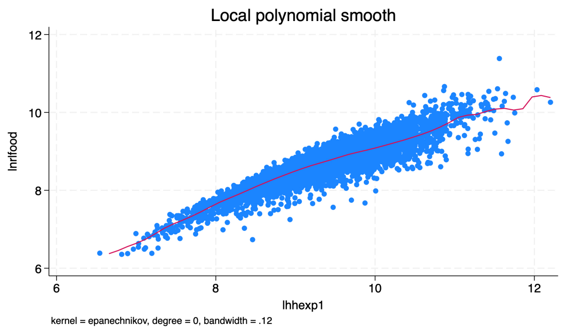
In general, the bandwith is more important than the kernel. We want it to be smooth enough but not too smooth that we miss important features of the data.

Problem 3

Obtain the Nadaraya-Watson estimator of the conditional expectation of lnrlfood, the log household expenditure on food, given lhhexp1. Does the result suggest that this conditional expectation is linear?

Solution

We wan to estimate $E[y_i|x_i] = m(x_i)$ where y_i is lnrlfood.



lpoly lnrlfood lhhexp1
Blue dots are the data points, the red line is the estimated conditional expectation (the function). Could be linear looking at the graph, but not exactly linear.

Problem 4

Using OLS, estimate a regression of lnrlfood on lhhexp1, age and its square. Is the effect of age linear?

Solution

First generate the new variable age squared.
gen age2 = age^2
Then run the OLS regression using robust standard errors to control for heteroskedasticity.
reg lnrlfood lhhexp1 age age2, r
We want to test the null hypothesis that the effect of age is linear, which is equivalent to testing if the coefficient of age2 is zero.

$$H_0 : \beta_{\text{age2}} = 0 \text{ vs. } H_a : \beta_{\text{age2}} \neq 0$$

From the regression output, we see that the p-value is 0.000, which is less than 0.05, so we reject H_0 in favor of H_1 at 5% confidence level.

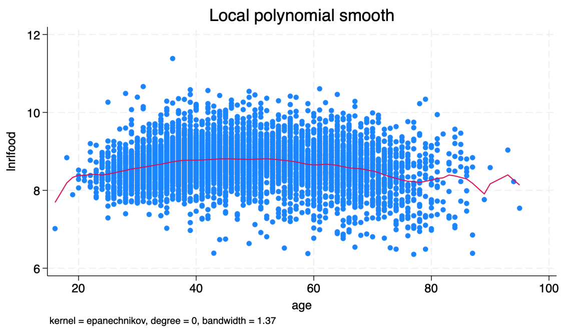
Problem 5

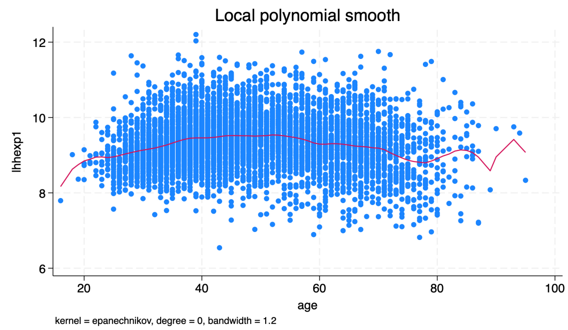
Use the partially linear estimator to estimate in the following model: $\text{lnrlfood}_i = \beta \text{lhhexp1}_i + g(\text{age}_i) + u_i$

Solution

$G(\cdot)$ is now an unknown function.
Get the estimates

- $E[y_i|\text{age}_i] \Rightarrow$ foodage. lpoly lnrlfood age, generate(foodage) at (age)
- $E[x_i|\text{age}_i] \Rightarrow$ expage. lpoly lhhexp1 age, generate(expage) at (age)





Now we compute the differences.

- gen y= lnrlfood- foodage
- gen x= lhhexp1- expage

Finally, we run the regression of y on x, not assuming anything on the distribution of the data.

```
      Source |       SS           df       MS      Number of obs   =
> 5,999
-------------+----------------------------------   F(1, 5998)      =  235
> 64.09
       Model |  1274.81681         1  1274.81681   Prob > F        =    0
> .0000
    Residual |  324.491644     5,998  .054099974   R-squared       =    0
> .7971
-------------+----------------------------------   Adj R-squared   =    0
> .7971
       Total |  1599.30845     5,999  .266595841   Root MSE        =    .
> 23259

------------------------------------------------------------------------------
> ------
           y | Coefficient  Std. err.      t    P>|t|     [95% conf. inte
> rval]
-------------+----------------------------------------------------------------
> ------
           x |    .695743   .0045324   153.51   0.000      .686858     .70
> 46281
------------------------------------------------------------------------------
```

reg y x, noconstant

Problem 6

Compare the results.

```
Linear regression                               Number of obs     =      5,999
                                                F(3, 5995)        =    5484.44
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8112
                                                Root MSE          =      .2333
```

lnrlfood	Coefficient	Robust std. err.	t	P>|t|	[95% conf. interval]	
lhhexp1	.6942858	.005711	121.57	0.000	.6830901	.7054815
age	.005095	.0016213	3.14	0.002	.0019168	.0082733
age2	-.0000716	.0000161	-4.45	0.000	-.0001031	-.00004
_cons	2.1277	.0561216	37.91	0.000	2.017681	2.237718

Actually quite close.