Problem Set IV - Sharp Regression Discontinuity

Enrijeta Shino

June 22, 2018

1 (Sharp) Regression Discontinuity

Framework

The dataset that I use in this part is the smoking data. The cut-off point is the $age = 45$ and I create a placebo treatment by applying the following rule:

Decision rule:

Let $\tilde{Y}$ = consumption of cigarettes and C = age of consumers.

$$\tilde{Y}_i = \begin{cases} \tilde{Y}_i(0) & if \quad c > 45 \\ \tilde{Y}_i(1) & otherwise \end{cases} \tag{1}$$

where we manipulate the outcome Y to create the sharp discontinuity design framework. We assume that consumers above the age of 45 receive a treatment in order to reduce consumption of cigarettes.

Randomly generated outcomes:

We create the outcomes $(\tilde{Y}_i(1), \tilde{Y}_i(0))$ as follows:

$$\tilde{Y}_i(1) = Y_i + \sigma_{Y_i} + \epsilon_i \tag{2}$$

$$\tilde{Y}_i(0) = Y_i + \sigma_{Y_i} + \eta_i \tag{3}$$

where $\epsilon_i \sim U(0,1)$ and $\eta_i \sim U(0,1)$ are both randomly generated numbers in R.

Testing and visual inspection:

We apply McCrary (2008) to test whether the discontinuity in the density of the running variable (age) at the cut-off point (age = 45) is equal to zero. The p-value is equal to 0.0077, hence we reject the null hypothesis.

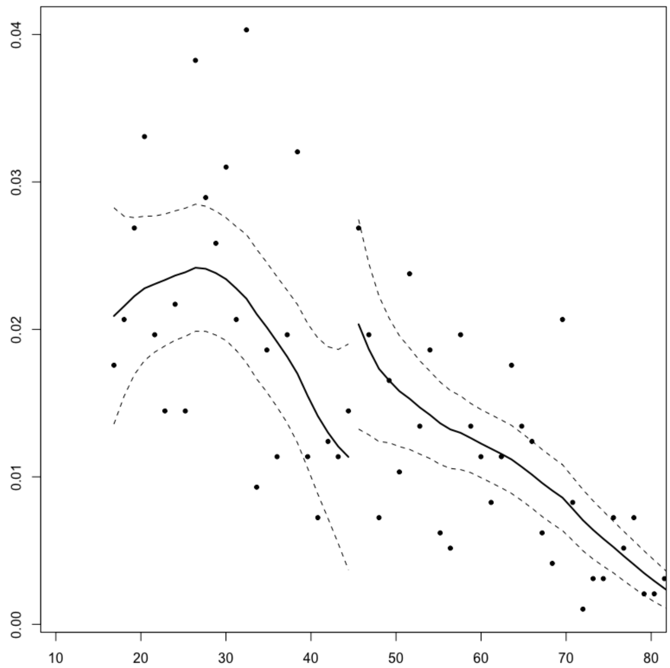

Figure 1: McCrary (2008) test

2.1 Plot the outcome by forcing variable (the standard graph showing the discontinuity)

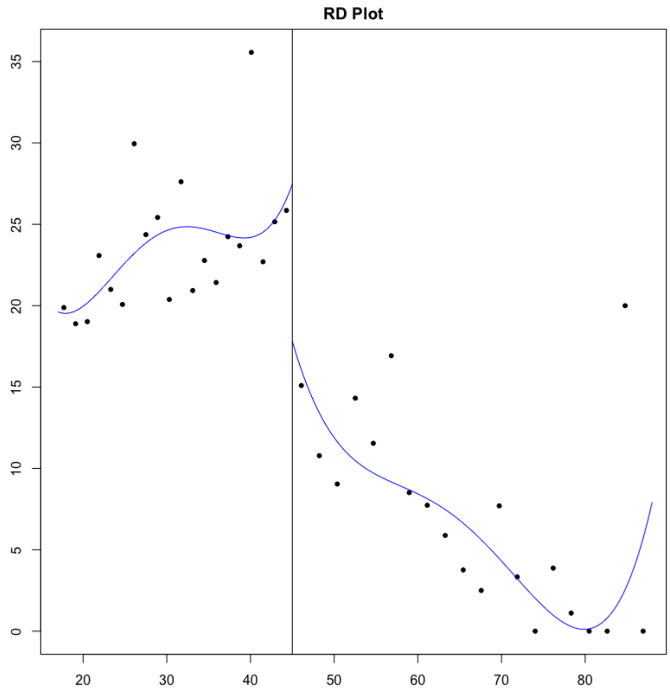

Figure 2: Consumption of cigarettes by age

2.2. Plot the density of the forcing variable

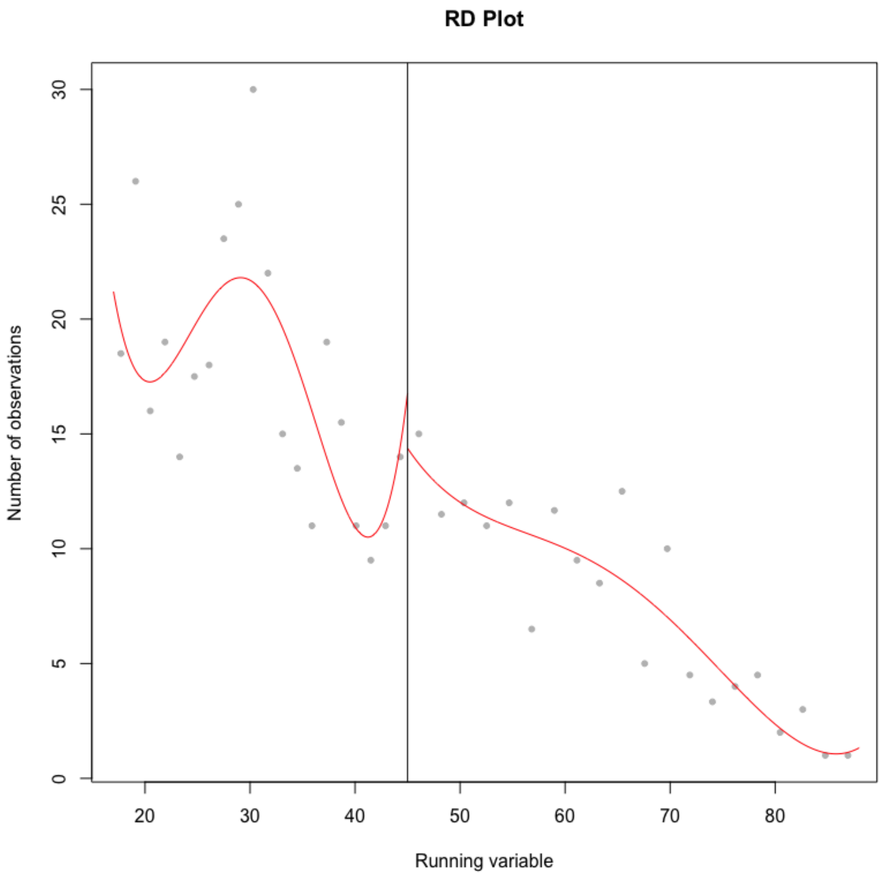

Figure 3: Density of the Forcing Variable

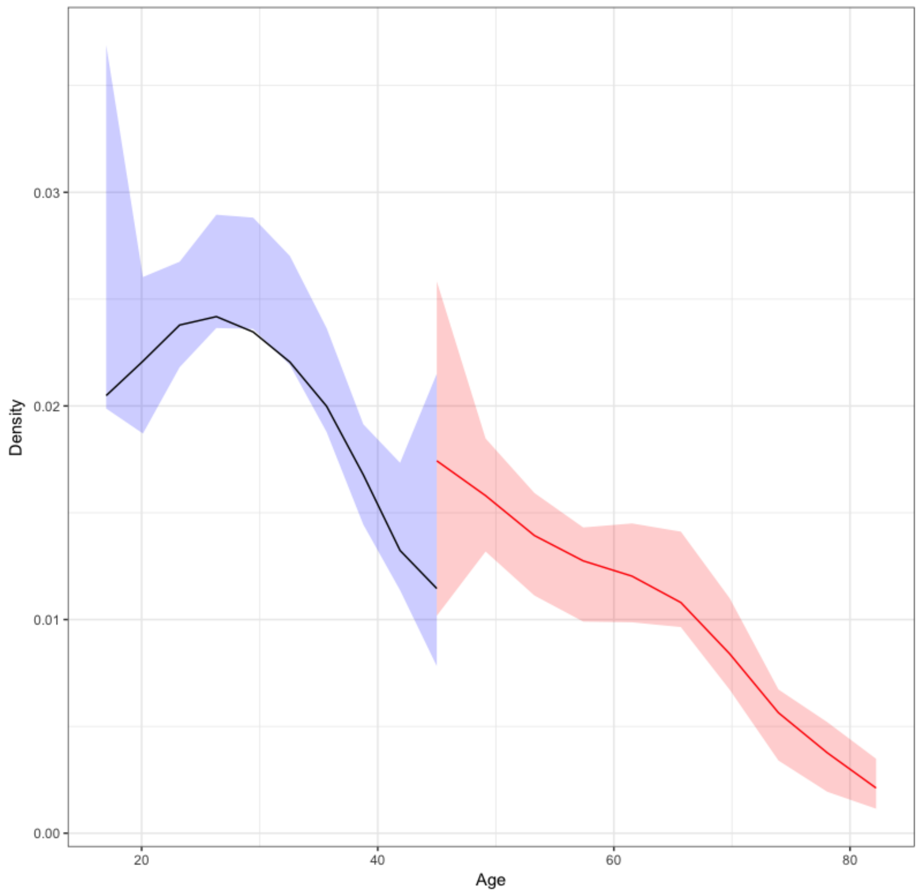

Figure 4: Density of the Forcing Variable

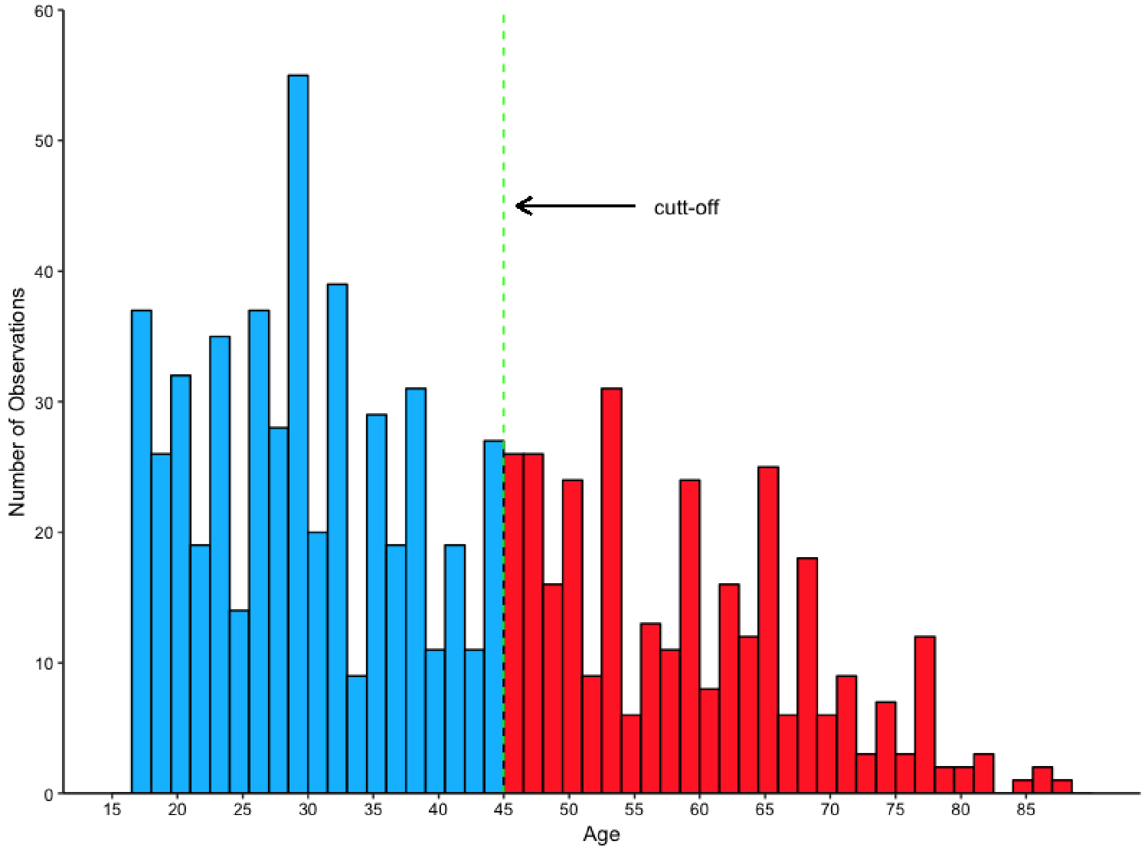

Figure 5: Distribution of the Forcing Variable (Not Sorted)

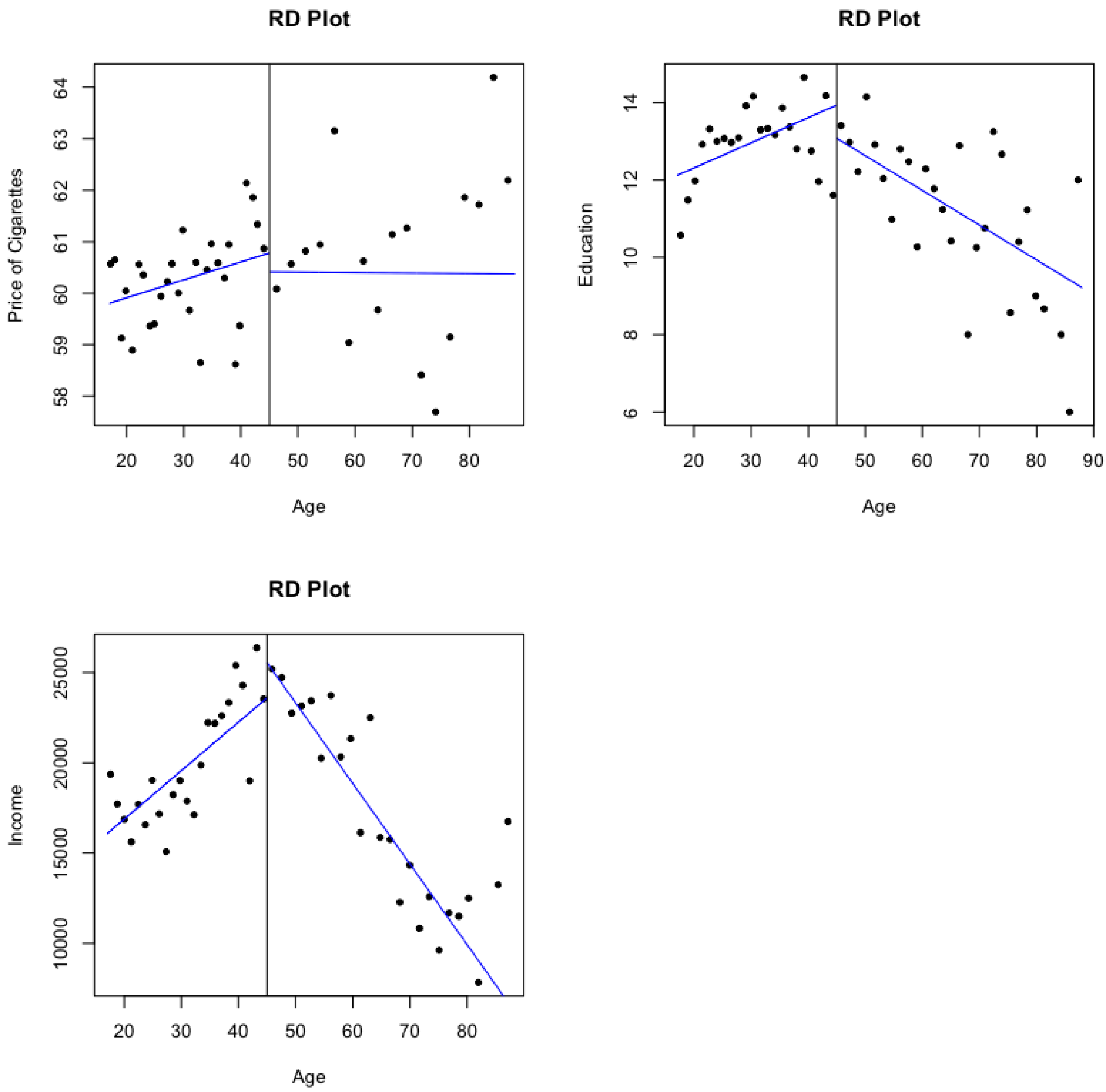

Figure 6: Continuity of covariates

2.3. Estimate the effect using a local linear regression

- Local linear regression using a uniform kernel

Table 1: **RDD results for local linear regression (using the package & uniform kernel)**

Method	Coef.	Std. Err.	z	p-value	95% C.I.
Convential	**-5.580**	5.946	-0.938	0.348	[-17.234 , 6.074]
Robust			-0.486	0.627	[-16.573 , 9.985]

Table 2: RDD results for local linear regression (uniform kernel)

	Dependent variable:	
	cigs_new	
	(LHS)	(RHS)
I(age - 45)	−0.232	−2.251**
	(1.182)	(0.887)
Constant	**24.600*****	**19.020*****
	(4.579)	(3.151)
Observations	68	92
R^2	0.001	0.067
Adjusted R^2	−0.015	0.056
Residual Std. Error	17.632 (df = 66)	17.047 (df = 90)
F Statistic	0.038 (df = 1; 66)	6.443** (df = 1; 90)
Note:		*p<0.1; **p<0.05; ***p<0.01

$$\textbf{ATE} = 19.02 - 24.6 = \textbf{-5.58} \tag{4}$$

2.4. Estimate the effect using a local polynomial (of order 2 and 3) regression

- **2nd order local polynomial regression**

Table 3: **RDD results for 2nd order local polynomial regression (using the package & uniform kernel)**

Method	Coef.	Std. Err.	z	p-value	95% C.I.
Conventional	**3.495**	8.716	0.401	0.688	[-13.589 , 20.578]
Robust			0.584	0.559	[-12.988 , 24.019]

Table 4: RDD results for 2nd order local polynomial regression (uniform kernel)

	Dependent variable:	
	cigs_new	
	(LHS)	(RHS)
I(age - 45)	-1.346	-11.030^{***}
	(4.383)	(2.611)
I((age - 45)^2)	-0.142	1.445^{***}
	(0.529)	(0.367)
Constant	**23.023*****	**26.518*****
	(7.434)	(3.762)
Observations	86	101
R^2	0.001	0.156
Adjusted R^2	-0.023	0.138
Residual Std. Error	17.265 (df = 83)	16.104 (df = 98)
F Statistic	0.061 (df = 2; 83)	9.027^{***} (df = 2; 98)
Note:		*p<0.1; **p<0.05; ***p<0.01

$$\mathbf{ATE} = 26.518 - 23.023 = \mathbf{3.495} \tag{5}$$

- **3rd order local polynomial regression**

Table 5: **RDD results for 3nd order local polynomial regression (using the package & uniform kernel)**

Method	Coef.	Std. Err.	z	p-value	95% C.I.
Conventional	**3.766**	9.546	0.395	0.693	[-14.944 , 22.476]
Robust			0.484	0.629	[-15.020 , 24.861]

Table 6: RDD results for 2nd order local polynomial regression (uniform kernel)

	Dependent variable:	
	cigs_new	
	(LHS)	(RHS)
I(age - 45)	-2.035	-11.231^{***}
	(5.740)	(3.319)
I((age - 45)^2)	-0.334	1.941^{**}
	(1.093)	(0.745)
I((age - 45)^3)	-0.013	-0.091^{*}
	(0.060)	(0.046)
Constant	**22.496** ***	**26.262*****
	(8.007)	(3.879)
Observations	143	143
R^2	0.006	0.108
Adjusted R^2	-0.016	0.089
Residual Std. Error (df = 139)	16.181	16.002
F Statistic (df = 3; 139)	0.269	5.613^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01	

$$\mathbf{ATE} = 26.262 - 22.496 = \mathbf{3.766} \tag{6}$$

Bandwidth is computed using Imbens-Kalyanaraman Optimal Bandwidth and using different kernels:

Table 7: Bandwidths under different kernels using Imbens-Kalyanaraman Optimal Bandwidth

	Rectangular	Triangular	Epanechnikov	Quatric
Bandwidth	10.23	6.51	6.51	6.92

Table 8: Compare

Estimate	Rectangular	Triangular	Epanechnikov	Quatric
ATE (local linear)	-10.62	-5.58	-5.58	-5.58
ATE (local 2nd order poly)	-1.62	4.62	4.62	4.62
ATE (local 3rd order poly)	5.05	-19.33	-19.33	-19.32

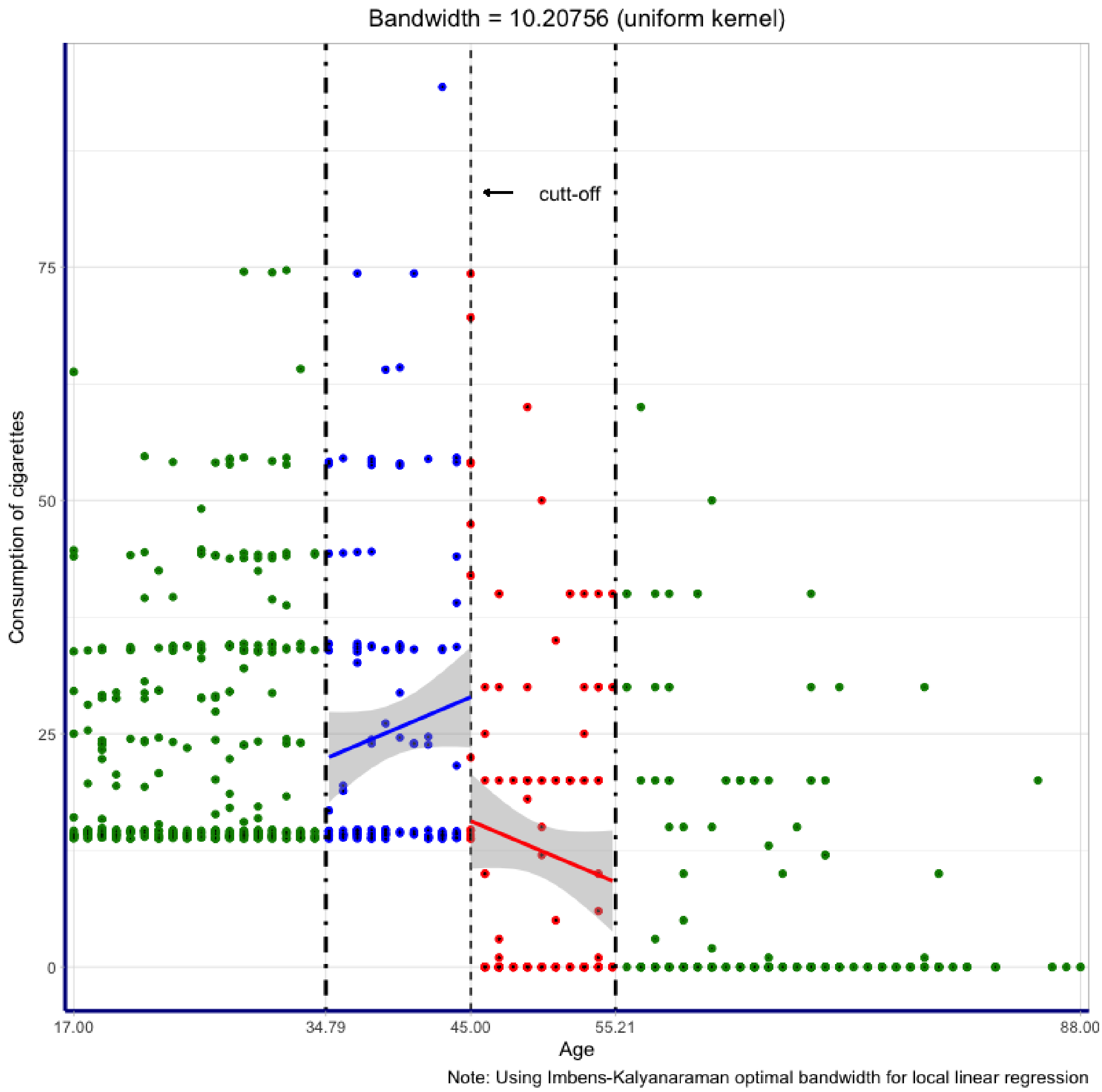

Figure 7: Local linear regression (uniform kernel)

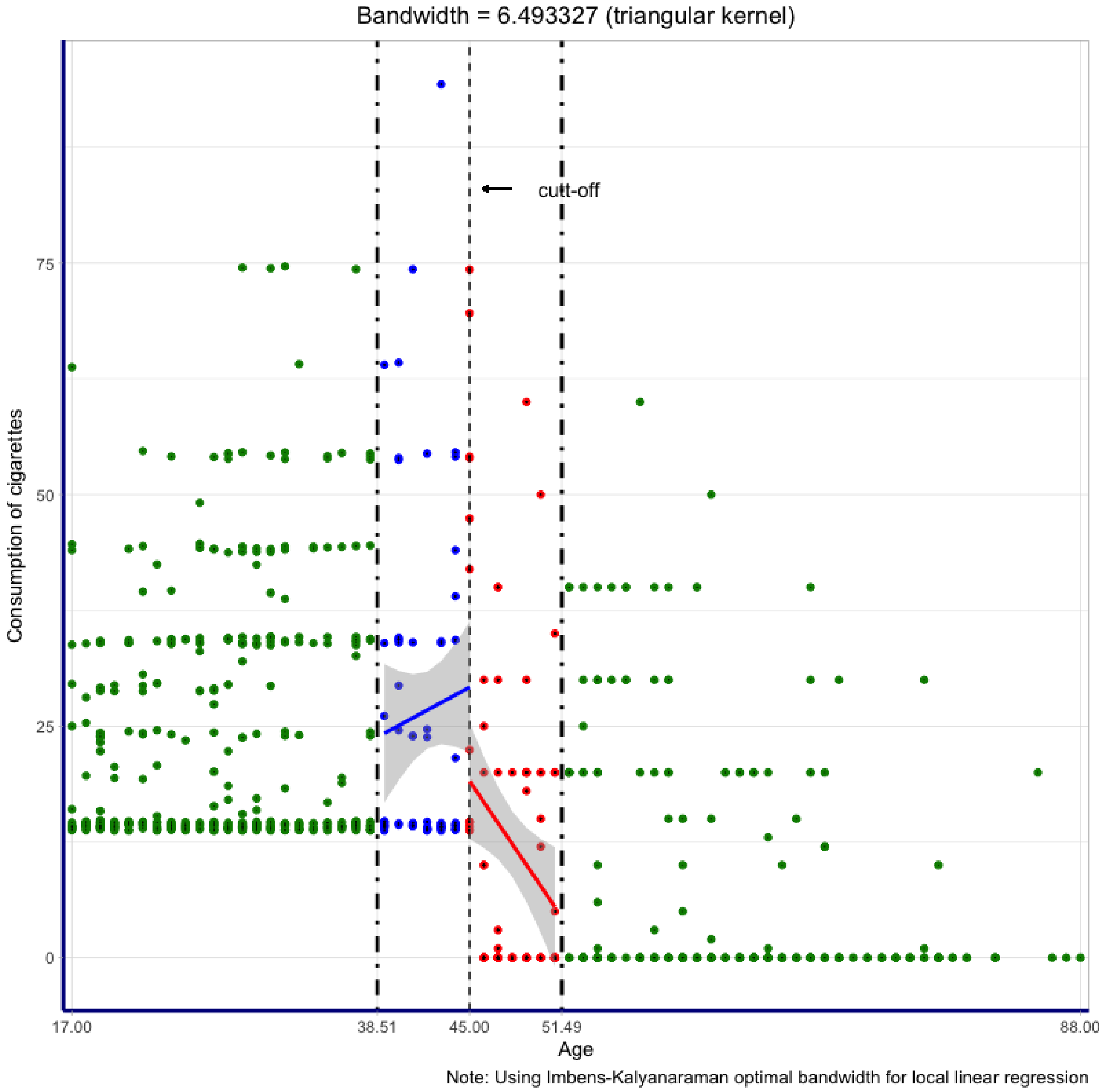

Figure 8: Local linear regression (triangular kernel)

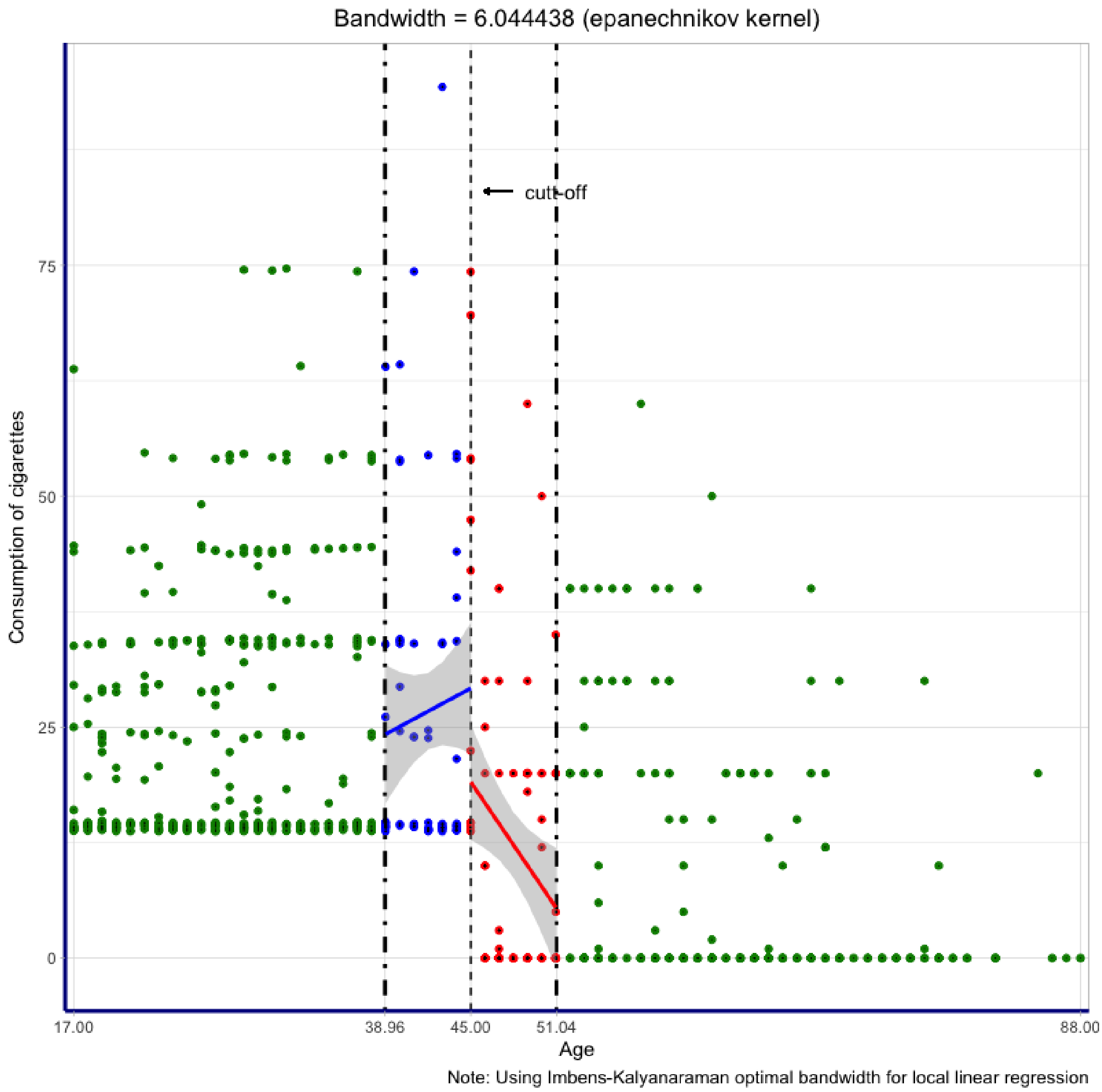

Figure 9: Local linear regression (epanechnikov kernel)

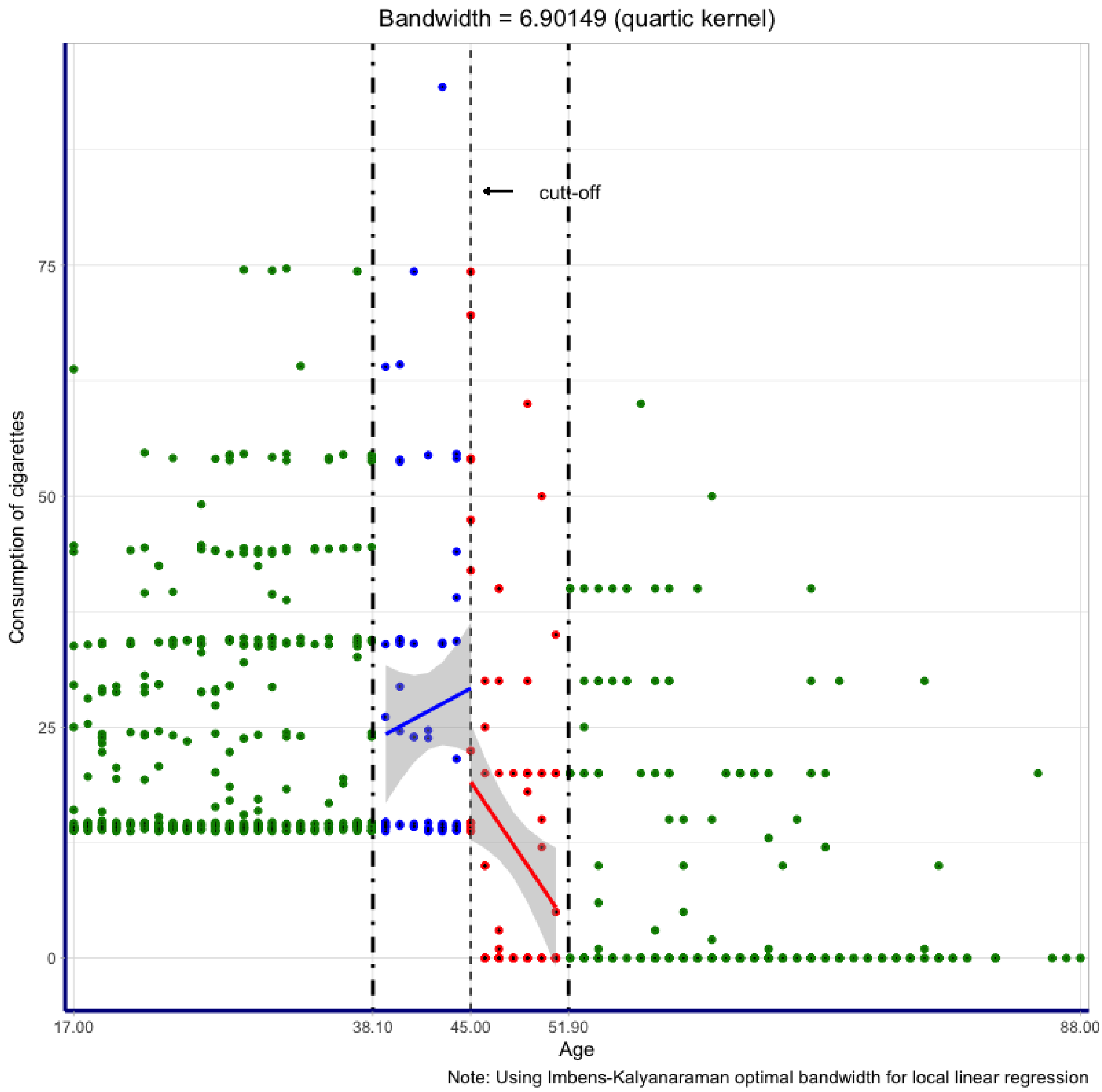

Figure 10: Local linear regression (quatric kerne)