12) Sharp Regression Discontinuity Design

Vitor Kamada

December 2018

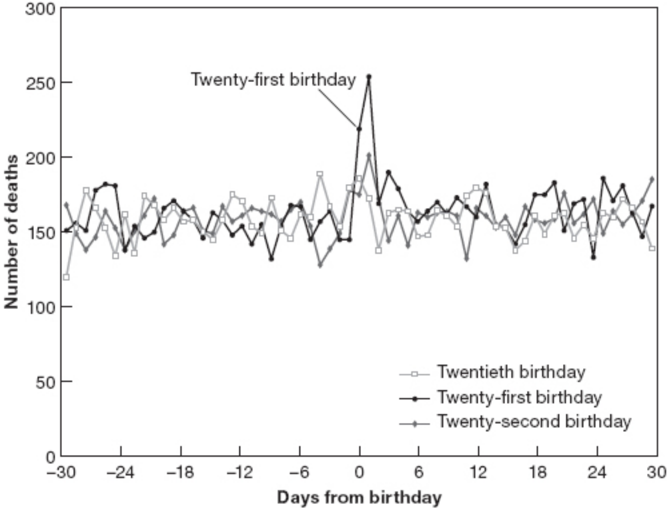
Reference

Tables, Graphics, and Figures from:

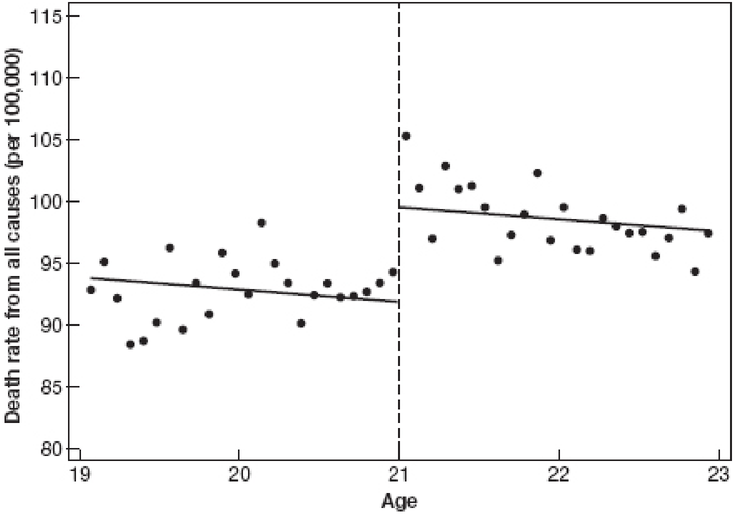
1) Angrist & Pischke (2014)**. Mastering 'Metrics: The Path from Cause to Effect.** Chapter 4.

2) Gertler et al. (2016). **Impact Evaluation in Practice**. Chapter 6.

Birthdays and Funerals from 1997 to 2003



A Sharp RD Estimate of MLDA Mortality Effects



Sharp RD

$$D_a = \begin{cases} 1 \;\; if \;\; a \geq 21 \\ 0 \;\; if \;\; a < 21 \end{cases}$$

$$\bar{M}_a = \alpha + \rho D_a + \gamma a + e_a$$

$\bar{M}_a$: death rate in month a

$$\hat{\rho} = 7.7$$

Carpenter and Dobkin (2011)

file = ("https://github.com/VitorKamada/ECO5100/raw/master/Data/AEJfigs.dta")

import pandas as pd

df = pd.read_stata(file)

df.head()

	agecell	all	allfitted	internal	internalfitted	external
0	19.068493	92.825401	91.706146	16.617590	16.738131	76.207817
1	19.150684	95.100739	91.883720	18.327684	16.920654	76.773056
2	19.232876	92.144295	92.049065	18.911053	17.098843	73.233238
3	19.315069	88.427757	92.202141	16.101770	17.272680	72.325981
4	19.397261	88.704941	92.342918	17.363520	17.442156	71.341415

	externalfitted	alcohol	alcoholfitted	homicide	homicidefitted	\
0	74.968010	0.639138	0.794344	16.316818	16.284573	
1	74.963066	0.677409	0.837575	16.859964	16.270697	
2	74.950226	0.866443	0.877835	15.219254	16.262882	
3	74.929466	0.867308	0.915115	16.742825	16.261148	
4	74.900757	1.019163	0.949407	14.947726	16.265511	

df.describe()

```
       agecell         all   allfitted   internal  internalfitted  \
count  50.000000   48.000000   50.000000  48.000000       50.000000
mean   21.000000   95.672722   95.802841  20.285294       20.281301
std     1.126957    3.831062    3.286415   2.253907        1.994682
min    19.068493   88.427757   91.706146  15.977087       16.738131
25%    20.075342   92.785929   93.040606  18.597654       18.674128
50%    20.999995   95.686272   95.178303  20.288866       20.537065
75%    21.924658   98.025751   97.786827  21.976349       21.658084
max    22.931507  105.268349  102.891762  24.372910       24.043783

        external  externalfitted    alcohol  alcoholfitted   homicide
count  48.000000       50.000000  48.000000      50.000000  48.000000
mean   75.387436       75.521538   1.257337       1.267447  16.912066
std     2.986008        2.269975   0.350312       0.259862   0.729982
min    71.341415       73.157860   0.639138       0.794344  14.947726
25%    73.042023       74.061251   0.996152       1.072381  16.611996
50%    74.813251       74.736385   1.211941       1.247127  16.985353
75%    77.242350       76.063623   1.470119       1.445450  17.288067
max    83.330986       81.783722   2.519309       1.817361  18.410973
```

Data Manipulation

```
import numpy as np
df['const'] = 1
df['age'] = df['agecell'] - 21
df['over21'] = np.where(df['agecell'] >=21, 1, 0)
df['over21_age'] = df['age']*df['over21']
df['age2'] = df['age']*df['age']
df['over21_age2'] = df['age2']*df['over21']
```

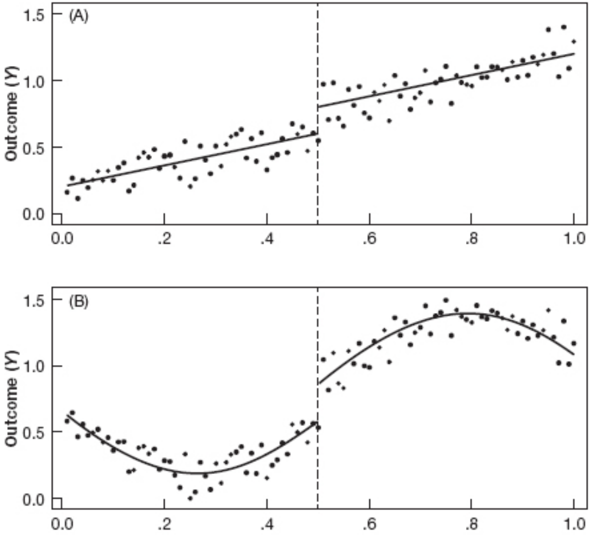
Index	const	age	over21	over21_age	age2	over21_age2
0	1	-1.9315071	0	-0	3.7307198	0
1	1	-1.8493156	0	-0	3.4199684	0
2	1	-1.7671242	0	-0	3.1227279	0
3	1	-1.6849308	0	-0	2.8389919	0

$\bar{M}_a = \alpha + \rho D_a + \gamma a + e_a$

```
import statsmodels.api as sm
result1 = sm.OLS(df['all'], df[['const', 'age', 'over21']],
                 missing='drop').fit()
print(result1.summary())
```

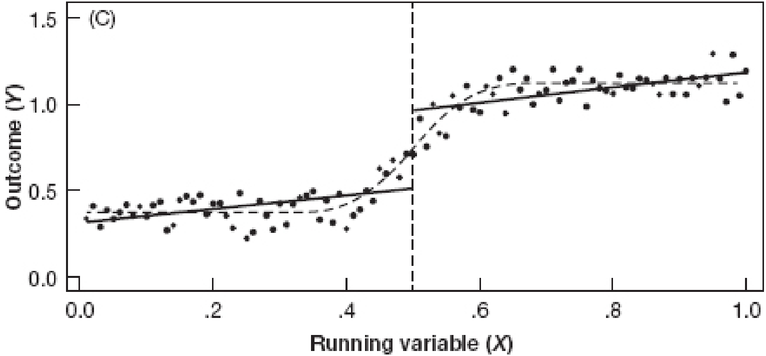
```
Dep. Variable:                    all   R-squared:                       0.595
Model:                            OLS   Adj. R-squared:                  0.577
Method:                 Least Squares   F-statistic:                     32.99
Date:                Sat, 21 Jul 2018   Prob (F-statistic):           1.51e-09
Time:                        19:27:22   Log-Likelihood:                -110.41
No. Observations:                  48   AIC:                             226.8
Df Residuals:                      45   BIC:                             232.4
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         91.8414      0.805    114.083      0.000      90.220      93.463
age           -0.9747      0.632     -1.541      0.130      -2.249       0.299
over21         7.6627      1.440      5.320      0.000       4.762      10.564
```

Valid RD



(A)
Outcome (Y)
1.5
1.0
0.5
0.0
0.0
.2
.4
.6
.8
1.0
(B)
Outcome (Y)
1.5
1.0
0.5
0.0
0.0
.2
.4
.6
.8
1.0

Nonlinear Trend with no Discontinuity



Nonlinearities in an RD

$$\bar{M}_a = \alpha + \rho D_a + \gamma_1 a + \gamma_2 a^2 + e_a$$

$$\bar{M}_a = \alpha + \rho D_a + \gamma(a - a_0) + \delta[(a - a_0)D_a] + e_a$$

$$[\alpha + \rho + (\gamma + \delta)(a - a_0)] - [\alpha + \gamma(a - a_0)]$$

$$\text{TE} = \rho + \delta[(a - a_0)]$$

$\bar{M}_a = \alpha + \rho D_a + \gamma_1 a + \gamma_2 a^2 + e_a$

```
result3 = sm.OLS(df['all'],
                 df[['const', 'age', 'age2', 'over21']],
                 missing='drop').fit()
print(result3.summary())
```

```
Dep. Variable:                    all   R-squared:                       0.657
Model:                            OLS   Adj. R-squared:                  0.634
Method:                 Least Squares   F-statistic:                     28.12
Date:                Sat, 21 Jul 2018   Prob (F-statistic):           2.61e-10
Time:                        19:52:01   Log-Likelihood:                -106.38
No. Observations:                  48   AIC:                             220.8
Df Residuals:                      44   BIC:                             228.2
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         92.9027      0.837    110.994      0.000      91.216      94.590
age           -0.9747      0.588     -1.657      0.105      -2.160       0.211
age2          -0.8187      0.289     -2.835      0.007      -1.401      -0.237
over21         7.6627      1.339      5.721      0.000       4.963      10.362
```

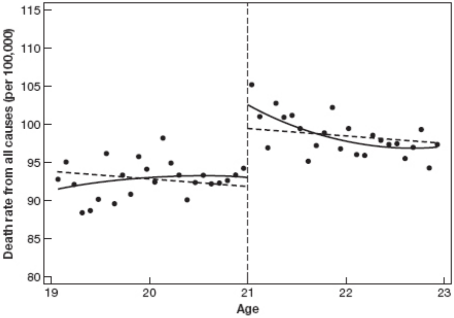
$$\bar{M}_a = \alpha + \rho D_a + \gamma(a - a_0) + \delta[(a - a_0)D_a] + e_a$$

```
result2 = sm.OLS(df['all'],
                 df[['const', 'age', 'over21', 'over21_age']],
                 missing='drop').fit()
print(result2.summary())
```

```
Dep. Variable:                    all   R-squared:                       0.668
Model:                            OLS   Adj. R-squared:                  0.645
Method:                 Least Squares   F-statistic:                     29.47
Date:                Sat, 21 Jul 2018   Prob (F-statistic):           1.33e-10
Time:                        19:54:34   Log-Likelihood:                -105.64
No. Observations:                  48   AIC:                             219.3
Df Residuals:                      44   BIC:                             226.8
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         93.6184      0.932    100.399      0.000      91.739      95.498
age            0.8270      0.819      1.010      0.318      -0.823       2.477
over21         7.6627      1.319      5.811      0.000       5.005      10.320
over21_age    -3.6034      1.158     -3.111      0.003      -5.937      -1.269
```

Concave to the Left and Convex Thereafter

$$\bar{M}_a = \alpha + \rho D_a + \gamma_1(a - a_0) + \gamma_2(a - a_0)^2 + (\delta_1[(a - a_0)D_a] + \delta_2[(a - a_0)^2 D_a] + e_a$$



TE: $\rho + \delta_1(a - a_0) + \delta_2(a - a_0)^2 = 9.5$

$\bar{M}_a = \alpha + \rho D_a + \gamma_1(a - a_0) + \gamma_2(a - a_0)^2 + (\delta_1[(a - a_0)D_a] + \delta_2[(a - a_0)^2 D_a] + e_a$

```python
result4 = sm.OLS(df['all'],
                 df[['const', 'age', 'age2',
                     'over21', 'over21_age', 'over21_age2']],
                 missing='drop').fit()
print(result4.summary())
```

	coef	std err	t	P>|t|	[0.025	0.975]
const	93.0729	1.404	66.301	0.000	90.240	95.906
age	-0.8306	3.290	-0.252	0.802	-7.470	5.809
age2	-0.8403	1.615	-0.520	0.606	-4.100	2.419
over21	9.5478	1.985	4.809	0.000	5.541	13.554
over21_age	-6.0170	4.653	-1.293	0.203	-15.407	3.373
over21_age2	2.9042	2.284	1.271	0.211	-1.706	7.514

Y = Mortality of Motor Vehicle Accidents

```
df1 = df[df['agecell'] >= 20]
df1 = df1[df1['agecell'] <= 22]
result5 = sm.OLS(df1['mva'],
                 df1[['const', 'age', 'age2',
                     'over21', 'over21_age', 'over21_age2']],
                 missing='drop').fit(cov_type='HC1')
print(result5.summary())
```

	coef	std err	z	P>|z|	[0.025	0.975]
const	30.1883	0.562	53.716	0.000	29.087	31.290
age	0.6801	3.816	0.178	0.859	-6.800	8.160
age2	4.4599	4.716	0.946	0.344	-4.783	13.702
over21	5.8925	1.329	4.433	0.000	3.287	8.498
over21_age	-15.1667	6.351	-2.388	0.017	-27.614	-2.720
over21_age2	6.9652	7.053	0.988	0.323	-6.858	20.789

Sharp RD Estimates of MLDA Effects on Mortality

Dependent variable	Ages 19–22 (1)	(2)	Ages 20–21 (3)	(4)
All deaths	7.66	9.55	9.75	9.61
	(1.51)	(1.83)	(2.06)	(2.29)
Motor vehicle accidents	4.53	4.66	4.76	5.89
	(.72)	(1.09)	(1.08)	(1.33)
Suicide	1.79	1.81	1.72	1.30
	(.50)	(.78)	(.73)	(1.14)
Homicide	.10	.20	.16	−.45
	(.45)	(.50)	(.59)	(.93)
Other external causes	.84	1.80	1.41	1.63
	(.42)	(.56)	(.59)	(.75)
All internal causes	.39	1.07	1.69	1.25
	(.54)	(.80)	(.74)	(1.01)
Alcohol-related causes	.44	.80	.74	1.03
	(.21)	(.32)	(.33)	(.41)
Controls	age	age, age^2, interacted with over-21	age	age, age^2, interacted with over-21
Sample size	48	48	24	24

RD Estimates of MLDA Effects on Mortality by Cause of Death