Lab8A : IV and RD in Stata

Introduction to Econometrics,Spring 2023

Jiayi Cheng

Nanjing University

12/5/2023

Section 1

IV estimator in Stata

Subsection 1

Review the Theory

IV estimator in Stata

- Review the Theory

Review Previous Lecture of Internal Validity

Threatens to Internal Validity

- Three endogenous in OLS regression are:
  - **Omitted Variable Bias**(a variable that is correlated with X but is unobserved)
  - **Simultaneity or reverse causality Bias** (X causes Y,Y causes X)
  - **Errors-in-Variables Bias** (X is measured with error)
- One easy way to deal with these endogeneity is using **Instrumental Variable** method.

- Review the Theory

Instrumental variables: 1 endogenous regressor & 1 instrument

- suppose a simple OLS regression like previous equation

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Because $E[u_i|X_i] \neq 0$, then we can use an instrumental variable(Z_i) to obtain an consistent estimate of coefficient.
- Intuitively, we want to split X_i into two parts:
  1. part that is correlated with the error term.
  2. part that is uncorrelated with the error term.
- If we can isolate the variation in X_i that is uncorrelated with u_i,then we can use this part to obtain a consistent estimate of the causal effect of X_i on Y_i.

IV estimator in Stata

- Review the Theory

Instrumental Variable Method

Instrumental variables: 1 endogenous regressor & 1 instrument

- An instrumental variable Z_i must satisfy the following 2 properties:

1. **Instrumental relevance**: Z_i should be **correlated** with the casual variable of interest, X_i (endogenous variable),thus

$$Cov(X_i, Z_i) \neq 0$$

.

2. **Instumental exogeneity**: Z_i is as good as randomly assigned and Z_i only affect on Y_i through X_i affecting Y_i channel.

$$Cov(Z_i, u_i) = 0$$

$$\begin{tikzcd} z \arrow[r] & x \arrow[r] & y \\ & u \arrow[u] \arrow[ur] & \end{tikzcd}$$

Zhaopeng Qu (Nanjing University) Lecture 9: Instrumental Variables 11/19/2020 9 / 101

IV estimator in Stata

- Review the Theory

IV estimator: Two Steps Least Square (2SLS)

- We can estimate the causal effect of X_i on Y_i in two steps
  1. **First stage**: Regress X_i on Z_i & obtain predicted values of $\hat{X}_i$,if $Cov(Z_i, u_i) = 0$, then $\hat{X}_i$ contains variation in X_i that is uncorrelated with u_i
  $$\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$$
  .
  2. **Second stage**: Regress Y_i on $\hat{X}_i$ to obtain the Two Stage Least Squares estimator $\hat{\beta}_{2SLS}$
  $$\hat{\beta}_{2SLS} = \frac{\sum(Y_i - \bar{Y})(\hat{X}_i - \bar{\hat{X}})}{\sum(\hat{X}_i - \bar{\hat{X}})^2}$$

IV estimator in Stata

- Review the Theory

IV with Heterogeneous Causal Effects

Instrument Variables:Constant-effect

- Instrumental Variable is a useful method to make causal inference. It can eliminate
  - Omitted Variable Bias
  - Measurement Error
  - Reverse Causality
- Two Assumptions
  - Relevance(Weak Instrument): It can be test by the first stage regression and F-statistic.
  - Exogeneity: Can't be test formally but argue it using professional knowledges.
- Estimation and Inference
  - When IV satisfy these two assumptions,the causal effect of coefficients of interest,TSLS estimator,β_{TSLS} can be NOT unbiased but **consistent**.
  - The sampling distribution of the TSLS estimator is also normal in large samples,so the general procedures for statistical inference in OLS can be used.

Zhaopeng Qu (Nanjing University) Lecture 9: Instrumental Variables 11/19/2020 63 / 101

IV estimator in Stata

- Review the Theory

Some Practical Guides by Angrist and Pischke(2012)

Practical Guides

1. Check IV relevance
   - Always report the first stage and think about whether it makes sense(Signs and magnitudes)
   - Always report the F-statistic on the excluded instruments. The bigger,the better. Don't forget the rule of thumb.($F > 10$)
2. Check exclusion restriction
   - The exclusion restriction cannot be tested directly, but it can be falsified
   - Run and examine the reduced form(regression of dependent variable on instruments) and look at the coefficients, t-statistics and F-statistics for excluded instruments.
   - Because the reduced form is proportional to the casual effect of interest and is unbiased(OLS), so we should see the causal relation in the reduced form.If you can't see the causal relation in the reduced form,it's probably not there

Zhaopeng Qu (Nanjing University) Lecture 9: Instrumental Variables 11/19/2020 80 / 101

IV estimator in Stata

- Review the Theory

Some Practical Guides by Angrist and Pischke(2012)

Practical Guides

3. Provide a substantive explanation for observed difference between 2SLS and OLS
   - How bid is the difference? What does this tell you?
   - Is the coefficient bigger when theory of endogeneity suggests it should be smaller? If so, why?
   - Measurement Error or heterogeneous effects?
4. If you have multiple instruments, report over-identification tests.
   - Pick your best single instrument and report just-identified estimates using this one only because just-identified IV is relatively unlikely to be subject to a weak bias.
   - Worry if it is substantially different from what you get using multiple instruments.
   - Check over-identified 2SLS estimates with LIML. LIML is less than precise than 2SIS but also less biased. If the results come out similar, be happy. If not, worry, and try to find stronger instruments.

Zhaopeng Qu (Nanjing University) Lecture 9: Instrumental Variables 11/19/2020 81 / 101

Subsection 2

Introduction

IV estimator in Stata

- Syntax

```
ivregress estimator depvar [varlist1] (varlist2 = varlist_iv)
^^I[if] [in] [weight] [, options]

help ivregress
```

IV estimator in Stata

- Options
  - estimator : 2sls/liml/gmm
  - depvar : dependent variable
  - varlist1 : exogenous variables
  - varlist2 : endogenous variables
  - varlist_iv : instrument variables

IV estimator in Stata

- Options
  - e.g. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$,
    in which, x_1 is exogenous variables,
    x_2 is endogenous variables,
    and z_1, z_2 are instrument variables.

```
ivregress 2sls y x1 (x2 = z1 z2)
ivregress 2sls y x1 (x2 = z1 z2), r first

/* r--异方差稳健标准误
   first--report first-stage regression */
```

Subsection 3

An Example

IV estimator in Stata

- An Example
  - Mincer(1958) first researched the positive correlation between **salary** and **years of education**, but omitted the **ability** variable.
  - Griliches(1976) addressed the problem of omitted variable bias with IV method.
  - Data：Young Men's Cohort of the National Longitudinal Survey (NLS-Y).
  - Blackburn and Neumark(1992) Updated data of Griliches(1976).
  - Two-period panel data：the initial period is the earliest year in which the above variables have data; The end period is 1980.

IV estimator in Stata

- An Example：Data
  - lw (工资对数)
  - s (受教育年限)
  - age (年龄)
  - expr (工龄)
  - tenure (在现单位的工作年数)
  - iq (智商)
  - med (母亲的受教育年限)
  - kww (在 "knowledge of the World of Work" 测试中的成绩)
  - mrt (=1，已婚)
  - rns (=1，住在美国南方)
  - smsa (=1，住在大城市)
  - year (有数据的最早年份，1966-1973 中的某一年)

IV estimator in Stata

1. Data Summary.

```
. use grilic, clear
. sum
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         rns |        758    .2691293    .4438001          0          1
       rns80 |        758     .292876    .4553825          0          1
         mrt |        758    .5145119    .5001194          0          1
       mrt80 |        758    .8984169    .3022988          0          1
        smsa |        758    .7044855     .456575          0          1
-------------+---------------------------------------------------------
      smsa80 |        758    .7124011     .452942          0          1
         med |        758    10.91029     2.74112          0         18
          iq |        758    103.8562    13.61867         54        145
         kww |        758    36.57388    7.302247         12         56
        year |        758    69.03166    2.631794         66         73
-------------+---------------------------------------------------------
         age |        758    21.83509    2.981756         16         30
       age80 |        758    33.01187    3.085504         28         38
           s |        758    13.40501    2.231828          9         18
         s80 |        758    13.70712    2.214693          9         18
        expr |        758    1.735429    2.105542          0     11.444
-------------+---------------------------------------------------------
      expr80 |        758    11.39426    4.210745       .692     22.045
      tenure |        758    1.831135     1.67363          0         10
    tenure80 |        758    7.362797     5.05024          0         22
          lw |        758    5.686739    .4289494      4.605      7.051
        lw80 |        758    6.826555    .4099268      4.749      8.032
```

IV estimator in Stata

2. Test the correlation between ig(智商) and s(受教育年限).

```
. pwcorr iq s, star(.01)
             |       iq        s
-------------+------------------
          iq |   1.0000
           s |   0.5131*  1.0000


 * Significant positive correlation at the level of 1%,
 * correlation coefficient = 0.51.
```

IV estimator in Stata

3. Run an OLS regression.

- With robust standard error
- control variables : expr tenure rns smsa
- We are interested in s.

```
. reg lw s expr tenure rns smsa, r

Linear regression                               Number of obs     =        758
                                                F(5, 752)         =      84.05
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3521
                                                Root MSE          =     .34641

------------------------------------------------------------------------------
             |               Robust
          lw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           s |    .102643   .0062099    16.53   0.000     .0904523    .1148338
        expr |   .0381189   .0066144     5.76   0.000      .025134    .0511038
      tenure |   .0356146   .0079988     4.45   0.000     .0199118    .0513173
         rns |  -.0840797    .029533    -2.85   0.005    -.1420566   -.0261029
        smsa |   .1396666    .028056     4.98   0.000     .0845893     .194744
       _cons |   4.103675   .0876665    46.81   0.000     3.931575    4.275775
------------------------------------------------------------------------------
```

IV estimator in Stata

3. Run an OLS regression.

- Annual return on investment in education : 10.26% (Significant at the 1% level).
- **Overestimation** coefficient : omitted the **ability** variable.
- Ability is positively correlated with years of education.
- The contribution of ability to wages is included into the contribution of education.

IV estimator in Stata

4. Using iq(智商) as a **proxy variable** for ability, run an OLS regression.
   - Other **proxy variables**: High school test scores; Armed Forces Qualification Test(美国参军资格考试), etc.

```
. reg lw s iq expr tenure rns smsa, r
Linear regression                               Number of obs     =        758
                                                F(6, 751)         =      71.89
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3600
                                                Root MSE          =     .34454

------------------------------------------------------------------------------
             |               Robust
          lw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           s |   .0927874   .0069763    13.30   0.000     .0790921    .1064826
          iq |   .0032792   .0011321     2.90   0.004     .0010567    .0055016
        expr |   .0393443   .0066603     5.91   0.000     .0262692    .0524193
      tenure |    .034209   .0078957     4.33   0.000     .0187088    .0497092
         rns |  -.0745325   .0299772    -2.49   0.013    -.1333815   -.0156834
        smsa |   .1367369   .0277712     4.92   0.000     .0822186    .1912553
       _cons |   3.895172   .1159286    33.60   0.000     3.667589    4.122754
------------------------------------------------------------------------------
```

IV estimator in Stata

4. Using iq(智商) as a **proxy variable** for ability, run an OLS regression.
   - Annual return on investment in education reduced to 9.28%.
   - More reasonable, but still too large.
   - So iq(智商) is an **endogenous variable**.

IV estimator in Stata

5. Run 2SLS regression.
   - Consider using med (母亲的受教育年限), kww (在 "knowledge of the World of Work" 测试中的成绩), mrt (=1, 已婚), age (年龄) as IVs of iq(智商).
   - With robust standard error

5. Run 2SLS regression.

```
Instrumental variables (2SLS) regression          Number of obs   =        758
                                                  Wald chi2(6)    =     355.73
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.2002
                                                  Root MSE        =     .38336

------------------------------------------------------------------------------
             |               Robust
          lw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          iq |  -.0115468   .0056376    -2.05   0.041    -.0225962   -.0004974
           s |   .1373477   .0174989     7.85   0.000     .1030506    .1716449
        expr |   .0338041   .0074844     4.52   0.000      .019135    .0484732
      tenure |    .040564   .0095848     4.23   0.000     .0217781      .05935
         rns |  -.1176984   .0359582    -3.27   0.001    -.1881751   -.0472216
        smsa |    .149983   .0322276     4.65   0.000     .0868182    .2131479
       _cons |   4.837875   .3799432    12.73   0.000       4.0932     5.58255
------------------------------------------------------------------------------
Instrumented:  iq
Instruments:   s expr tenure rns smsa med kww mrt age
```

IV estimator in Stata

5. Run 2SLS regression.

- Annual return on investment in education increased to 13.73%? (incredible)
- The contribution of iq(智商) to wages is negative? (incredible)
- We should check **instrument validity**.

IV estimator in Stata

6. Overidentification test.

- the number of instruments(4) $>$ the number of endogenous regressors(1)
- To test instrument exogeneity, thus overidentification test.

```
. estat overid
  Test of overidentifying restrictions:
  Score chi2(3)          =  51.5449  (p = 0.0000)
```

- Compute J-Statistic, some (or one) of the instrumental variables are invalid.
- We suspect mrt($=1$, 已婚), age(年龄) are invalid.
- Compute C-Statistic(检验部分工具变量不满足外生性) using *-ivreg2-*.

```
ssc install ivreg2
```

IV estimator in Stata

6. Overidentification test.

```
/* ivreg2默认估计量为2SLS
   orthog(mrt age):检验(mrt,age)是否满足外生性 */

. ivreg2 lw s expr tenure rns smsa (iq = med kww mrt age), r orthog (mrt age)
IV (2SLS) estimation

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity
                                                      Number of obs =       758
                                                      F(  6,    751) =     58.74
                                                      Prob > F      =    0.0000
Total (centered) SS     =  139.2861498                Centered R2   =    0.2002
Total (uncentered) SS   =  24652.24662                Uncentered R2 =    0.9955
Residual SS             =    111.39959                Root MSE      =     .3834

------------------------------------------------------------------------------
             |               Robust
          lw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          iq |  -.0115468   .0056376    -2.05   0.041    -.0225962   -.0004974
           s |   .1373477   .0174989     7.85   0.000     .1030506    .1716449
        expr |   .0338041   .0074844     4.52   0.000      .019135    .0484732
      tenure |    .040564   .0095848     4.23   0.000     .0217781      .05935
         rns |  -.1176984   .0359582    -3.27   0.001    -.1881751   -.0472216
        smsa |    .149983   .0322276     4.65   0.000     .0868182    .2131479
       _cons |   4.837875   .3799432    12.73   0.000       4.0932     5.58255
------------------------------------------------------------------------------
```

IV estimator in Stata

6. Overidentification test.

```
. ivreg2 lw s expr tenure rns smsa (iq = med kww mrt age), r orthog (mrt age)
Underidentification test (Kleibergen-Paap rk LM statistic):              33.294
                                                    Chi-sq(4) P-val =    0.0000
-------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):                10.538
                         (Kleibergen-Paap rk Wald F statistic):           9.585
Stock-Yogo weak ID test critical values:  5% maximal IV relative bias    16.85
                                         10% maximal IV relative bias    10.27
                                         20% maximal IV relative bias     6.71
                                         30% maximal IV relative bias     5.34
                                         10% maximal IV size             24.58
                                         15% maximal IV size             13.96
                                         20% maximal IV size             10.26
                                         25% maximal IV size              8.31
Source: Stock-Yogo (2005).  Reproduced by permission.
NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.
-------------------------------------------------------------------------------
Hansen J statistic (overidentification test of all instruments):         51.545
                                                    Chi-sq(3) P-val =    0.0000
-orthog- option:
Hansen J statistic (eqn. excluding suspect orthog. conditions):           0.116
                                                    Chi-sq(1) P-val =    0.7333
C statistic (exogeneity/orthogonality of suspect instruments):           51.429
                                                    Chi-sq(2) P-val =    0.0000
Instruments tested:   mrt age
-------------------------------------------------------------------------------
Instrumented:         iq
Included instruments: s expr tenure rns smsa
Excluded instruments: med kww mrt age
-------------------------------------------------------------------------------
```

IV estimator in Stata

6. Overidentification test.

- 使用-*ivreg2*-得到的回归系数和稳健标准误与-*ivregress*-相同；
- 拒绝（mrt age）满足外生性的原假设;
- 考虑仅使用（med kww）作为iq的工具变量。

IV estimator in Stata

7. Run 2SLS regression again

```
. ivregress 2sls lw s expr tenure rns smsa (iq=med kww), r first

First-stage regressions
-----------------------

                                                Number of obs     =        758
                                                F(   7,    750)   =      47.74
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3066
                                                Adj R-squared     =     0.3001
                                                Root MSE          =    11.3931

------------------------------------------------------------------------------
             |               Robust
          iq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           s |   2.467021   .2327755    10.60   0.000     2.010052     2.92399
        expr |  -.4501353   .2391647    -1.88   0.060    -.9196471    .0193766
      tenure |   .2059531    .269562     0.76   0.445    -.3232327    .7351388
         rns |  -2.689831   .8921335    -3.02   0.003    -4.441207    -.938455
        smsa |   .2627416   .9465309     0.28   0.781    -1.595424    2.120907
         med |   .3470133   .1681356     2.06   0.039     .0169409    .6770857
         kww |   .3081811   .0646794     4.76   0.000     .1812068    .4351553
       _cons |   56.67122   3.076955    18.42   0.000     50.63075    62.71169
------------------------------------------------------------------------------
```

7. Run 2SLS regression again

```
Instrumental variables (2SLS) regression          Number of obs   =        758
                                                  Wald chi2(6)    =     370.04
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.2775
                                                  Root MSE        =     .36436

------------------------------------------------------------------------------
             |               Robust
          lw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          iq |   .0139284   .0060393     2.31   0.021     .0020916    .0257653
           s |   .0607803   .0189505     3.21   0.001      .023638    .0979227
        expr |   .0433237   .0074118     5.85   0.000     .0287968    .0578505
      tenure |   .0296442    .008317     3.56   0.000     .0133432    .0459452
         rns |  -.0435271   .0344779    -1.26   0.207    -.1111026    .0240483
        smsa |   .1272224   .0297414     4.28   0.000     .0689303    .1855146
       _cons |   3.218043   .3983683     8.08   0.000     2.437256    3.998831
------------------------------------------------------------------------------
Instrumented:  iq
Instruments:   s expr tenure rns smsa med kww
```

IV estimator in Stata

7. Run 2SLS regression again
   - Annual return on investment in education reduced to 6.08%, which is reasonable.
   - The contribution of iq(智商) to wages turns to positive again.

IV estimator in Stata

8. Check instrument validity

- Check **IV relevance** : report the first stage.
  - Instrument perform well in the first stage.
  - A more formal test : F-statistic exceeds 10 (13.40), no Weak Instruments.

```
. estat firststage, all forcenonrobust
  First-stage regression summary statistics
  --------------------------------------------------------------------------
               |            Adjusted      Partial       Robust
      Variable |   R-sq.       R-sq.        R-sq.     F(2,750)   Prob > F
  -------------+------------------------------------------------------------
            iq |  0.3066      0.3001       0.0382      13.4028    0.0000
  --------------------------------------------------------------------------

  Shea's partial R-squared
  --------------------------------------------------
               |     Shea's              Shea's
      Variable |  Partial R-sq.   Adj. Partial R-sq.
  -------------+------------------------------------
            iq |     0.0382              0.0305
  --------------------------------------------------

  Minimum eigenvalue statistic = 14.9058
  Critical Values                      # of endogenous regressors:    1
  Ho: Instruments are weak             # of excluded instruments:     2
  ---------------------------------------------------------------------
                                     |    5%     10%     20%     30%
  2SLS relative bias                 |         (not available)
  -----------------------------------+---------------------------------
                                     |   10%     15%     20%     25%
  2SLS Size of nominal 5% Wald test  | 19.93   11.59    8.75    7.25
```

IV estimator in Stata

8. Check instrument validity

- Run and examine the **reduced form**.

```
qui reg lw s expr tenure rns smsa med, r
est store m1

qui reg lw s expr tenure rns smsa kww, r
est store m2

qui reg lw s expr tenure rns smsa med kww, r
est store m3

esttab m1 m2 m3,
    mtitle("reduced form:med" "reducedform:kww" "reducedform:med,kww")
    b(%6.3f) nogap compress
    star(* 0.1 ** 0.05 *** 0.01)
    ar2 order(med kww)
```

IV estimator in Stata

8. Check instrument validity

- Run and examine the **reduced form**.

	(1) reduced~d	(2) reduced~w	(3) reduced~w
med	0.007		0.007
	(1.55)		(1.38)
kww		0.004**	0.004*
		(2.07)	(1.96)
s	0.100***	0.097***	0.095***
	(15.36)	(15.01)	(14.06)
expr	0.039***	0.037***	0.037***
	(5.88)	(5.42)	(5.52)
tenure	0.036***	0.032***	0.033***
	(4.49)	(4.04)	(4.10)
rns	-0.079***	-0.085***	-0.080***
	(-2.62)	(-2.87)	(-2.66)
smsa	0.138***	0.132***	0.131***
	(4.92)	(4.65)	(4.61)
_cons	4.058***	4.039***	4.002***
	(44.13)	(42.70)	(40.98)
N	758	758	758
adj. R-sq	0.349	0.351	0.352

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

IV estimator in Stata

8. Check instrument validity

- Run and examine the **reduced form**.
  - Reduced form is proportional to the casual effect of interest and is unbiased(OLS), so we should see the causal relation in the reduced form.
  - If you can't see the causal relation in the reduced form, it's probably not there.
  - **Notice!** Probably med is not exogenous enough (or not a very good IV).

IV estimator in Stata

8. Check instrument validity

- Check **exclusion restriction** : Overidentification test.
  - Instrumental variables (med kww) satisfy exogeneity.

```
. qui ivregress 2sls lw s expr tenure rns smsa (iq=med kww), r
. estat overid
  Test of overidentifying restrictions:
  Score chi2(1)          =   .151451   (p = 0.6972)
```

IV estimator in Stata

9. Check 2SLS estimates with LIML.
   - LIML is less than precise than 2SIS but also less biased.
   - If the results come out similar, be happy.

```
. ivregress liml lw s expr tenure rns smsa (iq=med kww), r
Instrumental variables (LIML) regression        Number of obs   =        758
                                                Wald chi2(6)    =     369.62
                                                Prob > chi2     =     0.0000
                                                R-squared       =     0.2768
                                                Root MSE        =     .36454

------------------------------------------------------------------------------
             |               Robust
          lw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          iq |   .0139764   .0060681     2.30   0.021     .0020831    .0258697
           s |   .0606362    .019034     3.19   0.001     .0233303    .0979421
        expr |   .0433416   .0074185     5.84   0.000     .0288016    .0578816
      tenure |   .0296237    .008323     3.56   0.000     .0133109    .0459364
         rns |  -.0433875    .034529    -1.26   0.209    -.1110631    .0242881
        smsa |   .1271796   .0297599     4.27   0.000     .0688512     .185508
       _cons |   3.214994   .4001492     8.03   0.000     2.430716    3.999272
------------------------------------------------------------------------------
Instrumented:  iq
Instruments:   s expr tenure rns smsa med kww
```

IV estimator in Stata

10. Put all estimates into one table.

```
qui reg lw s expr tenure rns smsa,r
est sto ols_noiq
qui reg lw iq s expr tenure rns smsa,r
est sto ols_iq
qui ivreg2 lw s expr tenure rns smsa (iq=med kww), r
est sto tsls
qui ivreg2 lw s expr tenure rns smsa (iq=med kww), r liml
est sto liml

esttab ols_noiq ols_iq tsls liml, mtitle    ///
    star(* 0.1 ** 0.05 *** 0.01)            ///
    b(%6.3f) nogap compress order(s iq)     ///
    stats(rkf j jp N r2_a, labels("First-stage F-statistic" ///
    "Overidentifying restrictions J-test and P-value" N r2_a) layout(@ `""@ (@)""' @ @) )
```

IV estimator in Stata

10. Put all estimates into one table.

	(1) ols_noiq	(2) ols_iq	(3) tsls	(4) liml
s	0.103***	0.093***	0.061***	0.061***
	(16.53)	(13.30)	(3.21)	(3.19)
smsa	0.140***	0.137***	0.127***	0.127***
	(4.98)	(4.92)	(4.28)	(4.27)
iq		0.003***	0.014**	0.014**
		(2.90)	(2.31)	(2.30)
expr	0.038***	0.039***	0.043***	0.043***
	(5.76)	(5.91)	(5.85)	(5.84)
tenure	0.036***	0.034***	0.030***	0.030***
	(4.45)	(4.33)	(3.56)	(3.56)
rns	-0.084***	-0.075**	-0.044	-0.043
	(-2.85)	(-2.49)	(-1.26)	(-1.26)
_cons	4.104***	3.895***	3.218***	3.215***
	(46.81)	(33.60)	(8.08)	(8.03)
First-st~c			13.403	13.403
Overiden~t			0.151 (0.697)	0.151 (0.697)
N	758.000	758.000	758.000	758.000
r2_a	0.348	0.355	0.272	0.271

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

IV estimator in Stata

- Followed research
  - Return on investment in education is the core issue of labor economics.
  - Behrman et al(1980) compared **identical twins** with different years of education to controll the factors such as genetics and family background.
  - Angrist and Krueger(1991) used **the quarter of birth** as the instrumental variable of years of education.
  - Bound et al(1995) found the quarter of birth is a **weak** instrument.
  - Buckles and Hungerman(2012)'s latest research showed that the quarter of birth is **not** independent of family background.

Section 2

RDD in Stata

Subsection 1

Review the Theory

RDD in Stata

- Review the Theory : Summary

RDD in the toolkit of Causal Inference

- It is so called the **nearest** method to RCT which identify causal effect of treatment on outcome.
- RDD needs a arbitrary cut-off and agents can **imperfect** manipulate the treatment.
- Two types
  - Sharp RD
  - Fuzzy RD
- Assumption: continued at the cut-off
- Concerns:
  - Functional form
  - Bandwidth selection
  - Bin selection

RDD in Stata

- Review the Theory : Main idea

Main Idea of Regression Discontinuity Design

- Regression Discontinuity Design (RDD) exploits the facts that:
  - Some rules are *arbitrary* and generate a *discontinuity* in treatment assignment.
  - The treatment assignment is determined based on whether a unit exceeds some threshold on a variable (**assignment variable**, **running variable** or **forcing variable**)
  - Assume other factors *do NOT change* abruptly at threshold.
  - Then any change in outcome of interest can be attributed to the assigned treatment.

RDD in Stata

- Review the Theory : Two types

Sharp RDD and Fuzzy RDD

- In general, depending on enforcement of treatment assignment, RDD can be categorized into two types:
  1. **Sharp RDD**: nobody below the cutoff gets the "treatment", everybody above the cutoff gets it
     - Everyone follows treatment assignment rule (all are compliers).
     - Local randomized experiment with perfect compliance around cutoff.
  2. **Fuzzy RDD**: the probability of getting the treatment jumps discontinuously at the cutoff (NOT jump from 0 to 1)
     - Not everyone follows treatment assignment rule.
     - Local randomized experiment with partial compliance around cutoff.
     - Using initial assignment as an instrument for actual treatment.

RDD in Stata

- Review the Theory : Assumption
  - **Deterministic Assumption**

$$D_i = 1(X_i \geq c)$$

  - Treatment assignment is a deterministic function of the assignment variable X_i and the threshold c.
  - **Continuity Assumption**
    - $E[Y_{1i}|X_i]$ and $E[Y_{0i}|X_i]$ are continuous at $X_i = c$
  - Assume potential outcomes do not change at cutoff.
  - This means that except treatment assignment, all other unobserved determinants of Y_i are continuous at cutoff c.
  - This implies no other confounding factor affects outcomes at cutoff c.
  - Any observed discontinuity in the outcome can be attributed to treatment assignment.

RDD in Stata

- Review the Theory : Identification in Sharp RD

Sharp RDD specification

- A simple RD regression is

$$Y_i = \alpha + \rho D_i + \gamma(X_i - c) + u_i$$

  - Y_i is the outcome variable
  - D_I is the the treatment variable(indepent variable)
  - X_i is the running variable
  - c is the value of cut-off
  - u_i is the error term including other factors
- **Qustion**: Which parameter do we care about the most?

- Review the Theory : Identification in Sharp RD
  - More generally,we could also estimate two separate regressions for each side respectively.

$$Y_i^b = \beta^b + f(X_i^b - c) + u_i^b$$
$$Y_i^a = \beta^a + g(X_i^a - c) + u_i^a$$

  - Can do all in one step; just use all the data at once and estimate:

$$Y_i = \alpha + \rho D_i + f(X_i - c) + D_i \times h(X_i - c) + u_i$$

where D_i is a dummy variable for treated status.

RDD in Stata

- Review the Theory : Identification in Fuzzy RD
  - Encourage Variable:

$$Z_i = 1 \; if \; assign \; to \; treatment \; group$$
$$Z_i = 0 \; if \; assign \; to \; control \; group$$

  - Then the **First Stage** of FRD regression:

$$P(D_i = 1|x_i) = \alpha_1 + \phi Z_i + f(x_i - c) + Z_i \times g(x_i - c) + \eta_{1i}$$

  - The **second stage** regression is

$$Y_i = \alpha_2 + \delta \hat{D}_i + f(x_i - c) + \hat{D}_i \times g(x_i - c) + \eta_{2i}$$

  - The **reduced form** regression in FRD is

$$Y_i = \alpha_3 + \beta Z_i + f(x_i - c) + Z_i \times g(x_i - c) + \eta_{3i}$$

Subsection 2

Introduction : Package & Commands

RDD in Stata

- Introduction : Package & Commands
  - Package : [Install Link]
  - **rdrobust** package : inference and graphical procedures using local polynomial and partitioning regression methods.
    - -*rdrobust*- : Local Polynomial Regression Discontinuity Estimation with Robust Bias-Corrected Confidence Intervals and Inference Procedures.
    - -*rdbwselect*- : Data-driven Bandwidth Selection,
    - -*rdplot*- : Data-Driven Regression Discontinuity Plots.
  - **rddensity** package : manipulation testing using local polynomial density methods.
  - **Others** : -*cmogram*- ; -*rd*- ; -*rdcv*- ; -*DCdensity*- ; ...

Subsection 3

Example for Sharp RDD

- Example for Sharp RDD
  - Data : the dataset comes from a study on **party advantages** in **U.S. Senate elections** for the period 1914–2010.
  - We focus here on the RD effect of **the Democratic party winning a U.S. Senate seat** on **the vote share obtained in the following election for that same seat**.
  - The unit of observation is the **state**.
  - Main variables :

    demmv : ranges from -100 to 100 and records the Democratic party's margin of victory in the statewide election for a given U.S. Senate seat (the vote share of the Democratic party − the vote share of its strongest opponent).

    demvoteshfor2 : ranges from 0 to 100 and records the Democratic vote share in the following election for the same seat.
  - To estimate the **incumbency advantage of parties** with an RD design.

RDD in Stata

- Example for Sharp RDD
  - Re-labeling the three main variables

```
. use senate, clear

. sum

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       state |      1,390    40.01367    21.99304          1         82
        year |      1,390     1964.63    28.05466       1914       2010
       dopen |      1,380    .2471014    .4314826          0          1
  population |      1,390     3827919     4436950      78000   3.73e+07
presdemvot~1 |      1,387    46.11975    14.31701          0   97.03408
-------------+---------------------------------------------------------
       demmv |      1,390    7.171159    34.32488       -100        100
demvoteshl~1 |      1,349    52.69048     18.2706          0        100
demvoteshl~2 |      1,308    52.86918    18.23913          0        100
demvoteshf~1 |      1,341    52.41856    18.36641          0        100
demvoteshf~2 |      1,297    52.66627    18.12219          0        100
-------------+---------------------------------------------------------
  demwinprv1 |      1,349    .5441067    .4982355          0          1
  demwinprv2 |      1,308     .543578    .4982879          0          1
    dmidterm |      1,390    .5136691     .499993          0          1
    dpresdem |      1,390    .3884892    .4875822          0          1
```

RDD in Stata

- Example for Sharp RDD
  - Re-labeling the three main variables

    Assignment variable (running variable) : X

    Outcome variable: Y

    Treatment variable : T

    Threshold (cutoff) for treatment assignment : c=0

RDD in Stata

- Example for Sharp RDD
  - Re-labeling the three main variables

```
. rename demmv X                //X--民主党获胜的差额
. rename demvoteshfor2 Y        //Y--t+2期民主党得票数

. gen T=.
. replace T=0 if X<0 & X!=.
. replace T=1 if X>=0 & X!=.
 /* the Democratic party wins the election for that seat.*/

. label var T "Democratic Win at t"
```

RDD in Stata

- Example for Sharp RDD
  - Check RD's type

```
. gen ranwin=(X>=0)

. tab ranwin T

           |  Democratic Win at t
    ranwin |         0          1 |     Total
-----------+----------------------+----------
         0 |       640          0 |       640
         1 |         0        750 |       750
-----------+----------------------+----------
     Total |       640        750 |     1,390
```

RDD in Stata

- Example for Sharp RDD
  - Show the scatter plot of the raw data (where each point is an observation).

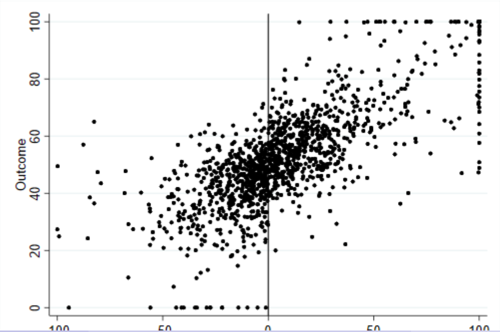
```
. twoway (scatter Y X, msize(vsmall)                 ///
      mcolor(black) xline(0, lcolor(black))),       ///
      graphregion(color(white)) ytitle(Outcome)     ///
      xtitle(Score)

. graph export fig1.png, width(500) replace
(file fig1.png written in PNG format)
```

RDD in Stata

- Example for Sharp RDD
  - Show the scatter plot of the raw data (where each point is an observation).
  - Often hard to see "jumps" or discontinuities in the outcome-score relationship by simply looking at the raw data
  - Two problems :

    样本太多时不够直观；

    实际分析时中跳跃现象可能不那么清晰。



RDD in Stata

- Example for Sharp RDD
  - Three Steps:
    1. Graph the data for visual inspection
    2. Estimate the treatment effect using regression methods
    3. Run checks on assumptions underlying research design

Subsection 4

Example for Sharp RDD : Step 1

Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-
  - A more useful approach is to aggregate or "smooth" the data before plotting.
  - The typical RD plot presents two ingredients :

    (i) a global polynomial fit, represented by a **solid line**, using the original **raw** data.

    (ii) local sample means, represented by **dots**, choosing bins of the score, calculating the mean of the outcome for the observations falling within each bin, and then plotting the **average outcome in each bin** against the mid point of the bin.

Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-

```
preserve
rdplot Y X, nbins(20 20) genvars support(-100 100)
gen obs = 1
collapse (mean) rdplot_mean_x rdplot_mean_y (sum) obs, by (rdplot_id)
order rdplot_id
tabstat rdplot_mean_x rdplot_mean_y obs,by(rdplot_id)
restore
```

Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-
  - Bin selection (1) : Choosing the Location of Bins
    1. **Evenly-spaced bins** : bins that have equal length.

```
. rdplot Y X, nbins(20 20) binselect(es)        ///
      graph_options(graphregion(color(white))   ///
      xtitle(Score) ytitle(Outcome))


RD Plot with RD plot with manually set number of bins.
         Cutoff c = 0 | Left of c  Right of c        Number of obs  =       1297
----------------------+-----------------------        Kernel         =    Uniform
        Number of obs |       595         702
   Eff. Number of obs |       595         702
  Order poly. fit (p) |         4           4
     BW poly. fit (h) |   100.000     100.000
 Number of bins scale |     1.000       1.000
```

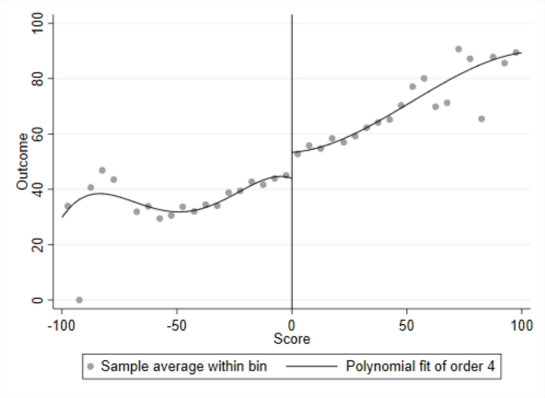
Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-
  - Bin selection (1) : Choosing the Location of Bins
    1. **Evenly-spaced bins** : bins that have equal length.

```
Outcome: Y. Running variable: X.
---------------------------------------------------
                        |  Left of c   Right of c
---------------------------------------------------
          Bins selected |         20           20
     Average bin length |      5.000        5.000
      Median bin length |      5.000        5.000
---------------------------------------------------
      IMSE-optimal bins |          8            9
    Mimicking Var. bins |         15           35
---------------------------------------------------
Rel. to IMSE-optimal:   |
          Implied scale |      2.500        2.222
      WIMSE var. weight |      0.060        0.084
     WIMSE bias weight |      0.940        0.916
---------------------------------------------------

. graph export fig2.png,width(500) replace
(note: file fig2.png not found)
(file fig2.png written in PNG format)
```

Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-
  - Bin selection (1) : Choosing the Location of Bins
    1. **Evenly-spaced bins** : bins that have equal length.



Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-
  - Bin selection (1) : Choosing the Location of Bins
    2. **Quantile-spaced bins** : bins that contain (roughly) the same number of observations.

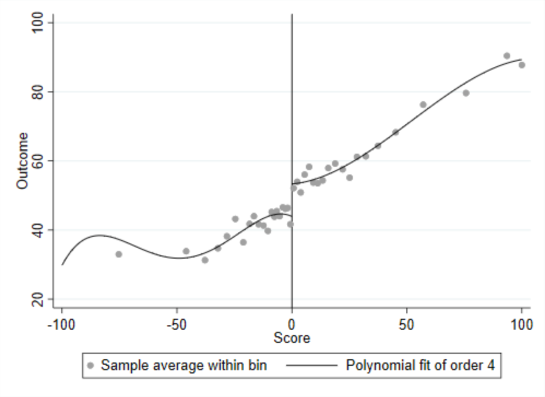
```
.   rdplot Y X, nbins(20 20) binselect(qs)       ///
      graph_options(graphregion(color(white))    ///
      xtitle(Score) ytitle(Outcome))
RD Plot with RD plot with manually set number of bins.
         Cutoff c = 0 | Left of c  Right of c        Number of obs  =       1297
----------------------+-----------------------       Kernel         =    Uniform
        Number of obs |       595         702
   Eff. Number of obs |       595         702
  Order poly. fit (p) |         4           4
     BW poly. fit (h) |   100.000     100.000
 Number of bins scale |     1.000       1.000
```

- RDD graphical analysis : -*rdplot*-
  - Bin selection (1) : Choosing the Location of Bins
    2. **Quantile-spaced bins** : bins that contain (roughly) the same number of observations.

```
Outcome: Y. Running variable: X.
─────────────────────────────────────────────
                       │ Left of c  Right of c
─────────────────────────────────────────────
          Bins selected│        20          20
     Average bin length│     5.000       5.000
      Median bin length│     1.912       2.771
─────────────────────────────────────────────
      IMSE-optimal bins│        21          16
    Mimicking Var. bins│        28          49
─────────────────────────────────────────────
Rel. to IMSE-optimal:  │
          Implied scale│     0.952       1.250
      WIMSE var. weight│     0.537       0.339
     WIMSE bias weight │     0.463       0.661
─────────────────────────────────────────────

. graph export fig3.png,width(500) replace
(note: file fig3.png not found)
(file fig3.png written in PNG format)
```

Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-
  - Bin selection (1) : Choosing the Location of Bins
    2. **Quantile-spaced bins** : bins that contain (roughly) the same number of observations.



Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-
  - Bin selection (2) : Choosing the Number of Bins
    1. Integrated Mean Squared Error (IMSE) Method
       If we choose a large number of bins (narrower) :
       small bias – the bins are smaller and the **local constant** fit is better.
       less precisely – less observations **per bin**, thus more variability within bin.
       balance squared-bias and variance so that the IMSE is (approximately) minimized.

Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-
  - Bin selection (2) : Choosing the Number of Bins
    1. Integrated Mean Squared Error (IMSE) Method

```
. rdplot Y X, binselect(es)                         ///
      graph_options(graphregion(color(white))  ///
      xtitle(Score) ytitle(Outcome))

/* The IMSE criterion leads to different numbers of ES bins above and
below the cutoff.*/

RD Plot with evenly spaced number of bins using spacings estimators.
         Cutoff c = 0 | Left of c  Right of c        Number of obs  =       1297
----------------------+----------------------        Kernel         =    Uniform
        Number of obs |       595         702
   Eff. Number of obs |       595         702
  Order poly. fit (p) |         4           4
     BW poly. fit (h) |   100.000     100.000
 Number of bins scale |     1.000       1.000
```

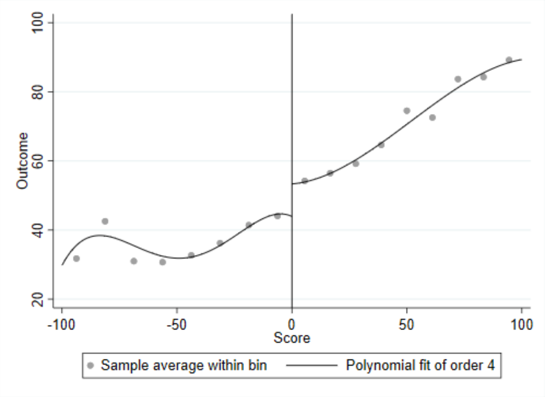
Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-
  - Bin selection (2) : Choosing the Number of Bins
    1. Integrated Mean Squared Error (IMSE) Method

```
Outcome: Y. Running variable: X.
---------------------------------------------------
                          |  Left of c  Right of c
---------------------------------------------------
           Bins selected  |          8           9
      Average bin length  |     12.500      11.111
       Median bin length  |     12.500      11.111
---------------------------------------------------
       IMSE-optimal bins  |          8           9
     Mimicking Var. bins  |         15          35
---------------------------------------------------
Rel. to IMSE-optimal:     |
           Implied scale  |      1.000       1.000
       WIMSE var. weight  |      0.500       0.500
      WIMSE bias weight   |      0.500       0.500
---------------------------------------------------

. graph export fig4.png,width(500) replace
(note: file fig4.png not found)
(file fig4.png written in PNG format)
```

Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-
  - Bin selection (2) : Choosing the Number of Bins
    1. Integrated Mean Squared Error (IMSE) Method



Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-
  - Bin selection (2) : Choosing the Number of Bins

1. Integrated Mean Squared Error (IMSE) Method

```
rdplot Y X, binselect(qs)                          ///
    graph_options(graphregion(color(white))  ///
    xtitle(Score) ytitle(Outcome))
```

Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-
  - Bin selection (2) : Choosing the Number of Bins
    2. Mimicking Variance (MV) Method
       "mimics" the overall variability in the raw scatter plot of the data.
       MV method leads to a larger number of bins than the IMSE method.
       More dots representing local means, thus giving a better sense of the variability of the data.

Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-
  - Bin selection (2) : Choosing the Number of Bins
    2. Mimicking Variance (MV) Method

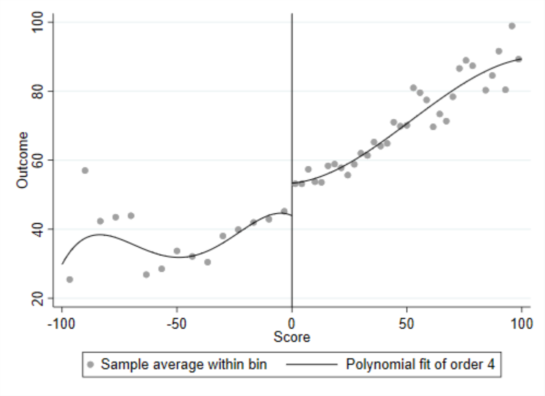
```
.   rdplot Y X, binselect(esmv)                   ///Default
      graph_options(graphregion(color(white))  ///
      xtitle(Score) ytitle(Outcome))

RD Plot with evenly spaced mimicking variance number of bins using spacings estimato
         Cutoff c = 0 | Left of c  Right of c        Number of obs  =       1297
----------------------+----------------------        Kernel         =    Uniform
        Number of obs |       595         702
   Eff. Number of obs |       595         702
  Order poly. fit (p) |         4           4
     BW poly. fit (h) |   100.000     100.000
 Number of bins scale |     1.000       1.000
```

- RDD graphical analysis : -*rdplot*-
  - Bin selection (2) : Choosing the Number of Bins
    2. Mimicking Variance (MV) Method

```
Outcome: Y. Running variable: X.
─────────────────────────────────────────────
                        │  Left of c  Right of c
────────────────────────┼────────────────────
           Bins selected │         15          35
      Average bin length │      6.667       2.857
       Median bin length │      6.667       2.857
────────────────────────┼────────────────────
       IMSE-optimal bins │          8           9
     Mimicking Var. bins │         15          35
────────────────────────┼────────────────────
Rel. to IMSE-optimal:    │
           Implied scale │      1.875       3.889
       WIMSE var. weight │      0.132       0.017
      WIMSE bias weight  │      0.868       0.983
─────────────────────────────────────────────

. graph export fig5.png,width(500) replace
(file fig5.png written in PNG format)
```

Example for Sharp RDD

- RDD graphical analysis : -*rdplot*-
  - Bin selection (2) : Choosing the Number of Bins
    2. Mimicking Variance (MV) Method



Example for Sharp RDD

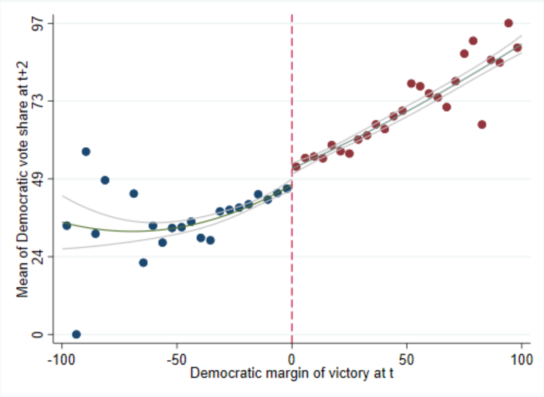
- RDD graphical analysis : -*rdplot*-
  - Bin selection (2) : Choosing the Number of Bins
    2. Mimicking Variance (MV) Method

```
rdplot Y X, binselect(qsmv)                          ///
    graph_options(graphregion(color(white))  ///
    xtitle(Score) ytitle(Outcome))
```

Example for Sharp RDD

- RDD graphical analysis : -*cmogram*-

```
. cmogram Y X, cut(0) scatter lineat(0) qfitci

. graph export fig6.png,width(500) replace
(note: file fig6.png not found)
(file fig6.png written in PNG format)
```



Subsection 5

Example for Sharp RDD : Step 2

Example for Sharp RDD

- Estimate the treatment effect using regression methods
    - There are 2 types of strategies for correctly specifying the functional form in a RDD:
        1. **Parametric**/global method: Use all available observations and Estimate treatment effects based on a specific functional form for the outcome and assignment variable relationship.
        2. **Nonparametric**/local method:Use the observations around cutoff: Compare the outcome of treated and untreated observations that lie within specific bandwidth.

Example for Sharp RDD

- Estimate the treatment effect using regression methods
  - Parametric/Global Approach (全局多项式回归)

```
sum X
local hvalueR=r(max)
local hvalueL= abs(r(min))

rdrobust Y X, h(`hvalueL' `hvalueR')      //自动选择阶数
rdrobust Y X, h(`hvalueL' `hvalueR') p(2) //二阶拟合
rdrobust Y X, h(`hvalueL' `hvalueR') p(3) //三阶拟合
```

Example for Sharp RDD

- Estimate the treatment effect using regression methods
  - Nonparametric/Local Approach : local linear regression (局部线性回归)
  - 三种方法 (任选):

    方法一：standard least-squares estimation (OLS)

    方法二：-rdrobust-进行的非参数估计

    方法三：-rd-进行的非参数估计

Example for Sharp RDD

- Estimate the treatment effect using regression methods
  - Nonparametric/Local Approach : local linear regression (局部线性回归)

    方法一：standard least-squares estimation (OLS)

```
. rdbwselect Y X, c(0) kernel(uni) bwselect(mserd)  //选择最优带宽h
Bandwidth estimators for sharp RD local polynomial regression.
      Cutoff c = | Left of c  Right of c            Number of obs =       1297
-----------------+----------------------            Kernel        =    Uniform
                 |                                  VCE method    =         NN
    Number of obs |       595         702
         Min of X |  -100.000       0.036
         Max of X |    -0.079     100.000
    Order est. (p) |         1           1
    Order bias (q) |         2           2
Outcome: Y. Running variable: X.
-------------------------------------------------------------------------------
                 |        BW est. (h)            |        BW bias (b)
          Method | Left of c          Right of c | Left of c          Right of c
-----------------+-------------------------------+-------------------------------
           mserd |    11.597              11.597 |    22.944              22.944
-------------------------------------------------------------------------------
```

Example for Sharp RDD

- Estimate the treatment effect using regression methods
  - Nonparametric/Local Approach : local linear regression (局部线性回归)

    方法一：standard least-squares estimation (OLS)

```
preserve
keep if X>=-11.597 & X<=11.597

local i=1
    forvalues i=2/4
    gen X`i´=X^`i´
                                                    // 产生分配变量的平方、三次方、四次方

eststo x1 : qui reg Y 1.T, r
eststo x2 : qui reg Y T##c.X, r              //局部线性回归法，一阶
eststo x3 : qui reg Y T##c.(X X2), r         //局部线性回归法，选择2阶多项式
eststo x4 : qui reg Y T##c.(X X2 X3), r      //局部线性回归法，选择3阶多项式
eststo x5 : qui reg Y T##c.(X X2 X3 X4), r //局部线性回归法，选择4阶多项式

esttab x1 x2 x3 x4 x5,                ///
    star(* .1 ** .05  * .01)          ///
    nogap nonumber replace            ///
    drop(0.T*) se(%5.4f) ar2 aic(%10.4f) bic(%10.4f)
restore
```

Example for Sharp RDD

- Estimate the treatment effect using regression methods
  - Nonparametric/Local Approach : local linear regression (局部线性回归)

方法一：standard least-squares estimation (OLS)

	Y	Y	Y	Y	Y
1.T	9.762*	7.202*	8.832*	13.27*	16.08*
	(0.8349)	(1.6332)	(2.3939)	(3.1258)	(3.8665)
X		0.240	-1.670**	-2.793	-7.721**
		(0.2028)	(0.7598)	(1.8143)	(3.7073)
1.T#c.X		-0.0122	3.029*	0.566	6.207
		(0.2585)	(0.9956)	(2.3214)	(4.7669)
X2			-0.169**	-0.405	-2.285*
			(0.0667)	(0.3493)	(1.3548)
1.T#c.X2			0.0691	1.110**	2.708
			(0.0860)	(0.4670)	(1.7333)
X3				-0.0135	-0.265
				(0.0196)	(0.1819)
1.T#c.X3				-0.0341	0.256
				(0.0267)	(0.2288)
X4					-0.0109
					(0.0079)
1.T#c.X4					0.00921
					(0.0099)
_cons	44.28*	45.60*	41.91*	40.71*	37.51*
	(0.6010)	(1.2794)	(1.8869)	(2.5757)	(3.1696)
N	506	506	506	506	506
adj. R-sq	0.209	0.211	0.224	0.230	0.231
AIC	3709.6679	3710.3097	3704.0478	3702.2052	3703.7348
BIC	3718.1210	3727.2159	3729.4070	3736.0175	3746.0002

Standard errors in parentheses
* p<.1, ** p<.05, * p<.01

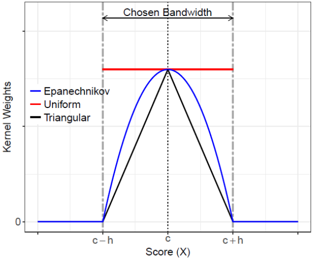
Example for Sharp RDD

- Estimate the treatment effect using regression methods
  - Nonparametric/Local Approach : local linear regression (局部线性回归)

方法二：-*rdrobust*-进行的非参数估计

p : set the order of the polynomial. Default is p(1).

kernel : set the kernel. Default is kernel(triangular).



h : choose the bandwidth manually.

c : sets the RD cutoff. Default is c(0).

Example for Sharp RDD

- Estimate the treatment effect using regression methods
  - Nonparametric/Local Approach : local linear regression (局部线性回归)

    方法二：-*rdrobust*-进行的非参数估计

    bwselect() : bandwidth selection procedure to be used. Default is bwselect(mserd).

    If a smaller h :

    fewer observations—increase the variance of the estimated coefficients.

    local polynomial approximation—will reduce treatment effect biase.

    MSE : bias-variance trade-off.

Example for Sharp RDD

- Estimate the treatment effect using regression methods
  - Nonparametric/Local Approach : local linear regression (局部线性回归)

    方法二：-*rdrobust*-进行的非参数估计

```
rdrobust Y X, kernel(uniform) p(1)
rdrobust Y X, c(0) kernel(uni) bwselect(mserd) p(2) h(11.597) all
rdrobust Y X, c(0) kernel(uni) bwselect(mserd) p(3) h(11.597) all
rdrobust Y X, c(0) kernel(uni) bwselect(mserd) p(4) h(11.597) all
```

Example for Sharp RDD

- Estimate the treatment effect using regression methods
  - Nonparametric/Local Approach : local linear regression (局部线性回归)

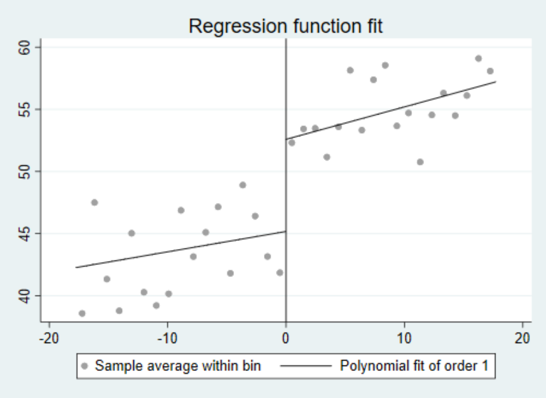
    方法二：-*rdrobust*-进行的非参数估计

```
* Using rdrobust and showing the associated rdplot

. rdrobust Y X, p(1) kernel(triangular) bwselect(mserd)
. eret list
. local bandwidth = e(h_l)
. rdplot Y X if abs(X) <= `bandwidth´, p(1) h(`bandwidth´) kernel(triangular)

. graph export fig7.png,width(500) replace
```

Example for Sharp RDD

- Estimate the treatment effect using regression methods
  - Nonparametric/Local Approach : local linear regression (局部线性回归)

方法二：-*rdrobust*-进行的非参数估计



Example for Sharp RDD

- Estimate the treatment effect using regression methods
  - Nonparametric/Local Approach : local linear regression (局部线性回归)

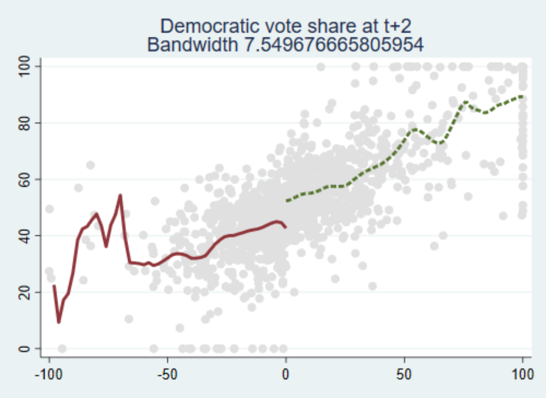
方法三：-*rd*-进行的非参数估计

```
. rd Y X, mbw(100) gr z0(0) kenel(tri)   //给出了带宽取最优带宽50%和200%的回归结果

Two variables specified; treatment is
assumed to jump from zero to one at Z=0.
 Assignment variable Z is X
 Treatment variable X_T unspecified
 Outcome variable y is Y
Command used for graph: lpoly; Kernel used: triangle (default)
Bandwidth: 7.5496767; loc Wald Estimate: 9.6449759
(93 missing values generated)
(93 missing values generated)
(93 missing values generated)
Estimating for bandwidth 7.549676665805968
------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwald |   9.644976     2.1155     4.56   0.000     5.498673    13.79128
------------------------------------------------------------------------------

. graph export fig8.png,width(500) replace
(note: file fig8.png not found)
(file fig8.png written in PNG format)
```

Example for Sharp RDD

- Estimate the treatment effect using regression methods
  - Nonparametric/Local Approach : local linear regression (局部线性回归)

方法三：-rd-进行的非参数估计



Subsection 6

Example for Sharp RDD : Step 3

Example for Sharp RDD

- Testing the Validity of the RDD
  1. **Test involving covariates(Nonoutcome Variable)** : Test whether other covariates exhibit a jump at the discontinuity
  2. **Test sorting behavior** : Testing discontinuity in the density of assignment variable X
  3. **Falsification Tests** :
     1) Placebo Cutoffs
     2) Sensitivity to Observations near the Cutoffs
     3) Sensitivity to Bandwidth Choice

Example for Sharp RDD

- Testing the Validity of the RDD

  1. **Test involving covariates(Nonoutcome Variable)**
     Using rdbwselect with covariates.

```
. global covariates "presdemvoteshlag1 demvoteshlag1 demvoteshlag2 demwinprv1 demwinprv2 dmidte rm dpresder
. rdrobust Y X, covs($covariates) p(1) kernel(tri) bwselect(mserd)

Covariate-adjusted sharp RD estimates using local polynomial regression.
      Cutoff c = 0 | Left of c  Right of c            Number of obs =       1213
-------------------+----------------------            BW type       =      mserd
     Number of obs |       555         658            Kernel        = Triangular
Eff. Number of obs |       326         295            VCE method    =         NN
    Order est. (p) |         1           1
    Order bias (q) |         2           2
       BW est. (h) |    17.266      17.266
       BW bias (b) |    27.178      27.178
         rho (h/b) |     0.635       0.635
Outcome: Y. Running variable: X.
--------------------------------------------------------------------------------
            Method |   Coef.    Std. Err.    z     P>|z|    [95% Conf. Interval]
-------------------+------------------------------------------------------------
      Conventional |  7.0876     1.4767   4.7995   0.000   4.19327      9.98194
            Robust |     -          -     4.0449   0.000   3.67067       10.572
--------------------------------------------------------------------------------
Covariate-adjusted estimates. Additional covariates included: 7
```

Example for Sharp RDD

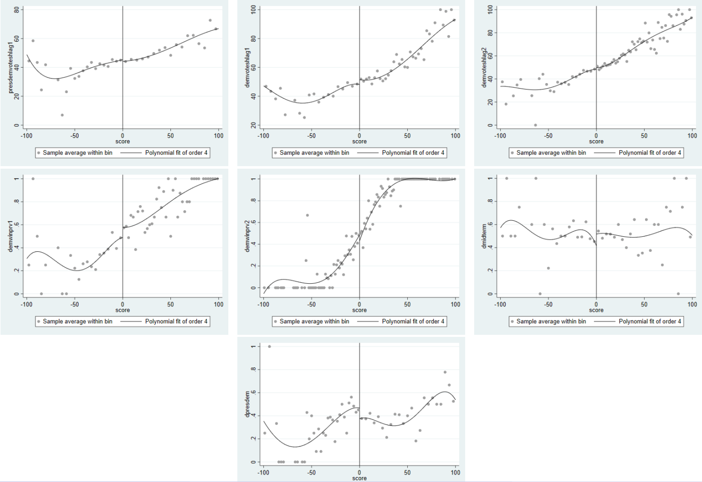
- Testing the Validity of the RDD

  1. **Test involving covariates(Nonoutcome Variable)**
     Test whether other covariates exhibit a jump at the discontinuity.

```
 * There should be no jump in other covariates.
 * 从图形，似乎是存在跳跃的，但这并不严格，要看回归结果

^^Iforeach y of global covariates {
^^I^^Iqui rdplot `y´ X, graph_options(xtitle("score")) saving(`y´)
^^I
^^I^^Igraph export fig_`y´.png, width(500) replace
^^I}
```

Example for Sharp RDD

- Testing the Validity of the RDD
  1. **Test involving covariates(Nonoutcome Variable)**
     Test whether other covariates exhibit a jump at the discontinuity.



Example for Sharp RDD

- Testing the Validity of the RDD

  1. **Test involving covariates(Nonoutcome Variable)**
     Test whether other covariates exhibit a jump at the discontinuity.

```
 * 估计具体系数看是否显著

. est clear
^^Iforeach y of global covariates {
^^I^^Ieststo : qui rdrobust `y´ X, all
^^I}

. esttab est1 est2 est3 est4 est5 est6 est7 , ///
      se r2 mtitle star(* 0.1 ** 0.05 *** 0.01) compress
```

Example for Sharp RDD

- Testing the Validity of the RDD

1. **Test involving covariates(Nonoutcome Variable)**
   Test whether other covariates exhibit a jump at the discontinuity.

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5	(6) est6	(7) est7
Conventi~l	-1.363	2.459	1.001	0.0728	-0.0270	0.0696	-0.1000
	(1.383)	(2.052)	(1.918)	(0.0724)	(0.0712)	(0.0662)	(0.0714)
Bias-cor~d	-1.193	2.898	1.495	0.0773	-0.0386	0.0828	-0.102
	(1.383)	(2.052)	(1.918)	(0.0724)	(0.0712)	(0.0662)	(0.0714)
Robust	-1.193	2.898	1.495	0.0773	-0.0386	0.0828	-0.102
	(1.633)	(2.454)	(2.246)	(0.0866)	(0.0845)	(0.0772)	(0.0854)
N	1387	1349	1308	1349	1308	1390	1390
R-sq							

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Example for Sharp RDD

- Testing the Validity of the RDD

  2. **Test sorting behavior**

     Testing discontinuity in the density of assignment variable X

     -*rddensity*-

```
. rdrobust Y X
. local h = e(h_l)   //获取最优带宽
. rddensity X, p(1) h(`h´ `h´) plot

RD Manipulation Test using local polynomial density estimation.
      Cutoff c = 0 |Left of c   Right of c          Number of obs =        1390
-------------------+----------------------          Model         = unrestricted
     Number of obs |      640          750          BW method     =       manual
Eff. Number of obs |      377          346          Kernel        =   triangular
    Order est. (p) |        1            1          VCE method    =    jackknife
    Order bias (q) |        2            2
       BW est. (h) |   17.754       17.754
Running variable: X.
-----------------------------------------------
            Method |          T          P>|T|
-------------------+---------------------------
            Robust |    -1.5083         0.1315
-----------------------------------------------
```

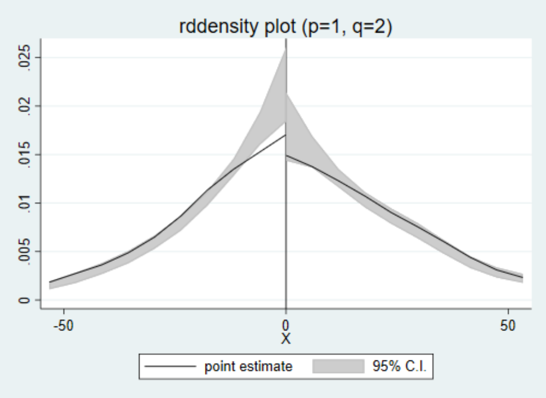
Example for Sharp RDD

- Testing the Validity of the RDD

  2. **Test sorting behavior**

     Testing discontinuity in the density of assignment variable X

     -*rddensity*-



Example for Sharp RDD

- Testing the Validity of the RDD

  2. **Test sorting behavior**

     Testing discontinuity in the density of assignment variable X
     Histogram(直方图)

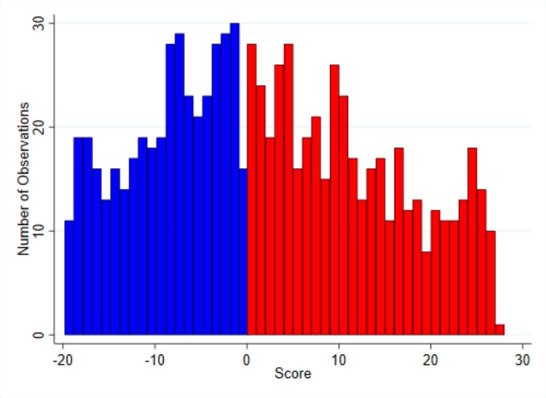
```
. qui rddensity X
. local bandwidth_left = e(h_l)
. local bandwidth_right = e(h_r)

. twoway (histogram X if X >= -`bandwidth_left´ & X < 0, freq width(1) color(blue) lcolor(black) lwidth(vthin)) ///
      (histogram X if X >= 0 & X <= `bandwidth_right´, freq width(1) color(red) lcolor(black) lwidth(vthin)),   ///
      xlabel(-20(10)30) graphregion(color(white)) xtitle(Score) ytitle(Number of Observations) legend(off)

. graph export fig10.png, width(500) replace
(file fig10.png written in PNG format)
```

Example for Sharp RDD

- Testing the Validity of the RDD

  2. **Test sorting behavior**

     Testing discontinuity in the density of assignment variable X
     Histogram(直方图)



Example for Sharp RDD

- Testing the Validity of the RDD

  2. **Test sorting behavior**

     Testing discontinuity in the density of assignment variable X
     a more formal test : **McCrary(2008) test -*DCdensity*-**

```
. preserve
. DCdensity X, breakpoint(0) generate(Xj Yj r0 fhat se_fhat) // McCracy test

Using default bin size calculation, bin size = 1.84133021
Using default bandwidth calculation, bandwidth = 25.8493835
Discontinuity estimate (log difference in height): -.100745626
                                                   (.117145041)
Performing LLR smoothing.
110 iterations will be performed
...........
. gen t= -.100745626/.117145041  // 产生t值，这个需要你根据系数提取出来
. display 2*ttail(50, t)         // 得到p值，50是自由度
1.6061102

. graph export fig11.png, width(500) replace
(note: file fig11.png not found)
(file fig11.png written in PNG format)
. restore
. /**可以看出在5%显著性水平下实际上Mccrary检验是通不过的，证明没有操纵**/
```

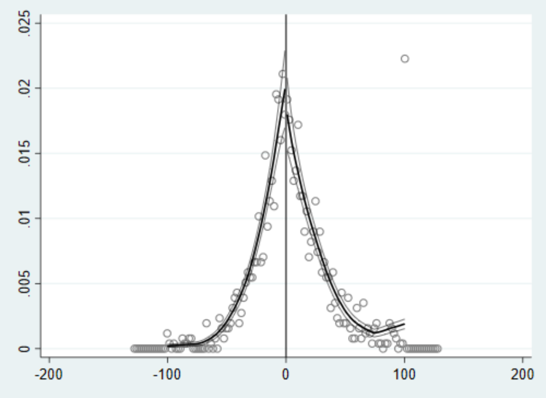
Example for Sharp RDD

- Testing the Validity of the RDD

  2. **Test sorting behavior**

     Testing discontinuity in the density of assignment variable X

     a more formal test : **McCrary(2008) test -*DCdensity*-**



Example for Sharp RDD

- Testing the Validity of the RDD

  3. **Falsification Tests**

     - Check 1：Placebo Cutoffs
     - 选择一个不同于断点的值作为安慰剂断点 (placcebo cutoff points)，分别取真实断点两侧 25%、50%、75% 样本分位数处作为断点。

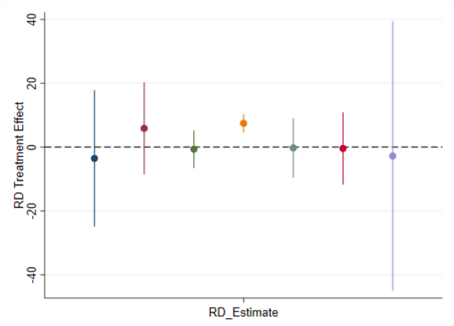
```
. sum X
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
           X |      1,390    7.171159    34.32488       -100        100
. local xmax=r(max)
. local xmin=r(min)

^^Iforvalues i=1(1)3{
^^I^^Ilocal jr=`xmax´/(4/(4-`i´))
^^I^^Ilocal jl=`xmin´/(4/(4-`i´))
^^I^^Iqui rdrobust Y X if X>0, c(`jr´)
^^I^^Iest store jl`i´
^^I^^Iqui rdrobust Y X if X<0, c(`jl´)
^^I^^Iest store jr`i´
^^I}

. qui rdrobust Y X ,c(0)   //加上真实断点的回归结果，作为benchmark结果
. est store jbaseline
```

- Testing the Validity of the RDD

  3. **Falsification Tests**

     - Check 1 : Placebo Cutoffs

```
. local vlist "jl1 jl2 jl3 jbaseline jr3 jr2 jr1 "
. coefplot `vlist´,  yline(0, lcolor(black) lpattern(dash)) drop(_cons) vertical ///
     graphregion(color(white)) ytitle("RD Treatment Effect") legend(off)

. graph export fig12.png, width(500) replace
(file fig12.png written in PNG format)
```



Example for Sharp RDD

- Testing the Validity of the RDD

  3. **Falsification Tests**

     - Check 2 : Sensitivity to Observations near the Cutoffs

       由于越接近断点的样本，越有动机去人为操控，删除最接近断点的样本，来观察回归是否显著(甜甜圈效应，donut hole approach)。

       分别删除断点附近 1%，2%，3%，4% 和 5% 的样本，进行了 5 组稳健性检验。

       图形给出了回归系数和 95% 的置信区间。

```
. sum X
. local xmax=r(max)

^^Iforvalues i=1(1)5{
^^I^^Ilocal j=`xmax´*0.01*`i´
^^I^^Iqui rdrobust Y X if abs(X)>`j´
^^I^^Iest store obrob`i´
^^I}
```

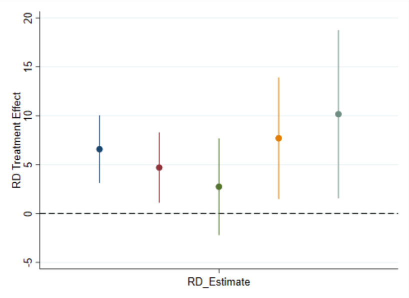
Example for Sharp RDD

- Testing the Validity of the RDD

  3. **Falsification Tests**

     - Check 2 : Sensitivity to Observations near the Cutoffs

```
. local vlist "obrob1 obrob2 obrob3 obrob4 obrob5"
. coefplot `vlist´,  yline(0, lcolor(black) lpattern(dash)) drop(_cons) vertical ///
      graphregion(color(white)) legend(off)  ytitle("RD Treatment Effect")

. graph export fig13.png, width(500) replace
(note: file fig13.png not found)
(file fig13.png written in PNG format)
```



- Testing the Validity of the RDD

  3. **Falsification Tests**

     - Check 3：Sensitivity to Bandwidth Choice

       带宽长度会显著影响回归结果，一个稳健的结果要求对带宽长度不那么敏感。
       提取最优带宽 h，然后分别手动设置带宽为 h 的 25%-400% 倍，看回归结果是否仍旧显著。
       图形给出了回归系数和 95% 的置信区间。

```
. qui rdrobust Y X   //自动选择最优带宽
. local h = e(h_l)   //获取最优带宽

^^Iforvalues i=1(1)8{
^^I^^Ilocal hrobust=`h´*0.25*`i´
^^I^^Iqui rdrobust Y X ,h(`hrobust´)
^^I^^Iest store hrob`i´
^^I}
```

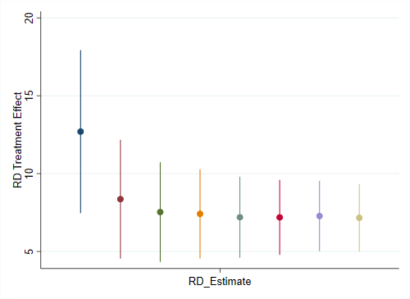
Example for Sharp RDD

- Testing the Validity of the RDD

  3. **Falsification Tests**

     - Check 3 : Sensitivity to Bandwidth Choice

```
. local vlist "hrob1 hrob2 hrob3 hrob4 hrob5 hrob6 hrob7 hrob8  "
. coefplot `vlist´,  yline(0, lcolor(black) lpattern(dash)) drop(_cons) vertical ///
      graphregion(color(white)) ytitle("RD Treatment Effect") legend(off)

. graph export fig14.png, width(500) replace
(note: file fig14.png not found)
(file fig14.png written in PNG format)
```



Subsection 7

Example for Fuzzy RDD

Example for Fuzzy RDD

- 三种方法（任选）：
  - 方法一：-*rd*-
  - 方法二：-*rdrobust*-
  - 方法三：IV 估计

Example for Fuzzy RDD

- 方法一：-*rd*-
  - Syntax

```
rd y d x, z0(real) strineq mbw(numlist) graph bdep oxline    ///
    kernel(rectangle) covar(varlist) x(varlist)

mbw(numlist)          //用来指定最优带宽的倍数，默认值为mbw(50 100 200)
z0(real)              //用来指定断点的位置，默认值为z0(0)，即断点为原点
*如果此处省去D，则为SRD，并根据分组变量X来计算处理变量
graph                 //根据每一带宽，画出局部线性回归图
bdep                  //根据画图来考察断点回归估计量对带宽的依赖性
oxline                //在此图的默认带宽上画出一条直线，以便识别
kernel(rectangle)     //使用均匀核（uniform），默认triangle
covar(varlist)        //用来指定加入局部线性回归的协变量
x(varlist)            //检验这些协变量在断点处是否存在跳跃（估计跳跃值和
```

Example for Fuzzy RDD

- 方法一：-*rd*-
  - Example background

    在美国国会，有一个民主党代表可能是被认为是对国会选区的一种 treatment。

    美国国会选区，如果有民主党众议员，对该选区的联邦政府的开支具有一定影响。

    传统意义上，民主党会更倾向于政府，如果当选，会加大对联邦政府的开支。

    然而直接对二者进行回归，可能会遗漏变量问题或者双向因果关系。

    为此选择该民主党候选人的得票比例作为分组变量，Z 是民主党候选人获得的选票份额。

    以 0.5 为断点（在民主党与共和党的政治中，$Z>=0.5$，则当选，反之落选），进行 RDD。

Example for Fuzzy RDD

- 方法一：-*rd*-
  - Example Data

```
. ssc inst rd, replace
. net get rd
. use votex, clear

  * lne   //选取联邦政府开支的对数
  * d     //分组变量，民主党派候选人的得票比例减去0.5，以标准化
  * win   //民主党派候选人当选
  * 另外还包括一些协变量

. desc
```

- 方法一：-*rd*-
  - SRD

```
//OLS回归
reg lne win i votpop bla-vet
*回归结果虽然win表示当选了，会增加lne的支出，但是不显著

//选择默认的带宽以及triangle kernel进行RD
rd lne d, gr mbw(100)
*不显著，说明拥有民主党派候选人当选的选区并不能显著的增加联邦政府开支

//加入协变量进行RD，省略作图
rd lne d, mbw(100) cov(i votpop black blucllr farmer fedwrkr forborn manuf unemplyd union urban veterans)
*显示估计值虽然为正，但是依然不显著

//去掉协变量，同时估计三种带宽，并画出估计值对带宽的依赖性
rd lne d, gr bdep oxline
*改变带宽对估计值有一定的影响，但是三个估计值全部为负，且依然不显著。可以看出，各个断点回归估计量对带宽的依赖性不大。

//检验协变量在断点处是否存在跳跃
rd lne d, mbw(100) x(i votpop black  blucllr farmer fedwrkr  forborn manuf  unemplyd union urban veterans)
*farmer的P值为 0.036，其余的协变量的条件密度函数在断点处都是连续的，即只有farmer（农民占人口比例）存在跳跃。
```

Example for Fuzzy RDD

- 方法一：-*rd*-
  - FRD

```
. //生成一个新的处理变量randwin，使得randwin不完全由分组变量d决定。
. set seed 20181203
. g byte randwin=cond(uniform()<.1,1-win,win)
. tab randwin win

           |     Dem Won Race
   randwin |         0          1 |     Total
-----------+----------------------+----------
         0 |       123         14 |       137
         1 |         8        204 |       212
-----------+----------------------+----------
     Total |       131        218 |       349

 *结果显示randwin与win基本相同，但不完全相同，说明randwin不完全由分组变量d决定。
```

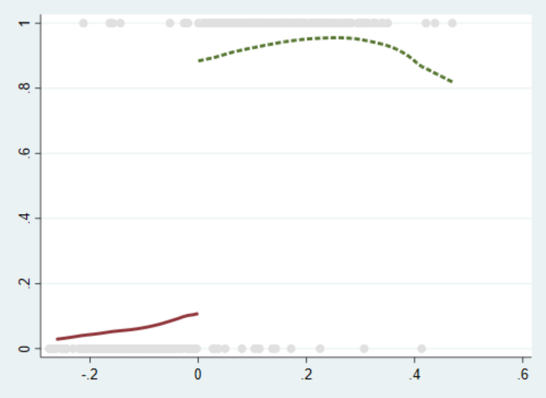
Example for Fuzzy RDD

- 方法一：-*rd*-
  - FRD

```
. //使用最优带宽与默认的triangle kernel进行FRD(含协变量)
. rd lne randwin d, gr mbw(100) cov(i votpop black blucllr farmer fedwrkr forborn manuf unemply
> d union urban veterans)

Three variables specified; jump in treatment
at Z=0 will be estimated. Local Wald Estimate
is the ratio of jump in outcome to jump in treatment.
 Assignment variable Z is d
 Treatment variable X_T is randwin
 Outcome variable y is lne
Command used for graph: lpoly; Kernel used: triangle (default)
Bandwidth: .29287776; loc Wald Estimate: -.09974965
Estimating for bandwidth .2928777592534943
A predicted value of treatment at cutoff lies outside feasible range;
switching to local mean smoothing for treatment discontinuity.
------------------------------------------------------------------------------
         lne |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       numer |   .0543733   .0900181     0.60   0.546    -.1220589    .2308055
       denom |   .8734363   .0301089    29.01   0.000      .814424    .9324487
       lwald |   .0622522   .1028807     0.61   0.545    -.1393903    .2638946
------------------------------------------------------------------------------

. graph export fig15.png, width(500) replace
(file fig15.png written in PNG format)
```

Example for Fuzzy RDD

- 方法一：-*rd*-
  - FRD



Example for Fuzzy RDD

- 方法二：-*rdrobust*-

```
 * options : fuzzy(fuzzyvar [sharpbw])

. rdrobust lne d, fuzzy(randwin)

Fuzzy RD estimates using local polynomial regression.
      Cutoff c = 0 | Left of c  Right of c            Number of obs =        349
-------------------+----------------------            BW type       =      mserd
     Number of obs |       131         218            Kernel        = Triangular
Eff. Number of obs |        73         105            VCE method    =         NN
    Order est. (p) |         1           1
    Order bias (q) |         2           2
       BW est. (h) |     0.142       0.142
       BW bias (b) |     0.186       0.186
         rho (h/b) |     0.762       0.762
First-stage estimates. Outcome: randwin. Running variable: d.
--------------------------------------------------------------------------------
            Method |   Coef.    Std. Err.    z     P>|z|    [95% Conf. Interval]
-------------------+------------------------------------------------------------
      Conventional |  .74178     .09171   8.0888   0.000    .562043      .921522
            Robust |     -         -      7.3050   0.000     .53871      .933789
--------------------------------------------------------------------------------
Treatment effect estimates. Outcome: lne. Running variable: d. Treatment Status: randwin.
--------------------------------------------------------------------------------
            Method |   Coef.    Std. Err.    z     P>|z|    [95% Conf. Interval]
-------------------+------------------------------------------------------------
      Conventional | -.12768     .20294   -0.6292  0.529   -.525433      .270072
            Robust |     -         -      -0.6411  0.521   -.665039        .3372
--------------------------------------------------------------------------------
```

Example for Fuzzy RDD

- 方法三：IV 估计

```
. use frd.dta, clear
. sum
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        anno |        120    1998.667    3.831171       1993       2004
      esse_m |        120           0    6.230853        -10         10
          mc |        120    9.792913    .1016675   9.522602   9.995189
         mcn |        120    9.726134     .093931    9.46739   9.921432
          mf |        120    6.100074    .1004518   5.892969   6.368767
-------------+---------------------------------------------------------
         pen |        120    .4036111    .3359836          0   .9861111
        obsc |        120      88.175      35.007         21        184
       obscn |        120      88.175      35.007         21        184
        obsf |        120    88.11667    34.99176         21        184
        obsp |        120      88.175      35.007         21        184
-------------+---------------------------------------------------------
    anno1993 |        120   -7.45e-09    .3742406   -.1666667   .8333333
    anno1995 |        120   -7.45e-09    .3742406   -.1666667   .8333333
    anno1998 |        120   -7.45e-09    .3742406   -.1666667   .8333333
    anno2000 |        120   -7.45e-09    .3742406   -.1666667   .8333333
    anno2002 |        120   -7.45e-09    .3742406   -.1666667   .8333333
-------------+---------------------------------------------------------
    anno2004 |        120   -7.45e-09    .3742406   -.1666667   .8333333
        elig |        120          .5    .5020964          0          1
     esse_m2 |        120        38.5    32.55583          1        100
     esse_m3 |        120           0    446.6576      -1000       1000
     esse_m4 |        120      2533.3    3231.493          1      10000
-------------+---------------------------------------------------------
         sel |        120           1           0          1          1
      mc_neg |         60    9.861339    .0171739    9.83221    9.88535
      mc_pos |         60    9.724487    .0307325   9.674244   9.769611
     mcn_neg |         60    9.787027    .0135329   9.763389   9.805106
     mcn_pos |         60    9.665241    .0282369   9.618672   9.706251
-------------+---------------------------------------------------------
      mf_neg |         60    6.159085    .0148918   6.135935   6.182206
```

Example for Fuzzy RDD

- 方法三：IV 估计
  - mcn, mf：非耐用品和食品支出 (Y)
  - ess_m：已经退休的年数，负数表示还未到退休年龄
  - pen：退休概率，内生变量，依赖于 ess_m，断点处存在不连续跳跃 (X)
  - elig：退休资格虚拟变量，若 ess_m>=0,elig=1，否则为 0(Z)

Example for Fuzzy RDD

- 方法三：IV 估计

```
. ivregress 2sls mcn (pen=elig) esse_m esse_m2 anno1995-anno2004, first robust
First-stage regressions
-----------------------

                                          Number of obs      =        120
                                          F(   8,     111)   =     177.06
                                          Prob > F           =     0.0000
                                          R-squared          =     0.9230
                                          Adj R-squared      =     0.9175
                                          Root MSE           =     0.0965

------------------------------------------------------------------------------
             |               Robust
         pen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      esse_m |   .0169943   .0031324     5.43   0.000     .0107873    .0232013
     esse_m2 |  -.0005738    .000294    -1.95   0.054    -.0011564    8.82e-06
    anno1995 |   .0230442   .0305492     0.75   0.452    -.0374911    .0835796
    anno1998 |   .0508942   .0336146     1.51   0.133    -.0157153    .1175038
    anno2000 |   .1172279   .0326428     3.59   0.000     .0525441    .1819118
    anno2002 |   .1400281   .0327296     4.28   0.000     .0751722     .204884
    anno2004 |   .1693907   .0333298     5.08   0.000     .1033455    .2354359
        elig |   .4349947   .0362406    12.00   0.000     .3631815    .5068078
       _cons |   .2082045    .018284    11.39   0.000     .1719736    .2444355
------------------------------------------------------------------------------
```

Example for Fuzzy RDD

- 方法三：IV 估计

```
Instrumental variables (2SLS) regression          Number of obs   =        120
                                                  Wald chi2(8)    =     332.96
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.6223
                                                  Root MSE        =     .05749

------------------------------------------------------------------------------
             |               Robust
         mcn |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         pen |  -.0983277   .0544996    -1.80   0.071    -.2051449    .0084895
      esse_m |  -.0055121   .0025899    -2.13   0.033    -.0105881    -.000436
     esse_m2 |   -.000288    .000145    -1.99   0.047    -.0005722   -3.84e-06
    anno1995 |   .0018884   .0179095     0.11   0.916    -.0332135    .0369903
    anno1998 |  -.0334648   .0181819    -1.84   0.066    -.0691007     .002171
    anno2000 |   .0121598    .019791     0.61   0.539    -.0266299    .0509495
    anno2002 |   .0210096   .0229841     0.91   0.361    -.0240384    .0660575
    anno2004 |   .0843976   .0194665     4.34   0.000     .0462439    .1225512
       _cons |    9.77691   .0244502   399.87   0.000     9.728988    9.824831
------------------------------------------------------------------------------
Instrumented:  pen
Instruments:   esse_m esse_m2 anno1995 anno1998 anno2000 anno2002 anno2004
               elig
```

Example for Fuzzy RDD

- 方法三：IV 估计

```
. ivregress 2sls mf (pen=elig) esse_m esse_m2 anno1995-anno2004, first robust
First-stage regressions
-----------------------

                                                Number of obs     =        120
                                                F(   8,     111)  =     177.06
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9230
                                                Adj R-squared     =     0.9175
                                                Root MSE          =     0.0965

------------------------------------------------------------------------------
             |               Robust
         pen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      esse_m |   .0169943   .0031324     5.43   0.000     .0107873    .0232013
     esse_m2 |  -.0005738    .000294    -1.95   0.054    -.0011564    8.82e-06
    anno1995 |   .0230442   .0305492     0.75   0.452    -.0374911    .0835796
    anno1998 |   .0508942   .0336146     1.51   0.133    -.0157153    .1175038
    anno2000 |   .1172279   .0326428     3.59   0.000     .0525441    .1819118
    anno2002 |   .1400281   .0327296     4.28   0.000     .0751722     .204884
    anno2004 |   .1693907   .0333298     5.08   0.000     .1033455    .2354359
        elig |   .4349947   .0362406    12.00   0.000     .3631815    .5068078
       _cons |   .2082045    .018284    11.39   0.000     .1719736    .2444355
------------------------------------------------------------------------------
```

Example for Fuzzy RDD

- 方法三：IV 估计

```
Instrumental variables (2SLS) regression          Number of obs   =        120
                                                  Wald chi2(8)    =     325.13
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.7124
                                                  Root MSE        =     .05365

------------------------------------------------------------------------------
             |               Robust
          mf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         pen |  -.1409689   .0523376    -2.69   0.007    -.2435487   -.0383891
      esse_m |  -.0027591   .0024779    -1.11   0.266    -.0076157    .0020975
     esse_m2 |  -.0000821   .0001355    -0.61   0.545    -.0003478    .0001835
    anno1995 |  -.0675309   .0144149    -4.68   0.000    -.0957835   -.0392782
    anno1998 |  -.1365082   .0154907    -8.81   0.000    -.1668694    -.106147
    anno2000 |    -.13855   .0177911    -7.79   0.000    -.1734199   -.1036801
    anno2002 |  -.1420846    .017727    -8.02   0.000     -.176829   -.1073403
    anno2004 |  -.1042948   .0179742    -5.80   0.000    -.1395236   -.0690661
       _cons |   6.160132   .0223924   275.10   0.000     6.116244     6.20402
------------------------------------------------------------------------------
Instrumented:  pen
Instruments:   esse_m esse_m2 anno1995 anno1998 anno2000 anno2002 anno2004
               elig
```