```
*    ------------------------------------------------

*
*        $$$$$ $$$$$$$$    $    $$$$$$$$    $
*       $$  $$ $  $$  $   $$    $  $$  $   $$
*       $$   $    $$      $$$      $$      $$$
*       $$$       $$     $ $$      $$     $ $$
*         $$$     $$     $  $$     $$     $  $$
*           $$$   $$     $$$$$     $$     $$$$$
*       $   $$    $$    $   $$     $$    $   $$
*       $$  $$    $$    $    $$    $$    $    $$
*       $$$$$    $$$$  $$$  $$$$  $$$$  $$$  $$$$
*
*    ------------------------------------------------

*
*        ____________________________________________
*        ____________________________________________

*
*
*                    断点回归设计
*              Regression Discontinuity Design
*
*        ____________________________________________
*        ____________________________________________

*
*
*
*                    主讲人：连玉君
*
*          单  位：中山大学岭南学院金融系
*          电  邮：arlionn@163.com
*          主  页：http://www.lingnan.sysu.edu.cn/lnshizi/faculty_vch.asp?tn=50
*          简  书：http://www.jianshu.com/u/69a30474ef33
*          微  博：http://weibo.com/arlionn
*          课  程：http://www.peixun.net/author/3.html
*          现场班：https://www.jianshu.com/p/af6fb0448297
*
*          微  信：lianyj45
*          公众号：Stata连享会(微信: StataChina)


*-注意：执行后续命令之前，请先执行如下命令
  global path "`c(sysdir_personal)'\Lian_RDD" //定义课程目录
  global D    "$path\data"      //范例数据
  global R    "$path\refs"      //参考文献
  global out  "$path\out"       //结果：图形和表格
  global ex   "$path\examples"  //范例论文和dofiles
  adopath +   "$path\adofiles"  //自编程序
  cd "$D"
  set scheme s2color


                                                        /*
                ====导言====

    ***********************************************

          史上最牛的断点效应！！！

     shellout "$R\连玉君_断点回归分析-RDD.ppt"

```

```
**************************************************


  - 在一个高度依赖规则的世界里，有些规则的出现十分随意，
    这种随意性为我们提供了性质良好的实验 (Angrist& Pischke，2009)

  - 断点回归设计(RDD)是一种仅次于随机实验的，
    能够有效利用现实约束条件分析变量之间因果关系的实证方法

  - Lee(2008) 认为，在随机实验不可得的情况下，
    断点回归能够避免参数估计的内生性问题，
    从而真实反映出变量之间的因果关系。
                                                        */


*--------------
*-A.1  RDD 简介
*--------------


*------
*-A.1.1 为何使用 RDD?


  *---------------------------------------
  *-例1: 就读岭南学院的收入效应如何估算 ?
  *---------------------------------------
  *-        一分之差，命运迥然?

                                                    /*
     2014-2016年岭南学院经济学(理科)广东录取分数
    ------------------------------------------------
        年度     最高分   最低分   平均分     一本线
        ----------------------------------------
        2016      668      643      650       508

        2015      680      671      674       577    (忽略)
        2014      672      658      662       560    (忽略)
    ------------------------------------------------
                                                    */

*-------
*-方案0: 自然实验 -- 抛硬币 或 摇塞子 决定谁上岭院.


*-------
*-方案1: OLS + dummy  |  treatment effect model
*
*     wage[i] = a0 + a1*Treat[i] + a2*x[i] + u[i]
*
*  Treat=1 (if mark>=643),
*  Treat=0 (if mark< 643)
*  x[i]: 高考分数; 性别; 是否一线城市; 父母EDU; 家庭收入等
*  u[i]: 其它不可观测因素 -- 能力; 父母社会资源;
*
* Sample: 全样本
* 内生性的来源: ?  ?  ?


*-------
*-方案2: PSM
*
* Sample: 子样本 -- 满足共同支撑和平行假设的样本
*
```

```
        * Q: 如何匹配? 从哪些维度匹配?


        *-------
        *-方案3: RDD
        *
        * 基本思想: 高考分数超过 643分 的同学入读中大岭院;
        *           然而, 高考分数为 642分 和 644分 的考生没什么本质差别,
        *           因此二者毕业时的[薪水差异]可以归结为岭院教育产生的效果
        * 可能的质疑: 运气? (两边都有! )
        *
        *-RDD 中的几个基本术语:
        *
        *  @  cut-point (分配点) = 643分
        *  @  assignment variable (分配变量): 高考分数
        *  @  Treat   group (实验组): Mark>=643 (cut-point)
        *  @  Control group (控制组): Mark< 643
        *
        * 典型特征: (random assignment, 随机分配原则)
        *
        *    对于在分数线附近的学生而言, 实验对象的选择具有随机性
        *
        *    在分数线公布之前他是不知道自己能否进入 Treat 组的;
        *
        * Sample: 子样本 -- 在断点(643分) 附近, 如 (643-5, 643+5)
        *                                        或 (643-2, 643+2)
        *         即 (643-h, 643+h), h 称为窗口宽度
        *
        *-Q1: 对样本有何要求 ?
        *-Q2: 有没有违反"随机分配原则"的实例 ?
        *
        *     ROE 6%;  及格线 60;


    *------
    *-A.1.2  图解 RDD


      *---
      *-1- 生成一份模拟数据

        clear
        set obs 4000
        set seed 123
        gen x = runiform()
        gen z = rnormal()*0.5  //其他影响 y 的因素
        gen T=0
        replace T=1 if x>0.5

        gen g0 = 0 + 3*log(x+1) + sin(x*6)/3
        gen g1 = T + 3*log(x+1) + sin(x*6)/3

        scatter g0 x, msize(*0.5)
        scatter g1 x, msize(*0.5)

        gen e = rnormal()/5       // noise
        gen y1 = g1 + 0.5*z + e
        gen y0 = g0 + 0.5*z + e

        gen xc = x-0.5

        label var y1 "Outcome variable (y)"
        label var y0 "Outcome variable (y)"
```

```
191       label var x  "Assignment variable (x)"
192       label var xc "Centered Assignment variable (x-c)"
193       label var T  "T=1 for x>0.5, T=0 otherwise"
194
195       save "RDD_simu_data0.dta", replace  //保存一份数据以备后用
196
197
198     *---
199     *-2- RDD 图示
200
201       use "RDD_simu_data0.dta", clear
202
203       *---------------------
204       *-Without Treat effect                              -------图1-----begin--
205         twoway (scatter y0 x, msymbol(+) msize(*0.4) mcolor(black*0.3)) ///
206                (qfit y0 x if T==0, lcolor(red)  msize(*0.4))  ///
207                (qfit y0 x if T==1, lcolor(blue) msize(*0.4)), ///
208                 xline(0.5, lpattern(dash) lcolor(gray))       ///
209                 text(3.5 0.3 "Control") text(3.5 0.7 "Treat") ///
210                 legend(off)  ///
211                 ylabel(-1 "10" 0 "12" 1 "14" 2 "16" 3 "18") ///
212                 xlabel(0 "650"  0.5 "CP(643)"  1 "668")  ///
213                 ytitle("毕业当年月薪(万元)")  ///
214                 xtitle("高考分数 (分配变量)",place(right)) ///
215                 xscale(titlegap(4))
216         graph export "$out\Fig_noeffect.png", replace
217         *                                                ---------------over---
218         *
219         rdplot y0 x, c(0.5)  //快捷命令，后续会详细介绍
220
221       *-----------------
222       *-With Treat effect                                 -------图2-----begin--
223         twoway (scatter y1 x, msymbol(+) msize(*0.4) mcolor(black*0.3)) ///
224                (qfit y1 x if T==0, lcolor(red)  msize(*0.4))  ///
225                (qfit y1 x if T==1, lcolor(blue) msize(*0.4)), ///
226                 xline(0.5, lpattern(dash) lcolor(gray))       ///
227                 text(3.5 0.3 "Control") text(3.5 0.7 "Treat") ///
228                 legend(off)  ///
229                 ylabel(-1 "10" 0 "12" 1 "14" 2 "16" 3 "18") ///
230                 xlabel(0 "650"  0.5 "CP(643)"  1 "668")  ///
231                 ytitle("毕业当年月薪(万元)")  ///
232                 xtitle("高考分数 (分配变量)",place(right)) ///
233                 xscale(titlegap(4))
234         graph export "$out\Fig_witheffect.png", replace
235         *                                                ---------------over---
236         *
237         rdplot y1 x, c(0.5)
238
239
240       *-------
241       *-反事实
242
243         *                                                ---------------begin--
244         do  "RDD_simudata02.do"
245         use "RDD_simu_data2.dta", clear
246         #d ;
247         twoway (scatter y1 x if T==1, msymbol(Oh) msize(*0.4) mcolor(black*0.6))
248                (scatter y1 x if T==0, msymbol(Oh) msize(*0.4) mcolor(black*0.6))
249                (scatter y0 x if T==1, msymbol(Dh) msize(*0.4) mcolor(black*0.2))
250                (qfit y1 x if T==0, lcolor(blue)    msize(*0.4) lw(*1.2))
251                (qfit y1 x if T==1, lcolor(red*1.3) msize(*0.4) lw(*1.2))
252                (qfit y0 x if T==1, lcolor(red*1.3) msize(*0.4) lw(*1.2) lp(dash))
253                ,
254                 xline(0.5, lpattern(dash) lcolor(gray))
```

```
                text(4.5 0.3 "Control") text(4.5 0.7 "Treat")
                legend(off)
                ylabel(-1 "10" 0 "12" 1 "14" 2 "16" 3 "18" 4 "20" 5 "22")
                xlabel(0 "650"  0.5 "CP(643)"  1 "668")
                ytitle("毕业当年月薪(万元)")
                xtitle("高考分数 (分配变量)",place(right))
                xscale(titlegap(4)) ;
        #d cr
        graph export "$out\Fig_antiFact.png", replace
        *                                                ----------------over---

    *---
    *-3- 传统估计方法(方案1)存在的问题

      * ------------------------------------------------------图3-----begin---
      *-简单的均值比较: 结果有偏   outcome = y1
        use "RDD_simu_data0.dta", clear
        sum y1 if T==0
        local y0: dis %4.2f r(mean)
        sum y1 if T==1
        local y1: dis %4.2f r(mean)
        twoway (scatter y1 x, msymbol(+) msize(*0.4) mcolor(black*0.3)) ///
               (function y=`y0' if T==0, lcolor(red)  msize(*0.4))  ///
               (function y=`y1' if T==1, lcolor(blue) msize(*0.4)), ///
                xline(0.5, lpattern(dash) lcolor(gray))       ///
                text(3.5 0.3 "Control") text(3.5 0.7 "Treat") ///
                legend(off)  ///
                ylabel(-1 "10" 0 "12" 1 "14" 2 "16" 3 "18") ///
                xlabel(0 "650"  0.5 "CP(643)"  1 "668")  ///
                ytitle("毕业当年月薪(万元)")  ///
                xtitle("高考分数 (分配变量)",place(right)) ///
                xscale(titlegap(4))
        graph export "$out\Fig_bias_mean.png", replace
        *-----------------------------------------------------------over---
        *-Q: 得到的 ATE (Average Treatment Effect) 是什么？


    * ------------------------------------------------------图3-----begin---
    *-针对全样本进行线性回归 (OLS): 结果有偏   outcome = y1
      use "RDD_simu_data0.dta", clear
      twoway (scatter y1 x, msymbol(+) msize(*0.4) mcolor(black*0.3)) ///
             (lfit y1 x if T==0, lcolor(red)  msize(*0.4))  ///
             (lfit y1 x if T==1, lcolor(blue) msize(*0.4)), ///
              xline(0.5, lpattern(dash) lcolor(gray))       ///
              text(3.5 0.3 "Control") text(3.5 0.7 "Treat") ///
              legend(off)  ///
              ylabel(-1 "10" 0 "12" 1 "14" 2 "16" 3 "18") ///
              xlabel(0 "650"  0.5 "CP(643)"  1 "668")  ///
              ytitle("毕业当年月薪(万元)")  ///
              xtitle("高考分数 (分配变量)",place(right)) ///
              xscale(titlegap(4))
      graph export "$out\Fig_bias01.png", replace
      *-----------------------------------------------------------over---
      *-Q: 得到的 ATE (Average Treatment Effect) 是无偏估计吗？原因何在？


    *-OLS 导致的偏误2：错误判断 ATE            outcome = y0
    * ------------------------------------------------------图4-----begin---
      use "RDD_simu_data0.dta", clear
      twoway (scatter y0 x, msymbol(+) msize(*0.4) mcolor(black*0.3)) ///
             (qfit y0 x if T==0, lcolor(red) msize(*0.6))   ///
             (qfit y0 x if T==1, lcolor(red) msize(*0.6))   ///
             (lfit y0 x if T==0, lcolor(blue) lp(dash) msize(*0.5))  ///
             (lfit y0 x if T==1, lcolor(blue) lp(dash) msize(*0.5)), ///
```

```
                xline(0.5, lpattern(dash) lcolor(gray))       ///
                text(3.5 0.3 "Control") text(3.5 0.7 "Treat") ///
                legend(off)  ///
                ylabel(-1 "10" 0 "12" 1 "14" 2 "16" 3 "18") ///
                xlabel(0 "650"  0.5 "CP(643)"  1 "668")  ///
                ytitle("毕业当年月薪(万元)")  ///
                xtitle("高考分数 (分配变量)",place(right)) ///
                xscale(titlegap(4))
        graph export "$out\Fig_bias02.png", replace

*-------------------------------------------------------------------over---
        *-由于 y=f(x) 是非线性的，因此，若在分界点两侧采用直线拟合，
        * 会错以为存在处理效应


    *---
    *-4- RDD 分析中的两种典型估计方法

      *-视角1: “断点” (Hahn,Todd, 和van der Klaauw, 1999)
      *        “discontinuity at the cut-point.”
      *-e.g.1 多项式回归：二次函数
      * ------------------------------------------------------图5-----begin---
        use "RDD_simu_data0.dta", clear
        twoway (scatter y1 x, msymbol(+) msize(*0.4) mcolor(black*0.3)) ///
               (qfit y1 x if T==0, lcolor(red) msize(*0.6))   ///
               (qfit y1 x if T==1, lcolor(red) msize(*0.6))   ///
               (lfit y1 x if T==0, lcolor(blue) lp(dash) msize(*0.5))  ///
               (lfit y1 x if T==1, lcolor(blue) lp(dash) msize(*0.5)), ///
                xline(0.5, lpattern(dash) lcolor(gray))       ///
                text(3.5 0.3 "Control") text(3.5 0.7 "Treat") ///
                legend(off)  ///
                ylabel(-1 "10" 0 "12" 1 "14" 2 "16" 3 "18") ///
                xlabel(0 "650"  0.5 "CP(643)"  1 "668")  ///
                ytitle("毕业当年月薪(万元)")  ///
                xtitle("高考分数 (分配变量)",place(right)) ///
                xscale(titlegap(4))
        graph export "$out\Fig_x2.png", replace

*-------------------------------------------------------------------over---
        *-评述: 重点在于分析 cut-point 处的跳跃(jump),
        *       跳跃的方向和幅度是评估处理效应的主要依据。

      *-e.g.2 核加权局部多项式平滑 (Kernel-weighted local polynomial smoothing)
        help lpoly
      * ------------------------------------------------------图5-----begin---
        twoway (scatter y1 x, msymbol(+) msize(*0.4) mcolor(black*0.3)) ///
               (lpoly y1 x if T==0, lcolor(red) msize(*0.6))   ///
               (lpoly y1 x if T==1, lcolor(red) msize(*0.6))   ///
               (lfit y1 x if T==0, lcolor(blue) lp(dash) msize(*0.5))  ///
               (lfit y1 x if T==1, lcolor(blue) lp(dash) msize(*0.5)), ///
                xline(0.5, lpattern(dash) lcolor(gray))       ///
                text(3.5 0.3 "Control") text(3.5 0.7 "Treat") ///
                legend(off)  ///
                ylabel(-1 "10" 0 "12" 1 "14" 2 "16" 3 "18") ///
                xlabel(0 "650"  0.5 "CP(643)"  1 "668")  ///
                ytitle("毕业当年月薪(万元)")  ///
                xtitle("高考分数 (分配变量)",place(right)) ///
                xscale(titlegap(4))
        graph export "$out\Fig_lpoly.png", replace
        *-------------------------------------------------------------over---

        *-快捷命令(随后还会详细讲解)
          rdplot y1 x, c(0.5)      //自动选择高次项的阶数
          rdplot y1 x, c(0.5) p(2) //自行设定, 只加入一次项和二次项
```

```
381           rdplot y1 x, c(0.5) p(1) //自行设定, 一次线性关系
382
383         *-视角2: “局部随机化” (Lee,2008)
384         *
385         *-e.g. 局部线性回归
386         * ------------------------------------------------------图6-----begin---
387           dropvars left right
388           local h=0.1           // width of window, double sides
389           local cL = 0.5 - `h'
390           local cR = 0.5 + `h'
391           gen left  = (x>0.5-`h')&(x<0.50)
392           gen right = (x>0.50)&(x<0.5+`h')
393           twoway (scatter y1 x, msymbol(+) msize(*0.4) mcolor(black*0.3))  ///
394                  (lfit y1 x if (T==0&left==1) , lcolor(red)  msize(*0.4))  ///
395                  (lfit y1 x if (T==1&right==1), lcolor(blue) msize(*0.4)), ///
396                   xline(0.5, lpattern(dash) lcolor(gray))       ///
397                   xline(`cL' `cR', lp(dash) lc(black*0.2))      ///
398                   text(3.5 0.3 "Control") text(3.5 0.7 "Treat") ///
399                   legend(off)  ///
400                   ylabel(-1 "10" 0 "12" 1 "14" 2 "16" 3 "18") ///
401                   xlabel(0 "650"  0.5 "CP(643)"  1 "668")  ///
402                   ytitle("毕业当年月薪(万元)")  ///
403                   xtitle("高考分数 (分配变量)",place(right)) ///
404                   xscale(titlegap(4))
405           graph export "$out\Fig_local01.png", replace
406           *-----------------------------------------------------------over---
407           *-评述:
408           * [1] 核心思想, 643分与642分的考生有何差别呢? 有! -- 643进入岭院了
409           * [2] 要点: 带宽的选择!
410           *      - 非线性的程度, y=f(x) 的非线性程度越高, h 越小还是越大?
411           *      - 可以借助图形进行分析
412           *      - 选择不同的带宽做稳健性测试
413
414           rdplot y1 x, c(0.5) h(0.1 0.1) p(1)  // Q: 参数的对应关系?
415
416         *-对比: 视角1 v.s. 视角2
417          * ------------------------------------------------------图7-----begin---
418           local method "qfit"   //二次函数
419           local method "lpoly"  //核加权多项式
420           local h=0.1
421           local cL = 0.5 - `h'
422           local cR = 0.5 + `h'
423           dropvars left right
424           gen left  = (x>0.5-`h')&(x<0.50)
425           gen right = (x>0.50)&(x<0.5+`h')
426           twoway (scatter y1 x, msymbol(+) msize(*0.4) mcolor(black*0.3)) ///
427                  (`method' y1 x if T==0, lcolor(red) msize(*0.4))  ///
428                  (`method' y1 x if T==1, lcolor(red) msize(*0.4))  ///
429                  (lfit y1 x if (T==0&left==1) , lcolor(blue) msize(*0.4))  ///
430                  (lfit y1 x if (T==1&right==1), lcolor(blue) msize(*0.4)), ///
431                   xline(0.5, lpattern(dash) lcolor(gray))       ///
432                   xline(`cL' `cR', lp(dash) lc(black*0.2)) ///
433                   text(3.5 0.3 "Control") text(3.5 0.7 "Treat") ///
434                   legend(off)  ///
435                   ylabel(-1 "10" 0 "12" 1 "14" 2 "16" 3 "18") ///
436                   xlabel(0 "650"  0.5 "CP(643)"  1 "668")  ///
437                   ytitle("毕业当年月薪(万元)")  ///
438                   xtitle("高考分数 (分配变量)",place(right)) ///
439                   xscale(titlegap(4))
440           graph export "$out\Fig_lpoly_local.png", replace
441           *-----------------------------------------------------------over---
442           *
443           *-Q: 将 h 分别修改为 0.05, 0.1, 0.2, 0.4, 结果有何变化?
444           *
```

```
445         *-Q: 何种情况下，两种方法是等价的？
446         *
447         *-Note: 对于局部线性回归而言，主要的问题就在于选择带宽，
448         *       对于全局非线性回归而言，主要的问题在于选择合适的函数形式
449         *       来拟合数据，通常会加入 x-c 的高阶项
450
451
452
453   *--------------------------
454   *-A.2  估计方法和统计推断
455   *--------------------------
456
457     *-A.2.1 局部线性回归
458     *-A.2.2 最优带宽的选择
459     *-A.2.3 局部线性回归的惯用模型
460     *-A.2.4 多项式回归
461
462
463   *------
464   *-A.2.1 局部线性回归 (Local Linear Regression)
465
466     *-图3 的 OLS 估计
467
468       use "RDD_simu_data0.dta", clear
469       cap drop xc
470       gen xc = x-0.5       //why?
471       reg y1 xc if xc<0    // Left
472       reg y1 xc if xc>0    // Right
473       dis 2.097-1.495
474       *-图示
475         rdplot  y1 xc, c(0) p(1)
476         cmogram y1 xc, scatter lfit cut(0)
477       *-Q: 上述两个回归中，常数项的含义是什么？
478
479       *-Note: 在 RDD 分析中，通常都会预先对分配变量进行中心化处理，即
480       *
481       *       xc = x - cut-point
482       *
483       * 这样有助于结果的解释
484       * 为此，多数 RDD 命令的 cut() 或 thres() 选项通常都默认为 0
485
486       *-Q: 上述针对全样本的分析有何问题 ?
487
488
489     *-图6 的 OLS 估计：局部线性回归
490
491       *-计量表述:
492         shellout "$R\Imbens_2008_JE_Guide.pdf"  //pp.624, sec 4.2
493
494       *-先指定一个带宽 h, 例如 h=0.1
495
496       *-在 [-h< xc < +h] 的窗口范围内，分别在 xc=0 左右两侧执行 OLS 估计
497
498         local h=0.1                       //带宽
499         reg y1 xc if (xc>-`h')&(xc<0)    // Left
500         reg y1 xc if (xc>0)&(xc<`h')     // Right
501         dis "ATE = " %4.3f  2.213-1.262
502
503         *-ATE 的假设检验
504         local h=0.1                       //带宽
505         qui reg y1 xc if (xc>-`h')&(xc<0)    // Left
506         est store Left
507         qui reg y1 xc if (xc>0)&(xc<`h')     // Right
508         est store Right
```

```
509         *-Seemingly Unrelated Estimation (SUR) test
510         suest Left Right
511         lincom [Right_mean]_cons - [Left_mean]_cons
512
513         *-上述分析的完整实现：结果的输出和呈现 (self-reading)
514           use "RDD_simu_data0.dta", clear
515             local h=0.1   //带宽
516             *-test ATE_RDD
517               reg y1 xc if (xc>-`h'&xc<0)
518               est store Left
519               reg y1 xc if (xc>0&xc<`h')
520               est store Right
521               suest Left Right
522               lincom [Right_mean]_cons - [Left_mean]_cons
523               global ATE = r(estimate)
524               global ATE_se = r(se)
525             *-分别估计两侧，存储结果
526               reg y1 xc if T==0
527                 est store Left
528               reg y1 xc if T==1
529                 estadd scalar ATE = $ATE
530                 estadd scalar ATE_se = $ATE_se
531                 estadd scalar h = `h'
532                 est store Right
533             *-呈现结果
534               esttab Left Right, nogap s(ATE ATE_se h r2 N)
535
536         *-敏感性测试：
537         *
538         *  更改带宽 h=0.2, 0.3, 看看结果有何变化 ？
539         *  你期望结果对带宽不敏感吗 ？
540         *  何种情况下 ATE 对带宽的选择不敏感.
541
542
543     *------
544     *-A.2.2 最优带宽的选择 (Boundary Bias problem)
545
546       *-RDD 文献中对此进行非常深入的讨论
547             shellout "$R\Jacob_2012_RDD_Guide.pdf" // pp.31-38
548             shellout "$R\rdrobust-SJ-17-2.pdf"
549             shellout "$R\rdrobust-Calonico-SJ14-2.pdf"
550
551       *-带宽的影响
552         use "RDD_simu_data0.dta", clear
553         rd y1 xc, cut(0) bwidth(0.2) mbw(100 200)
554
555         rdplot y1 xc, p(1) h(0.2)  ///
556            graph_options(xlabel(-0.4(0.2)0.4,format(%2.1f)) ///
557                          xline(-0.2 0.2,lp(dash) lc(black*0.4)))
558
559         rdplot y1 xc, p(1) h(0.4)  ///
560            graph_options(xlabel(-0.4(0.2)0.4,format(%2.1f)) ///
561                          xline(-0.4 0.4,lp(dash) lc(black*0.4)))
562
563           *-从上例可以看出，在采用局部回归时，我们需要权衡如下两个问题:
564           *  -   h 越大, 导致偏误的可能性越大;
565           *  -   h 越大, 参数估计越准确，因为包含的样本数更多一些
566
567           *-主要方法: mean squared error (MSE); plug-in; data-driven
568
569           *- Ludwig and Miller(2007) 提出的 Cross-validation (MSE) 以及
570           *  Imbens and Kalyanaraman(2012) 提出的最优带宽估计法
571           *  都仅适用于 Sharp RDD;
572           *- Calonico et al.(2014) 提出的 数据驱动 (data-driven)
```

```
573             *  最优带宽估计法同时适用于 Sharp RDD 和 Fuzzy RDD
574                shellout "$R\C-刘生龙_2016_义务教育法_RDD.pdf"    //pp.163
575
576             *-各种带宽设定下的结果 (略)
577
578                help rdcv   //Cross-validation, Ludwig and Miller(2007)
579
580                 *-Our data
581                 use "RDD_simu_data0.dta", clear
582                 set matsize 2000
583                 set seed 135
584                 sample 10          //随机抽取10%的观察值,否则非常耗时
585                 rdplot y1 xc, c(0) //检测一下，看看数据特征是否发生明显变化
586
587               *-Ludwig and Miller(2007)       Cross-validation
588                 rdcv y1 xc, thr(0) ci  // thr(0.5) 设定 cut-point=0
589                 * bw_L=0.212 ; bw_R=0.227;  Jump=0.979
590
591               *-Imbens and Kalyanaraman(2012) optimal bandwidth
592                 rdcv y1 xc, thr(0) ci ikbwidth
593                 * bw_L=bw_R=0.208;          Jump=0.973
594                 rd   y1 xc, c(0)
595                 * bw_L=bw_R=0.208;          Jump=0.982
596
597               *-ROT plug-in bandwidth
598                 rdcv y1 xc, thr(0) ci rotbwidth
599                 * bw_L=0.1498; bw_R=0.1436; Jump=0.949
600
601               *-Calonico et al.(2014) data-driven bandwidth selection
602                 rdrobust y1 xc, c(0)
603                 * bw_L=bw_R=0.187;          Jump=0.978
604
605
606             *----------
607             *-小结：
608
609             * 目前来看，主流的几个命令 -rd-, -rdcv-, -rdrobust- 都可以很好地
610             * 找到最优带宽，对于实际数据而言，可以同时使用上述2个或多个最优带宽
611             * 对比估计出的结果是否稳定。一个简单的处理方法是使用 -rd- 命令
612                 rd y1 xc
613                 rd y1 xc, mbw(25 50 100 200)
614
615                 *-一次性呈现多种带宽下的 LATE 及 95% CI (这个最实用！)
616                 rd y1 xc, mbw(40(20)200) bdep
617
618             * 这是论文稳健性检验的一个重要环节。
619             * 经过测试，如下命令的结果都差别很小，
620             * ----建议使用----：  rd 和 rdrobust
621                  help rdrobust    //bwselect() option, 通常使用默认值就可以了
622                  help rdbwselect
623                  help rd
624                  help rdcv
625                  help next
626
627             *-公式和依据
628                  help rdrobust
629                  help rdbwselect
630
631             *-描述和表述方式参见:
632                 shellout "$R\C-刘生龙_2016_义务教育法_RDD.pdf" //pp.163
633
634          *-关于最优带宽的实操建议
635          *-Cattaneo, M. D., 2016,
636          *  The choice of neighborhood in regression discontinuity designs,
```

```
637       *  Observational Studies, 2: 134-146.
638       shellout "$R\Cattaneo_2016_RDD_bw.pdf"
639     /*
640     1. Always employ RD optimal data-driven neighborhood (bandwidth or window)
641        selectors, at least as a benchmark or starting point.
642        This gives objectivity and robustness because it incorporates
643        explicitly empirical features such as density of observations,
644        variability of the data, or curvature of the unknown regression
645        functions, in a principled way.
646     2. Employ data-driven neighborhood (bandwidth or window) selectors
647        according to the specific goal and assumptions imposed,
648        which should also be explicitly stated and explained.
649        There is no one neighborhood selector appropriate for all objectives
650        when using local polynomial approximations,
651        and even for local randomization methods sensitivity analysis with
652        respect to the neighborhood used is very important.
653     3. Do not employ the same neighborhood for different outcome variables,
654        pre-intervention covariates (if conducting falsification testing),
655        estimation and inference procedures, or falsification methods.
656        Using the same neighborhood for different goals, outcomes or
657        samples disregards the specific empirical features
658        (e.g., number of observations near the cutoff, variability or curvature),
659        and will lead to unreliable empirical results due
660        to invalidity of the methods employed.
661     */
662
663
664     *------
665     *-A.2.3  局部线性回归的惯用模型 (重要!)
666
667       * - 上述分析中，我们分了左右两段分别执行 OLS 估计，
668       *   相当于允许左右两侧的斜率可以不同，即 dy/dxc(left) != dy/dxc(right)
669       *   显然，这是一个一般化的设定;
670       * - 文献中通常是假设左右两侧的斜率相同，因为 window 都长都比较窄
671       *   上述 -rd-, -rdrobust- 等命令都是使用的这个设定
672
673
674     *-实现方法:
675
676       use "RDD_simu_data0.dta", clear
677
678       *-选择最优带宽
679       rdrobust y1 xc      //自动选择最优带宽
680       global h = e(h_l)   //获取最优带宽，也可以用 -rdbdselect- 命令
681       *local h = 0.09885  //与上一行命令等价
682
683       *-估计
684       reg y1 T xc if (xc>=-$h)&(xc<=$h)
685       *           ------------------------
686       *            限制估计窗口(window)
687       *
688       *-Q: _cons=1.273 ;  _b[T]=0.963 , 这两个系数的含义是什么?
689
690       lincom _cons+T   // What is this?
691
692       *-Q: 能否在下图中找到对应的点?
693         rdplot y1 xc if (xc>=-$h)&(xc<=$h), h($h $h) p(1) //局部呈现
694         rdplot y1 xc, h($h $h) p(1)                       //全局呈现
695         *-Q: 此时为何设定 p(1) 选项?
696         *    h() 选项的作用是什么?
697
698       *-扩展：加入其它控制变量
699         reg y1 T xc z if (xc>=-$h)&(xc<=$h)  //回想一下我们的 DGP !
700
```

```
701
702         *-小结:
703
704         * ------------------------------------------------------------
705         *    y[i] = a0 + b1*Treat[i] + b2*XC[i] + b3*controls[i] + u[i]
706         * ------------------------------------------------------------
707
708         *      y[i] : outcome variable
709         *     XC[i] : Assignment variable (centered) = (X-cutpoint)
710         *  Treat[i] : Treat=1 if XC>0 (or X>C); Treat=0 if XC<0
711         *       ATE : Local ATE = b2 (Local! not Global!)
712         *    Sample :  -h < x < +h
713
714
715     *------
716     *-A.2.4 多项式回归 (全局/局部)
717
718
719       *-A.2.4.1 估计方法 (全局多项式回归)
720
721         *-基本思路: 加入 xc 的高阶项,
722         *           也可以进一步加入 xc 与 Treat 的交乘项及交乘项的高阶项
723
724         *-常用函数形式:
725            shellout "$R\Jacob_2012_RDD_Guide.pdf" // pp.21
726
727         *-加入 xc 的高阶项
728           use "RDD_simu_data0.dta", clear
729           gen xc2 = xc*xc
730           gen xc3 = xc*xc2
731           gen xc4 = xc*xc3
732           gen xc5 = xc*xc4
733           gen xc6 = xc*xc5
734           gen xc7 = xc*xc6
735           gen xc8 = xc*xc7
736
737           reg y1 T xc xc2 xc3 xc4
738           est store m2
739           reg y1 T xc xc2 xc3 xc4
740           est store m4
741           reg y1 T xc xc2 xc3 xc4 xc5 xc6
742           est store m6
743           reg y1 T xc xc2 xc3 xc4 z
744           est store m4z
745           reg y1 T xc xc2 xc3 xc4 xc5 xc6 z
746           est store m6z
747           local m "m2 m4 m6 m4z m6z"
748           esttab `m', mtitle(`m') b(%6.3f) t(%6.3f) s(N r2) ///
749                  nogap compress
750
751         *-阶数的选择 (信息准则: AIC 或 BIC 越小越好)
752           *----------------------------------myic----------
753            program define myic
754            version 13
755              qui estat ic
756              mat a = r(S)
757              estadd scalar AIC = a[1,5]
758              estadd scalar BIC = a[1,6]
759            end
760           *----------------------------------myic----------
761           *-Note: 调用方法
762           *-选中上述程序, 按快捷键 Ctrl+R, 将程序读入内存
763
764           #d ;
```

```
765             reg y1 T xc;              myic;   est store m1;
766             reg y1 T xc xc2;          myic;   est store m2;
767             reg y1 T xc xc2-xc3;      myic;   est store m3;
768             reg y1 T xc xc2-xc4;      myic;   est store m4;
769             reg y1 T xc xc2-xc5;      myic;   est store m5;
770             reg y1 T xc xc2-xc6;      myic;   est store m6;
771             reg y1 T xc xc2-xc8;      myic;   est store m8;
772             #d cr    // #d 表示 #delimit
773
774           *-对比结果
775             local m "m1 m2 m3 m4 m5 m6 m8"
776             esttab `m', mtitle(`m') b(%6.3f) t(%6.3f)  ///
777                    s(N r2 r2_a AIC BIC) nogap compress
778
779         *-评述:
780
781         * - 基于 AIC 和 BIC 准则，我们会选择 m3，但这个模型的结果似乎是有偏的;
782
783         * - 确定全局的模型形式并不是一件简单的事情;
784         * - 我们还可以在上述模型中进一步加入 xc 与 Treat 的交乘项;
785
786         * - 相对而言，局部线性回归反而效果更好 -- 可以结合起来
787
788
789       *-A.2.4.2 局部多项式回归  (文献中用的比较多)
790
791         global h = 0.2   //可以使用其他带宽
792         global window "(xc>=-$h)&(xc<=$h)"
793         reg y1 T xc          if $window
794         est store mx1
795         reg y1 T xc xc2      if $window
796         est store mx2
797         reg y1 T xc xc2 xc3 if $window
798         est store mx3
799         *-加入 Treat 与 xc 的交乘项
800         dropvars xc*T
801         gen xc1T =  xc*T
802         gen xc2T = xc2*T
803         gen xc3T = xc3*T
804         reg y1 T xc xc1T                    if $window
805         est store mTx1
806         reg y1 T xc xc2 xc1T xc2T           if $window
807         est store mTx2
808         reg y1 T xc xc2 xc3 xc1T xc2T xc3T  if $window
809         est store mTx3
810
811         local m "mx1 mx2 mx3  mTx1 mTx2 mTx3"
812         esttab `m', mtitle(`m') nogap compress s(N r2 r2_a)
813
814         *-可以确定 ATE 的大致范围: 0.96 - 0.99
815
816         *-快捷实现:
817           rd y1 xc
818           *-最优带宽:      BW est. (h) |    0.144
819           *-多项式阶数: Order est. (p) |        1
820           *- ATE:          Conventional |   0.96907
821
822           rd y1 xc,  covs(xc2) //加入 xc*xc
823           *                  BW est. (h) |    0.203
824
825         *-应用:
826           shellout "$R\Hoekstra_2009_RDD.pdf"  // pp.722, Table 1
827
828
```

```
829
830     *-A.2.4.3 核加权局部多项式平滑 (Kernel-weighted local polynomial smoothing)
831     *                                (self-reading)
832
833       *-计量表述和正式介绍:
834         shellout "$R\Imbens_2008_JE_Guide.pdf"  //pp.623, sec 4.1
835
836       help lpoly
837
838       use "RDD_simu_data0.dta", clear
839       gen cut=0 in 1
840       lpoly y1 xc if xc<0 , at(cut) gen(av_y0)
841       lpoly y1 xc if xc>=0, at(cut) gen(av_y1)
842       display "Estimate (jump): " av_y1[1]-av_y0[1]
843       *-see:
844         shellout "$R\Nichols_2007.pdf"  // pp.529
845
846       use "RDD_simu_data0.dta", clear
847       lpoly y1 xc if T==0, nograph kernel(triangle) gen(x0 sm_y0) bwidth(0.1)
848       lpoly y1 xc if T==1, nograph kernel(triangle) gen(x1 sm_y1) bwidth(0.1)
849       twoway (scatter sm_y1 x1, color(blue) msize(small))  ///
850              (scatter sm_y0 x0, color(blue) msize(small)), ///
851               xline(0, lp(dash)) legend(off) ///
852               xtitle("x-variable") ytitle("y-variable")
853
854       *-核估计 (self-reading)
855       use "RDD_simu_data0.dta", clear
856       global h = 0.1    //带宽,可以使用 -rd- 或 -rdrobust- 提供的带宽
857       gen dx1 = xc/$h   // (x-c)/h, h 是带宽(bandwidth), 通常使用最优带宽
858       gen kernal = (1-dx1*dx1)  // Epanechnikov Kernal
859       reg y1 T xc   [pweight=kernal] if ((xc>=-$h)&(xc<=$h))
860       reg y1 T xc z [pweight=kernal] if ((xc>=-$h)&(xc<=$h))
861       *-计量表述:
862         shellout "$R\Imbens_2008_JE_Guide.pdf"  //pp.623, sec 4.1
863
864
865   *------
866   *-A.2.4 小结: 估计 RDD 的两种主要方法
867
868       * [1] 局部线性回归 local regression; 关键点: 最优带宽的选择!
869       *
870       * [2] 非线性估计, 非参数估计
871       *     如, 分段多项式回归(fractional polynomial regression),
872       *         局部多项式回归(local polynomial regression), 各种 smoothing
873             shellout "$R\Ctaber_2012_RD.pdf"  // kernel function 介绍
874       *
875       *-Q: 二者的优劣?
876       *
877       *    @ 局部线性回归: 要求在断点附近有较多的观察值
878       *
879       *    @ 非参数估计需要选择合适的核函数及平滑方法来拟合非线性曲线
880       *
881       *    @ 数值上, 非参数方法等价于参数估计量 (Hahn et al., 1999, 2001)
882       *
883       *    详见 邹红和喻开志(2015)
884       * - 邹红, 喻开志, 2015,
885       *   退休与城镇家庭消费:基于断点回归设计的经验证据,
886       *   经济研究, (1): 124-139.
887           shellout "$R\C-邹红_2015_退休_RDD.pdf"
888
889       *-是否需要加入控制变量?
890
891         help rd  // see covar(varlist) ...
892
```

```
893         *-ref: Nichols, Austin. 2011.
894         *  rd 2.0: Revised Stata module for regression discontinuity estimation.
895         *  http://ideas.repec.org/c/boc/bocode/s456888.html
896
897         /*
898       It is generally a Very Bad Idea to add covariates to Local Wald Estimation.
899         It is possible that covariates could reduce residual variance
900         and improve efficiency, but estimation error in their coefficients
901         could also reduce efficiency, and any violations of the assumptions
902         that such covariates are exogenous and have a linear impact on mean
903         treatment and outcomes could greatly increase bias. */
904
905         *-参见如下文章 pp.1548 中的表述
906           shellout "$R\雷晓燕_CQJE_2010_退休会影响健康吗.pdf"
907
908
909
910     *--------------------
911     *-A.3 RDD 的假设条件
912     *--------------------
913
914     *-RDD 的有效性依赖于如下两个假设：
915
916         shellout "$R\Imbens_2008_JE_Guide.pdf"  // pp.618, 假设条件表述
917         shellout "$R\Imbens_2008_JE_Guide.pdf"  // pp.631, sec 7, 实现
918
919       *-Nichols, Austin. 2007.
920       *   Causal Inference with Observational Data."
921       *   The Stata Journal 7(4): 507-541
922         shellout "$R\Nichols_2007.pdf"  // pp.528, Section 5.1
923
924
925         *----A1------
926         *
927         *-局域平滑假设 (Local Smoothing)
928         * [or] 回归方程条件连续假设
929         *      (Continuity of Conditional Regression Functions, Imbens_2008)
930         *
931         *----A2------
932         *
933         *-分布函数条件连续假设
934         *      (Continuity of Conditional Distribution Functions)
935
936
937       *-直观解释：
938       *
939       *     若没有 Treatment，则在断点(cutoff)处，y 是 x 的连续函数;
940       *
941       *     只有这假设得到满足，RDD 才是一个真正的“局部实验设计”
942       *     我们观察到的 jump 才可以视为 Treat 产生的效果
943
944
945       *-检验方法：(无法直接检验，但可以从如下几个方面间接检验)
946
947         *-M1: 观察 Outcome 变量在除了 cut-point 以外的其他位置是否连续
948           use "RDD_simu_data0.dta", clear
949           cmogram y1 xc
950
951           *-see:
952             shellout "$R\Nichols_2007.pdf"  // pp.529
953
954         *-M2: 分配变量(margin)不受人为操控
955         * - 分配变量(本身)的在断点处的分布是连续的，不存在明显的断点;
956         * - 检验方法: 绘制分配变量的直方图或密度函数图
```

```
957             histogram xc, xline(0)
958             kdensity  xc, xline(0)
959
960         *-M3: 平滑性
961         * - 除 Outcome 变量外，其他影响变量在分界点两侧不应有明显的跳跃
962         *
963         * - 检验方法:
964         *   - 回归法, 使用 -rd- 或 -rdrobust- 命令分析协变量与x的关系
965         *   - 图形法, 使用 -rdplot- 观察协变量在 cut-point 处是否跳跃
966
967             use "RDD_simu_data0.dta", clear
968
969             cmogram z xc                // 似乎不连续
970
971             rdplot z xc                 // 似乎不连续
972             rdplot z xc, p(0)
973             rdplot z xc, p(1)
974             rdplot z xc if abs(xc)<0.1, p(2)
975
976             rdrobust z xc               // 其实是 ...
977
978
979       *-中文应用参见:
980         shellout "$R\C-刘生龙_2016_义务教育法_RDD.pdf" //pp.163
981
982       *-更具体实现方法参见下文 6.5 小节的[范例]: 美国参议院选举
983
984
985
986   *--------------------
987   *-A.6  RDD 总结
988   *--------------------
989
990   *------
991   *-A.6.1  RDD 的核心思想：局域实验设计(Local Experiment Design)
992
993     *----------------
994     *-何时应用 RDD ?
995     *
996     *  受益者/非受益者能够按一个可数量化的维度排序(如分数, 年龄) 【分配变量】
997     *
998     *  这一维度可以用来计算已定义的指数/参数
999     *
1000    *  这个指数/参数对目标人群的资格确定有一个【断点(cutoff)】(如重点线, 55岁)
1001    *
1002    *  这个指数值就是决定了把潜在的受益者分配到实验组或对照组(控制组)的临界值
1003    *    如，超过 55 岁即退休; 高考分数超过重点线即可上重点大学
1004
1005
1006    *--------------------
1007    *-对 RDD 的直观解释:
1008    *
1009    *  略高于界点的潜在受益者(单元)与略低于界点的潜在受益者(单元)是非常相似的
1010    *
1011    *  我们要比较略高于界点与略低于界点的两组人群(单元)的结果
1012    *
1013    *  LATE: Local Average Treatment Effects
1014
1015
1016      *-----------------------------------------------------begin---------
1017      *-一个简单的例子：退休对健康的影响
1018
1019        shellout "$R\雷晓燕_CQJE_2010_退休会影响健康吗.pdf"  // 写的比较细致
1020        shellout "$R\C-邹红_2015_退休_RDD.pdf" //姊妹篇: 退休对消费支出的影响
```

```
1021
1022          * Outcome: Y  健康状况(自评结果, 1,2,3,4)
1023
1024          *- Cutoff: c
1025          *
1026          *  男性：60岁；  女性：50 或 55岁
1027
1028          *-内生性问题的来源：
1029          * 其一，健康(Y)本身就是影响退休(T)的关键因素；
1030          *     这导致影响 Y 的因素(如 U) 也会影响 D，从而使 U 的 T 相关
1031          * 其二，一些不可观测的因素(个人偏好,身体素质,遗传因素等)
1032          *     都会同时影响健康和退休，这导致 U 和 T 是相关的
1033
1034          *-实验变量 T (Treatment variable)
1035          *  T=1 if Age>=55, 退休
1036          *  T=0 otherwise
1037
1038          *-分配变量 X (Assignment variable, or, Forced variable)
1039          *  X (Age), 其取值的大小决定了 T，或者说，决定了个体是否接受试验
1040          *       强制退休：Age 可以视为外生
1041          *       但个体可能虚报年龄。
1042          *  又如，
1043          *       高考分数(是否能被大学录取)；
1044          *       选票比例(是否能在选举中获胜)；
1045          *       月均收入(是否接受贫困救济)
1046          *-----------------------------------------------------over---------
1047
1048          *--------------
1049          *-SRDD v.s. FRDD (Sharp RDD, Fuzzy RDD)
1050
1051          *-SRDD (明确断点 RDD)
1052
1053          *  界点确切地决定了干预实验组
1054          *  等同于在一定邻近区域内的随机分配
1055          *  例如：议会选举 (得票超过 50% 获胜)
1056          *        高考分数更像一个 SRDD (电脑投档情况下，超过分数线被录取)
1057
1058          *-FRDD (模糊断点 RDD)
1059          *  界点（的值）与干预高度相关
1060          *  例如：规则决定了目标人群的资格界定，但是存在一定的管理误差
1061          *
1062          *   养老金(55岁);  高考分数(非电脑投档机制);
1063
1064
1065     *------
1066     *-A.6.2  RDD 的前世今生
1067
1068        shellout "$R\Jacob_2012_RDD_Guide.pdf" // pp.2
1069                                                                        /*
1070       - RD was first introduced by Thistlethwaite and Campbell (1960)
1071         as an alternative method for evaluating social programs.
1072       - Their work generated a flurry of related activity, which subsequently
1073         died out.
1074       - Economists revived the approach (Goldberger, 1972, 2008; van der Klaauw,
1075         1997, 2002; Angrist and Lavy, 1999),
1076         @ formalized it (Hahn, Todd, and van der Klaauw, 2001),
1077         @ strengthened its estimation methods (Imbens and Kalyanaraman, 2009),
1078           and began to apply it to many different research questions.
1079       - This renaissance culminated in a 2008 special issue on RD
1080         analysis in the Journal of Econometrics.
1081
1082       - Applications of RDD
1083         @ the impact of unionization (DiNardo and Lee, 2004),
1084         @ anti-discrimination laws (Hahn, Todd, and van der Klaauw, 1999),
```

```
1085        @ social assistance programs (Lemieux and Milligan, 2004),
1086        @ limits on unemployment insurance (Black, Galdo, and Smith, 2007),
1087        @ effect of financial aid offers on college enrollment(van der Klaauw,2002)
1088        @ impact of class size reduction (Angrist and Lavy, 1999),
1089        @ remedial education (Jacob and Lefgren, 2006),
1090        @ delayed entry to kindergarten (McEwan and Shapiro, 2008),
1091        @ impact of the Reading First program on instructional practice and student
1092          achievement (Gamse, Bloom, Kemple, and Jacob, 2008).                  */
1093          shellout "$R\余静文_2011_综述.pdf"  //对上述多数文献都进行了简要介绍
1094
1095      *- China Evidence
1096       *@ 退休对健康的影响
1097          shellout "$R\雷晓燕_CQJE_2010_退休会影响健康吗.pdf"
1098       *@ 退休对消费支出的影响
1099          shellout "$R\C-邹红_2015_退休_RDD.pdf"
1100       *@ 财政转移支付
1101          shellout "$R\C-刘畅_2015_财政转移支付_RDD.pdf"
1102       *@ 义务教育法与教育回报
1103          shellout "$R\C-刘生龙_2016_义务教育法_RDD.pdf"
1104       *@ 中文综述和介绍
1105          shellout "$R\C-罗胜_2016_综述_RDD.pdf"
1106          shellout "$R\余静文_2011_综述.pdf"
1107
1108      *-RDD 的局限
1109        shellout "$R\Regression_Discontinuity.pdf"
1110        *-断点回归衡量的是在临界值附近的局部平均效应，
1111        * 不是一个整体的平均效应，很难推广到整体研究中。
1112
1113
1114    *------
1115    *-A.6.3  RDD 参考资料
1116
1117      *-三本重要的参考书
1118
1119        view browse "https://www.jianshu.com/p/538ed1805004" // Stata连享会-简书文章
1120
1121        *-Cattaneo, M. D., N. Idrobo, and R. Titiunik (2018a):
1122        *  A Practical Introduction to Regression Discontinuity Designs: Part I.
1123        *  Cambridge Elements: Quantitative and Computational Methods for Social
       Science,
1124        *  Cambridge University
1125        shellout "$R\Cattane2018-RDD--V01.pdf"
1126
1127        *-Cattaneo, M. D., N. Idrobo, and R. Titiunik (2018a):
1128        *  A Practical Introduction to Regression Discontinuity Designs: Part II.
1129        *  Cambridge Elements: Quantitative and Computational Methods for Social
       Science,
1130        *  Cambridge University
1131        shellout "$R\Cattane2018-RDD--V02.pdf"
1132
1133        *-这本小书提供了 RDD 分析中的各种建议和实操指南
1134        *-Jacob, R., P. Zhu, M. A. Somers, H. Bloom, 2012,
1135        *   A practical guide to regression discontinuity, MDRC working paper.
1136            shellout "$R\Jacob_2012_RDD_Guide.pdf" //RDD 的前世今生, pp.2
1137
1138
1139      *-几份重要的 PPT
1140
1141        shellout "$R\Yang_2017_RDDa.pdf"     // pp.7  Sharp RDD
1142        shellout "$R\Yang_2017_RDDb.pdf"     // Fuzzy RDD
1143        shellout "$R\Yang_2017_RDD_PPT.pdf"  // 合集
1144        *-中文版
1145        shellout "$R\楊子霆-2017-断点回归方法的介绍与应用.pptx"
1146
```

```
1147        shellout "$R\Sylvia-2015-PPT-RDD.pdf"       //带有中文翻译，讲的很清楚
1148
1149        shellout "$R\lecture_4_-_rdd.pdf"           //Fabian Waldinger, 2010
1150
1151        shellout "$R\Regression_Discontinuity.pdf" //Jeremy Magruder, 2009
1152
1153
1154
1155
1156      *-几篇重要的文章
1157
1158        *-该文对各种内生性问题的处理方法进行了全面介绍，提供了范例
1159        *-Nichols, Austin. 2007.
1160        *   Causal Inference with Observational Data."
1161        *   The Stata Journal 7(4): 507–541
1162            shellout "$R\Nichols_2007.pdf"
1163
1164        *-Lee, D., T. Lemieux, 2010,
1165        *   Regression Discontinuity Designs in Economics,
1166        *   Journal of Economic Literature, 48: 281-355.
1167            shellout "$R\Lee_Lemieux_2010_JEL.pdf"
1168
1169        *-Imbens, G., T. Lemieux, 2008,
1170        *   Regression discontinuity designs: A guide to practice,
1171        *   Journal of Econometrics, 142 (2): 615-635.
1172            shellout "$R\Imbens_2008_JE_Guide.pdf"
1173
1174        *-Barrera-Osorio, F., D. Raju, 2010,
1175        * Evaluating a test-based public subsidy program for
1176        * low-cost private schools:
1177        * Regression-discontinuity evidence from Pakistan.
1178        * Working paper
1179            shellout "$R\Barrera_Raju_2010_RDD.pdf"  // RD 应用，写的很细致
1180
1181        * Lee, D. S., 2008, (Excellent paper)
1182        *   Randomized experiments from non-random selection
1183        *   in US House elections,
1184        *   Journal of Econometrics, 142 (2): 675-697.
1185            shellout "$R\Lee_2008_Selection.pdf"     // 已经被引用 300 多次
1186
1187        *-Fuji, D., G. Imbens, K. Kalyanaraman, 2009,
1188        *   Notes for matlab and stata regression discontinuity software,
1189        *   Working Paper.
1190            shellout "$R\rd_Imbens_procedure.pdf"
1191            adoedit "rdob.ado"
1192
1193
1194      *-经典应用
1195
1196        *-Hoekstra, M., 2009, The effect of attending the flagship state
1197        *  university on earnings: A discontinuity-based approach,
1198        *  Review of Economics and Statistics, 91 (4): 717-724.
1199           shellout "$R\Hoekstra_2009_RDD.pdf"   //名校的收入效应
1200           shellout "$R\Yang_2017_RDDa.pdf"      //pp.7对该文有介绍
1201
1202        *-荷兰-德国边境的房价差异
1203        *-Micheli, M., J. Rouwendal, J. Dekkers, 2014,
1204        *  Border effects in house prices,
1205        *  Real Estate Economics, working paper.
1206           shellout "$R\Micheli-2014-RDD-Border Effects in House Prices.pdf"
1207
1208
1209      *-RDD 相关的 Stata 命令
1210
```

```
1211        *-1-  rdrobust
1212
1213            help rdrobust
1214
1215            shellout "$R\rdrobust-SJ-17-2.pdf"
1216            shellout "$R\rdrobust-Calonico-SJ14-2.pdf"
1217
1218            help rdroubst
1219            help rdbwselect
1220            help rdplot
1221
1222        *-2- rd
1223
1224            help rd
1225
1226            shellout "$R\Nichols_2007.pdf"
1227            shellout "$R\rd_Imbens_procedure.pdf"  // -rd- 命令估计过程详解
1228
1229        *-3- 其他命令(请查看帮助文件中的范例和参考文献)
1230
1231            help next
1232                 shellout "$R\Long_2016_Next_RD.pdf"
1233
1234            help rdcv
1235            help cmogram
1236
1237
1238
1239    *------
1240    *-A.6.4  RDD 模型设定和估计方法
1241
1242
1243          *-See Imbens and Lemieux (2008), Lee and Lemieux(2008)
1244
1245          shellout "$R\Imbens_2008_JE_Guide.pdf"  // Section 2
1246          shellout "$R\Lee_Lemieux_2010_JEL.pdf"  // Section 3
1247
1248          shellout "$R\Yang_2017_RDD_PPT.pdf"     // Tzu-Ting Yang, 2017
1249
1250          shellout "$R\lecture_4_-_rdd.pdf"         //Fabian Waldinger, 2010
1251
1252          shellout "$R\Regression_Discontinuity.pdf" //Jeremy Magruder, 2009
1253
1254
1255        *----------
1256        *-Sharp RDD  // Lee_Lemieux_2010_JEL, p.293
1257
1258          * 参见 *-A.2.3 小节
1259
1260        * ------------------------------------------------------------
1261        *    y[i] = a0 + b1*Treat[i] + b2*XC[i] + b3*controls[i] + u[i]
1262        * ------------------------------------------------------------
1263
1264        *      y[i] : outcome variable
1265        *     XC[i] : Assignment variable (centered) = (X-cutpoint)
1266        *  Treat[i] : Treat=1 if XC>0 (or X>C); Treat=0 if XC<0
1267        *       ATE : Local ATE = b2 (Local! not Global!)
1268        *    Sample :  -h < x < +h
1269
1270        *----------
1271        *-Fuzzy RDD
1272
1273          * 参见 *-A.4 小节
1274
```

```
1275
1276    *----------
1277    *-Also see:
1278
1279      *-Quantile treatment effects RDD (分位数 RDD)
1280        help rdqtese
1281        shellout "$R\rdqtese_hlp.txt" //帮助文件
1282        *-Frandsen, B. R., M. Frolich, B. Melly, 2012,
1283        * Quantile treatment effects in the regression discontinuity design,
1284        * Journal of Econometrics, 168 (2): 382-395.
1285          shellout "$R\Frandsen-2012-QuantileRDD.pdf"
1286
1287    *-待更新
1288      help rddensity // Manipulation testing.
1289      help rdlocrand // Local randomization methods.
1290      help rdmulti   // RD plots, estimation, inference, and extrapolation with
        multiple cutoffs and multiple scores.
1291      help rdpower   // Power and sample size calculations.
1292
1293    *----------
1294    *-最后的话：   总而言，神秘的 RDD 其实就是升级版的 OLS + dummy !
1295    *----------
1296
1297
```