Applied Microeconometrics II
Assignment 3 Solutions

You are encouraged to work in groups of at most three students.

Please show your work. For Stata questions, **include the log files of your output.**

(1) (2X5p=10p) A bank offers a week long management training program to all loan officers who are at the end of the second year in the job. Participation in the program is voluntary. The bank is interested in knowing whether participation in the program makes it more likely that a loan officer is promoted to branch manager. You have data on all the bank's loan officers, their participation in the program, and whether they have been promoted between the third and fifth year on the job. You run a regression for the promotion decision on a constant and a dummy for participation in the training program.

(a) Now suppose the bank encourages participation in the training program by sending a personal letter from the CEO to selected employees who have been recommended by their supervisor as showing particular potential for a position in management. Would it be useful to use the CEO letter as an instrumental variable for participation in the training program? Explain why or why not.

This instrument fails the *independence assumption*. Presumably the letter would be sent to handpicked candidates who seem good candidates for promotion, e.g., candidates who have high ambition or strong managerial skills. So the letter instrument would be correlated with characteristics that drive the promotion decision.

(b) Suppose instead the bank encourages participation in the training program by sending a personal letter from the CEO to a randomly chosen set of employees. Would it be useful to use this CEO letter as an instrument for participation in the training programme? Explain why or why not.

Yes, as the randomization ensures that receiving the letter is not correlated with the outcome, so this instrument meets the *independence assumption* and it seems likely also to meet the *exclusion restriction*. Finally, receiving the letter will likely have an impact on training take-up, and if so this instrument would meet the first of our requirements—that the instrument has a non-zero impact on treatment take-up.

(2) (5X6=30p) Oregon Health Experiment

In 2008, a group of uninsured low-income individuals in Oregon was selected by lottery for a chance to enroll in Medicaid, a federal and state program that provides aid covering medical costs for low income individuals. An evaluation of the program[1] found that the treatment group selected by lottery was more likely to have health insurance than the control group, which did not win the lottery. The treatment group also reported better health outcomes, higher utilization of medical care, and lower out of pocket medical expenses.

(a) Using the Oregon data file, estimate the following regression by OLS, where Y_i is an indicator for whether individuals report they have received all the medical care they needed in the past six months (individuals were surveyed one year after the experiment began). In all the regressions in this question, cluster standard errors at the household level.

$$Y_i = \beta_0 + \beta_1 Lottery_i + \epsilon_i$$

Lottery is an indicator variable for whether an individual was selected by lottery to be eligible to enroll in Medicaid. Interpret the coefficient on Lottery.

See log file for output. Being selected to be eligible for Medicaid is associated with a 6 percentage point higher probability of reporting all medical needs are met. Note that this is an Intent to Treat (ITT) effect, in the sense that the treatment group is "winning the lottery", but this treatment group includes individuals we intended to treat, not all of whom actually received the treatment, as some did not sign up for Medicaid even if they won the lottery.

(b) Now suppose you are interested in estimating the effects of actually enrolling in Medicaid on whether individuals got all the medical care needed in the past six months (Y_i). Why do we believe the coefficient β_1 in the following regression will be biased?

$$Y_i = \beta_0 + \beta_1 Medicaid_i + \epsilon_i$$

The coefficient captures the outcomes for individuals who selected into treatment. In IV jargon, we are measuring the treatment effect for compliers, along with always takers, but we believe the potential outcomes for these groups may be fundamentally different than those of never takers or defiers, in which case

[1] Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and the Oregon Health Study Group. (2012). "The Oregon Health Insurance Experiment: Evidence from the First Year." *Quarterly Journal of Economics* 127(3): 1057-1106.

the β_1 coefficient will not show a causal effect of Medicare enrollment, capturing the bias resulting from selection into Medicaid.

(c) We will estimate the regression in part (b) using the random lottery assignment as an instrumental variable. Estimate the first stage equation and interpret the coefficient on the independent variable of interest.

Being selected for the lottery is associated with a 16.4 percentage point higher probability of enrolling in Medicaid. Note the effect is highly statistically significant.

(d) Find the 2SLS estimate of the effect of Medicaid on whether individuals received all the medical care needed in the past six months. Explain why we call this estimate a local average treatment effect.

The 2SLS effect is that enrolling in Medicaid as a result of the lottery results in a 36.5 percentage point higher probability of responding that all medical needs have been met in the past six months. This is a Local Average Treatment Effect (LATE), as it only pertains to those individuals whose enrollment in Medicaid was the result of winning the lottery.

(e) Verify numerically (a simple ratio) that the 2SLS effect is the ratio of the reduced form and the first stage coefficients.

.059786/.1638377 = .36490991

(3) (6*7=42p)Sharp Regression Discontinuity (RD) Design: Anderson, Michael L. (2014). "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion." *American Economic Review*, 104(9): 2763-96.

a) The author argues that the timing of the strike was exogenous. Consider a regression of average delay on the indicator for whether a strike was occurring. How could the date of the strike be endogenous to the problem at hand? Think about when unions choose to schedule strikes. Would the bias introduced by an endogenous strike date be positive or negative?

A union would likely wait for a time when traffic is congested, in order to increase its bargaining power. If that is the case, the before-after difference in average delays would include the actual effect of the strike plus the bias introduced by the fact that the strike was scheduled at a time of increasing traffic. The bias would be positive.

b) How does the regression model described in the paper (equation (8)) address the potential endogeneity of the strike date?

The author states: "Identification in the RD model comes from assuming that the underlying, potentially endogenous relationship between the error term and the date of the strike is eliminated by the flexible function f (·)." In other words, we are controlling for the evolution of traffic delays through a flexible polynomial, to account for the fact that traffic might have been increasing specifically at the time of the strike for other reasons.

c) In which sense is the regression discontinuity "sharp"? Why does the regression model include the interaction of the running variable time and the strike indicator?

The regresssion is sharp in the sense that the treatment variable, strike, sees a discrete one-time change at the strike date.

The interaction between the running variable and the strike indicator is attempting to capture the fact that the underlying relationship between the running variable and average delays may be different after the strike.

d) Using the file strikedata.dta, reproduce the coefficient for **strike** in Table 4, column 1, as described in equation (8) in the paper.

- You will need to use the **areg** command instead of the familiar regress command. The areg command allows you to incorporate many indicator variables using the **absorb** option.
- You will also need to weight observations: this is accomplished by using the [**aw=weight**] syntax after specifying the list of explanatory variables.
- You will need to center the date variable around the day the strike begins, and interact the running variable with the strike variable. The day of the strike has the value 15992 in the dataset.
- You will need to create indicator variables for days of the week.
- You should cluster observations at the ID level (variable **vds**).

See the log file.

e) Run the same regression as in part d), but now restrict the analysis to only 10 days before and after the cutoff [dates 15977 through 16006, inclusive]. What happens to the observed effect? What happens if you restrict the analysis to only five days before and after the cut-off [dates 15984 through 15999, inclusive] ? Do you think the results you obtained in part e) are more informative about the effect of the policy than results in part f)? Why, or why not? Conceptually, what do you think happens to the Type II error probability as you restrict the bandwidth?

When analyzing regression discontinuity designs, it is tempting to restrict the bandwidth to improve internal validity: the shorter the period around the discontinuity window, the more likely it is that the quasi-treatment (observations a few days

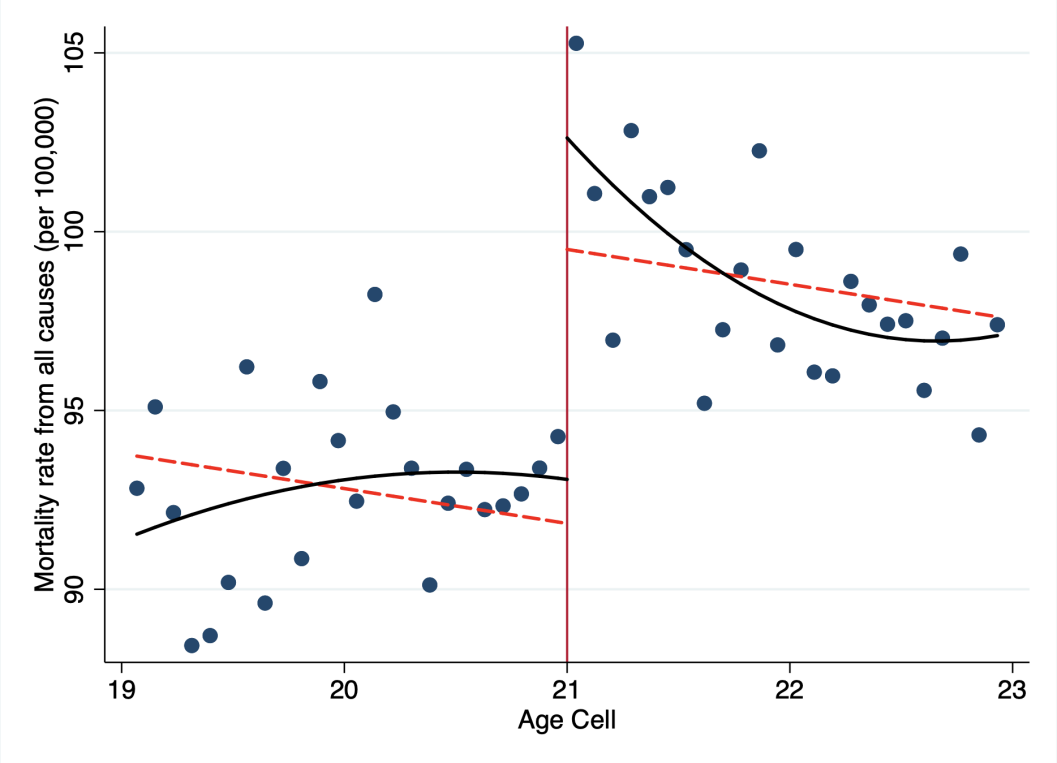
after the strike started) and quasi-control groups (observations a few days before) are comparable. There are downsides to this approach though: the smaller sample size means we will have a higher Type II error probability, meaning we may fail to identify a real treatment effect simply because we have too few observations.

Results when restricting the analysis to just ten days before and after the cutoff are similar to the 28 day bandwidth. However, results for just five days before and after have a different sign. Such a short time frame is however not informative, as it does not allow us to adequately capture the time trend and the fixed effects for day of the week. Five days might also too short a period since individuals may not switch immediately to driving after the strike commences (e.g. they may try walking, biking or carpooling first, and then rent a car or drive their own car).

f) How does Figure 6 help the interpretation of the results in Figure 2? Explain.

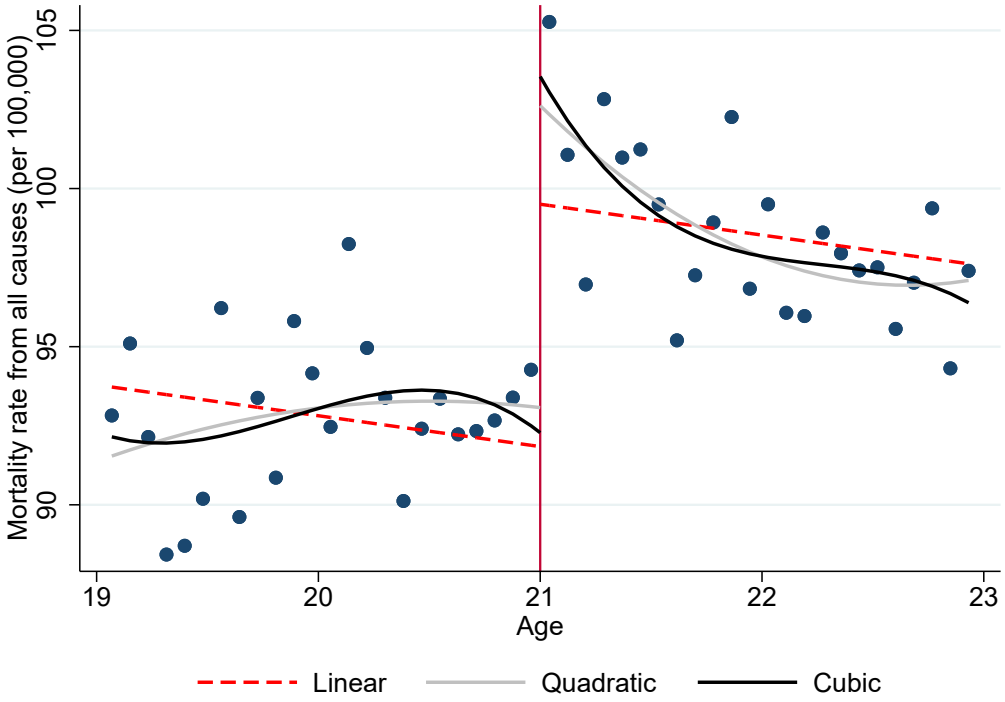
Figure 6 shows the functional relationship between time and average delays a year later. It highlights the fact that we see a clear discontinuity in Figure 2. If the two pictures looked similar, we would be concerned that the strike just happened to coincide with a period of congested traffic.

(4) (18 p) In the review session, you will replicate the figure below (Fig 4.4. in the "Mastering Metrics" textbook). Kimmo will guide you through the replication.



Your task is to Include a cubic function of age in the RD regression, instead of a quadratic. Does that make any meaningful difference in the results? Does it increase your confidence in the results? Show the figure you obtain and your log file.

The figure looks similar, and regression results (see log file) are also comparable: the main variable of interest continues to be highly statistically significant. Adding a cubic serves as a robustness check, strengthening our confidence in the results.



```
-------------------------------------------------------------------------------
> -------------------------------------------------
      name:  <unnamed>
       log:  C:\Users\sahlste1\OneDrive - Aalto University\jatko-opinnot\opetus\Applie
> d Microeconometrics 2\Assignment 3\assignment_3_lo
> g.log
  log type:  text
 opened on:  26 Nov 2023, 15:41:10

. 
. ******************************************************************************
. ** QUESTION 2 **
. 
. use "Oregon_data.dta", clear

. 
. ** 2 a) **
. 
. reg needmet_med_6m treatment, vce(cluster household_id)

Linear regression                               Number of obs     =     73,032
                                                F(1, 5048)        =      18.86
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0035
                                                Root MSE          =     .49255

                        (Std. err. adjusted for 5,049 clusters in household_id)
------------------------------------------------------------------------------
             |               Robust
needmet_m~6m | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   treatment |    .059786   .0137661     4.34   0.000     .0327984    .0867736
       _cons |   .5582936   .0085514    65.29   0.000      .541529    .5750581
------------------------------------------------------------------------------

. 
. ** 2 c) **
. 
. reg ohp_all_at_12m treatment, vce(cluster household_id)

Linear regression                               Number of obs     =     73,032
                                                F(1, 5048)        =     230.28
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0532
                                                Root MSE          =     .33564

                        (Std. err. adjusted for 5,049 clusters in household_id)
------------------------------------------------------------------------------
             |               Robust
ohp_all_~12m | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   treatment |   .1638377   .0107965    15.18   0.000     .1426719    .1850035
       _cons |   .0757817   .0045813    16.54   0.000     .0668003     .084763
------------------------------------------------------------------------------

. 
. ** 2 d) **
. 
. ivregress 2sls needmet_med_6m (ohp_all_at_12m=treatment), ///
>         vce(cluster household_id)

Instrumental variables 2SLS regression          Number of obs   =     73,032
                                                Wald chi2(1)    =      18.06
                                                Prob > chi2     =     0.0000
                                                R-squared       =          .
                                                Root MSE        =     .50029
```

```
                        (Std. err. adjusted for 5,049 clusters in household_id)
------------------------------------------------------------------------------
             |               Robust
needmet_med_6m | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
---------------+--------------------------------------------------------------
ohp_all_at_12m |   .3649099   .0858694     4.25   0.000     .1966089    .5332108
         _cons |   .5306401   .0136453    38.89   0.000     .5038959    .5573843
------------------------------------------------------------------------------
Endogenous: ohp_all_at_12m
Exogenous:  treatment

.
. * Note that you should NOT estimate the 2SLS regression above manually in two
. * steps by first regressing needmet_med_6m on treatment and then regressing
. * ohp_all_at_12m on the predicted values from the first regression. This is
. * because the resulting standard errors will be incorrect.
.
. ** 2 e) **
.
. display .059786 / .1638377
.36490991

.
. ******************************************************************************
. ** QUESTION 3 **
.
. use "strikedata.dta", clear

.
. ** 3 d) **
.
. * Create a date variable that is centered at the start of the strike
. gen date_centered = date - 15992

.
. * Method 1: create the date-strike interaction and day-of-week dummies yourself
. gen date_inter = date_centered * strike

. tab dayofwk, gen(day)

    dayofwk |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |     32,082       17.86       17.86
          2 |     34,383       19.14       36.99
          3 |     35,596       19.81       56.80
          4 |     37,630       20.94       77.74
          5 |     39,989       22.26      100.00
------------+-----------------------------------
      Total |    179,680      100.00

.
. areg deficit_60 strike date_centered date_inter day1-day5 [aw=weight], ///
>         vce(cluster vds) absorb(vds)
(sum of wgt is 493,344,753.30157)
note: day5 omitted because of collinearity.

Linear regression, absorbing indicators           Number of obs     = 178,549
Absorbed variable: vds                            No. of categories =     644
                                                  F(7, 643)         =   60.86
                                                  Prob > F          =  0.0000
                                                  R-squared         =  0.2198
                                                  Adj R-squared     =  0.2170
                                                  Root MSE          =  0.8144
```

```
                                    (Std. err. adjusted for 644 clusters in vds)
------------------------------------------------------------------------------
             |               Robust
  deficit_60 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      strike |   .1942413   .0177346    10.95   0.000     .1594166     .229066
date_centered |  -.0038735    .000665    -5.82   0.000    -.0051794   -.0025676
  date_inter |   .0065977   .0007967     8.28   0.000     .0050333    .0081621
        day1 |  -.2207996   .0209262   -10.55   0.000    -.2618915   -.1797076
        day2 |  -.0561812   .0199399    -2.82   0.005    -.0953365    -.017026
        day3 |  -.1313252   .0167051    -7.86   0.000    -.1641282   -.0985221
        day4 |  -.0836942   .0146311    -5.72   0.000    -.1124246   -.0549637
        day5 |           0  (omitted)
       _cons |     .44181   .0164223    26.90   0.000     .4095622    .4740577
------------------------------------------------------------------------------

.
. * Method 2: use the factor variable syntax
. areg deficit_60 strike##c.date_centered i.dayofwk [aw=weight], ///
>         vce(cluster vds) absorb(vds)
(sum of wgt is 493,344,753.30157)

Linear regression, absorbing indicators           Number of obs     = 178,549
Absorbed variable: vds                            No. of categories =     644
                                                  F(7, 643)         =   60.86
                                                  Prob > F          =  0.0000
                                                  R-squared         =  0.2198
                                                  Adj R-squared     =  0.2170
                                                  Root MSE          =  0.8144

                                       (Std. err. adjusted for 644 clusters in vd
> s)
------------------------------------------------------------------------------------------
> --
                       |               Robust
           deficit_60 | Coefficient  std. err.      t    P>|t|     [95% conf. interva
> l]
-----------------------+------------------------------------------------------------------
> --
             1.strike |   .1942413   .0177346    10.95   0.000     .1594166     .2290
> 66
        date_centered |  -.0038735    .000665    -5.82   0.000    -.0051794   -.00256
> 76
                      |
strike#c.date_centered |
                    1 |   .0065977   .0007967     8.28   0.000     .0050333    .00816
> 21
                      |
               dayofwk |
                    2 |   .1646183    .013882    11.86   0.000     .1373587    .19187
> 79
                    3 |   .0894744   .0114578     7.81   0.000     .0669753    .11197
> 36
                    4 |   .1371054   .0144675     9.48   0.000     .1086961    .16551
> 46
                    5 |   .2207996   .0209262    10.55   0.000     .1797076    .26189
> 15
                      |
                _cons |   .2210104   .0188926    11.70   0.000     .1839117    .25810
> 91
------------------------------------------------------------------------------------------
> --
```

```
.
. * Note: In the paper, the author clusters the standard errors along two
. * dimensions, namely "vds" and "date_centered". The original standard error
. * estimates can be replicated by using the "reghdfe" command, which allows for
. * multi-way clustering.
.
. reghdfe deficit_60 strike##c.date_centered i.dayofwk [aw=weight], ///
>         vce(cluster vds date_centered) absorb(vds)
(MWFE estimator converged in 1 iterations)

HDFE Linear regression                            Number of obs   =    178,549
Absorbing 1 HDFE group                            F(   7,     38) =      11.96
Statistics robust to heteroskedasticity           Prob > F        =     0.0000
                                                  R-squared       =     0.2198
                                                  Adj R-squared   =     0.2170
Number of clusters (vds)     =        644         Within R-sq.    =     0.0181
Number of clusters (date_centered) =         39   Root MSE        =     0.8144

                              (Std. err. adjusted for 39 clusters in vds date_centere
> d)
------------------------------------------------------------------------------------
> --
                        |               Robust
            deficit_60 | Coefficient  std. err.      t    P>|t|     [95% conf. interva
> l]
------------------------+-----------------------------------------------------------
> --
              1.strike |   .1942413   .0405455     4.79   0.000     .1121612    .27632
> 14
         date_centered |  -.0038735   .0016755    -2.31   0.026    -.0072653   -.00048
> 17
                       |
strike#c.date_centered |
                     1 |   .0065977   .0020662     3.19   0.003     .0024149    .01078
> 06
                       |
               dayofwk |
                     2 |   .1646183   .0302031     5.45   0.000     .1034753    .22576
> 14
                     3 |   .0894744   .0305268     2.93   0.006     .0276761    .15127
> 27
                     4 |   .1371054   .0321597     4.26   0.000     .0720014    .20220
> 94
                     5 |   .2207996   .0416246     5.30   0.000     .1365349    .30506
> 42
                       |
                 _cons |   .2210104   .0423838     5.21   0.000     .1352089     .3068
> 12
------------------------------------------------------------------------------------
> --

Absorbed degrees of freedom:
-----------------------------------------------------+
 Absorbed FE | Categories  - Redundant  = Num. Coefs |
-------------+---------------------------------------|
         vds |       644         644           0    *|
-----------------------------------------------------+
* = FE nested within cluster; treated as redundant for DoF computation

.
```

```
. ** 3 e) **
.
. areg deficit_60 strike##c.date_centered i.dayofwk [aw=weight] ///
>         if date >= 15977 & date <= 16006, vce(cluster vds) absorb(vds)
(sum of wgt is 256,182,067.46549)

Linear regression, absorbing indicators             Number of obs     = 91,752
Absorbed variable: vds                              No. of categories =    631
                                                    F(7, 630)         =  38.68
                                                    Prob > F          = 0.0000
                                                    R-squared         = 0.2392
                                                    Adj R-squared     = 0.2338
                                                    Root MSE          = 0.7547

                                          (Std. err. adjusted for 631 clusters in vd
> s)
------------------------------------------------------------------------------
> --
                     |               Robust
          deficit_60 | Coefficient  std. err.      t    P>|t|     [95% conf. interva
> l]
---------------------+--------------------------------------------------------------
> --
            1.strike |   .1315012   .0194233     6.77   0.000      .093359    .16964
> 34
       date_centered |   .0035745   .0012112     2.95   0.003     .0011961     .0059
> 53
                     |
strike#c.date_centered |
                   1 |  -.0032792   .0017187    -1.91   0.057    -.0066543    .00009
> 59
                     |
              dayofwk |
                   2 |   .1881145   .0152127    12.37   0.000     .1582408    .21798
> 82
                   3 |   .1273521   .0124958    10.19   0.000     .1028137    .15189
> 06
                   4 |   .1528586   .0155615     9.82   0.000     .1222999    .18341
> 73
                   5 |   .1964952   .0207615     9.46   0.000     .1557252    .23726
> 53
                     |
                _cons |   .2795772   .0176673    15.82   0.000     .2448832    .31427
> 12
------------------------------------------------------------------------------
> --

.
. areg deficit_60 strike##c.date_centered i.dayofwk [aw=weight] ///
>         if date >= 15984 & date <= 15999, vce(cluster vds) absorb(vds)
(sum of wgt is 138,990,414.81232)

Linear regression, absorbing indicators             Number of obs     = 49,332
Absorbed variable: vds                              No. of categories =    618
                                                    F(7, 617)         =  39.36
                                                    Prob > F          = 0.0000
                                                    R-squared         = 0.2623
                                                    Adj R-squared     = 0.2529
                                                    Root MSE          = 0.7511

                                          (Std. err. adjusted for 618 clusters in vd
> s)
------------------------------------------------------------------------------
> --
                     |               Robust
          deficit_60 | Coefficient  std. err.      t    P>|t|     [95% conf. interva
> l]
---------------------+--------------------------------------------------------------
> --
            1.strike |  -.1419216   .0380162    -3.73   0.000    -.2165785   -.06726
> 47
       date_centered |   .0412362   .0058645     7.03   0.000     .0297195    .05275
```

```
> 29
                        |
strike#c.date_centered |
                   1  |  -.0214901   .0069509    -3.09   0.002    -.0351403   -.00783
> 98
                        |
               dayofwk |
                   2  |   .2123428   .0195402    10.87   0.000     .1739695    .25071
> 62
                   3  |   .1440508   .0182285     7.90   0.000     .1082535    .17984
> 82
                   4  |   .1078216   .0212306     5.08   0.000     .0661285    .14951
> 46
                   5  |   .1861298   .0247714     7.51   0.000     .1374832    .23477
> 63
                        |
                 _cons |    .503851   .0429335    11.74   0.000     .4195375    .58816
> 44
------------------------------------------------------------------------------------------
> --

.
. ********************************************************************************
. ** QUESTION 4 **
.
. use "AEJfigs.dta", clear

.
. * Create an age variable centered around the treatment threshold
. gen age = agecell-21

.
. * Create the treatment dummy
. gen over = agecell>=21

.
. * Create interactions
. gen age2 = age^2

. gen age3 = age^3

. gen overXage = over*age

. gen overXage2 = over*age2

. gen overXage3 = over*age3

.
.
. * RDD regressions: estimate the RDD models and create new variables that contain
. * the predicted values
.
. * Linear fit with the same slope on both sides of the threshold
. reg all over age

      Source |       SS           df       MS      Number of obs   =        48
-------------+----------------------------------   F(2, 45)        =     32.99
       Model |  410.138151         2  205.069075   Prob > F        =    0.0000
    Residual |  279.682408        45  6.21516463   R-squared       =    0.5946
-------------+----------------------------------   Adj R-squared   =    0.5765
       Total |  689.820559        47  14.6770332   Root MSE        =     2.493

------------------------------------------------------------------------------
         all | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        over |   7.662709   1.440286     5.32   0.000     4.761824    10.56359
         age |  -.9746843   .6324613    -1.54   0.130    -2.248527    .2991581
       _cons |   91.84137   .8050394   114.08   0.000     90.21994     93.4628
------------------------------------------------------------------------------
```

```
. predict allfitlin
(option xb assumed; fitted values)

.
. * Quadratic fit with different slopes on each side of the threshold
. * Method 1:
. reg all over age age2 overXage overXage2

      Source |       SS           df       MS      Number of obs   =        48
-------------+----------------------------------   F(5, 42)        =     18.02
       Model |  470.512104         5  94.1024207   Prob > F        =    0.0000
    Residual |  219.308455        42  5.22162989   R-squared       =    0.6821
-------------+----------------------------------   Adj R-squared   =    0.6442
       Total |  689.820559        47  14.6770332   Root MSE        =    2.2851

------------------------------------------------------------------------------
         all | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        over |   9.547789   1.985277     4.81   0.000     5.541337    13.55424
         age |  -.8305828   3.290064    -0.25   0.802    -7.470202    5.809036
        age2 |  -.8402999   1.615268    -0.52   0.606    -4.100043    2.419443
    overXage |  -6.017014   4.652854    -1.29   0.203    -15.40685    3.372824
   overXage2 |   2.904189   2.284334     1.27   0.211    -1.705784    7.514162
       _cons |   93.07294   1.403803    66.30   0.000     90.23995    95.90593
------------------------------------------------------------------------------

. * Method 2:
. reg all c.over##c.age##c.age

      Source |       SS           df       MS      Number of obs   =        48
-------------+----------------------------------   F(5, 42)        =     18.02
       Model |  470.512103         5  94.1024205   Prob > F        =    0.0000
    Residual |  219.308457        42  5.22162992   R-squared       =    0.6821
-------------+----------------------------------   Adj R-squared   =    0.6442
       Total |  689.820559        47  14.6770332   Root MSE        =    2.2851

------------------------------------------------------------------------------
             all | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-----------------+----------------------------------------------------------------
            over |   9.547789   1.985277     4.81   0.000     5.541337    13.55424
             age |  -.8305827   3.290064    -0.25   0.802    -7.470202    5.809036
                 |
      c.over#c.age |  -6.017014   4.652854    -1.29   0.203    -15.40685    3.372825
                 |
       c.age#c.age |  -.8402999   1.615268    -0.52   0.606    -4.100043    2.419443
                 |
c.over#c.age#c.age |   2.904189   2.284334     1.27   0.211    -1.705784    7.514162
                 |
             _cons |   93.07294   1.403803    66.30   0.000     90.23995    95.90593
------------------------------------------------------------------------------

. predict allfitqi
(option xb assumed; fitted values)

.
. * Cubic fit with different slopes on each side of the threshold
. * Method 1:
. reg all over age age2 age3 overXage overXage2 overXage3

      Source |       SS           df       MS      Number of obs   =        48
-------------+----------------------------------   F(7, 40)        =     12.67
       Model |  475.400414         7  67.9143449   Prob > F        =    0.0000
    Residual |  214.420145        40  5.36050362   R-squared       =    0.6892
-------------+----------------------------------   Adj R-squared   =    0.6348
       Total |  689.820559        47  14.6770332   Root MSE        =    2.3153
```

```
------------------------------------------------------------------------------
         all | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        over |   11.26284    2.69992     4.17   0.000     5.806094    16.71958
         age |  -5.634237   8.397882    -0.67   0.506    -22.60699    11.33852
        age2 |  -6.922106   9.895116    -0.70   0.488    -26.92088    13.07667
        age3 |  -2.055425   3.298124    -0.62   0.537    -8.721183    4.610333
     overXage |  -6.790408    11.8764    -0.57   0.571     -30.7935    17.21269
    overXage2 |   16.04698   13.99381     1.15   0.258    -12.23556    44.32952
    overXage3 |  -.3309257   4.664252    -0.07   0.944    -9.757731    9.095879
        _cons |    92.2793   1.909132    48.34   0.000      88.4208     96.1378
------------------------------------------------------------------------------

. * Method 2:
. reg all c.over##c.age##c.age##c.age

      Source |       SS           df       MS      Number of obs   =        48
-------------+----------------------------------   F(7, 40)        =     12.67
       Model |  475.400405         7  67.9143435   Prob > F        =    0.0000
    Residual |  214.420154        40  5.36050386   R-squared       =    0.6892
-------------+----------------------------------   Adj R-squared   =    0.6348
       Total |  689.820559        47  14.6770332   Root MSE        =    2.3153

------------------------------------------------------------------------------
> ----
                     all | Coefficient  Std. err.      t    P>|t|     [95% conf. inter
> val]
-------------------------+----------------------------------------------------------
> ----
                    over |   11.26283    2.69992     4.17   0.000     5.806093    16.7
> 1958
                     age |  -5.634231   8.397881    -0.67   0.506    -22.60698    11.3
> 3852
                         |
             c.over#c.age |  -6.790409    11.8764    -0.57   0.571     -30.7935    17.2
> 1268
                         |
              c.age#c.age |  -6.922098   9.895115    -0.70   0.488    -26.92087    13.0
> 7667
                         |
      c.over#c.age#c.age |   16.04696   13.99381     1.15   0.258    -12.23557     44.
> 3295
                         |
       c.age#c.age#c.age |  -2.055422   3.298124    -0.62   0.537    -8.721179    4.61
> 0334
                         |
c.over#c.age#c.age#c.age |  -.3309261   4.664251    -0.07   0.944     -9.75773    9.09
> 5878
                         |
                   _cons |    92.2793   1.909132    48.34   0.000      88.4208     96.
> 1378
------------------------------------------------------------------------------
> ----

. predict allfitci
(option xb assumed; fitted values)

.
. label variable all       "Mortality rate from all causes (per 100,000)"
```

```
. label variable allfitlin "Mortality rate from all causes (per 100,000)"

. label variable allfitqi  "Mortality rate from all causes (per 100,000)"

. label variable allfitci  "Mortality rate from all causes (per 100,000)"

.
. * Figure 4.4.
. twoway scatter all agecell || ///
>         line allfitlin allfitqi allfitci agecell if age < 0, ///
>                 lcolor(red gs12 black) lwidth(medthick medthick medthick) ///
>                 lpattern(dash) || ///
>         line allfitlin allfitqi allfitci agecell if age >= 0, ///
>                 lcolor(red gs12 black) lwidth(medthick medthick medthick) ///
>                 lpattern(dash) ///
>         xline(21, lcolor(cranberry)) graphr(c(white)) ///
>         legend(region(c(white)) cols(3) order(2 3 4) lab(2 "Linear") ///
>                 lab(3 "Quadratic") lab(4 "Cubic") position(6)) ///
>         xtitle("Age")

. graph export "fig44cubic.pdf", replace
file fig44cubic.pdf saved as PDF format

.
. ********************************************************************************
. log close
      name:  <unnamed>
       log:  C:\Users\sahlste1\OneDrive - Aalto University\jatko-opinnot\opetus\Applie
> d Microeconometrics 2\Assignment 3\assignment_3_lo
> g.log
  log type:  text
 closed on:  26 Nov 2023, 15:41:14
-------------------------------------------------------------------------------------
> --------------------------------------------------
```