Empirical Practices

1. Estimating the intensive margin labor supply elasticity (see ps2q1.do for the codes)

(a) Summarize all the variables in the data (no need to report the results). What is the average number of children in this data? What is the share of hispanic women in this data?

The average number of children is 0.65 and the share of Hispanic women is 15.5%.

```
. sum

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        year |     304382    1997.843     7.95684       1984       2010
    statefip |     304382    27.48188    15.49014          1         56
       metro |     304382    2.408306    .9994369          0          4
         age |     304382    29.91811    7.490796         16         44
     hrswork |     304382    27.50024    19.59052          0         99
-------------+---------------------------------------------------------
     uhrswork |    304382    30.97973    17.16204          0         99
      incwage |    304382    16714.82    21136.69          0     713263
        nKids |    304382    .6536786    1.046529          0         12
        mstat |    304382           1           0          1          1
   avgtax_15k |    302619    .1346659     .187795  -.3804371   .7400001
-------------+---------------------------------------------------------
      emp_ind |    304382    .7441964    .4363126          0          1
      edu_yrs |    304382    12.95306    2.505825          1         20
       hsDrop |    304382     .149976    .3570485          0          1
       hsGrad |    304382    .3530958    .4779329          0          1
     bachelor |    304382     .159428    .3660753          0          1
-------------+---------------------------------------------------------
     advanced |    304382    .0559166    .2297609          0          1
         hisp |    304382    .1551964    .3620924          0          1
     nonwhite |    304382    .2557477    .4362814          0          1
```

(b) In order to estimate the effect of the wage on labor supply we need a measure of the hourly wage. We will use the two variables incwage (which is yearly income) and uhrswork (hours worked per week) and define the hourly wage to be the ratio: wage = incwage/ (52* uhrswork) . Generate this variable and label it "Hourly wage". What is the mean and standard deviation of this variable? How many missing values does the variable wage have? Why does it have missing values?

The mean is 10.1 and the standard deviation is 13.3.
There are 55727 missing values (the difference in observations between a) and b)).

Since division by zero is not allowed, for all those cases where uhrswork==0, the wage variable is missing.

```
.
. gen wage = incwage / (52 * uhrswork)
(55727 missing values generated)

. label var wage "Hourly wage"

.
. sum wage

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        wage |    248655    10.09143    13.27755   .0003205   2292.289

.
. sum wage if uhrswork==0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        wage |          0

. sum wage if uhrswork>0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        wage |    248655    10.09143    13.27755   .0003205   2292.289
```

(c) There is a variable `nKids` which takes the number of children younger than age 18 and an indicator variable `emp_ind` which takes value 1 if women are in the labor force. Let's start by examining how labor force participation of these women depends on the number of their children. Use the `tab` and `sum` commands and answer the following questions:

- What is the overall rate of labor force participation in the sample? **74.4 percent**
- What is the rate of labor force participation among women who have no children? What is the rate for 1 or 2 children?

No children: 78.6 percent
1 child: 71.8 percent
2 children: 67.7 percent
The rate of labor force participation among women with 10 children seems very high. Can you think why that may be?
The sample size for this group is very small (only 10 individuals) so this could just be by chance.

(d) Now we will estimate the regression model:

- What is your estimate $\beta_1$hat? **0.061**
- What is the standard error and the t-statistic for $\beta_1$hat? **0.0015**
- Is the coefficient $\beta_1$hat statistically significant? **Yes, the t-statistic is 41**
- Based on your estimate of β_1 how much do the number of hours worked increase if the wage increases by $1?

If the wage increases by $1, the number of hours worked per week increase by around 0.06.

- **The elasticity is around 0.02.**

```
.
. reg uhrswork wage

      Source |       SS       df       MS              Number of obs =  248655
-------------+------------------------------           F(  1,248653) = 1698.02
       Model |    164008.8     1    164008.8           Prob > F      =  0.0000
    Residual |  24016935248653  96.5881571           R-squared     =  0.0068
-------------+------------------------------           Adj R-squared =  0.0068
       Total |  24180943.8248654  97.2473551           Root MSE      =  9.8279

------------------------------------------------------------------------------
    uhrswork |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wage |   .0611671   .0014844    41.21   0.000     .0582578    .0640765
       _cons |   37.30545   .0247554  1506.96   0.000     37.25693    37.35397
------------------------------------------------------------------------------
-

. sum wage

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |    248655    10.09143    13.27755   .0003205   2292.289

. local meanwage = r(mean)

. sum uhrswork

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    uhrswork |    304382    30.97973    17.16204          0         99

. local meanhours = r(mean)

.
.
. local e = _b[wage] * `meanwage'/`meanhours'

.
. di "Estimated elasticity of labor supply: e = `e'"
Estimated elasticity of labor supply: e = .0199247484741731
-
```

(e) Usually when estimating labor supply models we use the logarithm of hours worked and

$$\log(uhrswork_i) = \beta_0 + \beta_1 \times \log(wage_i) + \varepsilon_i$$

In order to estimate this first generate two new variables: loghours and logwage using the 'gen' command, by typing `gen loghours = log(uhrswork)` and similarly for logwage. Regress loghours on logwage. What is your estimate of the female labor supply elasticity?

the logarithm of the wage as variables and estimate the model:

```
. g loghours = log(uhrswork)
(55727 missing values generated)

. g logwage = log(wage)
(55727 missing values generated)

.
. reg loghours logwage

      Source |       SS       df       MS              Number of obs =  248655
-------------+------------------------------           F(  1,248653) =14213.68
       Model |  1652.49664     1  1652.49664           Prob > F      =  0.0000
    Residual |  28908.6528248653  .116261026           R-squared     =  0.0541
-------------+------------------------------           Adj R-squared =  0.0541
       Total |  30561.1495248654  .122906326           Root MSE      =  .34097

------------------------------------------------------------------------------
    loghours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logwage |   .0824506   .0006916   119.22   0.000     .0810952    .0838061
       _cons |   3.428861   .0015026  2281.93   0.000     3.425916    3.431806
------------------------------------------------------------------------------
```

- What is your estimate $\beta_1$hat? **0.082**
- What is the standard error and the t-statistic for $\beta_1$hat? **0.00069, 119**
- Is the coefficient $\beta_1$hat statistically significant? **Yes, very highly**
- **The elasticity is now 0.08, still very small but a little bit larger than the 0.02 from above.**
- **If the wage increases by 10%, hours worked would increase by 0.8 percent.**

f) Calculate the covariance between logwage and age and the variance of logwage. You can use the command: corr logwage age, covariance

```
. corr logwage age, covariance
(obs=248655)

             |  logwage      age
-------------+------------------
     logwage |  .977591
         age |  2.21926  55.1053
```

In the table the topleft number is the variance of logwage and the bottomleft the covariance.

Suppose that the effect of age on loghours is 0.0053. Using the omitted variable bias formula, how much do you think your estimate of the labor supply elasticity will change if you control for age?

The covariance is $Cov(logwage, age) = 2.22$. **And** $Var(logwage) = 0.98$

The omitted variable formula says:

$E\hat{\beta} = \beta + \delta\frac{Cov(Age,logwage)}{Var(logwage)}$

The trueβ **should therefore be:** $\beta = E\hat{\beta} - \delta\frac{Cov(Age,logwage)}{Var(logwage)}$. **Above we obtained a** $\hat{\beta} = 0.082$, **which should be close to** $E\hat{\beta}$.

Therefore $\beta \approx \hat{\beta} - \delta\frac{Cov(Age,logwage)}{Var(logwage)} = 0.082 - 0.0053\frac{2.22}{0.98} = 0.070$

g) Estimate the labor supply elasticity as in e) but controlling for age. Does this line up with your calculations in f)? **Thus we get almost the same coefficient as the corrected coefficient from part g)**

```
. reg loghours logwage  age
```

Source	SS	df	MS
Model	2008.12203	2	1004.06102
Residual	28553.0274	248652	.11483128
Total	30561.1495	248654	.122906326

Number of obs =	248655
F(2,248652) =	8743.79
Prob > F =	0.0000
R-squared =	0.0657
Adj R-squared =	0.0657
Root MSE =	.33887

loghours	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logwage	.0703175	.0007211	97.52	0.000	.0689043	.0717308
age	.0053447	.000096	55.65	0.000	.0051564	.0055329
_cons	3.291758	.0028809	1142.61	0.000	3.286112	3.297405

h) Re-estimate the model in part e), but now control for years of education, age and age squared (you have to create a new variable for this), year, and the ethnicity variabile.

- Does the estimated labor supply elasticity stay the same as in part f)?
  The elasticity shrinks slightly to 0.066, but is still in a similar ballpark.
- Comment briefly on the coefficients on the education, age, year and ethnicity variables. **Individuals with more education and who are older tend to work more. The coefficient on year is negative, suggesting people work fewer hours over time.**

```
. g age2 = age^2

.
. reg  loghours logwage age age2 edu_yrs year hisp nonwhite
```

Source	SS	df	MS
Model	2409.9822	7	344.283171
Residual	28151.1673	248647	.113217402
Total	30561.1495	248654	.122906326

Number of obs =	248655
F(7,248647) =	3040.90
Prob > F =	0.0000
R-squared =	0.0789
Adj R-squared =	0.0788
Root MSE =	.33648

loghours	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logwage	.0658059	.0008012	82.13	0.000	.0642356	.0673763
age	.041429	.0008681	47.72	0.000	.0397275	.0431304
age2	-.0005776	.0000138	-41.80	0.000	-.0006047	-.0005505
edu_yrs	.0068864	.0003187	21.61	0.000	.0062618	.007511
year	-.0026475	.0000902	-29.34	0.000	-.0028244	-.0024707
hisp	.0087559	.0020383	4.30	0.000	.004761	.0127508
nonwhite	-.00431	.0016333	-2.64	0.008	-.0075113	-.0011087
_cons	7.966926	.1808884	44.04	0.000	7.612389	8.321462

i) How does the elasticity depend on whether women have children? Run the regression in h) but once restricting the sample to women with no children and once restricting the sample to women with more than one child. Why do you think the elasticity might be different for women with and without children?

```
. reg  loghours logwage age age2 edu_yrs year hisp nonwhite  if nKids==0

      Source |       SS       df       MS              Number of obs =  163045
-------------+------------------------------           F(  7,163037) = 2335.45
       Model |  1853.07037     7  264.724338           Prob > F      =  0.0000
    Residual |  18480.289163037  .113350276            R-squared     =  0.0911
-------------+------------------------------           Adj R-squared =  0.0911
       Total |  20333.3594163044  .124710872           Root MSE      =  .33668

------------------------------------------------------------------------------
    loghours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logwage |   .0607308   .0010316    58.87   0.000     .0587089    .0627528
         age |   .0522317   .0010944    47.73   0.000     .0500866    .0543768
        age2 |   -.000732   .0000176   -41.64   0.000    -.0007664   -.0006975
     edu_yrs |    .005671   .0003894    14.56   0.000     .0049077    .0064343
        year |  -.0024242   .0001111   -21.82   0.000     -.002642   -.0022064
        hisp |   .0070801   .0025673     2.76   0.006     .0020484    .0121119
    nonwhite |  -.0107246   .0021356    -5.02   0.000    -.0149103    -.006539
       _cons |   7.382588   .2228902    33.12   0.000     6.945728    7.819448
------------------------------------------------------------------------------
. reg  loghours logwage age age2 edu_yrs year hisp nonwhite  if nKids>0

      Source |       SS       df       MS              Number of obs =   85610
-------------+------------------------------           F(  7, 85602) =  798.50
       Model |  626.592069     7  89.5131527           Prob > F      =  0.0000
    Residual |  9596.09849 85602  .112101335           R-squared     =  0.0613
-------------+------------------------------           Adj R-squared =  0.0612
       Total |  10222.6906 85609    .1194114           Root MSE      =  .33482

------------------------------------------------------------------------------
    loghours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logwage |   .0703747   .0012849    54.77   0.000     .0678563    .0728932
         age |    .022317   .0016015    13.93   0.000      .019178     .025456
        age2 |  -.0002937   .0000247   -11.91   0.000     -.000342   -.0002453
     edu_yrs |   .0048833   .0006024     8.11   0.000     .0037025     .006064
        year |  -.0025308   .0001587   -15.94   0.000    -.0028419   -.0022197
        hisp |   .0108243   .0033467     3.23   0.001     .0042648    .0173839
    nonwhite |   .0092552   .0025638     3.61   0.000     .0042301    .0142802
       _cons |   8.038763   .3198253    25.13   0.000     7.411908    8.665618
------------------------------------------------------------------------------
```

The elasticity of labor supply for women with children is slightly larger, though the difference is small. For women with children, being able to pay for childcare may be an important obstacle to working and having a higher wage may be particularly important for them to determine whether it is worthwhile working more.

j) There are many possibilities here, the important thing is that you find a variable that is plausible and that you correctly figure out the direction of the bias.

One example for an omitted variable may be how monotonous a job is. Suppose some jobs are inherently more monotonous and therefore boring than others. E.g. working at an assembly line vs. working as a photographer. Suppose monotonous jobs pay less. Cov(Monotonous,Wage)<0, and people want to work less in monotonous jobs, so that the direct effect of monotonous Cov(hours, Monotonous) is negative. From the omitted variable bias formula, we can see that the bias term would therefore be positive and therefore the estimated elasticity would likely be larger than the true elasticity.

2. Minimum Wage Application (see ps2q3.do for the code)

a) What is the smallest and what is the highest federal minimum wage in this time period?

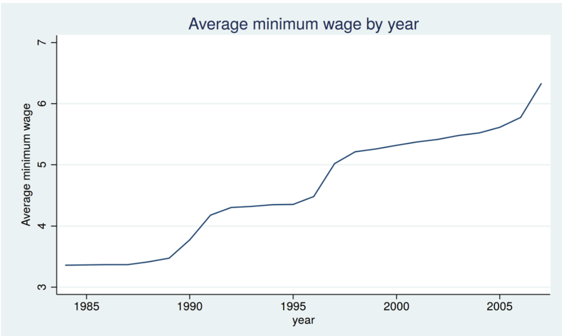
b) What is the smallest and what is the highest state minimum wage in this time period?

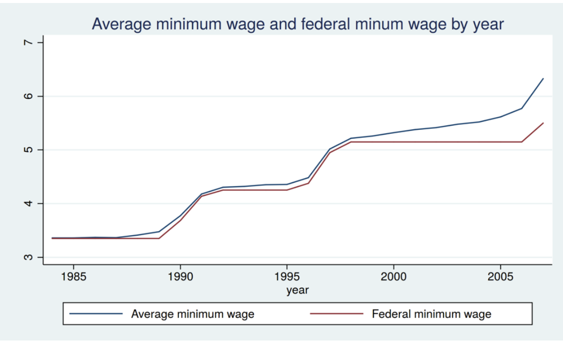
```
- sum st_mw fed_mw minwage

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       st_mw |        992    5.578992    1.199803       3.35       7.93
      fed_mw |       4896    4.420833    .7674355       3.35       5.85
     minwage |       4896    4.601632    .9969364       3.35       7.93
```

The smallest state minimum range is 3.35, the same as the federal. The highest state minimum wage was 7.93, quite a bit higher than the federal minimum wage.

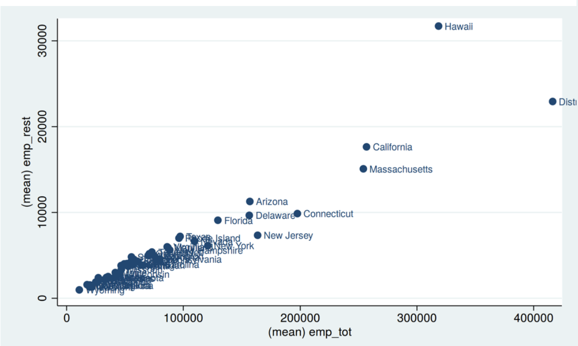
c) In which years did the average minimum wage increase the most? Why do you think it increases so dramatically in those years?

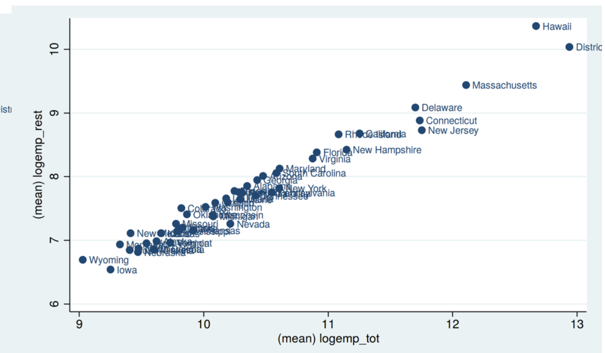
The federal minimum wage increased in 1990 and 1997. Since in every county the minimum wage is at least the federal minimum wage, these federal minimum wage increases raised the minimum wage in most places.





d) Plotting restaurant vs. total employment:





States that have more total employment are of course states with more population and thus, they also have more people working in restaurants. In both versions (log and levels) the relationship looks pretty linear and there is a high correlation.

e) Reload the employment data dubelesterreich_empdata.dta and but this time without restricting it to the year 2000. Estimate the following two regression models:

$$emp_{rest} = \alpha + \beta emp_{tot} + \epsilon$$

And

$$\log(emp_{rest}) = \gamma + \delta \log(emp_{tot}) + \varepsilon$$

```
      Source |       SS       df       MS              Number of obs =   91061
-------------+------------------------------           F(  1, 91059) =       .
       Model |  9.6427e+12     1  9.6427e+12           Prob > F      =  0.0000
    Residual |  4.3836e+11 91059  4814064.96           R-squared     =  0.9565
-------------+------------------------------           Adj R-squared =  0.9565
       Total |  1.0081e+13 91060   110708325           Root MSE      =  2194.1

------------------------------------------------------------------------------
    emp_rest |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     emp_tot |   .0597171   .0000422  1415.29   0.000     .0596344    .0597998
       _cons |   526.8174   7.796276    67.57   0.000     511.5368     542.098
------------------------------------------------------------------------------
    - reg logemp_rest logemp_tot

      Source |       SS       df       MS              Number of obs =   91061
-------------+------------------------------           F(  1, 91059) =       .
       Model |  136009.295     1  136009.295           Prob > F      =  0.0000
    Residual |  8037.82841 91059  .088270554           R-squared     =  0.9442
-------------+------------------------------           Adj R-squared =  0.9442
       Total |  144047.124 91060  1.58189242           Root MSE      =   .2971

------------------------------------------------------------------------------
 logemp_rest |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  logemp_tot |   .9515609   .0007666  1241.30   0.000     .9500584    .9530634
       _cons |  -2.110078   .0077914  -270.82   0.000    -2.125349   -2.094807
------------------------------------------------------------------------------
```

We have that βhat = 0.059, so for each additional job in a county there are about 0.059 additional restaurant jobs. The coefficient is (very) highly statistically significant, which you can see from the t-statistic of 1415 which is obviously much larger than 1.96.

When estimating the model in logs we obtain a coefficient of 0.95. Thus, suggests that if total employment goes up by one percent restaurant employment goes up by just slightly less than 1 percent. It makes sense that the elasticity is close to 1, since we would think that the number of restaurants and the number of jobs in restaurants is roughly proportional to total employment.

f) No write-up necessary. See do file for code.

2. Estimating the Effect of Minimum Wages

We will first focus on log(Earnings) as an outcome variable (the very top panel) and then look at employment later.

a) Use the dubelesterreich_empdata_minwage.dta that you created before. Look at equation (1) in the paper. Let's first ignore the ϕ_i *and* τ_t. and just estimate:

$$\log(earnings\_\_rest) = \alpha + \eta \log(minwage) + \delta \log(earnings\_tot) + \epsilon$$

<u>Estimate this regression using the regress command.</u>

<u>What is your estimate η? Is it statistically significant? Is it economically a large or a small number?</u>

```
- use ./dubelesterreich_empdata_minwage.dta, clear
- regress logearnings_rest logminwage logearnings_tot

      Source |       SS       df       MS              Number of obs =   91061
-------------+------------------------------           F(  2, 91058) =61240.45
       Model |  3151.30131     2  1575.65066           Prob > F      =  0.0000
    Residual |  2342.82398 91058  .02572892            R-squared     =  0.5736
-------------+------------------------------           Adj R-squared =  0.5736
       Total |  5494.12529 91060  .060335222           Root MSE      =   .1604

------------------------------------------------------------------------------
logearnings_r~t |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
     logminwage |   .5460275   .0047481   115.00   0.000     .5367212    .5553338
logearnings_tot |   .5181051    .002427   213.47   0.000     .5133482     .522862
          _cons |   1.072268   .0124983    85.79   0.000     1.047771    1.096765
------------------------------------------------------------------------------
```

The estimate η hat is 0.55. This would mean that if the minimum wage goes up by one percent average earnings in the restaurant sector go up by 0.55 percent. This suggests that the minimum wage has a big effect on people's earnings. Note that there are definitely people who are making more than the minimum wage who are not affected by a minimum wage increase, so it makes sense that the coefficient is less than 1. The coefficient is highly significant.

b) Estimating the model with period effects:

Note that controlling for period fixed effects does not make a big difference.

```
    - regress logearnings_rest logminwage logearnings_tot _I*
note: _Iperiod64 omitted because of collinearity

      Source |       SS       df       MS              Number of obs =   91061
-------------+------------------------------           F( 67, 90993) = 1930.66
       Model |  3225.31112    67   48.138972           Prob > F      =  0.0000
    Residual |  2268.81417 90993  .024933942           R-squared     =  0.5870
-------------+------------------------------           Adj R-squared =  0.5867
       Total |  5494.12529 91060  .060335222           Root MSE      =   .1579

------------------------------------------------------------------------------
logearnings_r~t |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
     logminwage |   .5324718   .0086223    61.76   0.000     .5155721    .5493715
logearnings_tot |   .5021017   .0026023   192.94   0.000     .4970012    .5072022
      _Iperiod1 |   .0034737   .0072945     0.48   0.634    -.0108234    .0177708
      _Iperiod2 |  -.0201377   .0068935    -2.92   0.003    -.0336488   -.0066265
      _Iperiod3 |   .0233379   .0068924     3.39   0.001     .0098289     .036847
      _Iperiod4 |  -.0183083   .0068534    -2.67   0.008     -.031741   -.0048757
      _Iperiod5 |  -.0297064   .0068481    -4.34   0.000    -.0431286   -.0162843
      _Iperiod6 |  -.0392276   .0065386    -6.00   0.000    -.0520431   -.0264121
      _Iperiod7 |   .0110044    .006535     1.68   0.092    -.0018041     .023813
      _Iperiod8 |  -.0385507   .0064962    -5.93   0.000    -.0512832   -.0258182
      _Iperiod9 |  -.0471605   .0065057    -7.25   0.000    -.0599117   -.0344093
```

Estimating the model with period and county fixed effects:

```
    - reghdfe logearnings_rest logminwage logearnings_tot , absorb(period county) vce(cluster state)
(converged in 4 iterations)

HDFE Linear regression                            Number of obs   =     91,061
Absorbing 2 HDFE groups                           F(   2,      50) =      80.68
Statistics robust to heteroskedasticity           Prob > F        =     0.0000
                                                  R-squared       =     0.9363
                                                  Adj R-squared   =     0.9352
                                                  Within R-sq.    =     0.0489
Number of clusters (state)   =         51         Root MSE        =     0.0625

                                    (Std. Err. adjusted for 51 clusters in state)
---------------------------------------------------------------------------------
                |               Robust
logearnings_r~t |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
     logminwage |   .2165586   .0278701     7.77   0.000     .1605798    .2725375
logearnings_tot |   .1940754   .0189192    10.26   0.000     .1560751    .2320756
---------------------------------------------------------------------------------

Absorbed degrees of freedom:
-----------------------------------------------------+
 Absorbed FE | Num. Coefs.  =   Categories  -  Redundant     |
----------------+------------------------------------|
         period |          66             66            0    |
         county |           0           1380         1380 *  |
-----------------------------------------------------+
```

We get exactly the coefficient and standard error as in Table 2 in the paper! Note that the coefficient on the minimum wage is quite a bit smaller now. If the minimum wage goes up by 1 percent, earnings for restaurant workers increase by around0.22 percent. It still sizable and clearly the minimum wage affects earnings but less than what we had before. Also, the coefficient is highly statistically significant (t = 7:77).

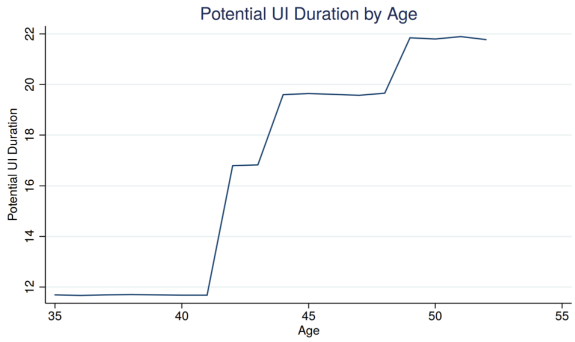
3. Unemployment Insurance Extension (See ps2q3.do)

Descriptive Analysis

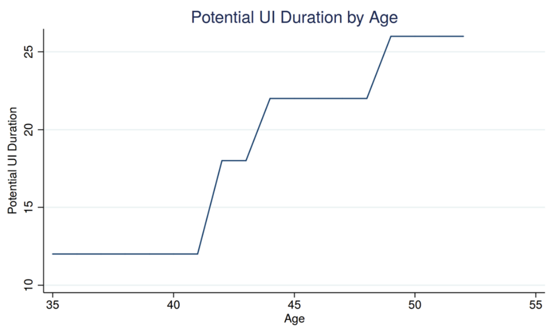
a) Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
birthdate	420032	-2008.471	3080.602	-9490	5457
begindate	420032	14059.61	2424.765	9862	18262
year	420032	1997.993	6.632183	1987	2009
unemp	420032	9.196776	1.578365	5.866667	11.725
chgunemp	420032	.0112775	.8784827	-1.75	1.875
realgdpgro~1	401899	1.745831	2.082687	-4.834638	5.121585
age	420032	43.4931	5.191032	35	52
agedays	420032	16068.08	1898.853	12783	19358
expbaseline	420032	51.20519	19.55438	12	148
P	420032	16.70799	5.633558	6	26
durnonemp	420032	8.694307	6.44661	-15.05626	164.825
duruib	420032	7.844934	4.60183	-15.05626	26
edyrs	420032	11.4976	2.291916	8	15
female	420032	.3507661	.4772104	0	1
nonger	420032	.0794201	.2703937	0	1
tenure	420032	6.002205	6.004541	2.66e-06	86.49593

b) Note that Potential UI durations are a bit lower than in the paper, that is because only people who worked for 52 months in the previous 7 years were eligible to the maximum potential benefit amounts, that is why later we will restrict the sample to individuals with at least 52 months experience in the baseperiod (past 7 years).



c) Here is the same graph with the restriction expbaseline >= 52. Now the potential UI durations line up with how we defined the policy in class: 12 if age < 42, then 18, then 22, then 26 for age >= 49.



1. **OLS Analysis**

Note that for this next part you have to reload the dataset.

a) You could have picked other control variables, but here is an example

	(1) durnonemp	(2) durnonemp	(3) durnonemp	(4) durnonemp
P	0.0306***	-0.206***	-0.206***	-0.205***
	(0.00177)	(0.00244)	(0.00244)	(0.00248)
agedays		0.000991***	0.000991***	0.000989***
		(0.00000725)	(0.00000725)	(0.00000735)
edyrs			0.00371	0.00395
			(0.00425)	(0.00431)
female			-0.00262	-0.000828
			(0.0204)	(0.0207)
nonger			-0.00411	0.00252
			(0.0360)	(0.0365)
tenure			-0.00187	-0.00179
			(0.00162)	(0.00164)
realgdpgro~1				0.0803***
				(0.00475)
unemp				0.653***
				(0.00620)
_cons	8.183***	-3.795***	-3.826***	-10.04***
	(0.0311)	(0.0928)	(0.106)	(0.122)
N	420032	420032	420032	401899
R-sq	0.001	0.043	0.043	0.069

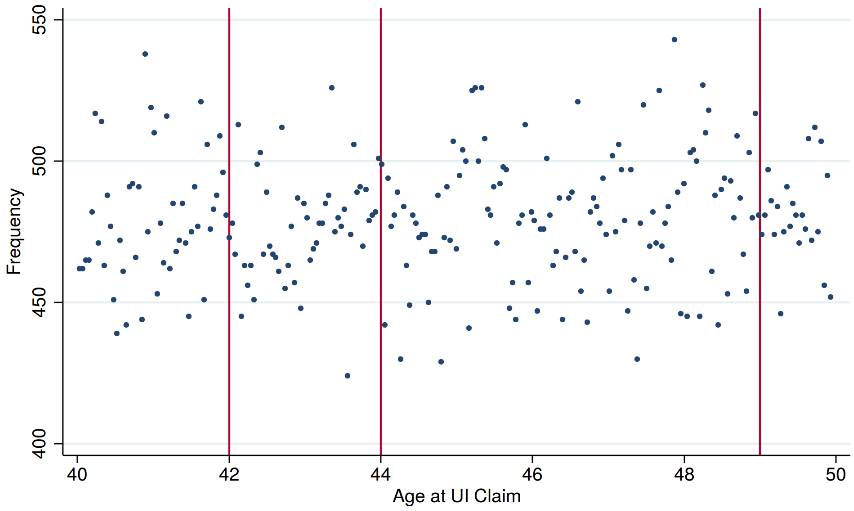
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

b) Controlling for age makes a big difference and the effect of potential UI durations goes from positive to -0.2, suggesting that an additional month of UI would reduce unemployment duration by 0.2 months, this is the opposite of what theory would predict (we think more generous UI benefits should lead people to search less hard for a job).

c) Note that potential UI duration might be related to positive worker characteristics. Worker who are older and had more employment in the past 7 years (one of the criteria to be eligible for higher P) might be more employable in general and might have an easier time finding a job.

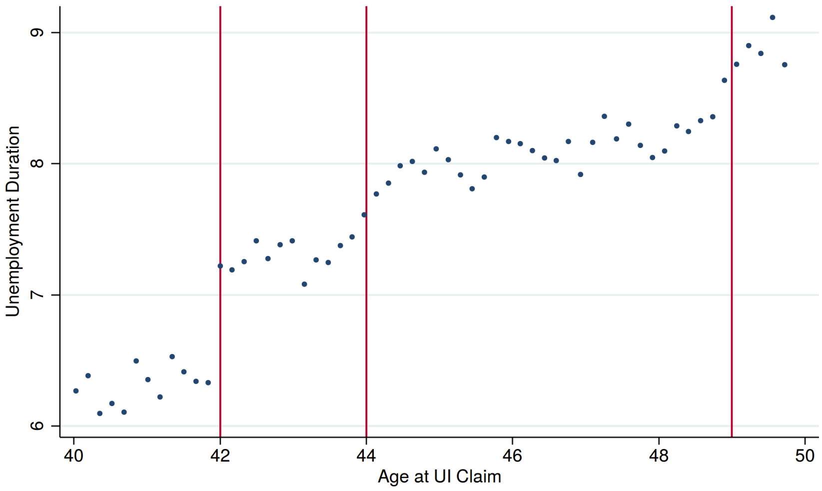
2. Regression Discontinuity Design

a) The following shows the density (number of UI claims) by 15-day age bins. The figure looks noisy because the y-axis is pretty narrow, but there are no obvious jumps in the density at the threshold. This is an important test that the RD design is likely valid since it does not seem like people are manipulating the age when they enter UI or that some people are more likely to get laid off if they are eligible to more benefits.



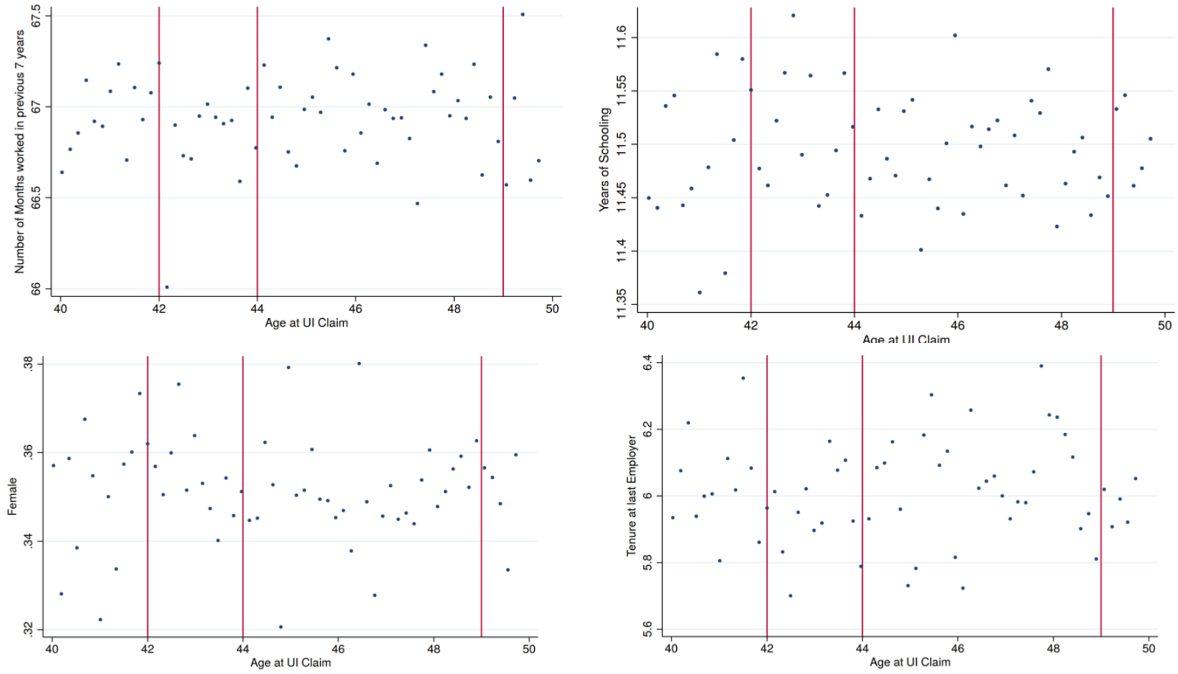
Validity Test: Smoothness of Density around Cutoffs

b) The following figure shows that there is a clear increase in unemployment durations at the thresholds. This is the main outcome of interest and suggests that potential UI durations do have a positive effect on unemployment durations. At the 42 cutoff it looks like the 6 additional months of UI increase unemployment duration by a bit less than a year.



Outcome of Interest: Unemployment Duration

The following figure shows that the predetermined variables (predetermined means that the characteristics are not affected by the potential UI durations themselves) are pretty smooth around the cutoff (that is they do not exhibit jumps at the cutoffs) and therefore this suggests that people on both sides of the cutoffs are similar to each other. This together with the test in 3a) suggests that the RD design is valid and that the estimates in figure 3b) and below are meaningful.



Validity Test: Smoothness of Predetermined Variables

From here on we use regression analysis. To make things a bit easier, focus only on the age 42 cutoff. Simply drop all observations where age<40 or age>=44.

d) Here is the simple regression (after restricting the sample to individuals age 40 to <44:

reg durnonemp P a0 a1

Note that the effect of P is now positive and quite similar to the estimates in the paper.

Source	SS	df	MS
Model	11297.0092	3	3765.66972
Residual	1460547.72	46527	31.3914011
Total	1471844.73	46530	31.632167

Number of obs =	46531
F(3, 46527) =	119.96
Prob > F =	0.0000
R-squared =	0.0077
Adj R-squared =	0.0076
Root MSE =	5.6028

durnonemp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
P	.1379493	.0173	7.97	0.000	.104041	.1718576
a0	.000282	.000174	1.62	0.105	-.0000591	.000623
a1	-.0001433	.0002462	-0.58	0.560	-.0006259	.0003392
_cons	4.756454	.2642429	18.00	0.000	4.238534	5.274374

e) Here are the regressions when restricting the sample to be within 2, 1, 0.5 and 0.2 years around the age 42 cutoffs. The standard errors get larger as the window becomes smaller.

	(1) durnonemp	(2) durnonemp	(3) durnonemp	(4) durnonemp
P	0.138***	0.125***	0.132***	0.111*
	(0.0173)	(0.0247)	(0.0343)	(0.0536)
a0	0.000282	0.000105	0.000754	0.00922
	(0.000174)	(0.000494)	(0.00137)	(0.00540)
a1	-0.000143	0.000776	-0.000947	-0.0135
	(0.000246)	(0.000703)	(0.00195)	(0.00763)
_cons	4.756***	4.888***	4.847***	5.338***
	(0.264)	(0.376)	(0.524)	(0.819)
N	46531	23223	11730	4704
R-sq	0.008	0.007	0.006	0.007

Alternatively you can show the unscaled effect, which in the paper is labeled D(Age>=42):

	(1) durnonemp	(2) durnonemp	(3) durnonemp	(4) durnonemp
RD	0.828***	0.751***	0.790***	0.669*
	(7.97)	(5.08)	(3.84)	(2.08)
a0	0.000282	0.000105	0.000754	0.00922
	(1.62)	(0.21)	(0.55)	(1.71)
a1	-0.000143	0.000776	-0.000947	-0.0135
	(-0.58)	(1.10)	(-0.49)	(-1.77)
_cons	6.412***	6.390***	6.426***	6.676***
	(87.70)	(61.46)	(44.32)	(29.37)
N	46531	23223	11730	4704

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

f) In your own words, how do the results from the OLS and RD analysis compare? What are the advantages of RD vs. OLS? Can you think of disadvantages?

The coefficients in the RD analysis have the opposite sign. The assumption for OLS regression to work is that there are no omitted variables that are correlated with the right-hand side variable and that have an effect on the outcome variable. The RD design on the other hand assumes that individuals to the left and the right of the threshold are similar and that the change in policy is the only reason for a change in the outcome variable.

Unlike the assumption for OLS, the RD assumptions can actually be tested (to some degree) by doing the validity checks: smoothness of density and smoothness of predetermined characteristics around the threshold.

One disadvantage of the RD is that we need a policy discontinuity. We also need a lot of data. Another disadvantage is that we can only find out the effect for people at the discontinuity. For example, in the case here the RD design does not tell out how people age 30 would react to a UI extension (though it may seem plausible that they react in a similar way as the 42 year olds).