Final File _V3

Kun, Giacomo, Francesca, Neeharika and Noor

21/03/2022

```
rm(list = ls())

library(ggplot2)
library(foreign)

## Warning: package 'foreign' was built under R version 4.1.2

library(stargazer)

## Warning: package 'stargazer' was built under R version 4.1.2

##
## Please cite as:

##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary
Statistics Tables.

##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

library(haven)
library(rlang)

## Warning: package 'rlang' was built under R version 4.1.2

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse
1.3.1 ──

## ✓ tibble  3.1.6     ✓ dplyr   1.0.8
## ✓ tidyr   1.2.0     ✓ stringr 1.4.0
## ✓ readr   2.1.2     ✓ forcats 0.5.1
## ✓ purrr   0.3.4

## Warning: package 'tidyr' was built under R version 4.1.2

## Warning: package 'readr' was built under R version 4.1.2

## Warning: package 'dplyr' was built under R version 4.1.2

## ── Conflicts ────────────────────────────────────────── ──
tidyverse_conflicts() ──
## x purrr::%@%()         masks rlang::%@%()
## x purrr::as_function() masks rlang::as_function()
```

```
## x dplyr::filter()      masks stats::filter()
## x purrr::flatten()     masks rlang::flatten()
## x purrr::flatten_chr() masks rlang::flatten_chr()
## x purrr::flatten_dbl() masks rlang::flatten_dbl()
## x purrr::flatten_int() masks rlang::flatten_int()
## x purrr::flatten_lgl() masks rlang::flatten_lgl()
## x purrr::flatten_raw() masks rlang::flatten_raw()
## x purrr::invoke()      masks rlang::invoke()
## x dplyr::lag()         masks stats::lag()
## x purrr::splice()      masks rlang::splice()
```

```
library(finalfit)
library(rdrobust)
library(ggplot2)
library(binsreg)
```

```
library(rddtools)
```

```
## Warning: package 'rddtools' was built under R version 4.1.2
```

```
## Loading required package: AER
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.1.2
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
## Loading required package: lmtest
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Loading required package: sandwich
```

```
## Loading required package: survival
```

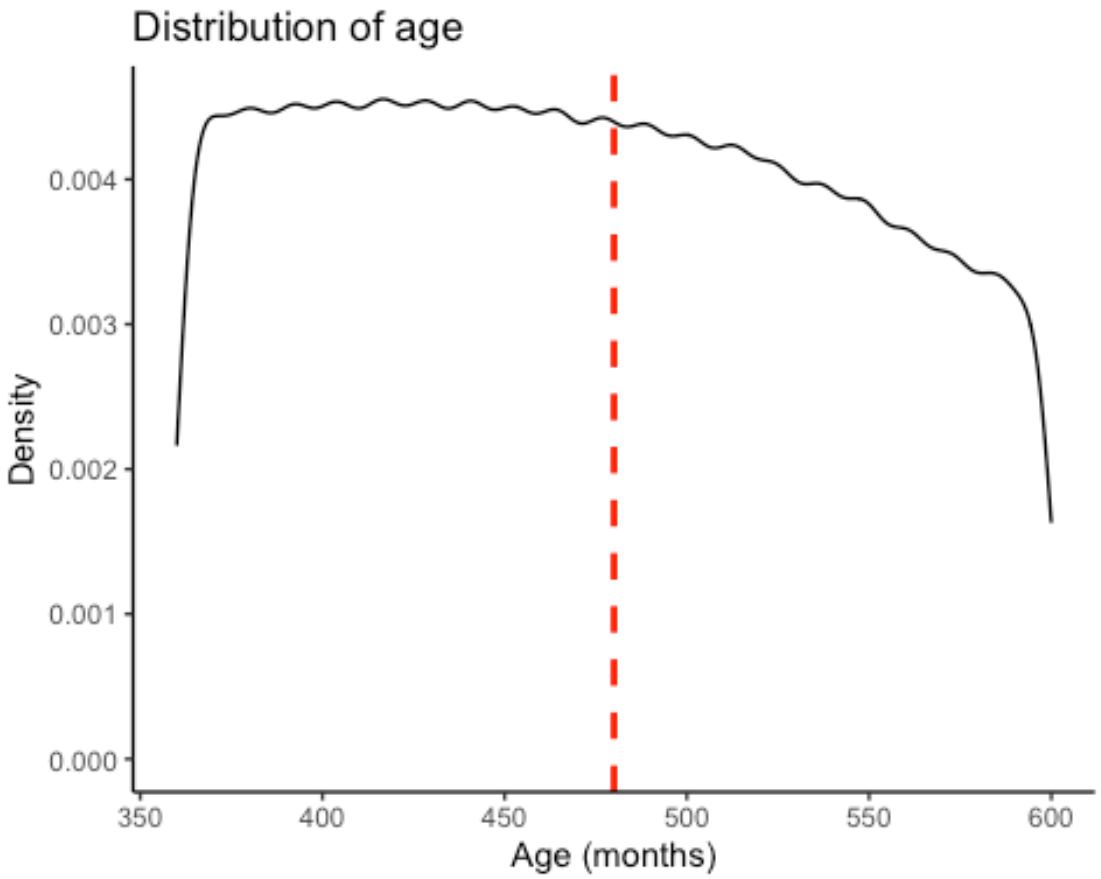
```
## Loading required package: np

## Nonparametric Kernel Methods for Mixed Datatypes (version 0.60-11)
## [vignette("np_faq",package="np") provides answers to frequently asked
questions]
## [vignette("np",package="np") an overview]
## [vignette("entropy_np",package="np") an overview of entropy-based methods]

##
## Please consider citing R and rddtools,
## citation()
## citation("rddtools")

data = read_dta("data_ps4.dta")

##Data screening

data$age_month = data$age*12
summary(data)

##       age            lwage0            nonemp          jobfind
##  Min.   :30.00   Min.   :-0.8335   Min.   :  1.0    Min.   :0.0000
##  1st Qu.:34.86   1st Qu.: 7.3066   1st Qu.: 49.0    1st Qu.:1.0000
##  Median :39.60   Median : 7.5786   Median : 82.0    Median :1.0000
##  Mean   :39.72   Mean   : 7.5169   Mean   :114.3    Mean   :0.8375
##  3rd Qu.:44.49   3rd Qu.: 7.7937   3rd Qu.:125.0    3rd Qu.:1.0000
##  Max.   :50.00   Max.   :13.8387   Max.   :729.0    Max.   :1.0000
##                  NA's   :3054      NA's   :149609
##      lwage1         age_month
##  Min.   : 1.1     Min.   :360.0
##  1st Qu.: 7.2     1st Qu.:418.4
##  Median : 7.5     Median :475.2
##  Mean   : 7.5     Mean   :476.6
##  3rd Qu.: 7.7     3rd Qu.:533.9
##  Max.   :11.7     Max.   :600.0
##  NA's   :549341

#Let's observe how NAs match inside the dataset
table(is.na(data$lwage1), is.na(data$nonemp)) #All nonemp NAs are also lwage1
NAs

##
##           FALSE    TRUE
##   FALSE 1189446       0
##   TRUE   399732  149609

table(is.na(data$lwage1), is.na(data$lwage0)) #Not all lwage0 NAs are also
lwage1 NAs

##
##           FALSE    TRUE
```

```
##   FALSE 1187476    1970
##   TRUE   548257    1084

#Observing if lwage1 NAs (the most numerous) are distributed uniformely
across the dataset

Diff_mean=mean(data[is.na(data$lwage1),]$lwage0, na.rm=T)-
mean(data[!is.na(data$lwage1),]$lwage0, na.rm=T)
Var=sqrt(var(data[is.na(data$lwage1),]$lwage0,
na.rm=T)+var(data[!is.na(data$lwage1),]$lwage0, na.rm=T))
2*(1-pnorm(abs(Diff_mean/Var)))

## [1] 0.9588284

#Mean of lwage0 in the part of population where we do not observe the lwage1
is not statistically different
#from the part of the population where we observe it

Diff_mean=mean(data[is.na(data$lwage1),]$age_month, na.rm=T)-
mean(data[!is.na(data$lwage1),]$age_month, na.rm=T)
Var=sqrt(var(data[is.na(data$lwage1),]$age_month,
na.rm=T)+var(data[!is.na(data$lwage1),]$age_month, na.rm=T))
2*(1-pnorm(abs(Diff_mean/Var)))

## [1] 0.9306214

#Distribution of age_month in the sample of population where we do not
observe the lwage1 is not statistically different
#from the sample of the population where we observe it

Diff_mean=mean(data[is.na(data$lwage1),]$nonemp, na.rm=T)-
mean(data[!is.na(data$lwage1),]$nonemp, na.rm=T)
Var=sqrt(var(data[is.na(data$lwage1),]$nonemp,
na.rm=T)+var(data[!is.na(data$lwage1),]$nonemp, na.rm=T))
2*(1-pnorm(abs(Diff_mean/Var)))

## [1] 0.5892496

#Distribution of nonemp in the part of population where we do not observe the
lwage1 is not statistically different
#from the part of the population where we observe it

Diff_mean=mean(data[is.na(data$lwage1),]$jobfind, na.rm=T)-
mean(data[!is.na(data$lwage1),]$jobfind, na.rm=T)
Var=sqrt(var(data[is.na(data$lwage1),]$jobfind,
na.rm=T)+var(data[!is.na(data$lwage1),]$jobfind, na.rm=T))
2*(1-pnorm(abs(Diff_mean/Var)))

## [1] 0.7625978
```

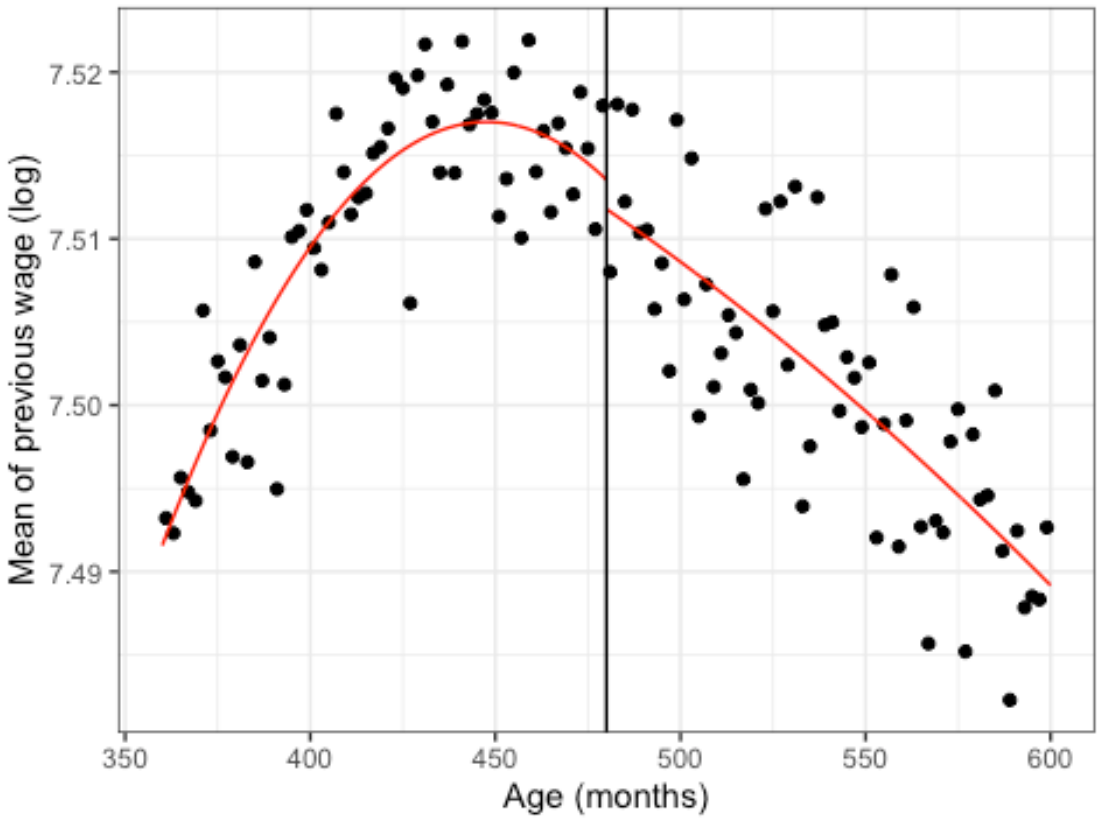
```r
#Distribution of jobfind in the part of population where we do not observe
the lwage1 is not statistically different
#from the part of the population where we observe it

#We can drop out NAs of lwage1 without excessive concerns

data = subset(data, !is.na(data$lwage1))
summary(data)
```

```
##       age            lwage0            nonemp         jobfind
##  Min.   :30.00   Min.   :-0.8335   Min.   :  1.0   Min.   :0.0000
##  1st Qu.:34.65   1st Qu.: 7.2954   1st Qu.: 60.0   1st Qu.:1.0000
##  Median :39.31   Median : 7.5595   Median : 95.0   Median :1.0000
##  Mean   :39.50   Mean   : 7.5065   Mean   :131.9   Mean   :0.8896
##  3rd Qu.:44.19   3rd Qu.: 7.7728   3rd Qu.:150.0   3rd Qu.:1.0000
##  Max.   :50.00   Max.   :13.8387   Max.   :729.0   Max.   :1.0000
##                  NA's   :1970
##      lwage1         age_month
##  Min.   : 1.094   Min.   :360.0
##  1st Qu.: 7.245   1st Qu.:415.9
##  Median : 7.521   Median :471.7
##  Mean   : 7.460   Mean   :474.0
##  3rd Qu.: 7.738   3rd Qu.:530.2
##  Max.   :11.737   Max.   :600.0
##
```

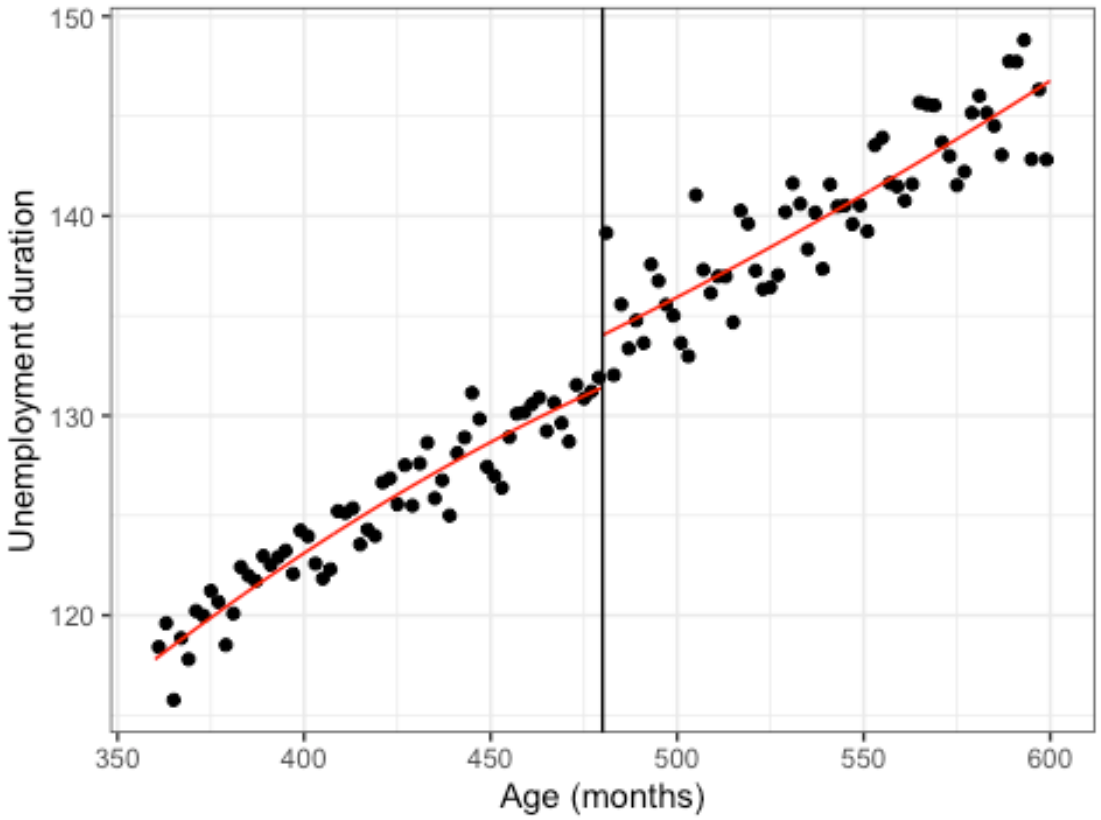
```r
##RDD diagnostics
Density <- ggplot(data, aes(x = age_month)) +
  geom_density() +
  geom_vline(aes(xintercept = 480),
             color = "red", linetype = "dashed", size = 1)+
  labs(x="Age (months)", y="Density", title="Distribution of age")+
  theme_classic()
Density
```



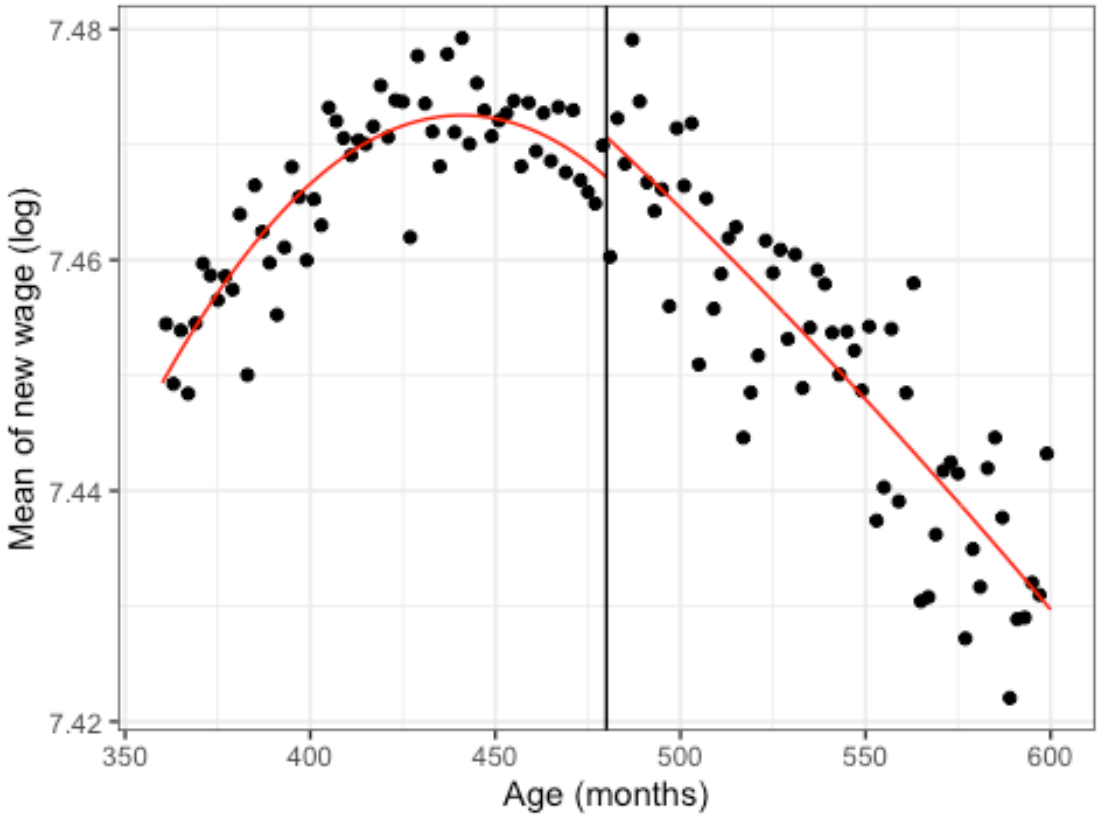

```
Prev_wage = rdplot(data$lwage0, data$age_month, c=480, p=2, nbins=60,
col.dots = "black",
                   x.label = "Age (months)", y.label="Mean of previous wage
(log)", title = "" )

## [1] "Mass points detected in the running variable."
```



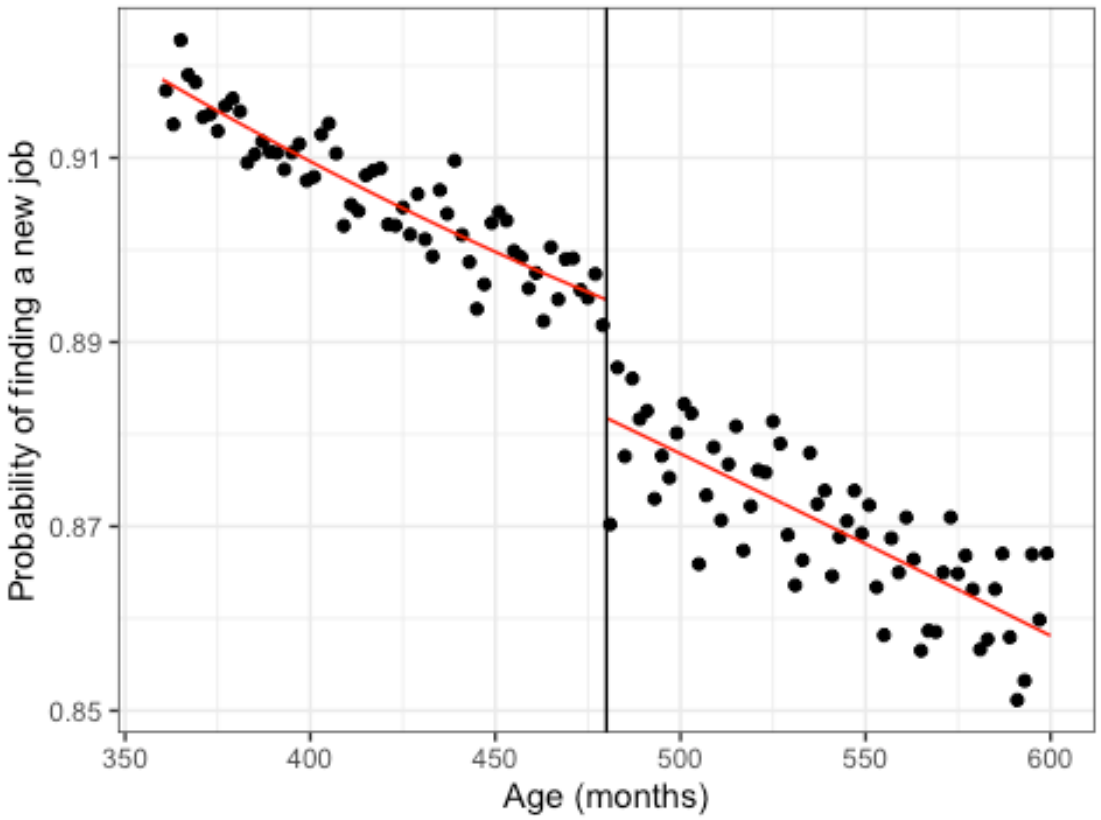

```
##RDD estimates graph

unemp_RDD <- rdplot(data$nonemp, data$age_month, c=480, p=2, nbins=60,
col.dots = "black",
                    x.label = "Age (months)", y.label="Unemployment
duration", title = "" )

## [1] "Mass points detected in the running variable."
```




```
lwage1_RDD <- rdplot(data$lwage1, data$age_month, c=480, p=2, nbins=60,
col.dots = "black",
                     x.label = "Age (months)", y.label="Mean of new wage
(log)", title = "" )

## [1] "Mass points detected in the running variable."
```




```
jobfind_RDD <- rdplot(data$jobfind, data$age_month, c=480, p=2, nbins=60,
col.dots = "black",
                      x.label = "Age (months)", y.label="Probability of
finding a new job", title = "" )

## [1] "Mass points detected in the running variable."
```




```
##Regression Table

data$cutoff = ifelse(data$age_month>=480, 1, 0)

reg1A<-lm(lwage1~cutoff, data = data)
reg2A<-lm(lwage1~cutoff, data = data, subset = data$age>=35&data$age<=45)
reg3A<-lm(lwage1~cutoff*age, data = data)
reg4A<-lm(lwage1~cutoff*age+cutoff*I(age^2)+cutoff*I(age^3)+cutoff*I(age^4),
data = data)
#reg5A<-rdrobust(data$lwage1, data$age, c=40, covs=data$lwage0)

reg1B<-lm(jobfind~cutoff, data = data)
reg2B<-lm(jobfind~cutoff, data = data, subset = data$age>=35&data$age<=45)
reg3B<-lm(jobfind~cutoff*age, data = data)
reg4B<-lm(jobfind~cutoff*age+cutoff*I(age^2)+cutoff*I(age^3)+cutoff*I(age^4),
data = data)
#reg5B<-rdrobust(data$jobfind, data$age, c=40, covs=data$lwage0)

reg1C<-lm(nonemp~cutoff, data = data)
reg2C<-lm(nonemp~cutoff, data = data, subset = data$age>=35&data$age<=45)
reg3C<-lm(nonemp~cutoff*age, data = data)
reg4C<-lm(nonemp~cutoff*age+cutoff*I(age^2)+cutoff*I(age^3)+cutoff*I(age^4),
```

```
data = data)
#reg5C<-rdrobust(data$nonemp, data$age, c=40, covs=data$lwage0)

stargazer(reg1A, reg2A, reg3A, reg4A, type = "text")

##
##
================================================================================
============================================================
##
Dependent variable:
##                     ---------------------------------------------------------
-----------------------------------------------------------
##
lwage1
##                                (1)                          (2)
(3)                          (4)
## ------------------------------------------------------------------------------
-----------------------------------------------------------
## cutoff                        -0.015***                   -0.010***
0.230***                     68.962
##                                (0.001)                     (0.001)
(0.011)                     (52.792)
##
## age
0.002***                     -1.772
##
(0.0002)                     (1.938)
##
## I(age2)
0.075
##
(0.083)
##
## I(age3)
-0.001
##
(0.002)
##
## I(age4)
0.00001
##
(0.00001)
##
## cutoff:age
-0.006***                     -5.786
##
(0.0003)                     (4.874)
##
## cutoff:I(age2)
```

```
0.178
##
(0.171)
##
## cutoff:I(age3)
-0.002
##
(0.003)
##
## cutoff:I(age4)
0.00001
##
(0.00002)
##
## Constant                       7.467***                    7.471***
7.405***                     22.969
##                                (0.001)                     (0.001)
(0.007)                     (16.837)
##
## ------------------------------------------------------------------------
-------------------------------------------------------
## Observations                  1,189,446                    619,039
1,189,446                    1,189,446
## R2                              0.0003                      0.0001
0.001                        0.001
## Adjusted R2                     0.0003                      0.0001
0.001                        0.001
## Residual Std. Error     0.433 (df = 1189444)        0.435 (df = 619037)
0.433 (df = 1189442)        0.433 (df = 1189436)
## F Statistic         338.892*** (df = 1; 1189444) 82.665*** (df = 1;
619037) 275.892*** (df = 3; 1189442) 98.132*** (df = 9; 1189436)
##
========================================================================
=======================================================
## Note:
*p<0.1; **p<0.05; ***p<0.01

stargazer(reg1B, reg2B, reg3B, reg4B, type = "text")

##
##
========================================================================
=============================================================
##
Dependent variable:
##                     ---------------------------------------------------
-------------------------------------------------------------
##
jobfind
##                                 (1)                           (2)
```

```
(3)                             (4)
## ------------------------------------------------------------------------
-------------------------------------------------------------
## cutoff                          -0.035***                       -0.024***
-0.013                          34.651
##                                  (0.001)                         (0.001)
(0.008)                         (38.165)
##
## age
-0.002***                         -0.128
##
(0.0001)                        (1.401)
##
## I(age2)
0.004
##
(0.060)
##
## I(age3)
-0.0001
##
(0.001)
##
## I(age4)
0.00000
##
(0.00001)
##
## cutoff:age
0.00003                          -3.141
##
(0.0002)                        (3.524)
##
## cutoff:I(age2)
0.107
##
(0.124)
##
## cutoff:I(age3)
-0.002
##
(0.002)
##
## cutoff:I(age4)
0.00001
##
(0.00001)
##
## Constant                        0.906***                        0.900***
0.990***                         2.348
```

```
##                                 (0.0004)                        (0.001)
(0.005)                       (12.172)
##
## ------------------------------------------------------------------------
-------------------------------------------------------------
## Observations                   1,189,446                       619,039
1,189,446                     1,189,446
## R2                                0.003                          0.001
0.004                         0.004
## Adjusted R2                       0.003                          0.001
0.004                         0.004
## Residual Std. Error      0.313 (df = 1189444)          0.315 (df = 619037)
0.313 (df = 1189442)          0.313 (df = 1189436)
## F Statistic         3,731.028*** (df = 1; 1189444) 884.104*** (df = 1;
619037) 1,432.560*** (df = 3; 1189442) 477.954*** (df = 9; 1189436)
##
========================================================================
=============================================================
## Note:
*p<0.1; **p<0.05; ***p<0.01
```

```
stargazer(reg1C, reg2C, reg3C, reg4C, type = "text")
```

```
##
##
========================================================================
=============================================================
##
Dependent variable:
##                     ---------------------------------------------------
-------------------------------------------------------------
##
nonemp
##                                 (1)                             (2)
(3)                           (4)
## ------------------------------------------------------------------------
-------------------------------------------------------------
## cutoff                         14.574***                       8.245***
5.427*                      -2,540.438
##                                (0.226)                         (0.314)
(3.206)                     (14,986.970)
##
## age
1.362***                       222.962
##
(0.053)                       (550.304)
##
## I(age2)
-8.977
##
```

```
(23.696)
##
## I(age3)
0.161
##
(0.452)
##
## I(age4)
-0.001
##
(0.003)
##
## cutoff:age
-0.091                        221.593
##
(0.079)                      (1,383.794)
##
## cutoff:I(age2)
-6.944
##
(48.653)
##
## cutoff:I(age3)
0.091
##
(0.776)
##
## cutoff:I(age4)
-0.0004
##
(0.005)
##
## Constant                       125.205***                    128.621***
77.528***                     -1,967.381
##                                (0.154)                       (0.218)
(1.875)                      (4,779.888)
##
## ------------------------------------------------------------------------
-------------------------------------------------------------
## Observations                   1,189,446                      619,039
1,189,446                     1,189,446
## R2                                0.003                         0.001
0.004                         0.004
## Adjusted R2                       0.003                         0.001
0.004                         0.004
## Residual Std. Error     122.907 (df = 1189444)        123.459 (df =
619037)        122.849 (df = 1189442)        122.848 (df = 1189436)
## F Statistic         4,157.099*** (df = 1; 1189444) 689.555*** (df = 1;
619037) 1,763.696*** (df = 3; 1189442) 588.803*** (df = 9; 1189436)
##
```

```
================================================================================
================================================================
## Note:
*p<0.1; **p<0.05; ***p<0.01

#reg5A[["coef"]]
#reg5A[["se"]]
#reg5A[["z"]]
#reg5A[["pv"]]
#reg5A[["bws"]]

#reg5B[["coef"]]
#reg5B[["se"]]
#reg5B[["z"]]
#reg5B[["pv"]]
#reg5B[["bws"]]

#reg5C[["coef"]]
#reg5C[["se"]]
#reg5C[["z"]]
#reg5C[["pv"]]
#reg5C[["bws"]]
```