Employment Effects of a Social Assistance Program: Analysis Using Canadian Census Data (1986 & 1991)

Bahareh Aghababaei

2023-04-17

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(haven)
```

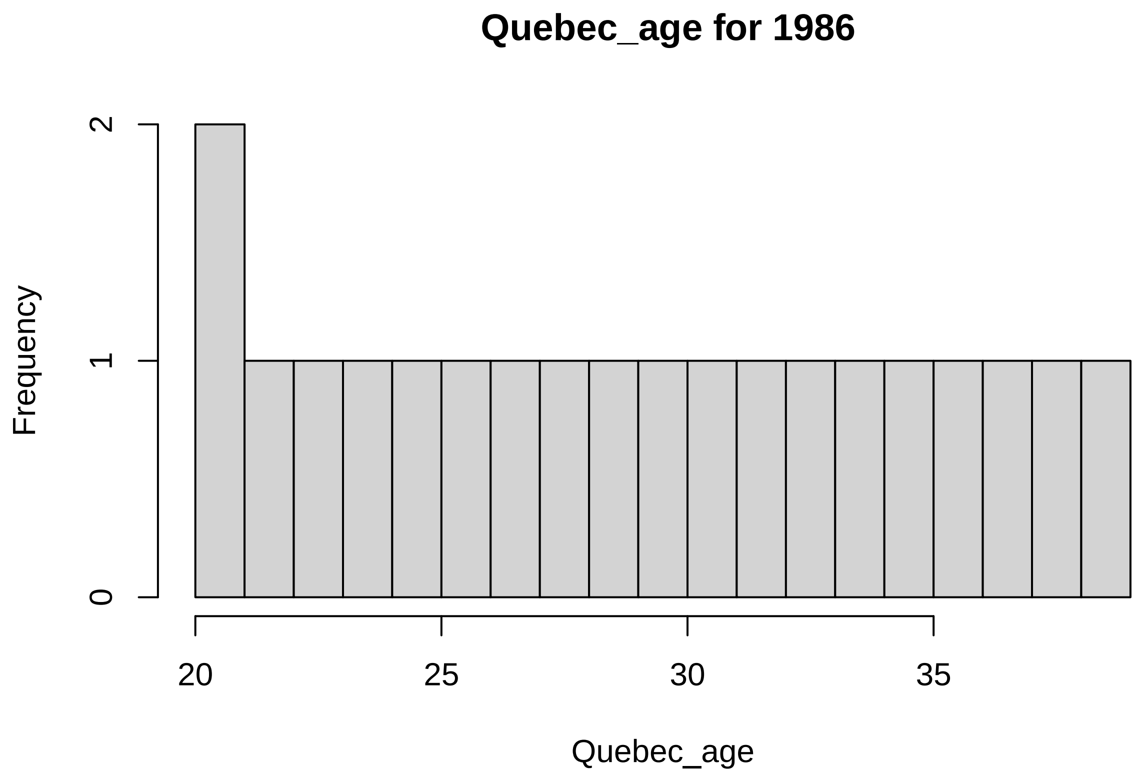
Dataset

```
dataset <- read_dta("C:/Users/bahar/Documents/MSc courses/ECON 562/Assignment 6/quebec-rd.dta")
head(dataset)
```

```
## # A tibble: 6 x 9
##   region  year   age number   emp sd_emp hours sd_hours   que
##   <chr>  <dbl> <dbl>  <dbl> <dbl>  <dbl> <dbl>    <dbl> <dbl>
## 1 Quebec  1986    20   2923 0.548  0.498  20.4     21.3     1
## 2 Quebec  1986    21   3055 0.591  0.492  22.4     21.7     1
## 3 Quebec  1986    22   3037 0.619  0.486  24.3     22.0     1
## 4 Quebec  1986    23   2909 0.657  0.475  25.9     21.9     1
## 5 Quebec  1986    24   2757 0.662  0.473  26.5     22.2     1
## 6 Quebec  1986    25   2730 0.676  0.468  26.6     21.5     1
```

Descriptive Analysis and Trends

```
Quebec_1986 <- subset(dataset, year=='1986' & region=='Quebec')
Quebec_age <- Quebec_1986$age
hist(Quebec_age,breaks = 20, main = " Quebec_age for 1986")
```



Yes, the distribution of x (age) is continuous around 30, and there is no jump in the value of x when crossing threshold.

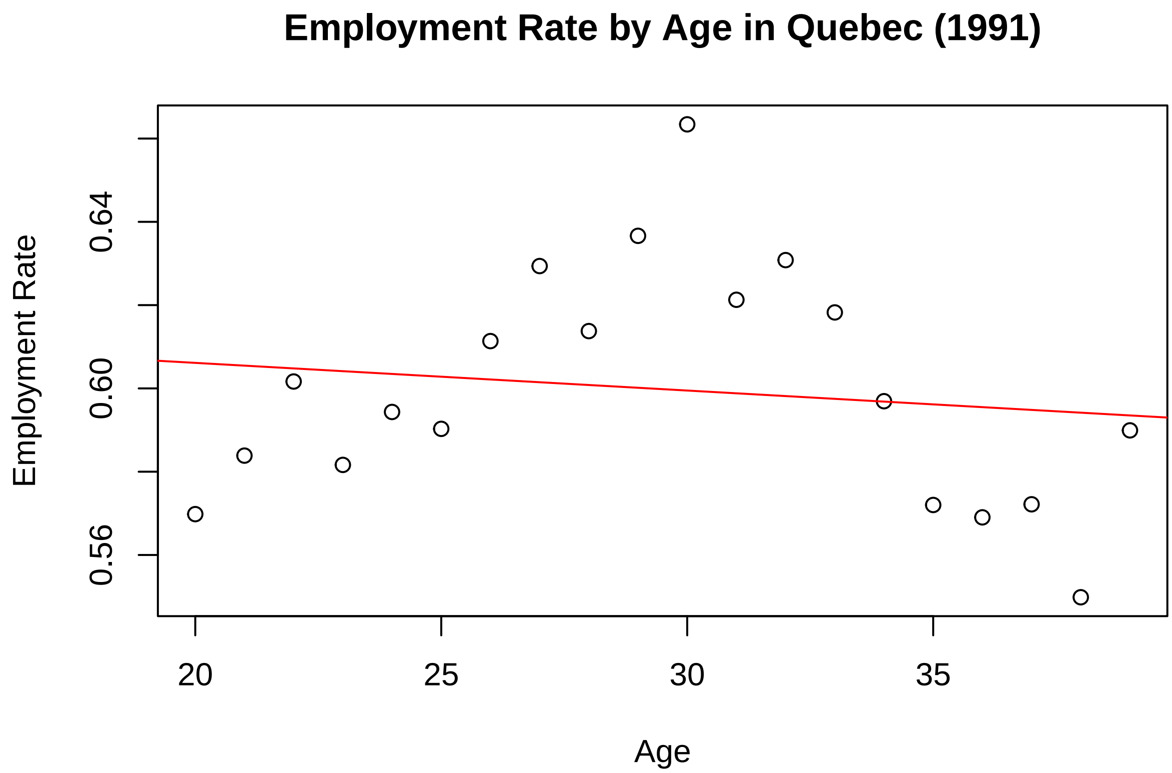
Employment rate trend with age (ROC, 1986)

```
ROC_1986 <- subset(dataset, year=='1986' & region=='RoC')
Roc_emp <- subset(ROC_1986,select=c(age, emp))
plot(Roc_emp$age, Roc_emp$emp, xlab="Age", ylab="Employment Rate",main="Employment Rate by Age in RoC (

fit <- lm(emp ~ age, data=ROC_1986)
abline(fit, col="red")
```

No, there is no jump in the value of employment rate at and after age 30.The overall trend is shown in the graph by red line, it is gradually increasing with a very low slope.

Employment rate trend with age (Quebec,1991)

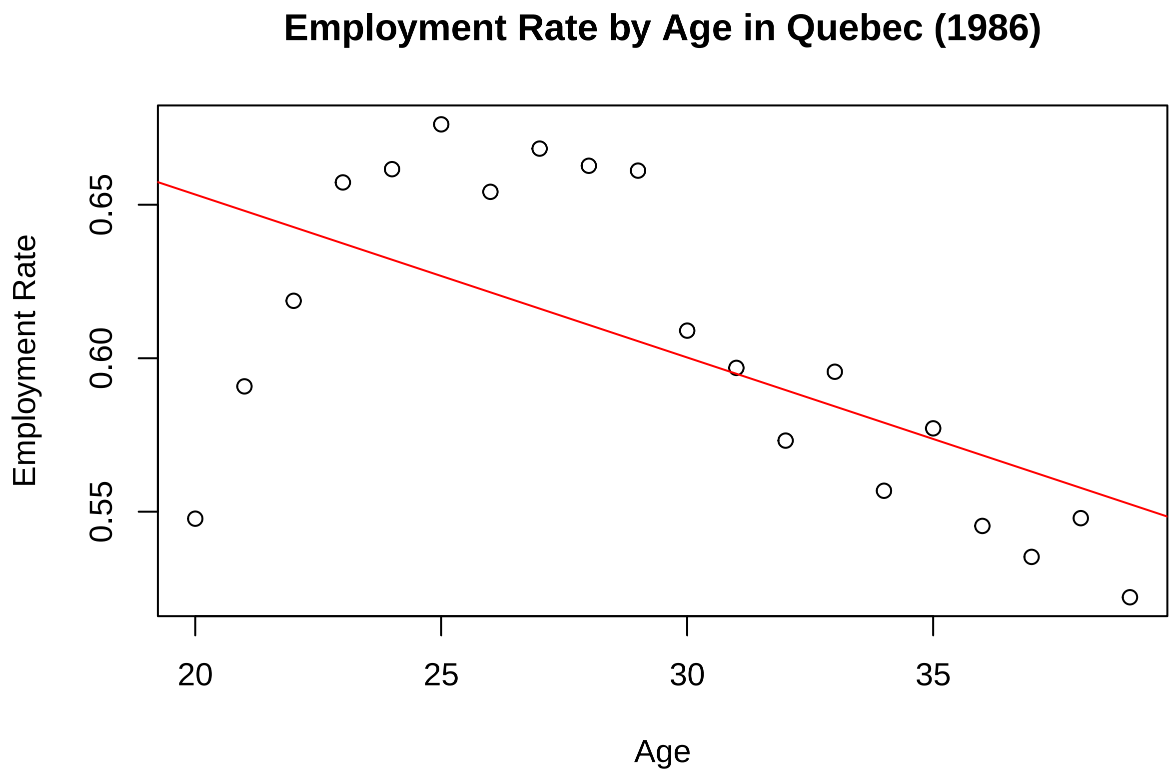
```
Quebec_1991 <- subset(dataset, year=='1991' & region=='Quebec')
Quebec_emp <- subset(Quebec_1991, select=c(age, emp))
plot(Quebec_1991$age, Quebec_1991$emp, xlab="Age", ylab="Employment Rate",main="Employment Rate by Age i

fit <- lm(emp ~ age, data=Quebec_emp)
abline(fit, col="red")
```



Yes, there is a jump at age 30, and the overall trend is downward.

Employment rate trend with age (Quebec,1986)

```
Quebec_emp <- subset(Quebec_1986, select=c(age, emp))
plot(Quebec_1986$age, Quebec_1986$emp, xlab="Age", ylab="Employment Rate",main="Employment Rate by Age i

fit <- lm(emp ~ age, data=Quebec_emp)
abline(fit, col="red")
```



Yes, there is a jump at and after age 30.

I believe we should remove the data for people under 25 because it's not really following the same pattern as the rest of the dataset. This can clearly be observed from the above graph. Removing those data points will help us get a more accurate estimate.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
```

Regression Discontinuity (RD) Analysis

Create a recentered running variable and polynomial transformations

```r
Quebec <- subset(dataset, region=='Quebec')
Quebec <- Quebec %>%
  mutate(running_var=age-30)


Quebec <- Quebec %>%
  mutate(runvar_sq = (running_var)^2)

Quebec <- Quebec %>%
  mutate(runvar_cub = (running_var)^3)

Quebec<- Quebec %>%
  mutate(disc_var=ifelse(age<30,1,0))
```

Estimate the policy effect using RD models (linear, quadratic, cubic)

```r
Quebec_RD <- subset(Quebec,age >= 25 )

lm1_RD <- lm(emp~ running_var + disc_var, data = Quebec_RD)
summary(lm1_RD)
```

```
##
## Call:
## lm(formula = emp ~ running_var + disc_var, data = Quebec_RD)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.065911 -0.017281 -0.001142  0.018006  0.045697
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.617712   0.010734  57.547  < 2e-16 ***
## running_var -0.007904   0.001976  -3.999 0.000443 ***
## disc_var    -0.001032   0.018113  -0.057 0.954998
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02688 on 27 degrees of freedom
## Multiple R-squared:  0.6366, Adjusted R-squared:  0.6097
## F-statistic: 23.65 on 2 and 27 DF,  p-value: 1.162e-06
```

```r
#Quebec_RD <- Quebec_RD %>%
#  mutate(age_sq = age^2)

lm2_RD <- lm(emp ~  running_var+runvar_sq + disc_var, data = Quebec_RD)
summary(lm2_RD)
```

```
##
## Call:
## lm(formula = emp ~ running_var + runvar_sq + disc_var, data = Quebec_RD)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.051130 -0.014349 -0.002176  0.018002  0.053262
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.6152951  0.0106968  57.521   <2e-16 ***
## running_var -0.0041115  0.0033521  -1.227    0.231
## runvar_sq   -0.0005140  0.0003702  -1.388    0.177
## disc_var     0.0184162  0.0226579   0.813    0.424
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02643 on 26 degrees of freedom
## Multiple R-squared:  0.6617, Adjusted R-squared:  0.6226
## F-statistic: 16.95 on 3 and 26 DF,  p-value: 2.636e-06
```

```r
#Quebec_RD <- Quebec_RD %>%
#  mutate(age_cub = age^3)

lm3_RD <- lm(emp~  running_var+runvar_sq+runvar_cub + disc_var, data = Quebec_RD)
summary(lm3_RD)
```

```
##
## Call:
## lm(formula = emp ~ running_var + runvar_sq + runvar_cub + disc_var,
##     data = Quebec_RD)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.038275 -0.019717  0.001374  0.017545  0.047626
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.289e-01  1.241e-02  50.674   <2e-16 ***
## running_var -9.028e-03  4.093e-03  -2.206   0.0368 *
## runvar_sq   -1.279e-03  5.322e-04  -2.403   0.0240 *
## runvar_cub   1.497e-04  7.804e-05   1.919   0.0665 .
## disc_var     5.212e-03  2.264e-02   0.230   0.8198
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02516 on 25 degrees of freedom
## Multiple R-squared:  0.7051, Adjusted R-squared:  0.6579
## F-statistic: 14.94 on 4 and 25 DF,  p-value: 2.305e-06
```

The estimated treatment effect is -0.001032 with a linear age variable, 0.0184162 with a quadratic age variable, and 5.212e-03 with cubic in age. None of them are statistically significant.

Perform robustness checks with different bandwidths (Age:20-39)

```
#windows 20-39

Quebec <- subset(dataset, region=='Quebec')
Quebec <- Quebec %>%
  mutate(running_var=age-30)


Quebec <- Quebec %>%
  mutate(runvar_sq = (running_var)^2)

Quebec <- Quebec %>%
  mutate(runvar_cub = (running_var)^3)

Quebec<- Quebec %>%
  mutate(disc_var=ifelse(age<30,1,0))


Quebec1_RD <- subset(Quebec, age >= 20 & age<=39)


lm1_RD <- lm(emp~ running_var + disc_var, data = Quebec1_RD)
summary(lm1_RD)
```

```
##
## Call:
## lm(formula = emp ~ running_var + disc_var, data = Quebec1_RD)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.074615 -0.030290 -0.005105  0.034550  0.079492
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.5839174  0.0123474  47.291   <2e-16 ***
## running_var -0.0003939  0.0020368  -0.193    0.848
## disc_var     0.0344838  0.0234894   1.468    0.151
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.037 on 37 degrees of freedom
## Multiple R-squared:  0.2263, Adjusted R-squared:  0.1845
## F-statistic:  5.41 on 2 and 37 DF,  p-value: 0.008685
```

```
lm2_RD <- lm(emp ~ running_var+runvar_sq + disc_var, data = Quebec1_RD)
summary(lm2_RD)
```

```
##
## Call:
## lm(formula = emp ~ running_var + runvar_sq + disc_var, data = Quebec1_RD)
##
```

```
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.041191 -0.016432 -0.003476  0.016414  0.058259
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.6121182  0.0099458  61.546  < 2e-16 ***
## running_var -0.0012484  0.0014578  -0.856   0.3974
## runvar_sq   -0.0008546  0.0001407  -6.075 5.52e-07 ***
## disc_var     0.0344838  0.0167334   2.061   0.0466 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02636 on 36 degrees of freedom
## Multiple R-squared:  0.618,  Adjusted R-squared:  0.5861
## F-statistic: 19.41 on 3 and 36 DF,  p-value: 1.173e-07
```

The treatment effect for windows 20-39 with linear and quadratic age are equal to each other, which is 0.0344838. ### Perform robustness checks with different bandwidths (Age:25-35)

```
#windows 25-35

Quebec <- subset(dataset, region=='Quebec')
Quebec <- Quebec %>%
  mutate(running_var=age-30)


Quebec <- Quebec %>%
  mutate(runvar_sq = (running_var)^2)

Quebec <- Quebec %>%
  mutate(runvar_cub = (running_var)^3)

Quebec<- Quebec %>%
  mutate(disc_var=ifelse(age<30,1,0))


Quebec2_RD <- subset(Quebec, age >= 25 & age<=35)


lm1_RD <- lm(emp~ running_var + disc_var, data = Quebec2_RD)
summary(lm1_RD)
```

```
##
## Call:
## lm(formula = emp ~ running_var + disc_var, data = Quebec2_RD)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.061968 -0.013821  0.001257  0.023018  0.047640
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  0.615769   0.012550  49.064   <2e-16 ***
## running_var -0.005933   0.003813  -1.556    0.136
## disc_var     0.006825   0.024216   0.282    0.781
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02828 on 19 degrees of freedom
## Multiple R-squared:  0.407,  Adjusted R-squared:  0.3445
## F-statistic: 6.519 on 2 and 19 DF,  p-value: 0.006986
```

```r
lm2_RD <- lm(emp ~ running_var+runvar_sq + disc_var, data = Quebec2_RD)
summary(lm2_RD)
```

```
##
## Call:
## lm(formula = emp ~ running_var + runvar_sq + disc_var, data = Quebec2_RD)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.041722 -0.019871  0.002889  0.017931  0.044179
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.6242048  0.0125045  49.918   <2e-16 ***
## running_var -0.0046673  0.0036215  -1.289   0.2138
## runvar_sq   -0.0012654  0.0006518  -1.941   0.0681 .
## disc_var     0.0161045  0.0231235   0.696   0.4950
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02642 on 18 degrees of freedom
## Multiple R-squared:  0.5096, Adjusted R-squared:  0.4279
## F-statistic: 6.236 on 3 and 18 DF,  p-value: 0.004315
```

The treatment effect for windows 25-35 with linear and quadratic age are 0.006825, and 0.0161045, respectively. ### Perform robustness checks with different bandwidths (Age:27-33)

```r
#windows 27-33

Quebec <- subset(dataset, region=='Quebec')
Quebec <- Quebec %>%
  mutate(running_var=age-30)


Quebec <- Quebec %>%
  mutate(runvar_sq = (running_var)^2)

Quebec <- Quebec %>%
  mutate(runvar_cub = (running_var)^3)

Quebec<- Quebec %>%
  mutate(disc_var=ifelse(age<30,1,0))
```

```r
Quebec3_RD <- subset(Quebec, age >= 27 & age<=33)


lm1_RD <- lm(emp~ running_var + disc_var, data = Quebec3_RD)
summary(lm1_RD)
```

```
##
## Call:
## lm(formula = emp ~ running_var + disc_var, data = Quebec3_RD)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.036994 -0.018738  0.001223  0.017086  0.039705
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.623704   0.013368  46.656 5.37e-14 ***
## running_var -0.006774   0.006684  -1.013    0.333
## disc_var     0.008060   0.027013   0.298    0.771
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02501 on 11 degrees of freedom
## Multiple R-squared:  0.3736, Adjusted R-squared:  0.2597
## F-statistic:  3.28 on 2 and 11 DF,  p-value: 0.07635
```

```r
lm2_RD <- lm(emp ~ running_var+runvar_sq + disc_var, data = Quebec3_RD)
summary(lm2_RD)
```

```
##
## Call:
## lm(formula = emp ~ running_var + runvar_sq + disc_var, data = Quebec3_RD)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.036994 -0.018086 -0.000969  0.018771  0.038207
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.6252018  0.0145736  42.900 1.14e-12 ***
## running_var -0.0060251  0.0072868  -0.827    0.428
## runvar_sq   -0.0007492  0.0021334  -0.351    0.733
## disc_var     0.0115563  0.0298669   0.387    0.707
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02607 on 10 degrees of freedom
## Multiple R-squared:  0.3812, Adjusted R-squared:  0.1956
## F-statistic: 2.053 on 3 and 10 DF,  p-value: 0.1703
```

The treatment effect for windows 27-33 with linear and quadratic age are 0.008060, and 0.0115563, respectively. ### Perform robustness checks with different bandwidths (Age: 28-32)

```r
#windows 28-32

Quebec <- subset(dataset, region=='Quebec')
Quebec <- Quebec %>%
  mutate(running_var=age-30)


Quebec <- Quebec %>%
  mutate(runvar_sq = (running_var)^2)

Quebec <- Quebec %>%
  mutate(runvar_cub = (running_var)^3)

Quebec<- Quebec %>%
  mutate(disc_var=ifelse(age<30,1,0))


Quebec4_RD <- subset(Quebec, age >= 28 & age<=32)


lm1_RD <- lm(emp~ running_var + disc_var, data = Quebec4_RD)
summary(lm1_RD)
```

```
##
## Call:
## lm(formula = emp ~ running_var + disc_var, data = Quebec4_RD)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.035562 -0.018754  0.002196  0.020843  0.036103
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.62731    0.01726  36.344  3.1e-09 ***
## running_var -0.01156    0.01275  -0.907    0.395
## disc_var    -0.00109    0.03680  -0.030    0.977
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0285 on 7 degrees of freedom
## Multiple R-squared:  0.3073, Adjusted R-squared:  0.1093
## F-statistic: 1.552 on 2 and 7 DF,  p-value: 0.2767
```

```r
lm2_RD <- lm(emp ~ running_var+runvar_sq + disc_var, data = Quebec4_RD)
summary(lm2_RD)
```

```
##
## Call:
## lm(formula = emp ~ running_var + runvar_sq + disc_var, data = Quebec4_RD)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.026410 -0.021804 -0.003905  0.020433  0.033053
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.630356   0.018397  34.264 4.12e-08 ***
## running_var -0.006980   0.014684  -0.475    0.651
## runvar_sq   -0.004576   0.006399  -0.715    0.501
## disc_var     0.014164   0.043711   0.324    0.757
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02955 on 6 degrees of freedom
## Multiple R-squared:  0.3617, Adjusted R-squared:  0.04252
## F-statistic: 1.133 on 3 and 6 DF,  p-value: 0.408
```

The treatment effect for windows 28-32 with linear and quadratic age are -0.00109, and 0.014164, respectively.

As we tighten our window around the discontinuity variable, the value of treatment estimation would be smaller. Still, it would be a lower bound for our treatment effect value, which ensures it is not overestimated.So it can better capture the behaviour of data point around the threshold, and the estimation would be more accurate. However, in some situations, when there is not enough data around the threshold, it may show us a zero value for treatment effect.

Add interaction variables (interaction of age with the discontinuity variable), Quebec

```r
#interaction_linear
Quebec_RD <- subset(Quebec,age >= 25 )

Quebec_RD <- Quebec_RD %>%
  mutate(inter_linear = running_var*disc_var)


lm1_RD <- lm(emp~ running_var + disc_var+inter_linear, data = Quebec_RD)
summary(lm1_RD)
```

```
##
## Call:
## lm(formula = emp ~ running_var + disc_var + inter_linear, data = Quebec_RD)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.042761 -0.018065 -0.000421  0.016457  0.049657
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.624026   0.010471  59.594  < 2e-16 ***
## running_var  -0.009307   0.001961  -4.745 6.59e-05 ***
## disc_var      0.027379   0.021419   1.278   0.2125
## inter_linear  0.012978   0.005965   2.175   0.0389 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0252 on 26 degrees of freedom
```

```
## Multiple R-squared:  0.6926, Adjusted R-squared:  0.6571
## F-statistic: 19.52 on 3 and 26 DF,  p-value: 7.754e-07
```

The estimated treatment effect here is 0.012978, while Question 2_b was -0.001032. So the treatment effect for a situation with a different linear impact on either side of the threshold variable is more significant than the situation when we assume age has the same effect on both sides of the discontinuity variable. In addition, this value is relatively significant in terms of statistics and has lower standard deviation. As a result, our estimated treatment effect is improved using this formula.

Add interaction variables (interactions of age and age squared with the discontinuity variable), Quebec

```r
#interaction_squared
Quebec_RD <- subset(Quebec,age >= 25 )

Quebec_RD <- Quebec_RD %>%
  mutate(inter_linear = running_var*disc_var)

Quebec_RD <- Quebec_RD %>%
  mutate(inter_sq = disc_var*runvar_sq)

lm2_RD <- lm(emp ~  running_var+runvar_sq + disc_var+inter_linear+inter_sq, data = Quebec_RD)
summary(lm2_RD)
```

```
##
## Call:
## lm(formula = emp ~ running_var + runvar_sq + disc_var + inter_linear +
##     inter_sq, data = Quebec_RD)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.042124 -0.021762  0.002169  0.017038  0.043778
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.6338136  0.0142645  44.433   <2e-16 ***
## running_var  -0.0166478  0.0073814  -2.255   0.0335 *
## runvar_sq     0.0008157  0.0007896   1.033   0.3119
## disc_var      0.0153599  0.0414437   0.371   0.7142
## inter_linear  0.0184062  0.0305580   0.602   0.5526
## inter_sq     -0.0011344  0.0049127  -0.231   0.8193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02566 on 24 degrees of freedom
## Multiple R-squared:  0.7057, Adjusted R-squared:  0.6444
## F-statistic: 11.51 on 5 and 24 DF,  p-value: 9.52e-06
```

The treatment effect is -0.0011344 in quadratic form. This is smaller and not statistically significant, but it has lower standard deviation compared to the previous one. The quadratic form is usually applied to better capture the true relationship between age and employment, but it risks outliers pulling around the estimates. Hence, I believe the previous one could better estimate the effect of policy on employment.

Add interaction variables (interactions of age and age squared with the discontinuity variable), ROC

```r
#interaction_squared
ROC_1986 <- subset(dataset, year=='1986' & region=='RoC')

ROC_1986 <- subset(ROC_1986,age >= 25 )

ROC_1986 <- ROC_1986 %>%
  mutate(running_var=age-30)


ROC_1986 <- ROC_1986 %>%
  mutate(runvar_sq = (running_var)^2)

ROC_1986<- ROC_1986 %>%
  mutate(disc_var=ifelse(age<30,1,0))



ROC_1986 <- ROC_1986 %>%
  mutate(inter_linear = running_var*disc_var)

ROC_1986 <- ROC_1986 %>%
  mutate(inter_sq = disc_var*runvar_sq)

Roc_RD <- lm(emp ~  running_var+runvar_sq + disc_var+inter_linear+inter_sq, data = ROC_1986)
summary(Roc_RD)
```

```
##
## Call:
## lm(formula = emp ~ running_var + runvar_sq + disc_var + inter_linear +
##     inter_sq, data = ROC_1986)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.009086 -0.002699 -0.000678  0.003351  0.011950
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.7077508  0.0057058 124.040 7.31e-16 ***
## running_var  -0.0053297  0.0029526  -1.805    0.105
## runvar_sq     0.0003780  0.0003158   1.197    0.262
## disc_var      0.0099813  0.0165775   0.602    0.562
## inter_linear  0.0033743  0.0122232   0.276    0.789
## inter_sq     -0.0004224  0.0019651  -0.215    0.835
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.007257 on 9 degrees of freedom
## Multiple R-squared:  0.8685, Adjusted R-squared:  0.7955
## F-statistic: 11.89 on 5 and 9 DF,  p-value: 0.0009455
```

The estimated treatment effect of ROC-1986 is -0.0004224 with the standard deviation 0.0019651, still it is not statistically significant. Using quadratic form shows insignificant value, around zero,for treatment effect.

Difference-in-Differences (DID) Analysis

Year:1986,Quebec/Roc vs over/under age 30

```
data_1986 <- subset(dataset, year=='1986' )

treat_gp <- subset(data_1986, region=='Quebec')
control_gp <- subset(data_1986, region=='RoC')

treat_post <- subset(treat_gp, age<30)
treat_pre <- subset(treat_gp, age>=30)

control_post <- subset(control_gp, age<30)
control_pre <- subset(control_gp, age>=30)

#treatment group
outcome_treat_after <- mean(treat_post$emp)
outcome_treat_before <- mean(treat_pre$emp)
dif_treat_gp <- outcome_treat_after-outcome_treat_before
dif_treat_gp
```

```
## [1] 0.07393721
```

```
#control group

outcome_control_after <- mean(control_post$emp)
outcome_control_before <- mean(control_pre$emp)
dif_control_gp <- outcome_control_after-outcome_control_before
dif_control_gp
```

```
## [1] 0.005154788
```

```
#Difference in Difference
DID_1986 <- dif_treat_gp-dif_control_gp
DID_1986
```

```
## [1] 0.06878242
```

```
data_1986 <- subset(dataset, year=='1986' )
data_1986$treat <- ifelse(data_1986$region=='Quebec',1,0)
data_1986$age <- ifelse(data_1986$age<30,1,0)

DID <- lm(emp~ age+treat+treat*age, data=data_1986 )

summary (DID)
```

```
##
## Call:
## lm(formula = emp ~ age + treat + treat * age, data = data_1986)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.09214 -0.01275  0.00840  0.02192  0.04306
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.694541   0.010196  68.117  < 2e-16 ***
## age          0.005155   0.014420   0.357  0.72282
## treat       -0.128617   0.014420  -8.919  1.2e-10 ***
## age:treat    0.068782   0.020393   3.373  0.00179 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03224 on 36 degrees of freedom
## Multiple R-squared:  0.7565, Adjusted R-squared:  0.7362
## F-statistic: 37.27 on 3 and 36 DF,  p-value: 3.894e-11
```

The estimated treatment effect is 0.06878. To provide a consistent estimate of the treatment effect, the parallel trends assumption must hold. This means that the trend in the outcome variable should be the same in the treatment and control groups in the absence of the policy intervention. It is essential to test this assumption before interpreting the results of the DID analysis. This means the trend of employment rate should be the same for people with age greater than 30 in Quebec and ROC in 1986.

Difference-in-Differences (DID) Analysis

Age<30, Quebec/Roc vs 1986/1991

```r
data_age30 <- subset(dataset, age<30 )

treat_gp <- subset(data_age30, region=='Quebec')

control_gp <- subset(data_age30, region=='RoC')

treat_post <- subset(treat_gp, year=='1991')
treat_pre <- subset(treat_gp, year=='1986')

control_post <- subset(control_gp, year=='1991')
control_pre <- subset(control_gp, year=='1986')

#treatment group
outcome_treat_after <- mean(treat_post$emp)
outcome_treat_before <- mean(treat_pre$emp)
dif_treat_gp <- outcome_treat_after-outcome_treat_before
dif_treat_gp
```

```
## [1] -0.03858711
```

```r
#control group

outcome_control_after <- mean(control_post$emp)
outcome_control_before <- mean(control_pre$emp)
dif_control_gp <- outcome_control_after-outcome_control_before
dif_control_gp
```

```
## [1] -0.05289531
```

```r
#Difference in Difference
DID_age13 <- dif_treat_gp-dif_control_gp
DID_age13
```

```
## [1] 0.0143082
```

```r
#regression
data_age30 <- subset(dataset, age<30 )

post <- ifelse(data_age30$year == "1991", 1, 0)
treat <- ifelse(data_age30$region == "Quebec", 1, 0)

DID_age <- lm(emp ~ post + treat + post * treat, data = data_age30)
summary(DID_age)
```

```
##
## Call:
## lm(formula = emp ~ post + treat + post * treat, data = data_age30)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.09214 -0.01797  0.01342  0.02447  0.05192
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.69970    0.01176  59.522  < 2e-16 ***
## post        -0.05290    0.01662  -3.182 0.003011 **
## treat       -0.05983    0.01662  -3.599 0.000953 ***
## post:treat   0.01431    0.02351   0.609 0.546618
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03717 on 36 degrees of freedom
## Multiple R-squared:  0.4972, Adjusted R-squared:  0.4553
## F-statistic: 11.87 on 3 and 36 DF,  p-value: 1.493e-05
```

The treatment effect is 0.0143082. The assumption required for this part is that the trend of employment rate should be the same for Quebec and ROC in 1986.

Difference-in-Differences (DID) Analysis

Quebec only, 1986/1991 vs over/under age 30

```
data_Quebce<- subset(dataset, region=='Quebec' )

treat_gp <- subset(data_Quebce, age <30)

control_gp <- subset(data_Quebce, age >=30)

treat_post <- subset(treat_gp, year=='1991')
treat_pre <- subset(treat_gp, year=='1986')

control_post <- subset(control_gp, year=='1991')
control_pre <- subset(control_gp, year=='1986')


#treatment group
outcome_treat_after <- mean(treat_post$emp)
outcome_treat_before <- mean(treat_pre$emp)
dif_treat_gp <- outcome_treat_after-outcome_treat_before
dif_treat_gp
```

```
## [1] -0.03858711
```

```
#control group

outcome_control_after <- mean(control_post$emp)
outcome_control_before <- mean(control_pre$emp)
dif_control_gp <- outcome_control_after-outcome_control_before
dif_control_gp
```

```
## [1] 0.0324426
```

```
#Difference in Difference
DID_Quebce <- dif_treat_gp-dif_control_gp
DID_Quebce
```

```
## [1] -0.07102971
```

```
#regression

Quebec <- subset(dataset, region=='Quebec')
post <- ifelse(Quebec$year == "1991", 1, 0)
treat <- ifelse(Quebec$age < 30, 1, 0)

DID_Que <- lm(emp ~ post + treat + post * treat, data = Quebec)
summary(DID_Que)
```

```
##
```

```
## Call:
## lm(formula = emp ~ post + treat + post * treat, data = Quebec)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.092136 -0.020712  0.003809  0.022851  0.065043
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.56592    0.01027  55.085  < 2e-16 ***
## post         0.03244    0.01453   2.233  0.03186 *
## treat        0.07394    0.01453   5.089 1.14e-05 ***
## post:treat  -0.07103    0.02055  -3.457  0.00142 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03249 on 36 degrees of freedom
## Multiple R-squared:  0.4196, Adjusted R-squared:  0.3712
## F-statistic: 8.676 on 3 and 36 DF,  p-value: 0.000183
```

The estimated treatment effect is -0.07103. To ensure a consistent estimate of the treatment effect, it is necessary to satisfy the parallel trends assumption, which requires that the employment trend for individuals in Quebec under the age of 30 and over the age of 30 is the same before the implementation of the policy intervention.

Triple-Difference (DDD) Analysis

1986/1991 vs over/under age 30 vs Que/RoC.

```r
# RUNNING REGRESSION

dataset <- read_dta("C:/Users/bahar/Documents/MSc courses/ECON 562/Assignment 6/quebec-rd.dta")
Treat <- ifelse(dataset$region=='Quebec',1,0)
post <- ifelse(dataset$year=='1991',1,0)
B <- ifelse(dataset$age<30 ,1,0)


Triple_est <- lm(emp~ B+ Treat+post+ Treat*B + Treat*post + post*B +Treat*B*post, data=dataset)
summary(Triple_est)
```

```
##
## Call:
## lm(formula = emp ~ B + Treat + post + Treat * B + Treat * post +
##     post * B + Treat * B * post, data = dataset)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.092136 -0.017278  0.005768  0.021367  0.065043
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)   0.694541   0.009899  70.165  < 2e-16 ***
## B             0.005155   0.013999   0.368 0.713783
## Treat        -0.128617   0.013999  -9.188 9.35e-14 ***
## post         -0.023554   0.013999  -1.683 0.096789 .
## B:Treat       0.068782   0.019797   3.474 0.000871 ***
## Treat:post    0.055997   0.019797   2.828 0.006054 **
## B:post       -0.029341   0.019797  -1.482 0.142686
## B:Treat:post -0.041689   0.027998  -1.489 0.140856
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0313 on 72 degrees of freedom
## Multiple R-squared:  0.6975, Adjusted R-squared:  0.6681
## F-statistic: 23.71 on 7 and 72 DF,  p-value: 2.239e-16
```

The estimated treatment effect is -0.0417. The assumption required here is that the city-specific trends should be the same for groups B and A, i.e., people above age 30 and below 30. The other assumption is that people over 30 are not affected by policy intervention (treatment). Based on these two assumptions, we can get a consistent estimator.

Comparison of RD vs. DID Approaches

In my opinion, RD is a more suitable estimator than DID because the identifying assumptions for DID may not hold in this situation. DID assumes that the outcome measure would have followed the same trend over time for the treatment and control groups in the absence of the treatment. However, when we examine the graphs for Quebec and ROC in 1986 for age above 30, we can see that they do not have the same trend before treatment. Even if this assumption holds, it also requires that the treatment and control groups are similar in all other respects except for their exposure to the treatment, and that any other factors that might affect the outcome measure are constant over time and affect both groups equally. This assumption is difficult to verify empirically, and there may be concerns about the potential for time-varying confounding factors that affect the outcome differently for the treatment and control groups over time. Therefore, DID may not be suitable for this problem.

On the other hand, the identifying assumption for RD is that the treatment assignment is based on a continuous variable that is closely related to the outcome measure, and that the relationship is discontinuous at the threshold value. The three assumptions for RD as follows: 1) distribution of x (running variable) should be continuous around cutoff point ,this means people are as good as randomly assigned to treatment. 2) Monotonicity assumption - movements in X only induce movements in d (treatment variable) in one direction.3) It should be a smooth relationship between x and y. So any jump in the expected value of y when x crosses the threshold must be due to the causal effect of policy intervention on the expected value of y. As we observed from the distribution of x, it was continuous around threshold 30 which implies that the required assumption of DID holds in this problem. Therefore, RD is a more appropriate estimator for this situation.