

[Project 1 based on Angrist and Krueger (1991, QJE)]

Contents

Question 1	1
Question 2 : Taches de réplication	2
Préliminaires pour la réalisation des figures et estimations	2
Réalisation de la figure 2	3
Réalisation de la figure 5	4
Présentation du tableau 5 (estimations OLS et IV)	4
Présentation du tableau 3.B (estimations de Wald)	11
Explication des figures et tables	11
Figure 1 : Années d'éducatons et saisonnalité des naissances de 1930 à 1940	11
Figure 2 : Années d'études et trimestre de naissance.	12
Figure 5 : Logarithme moyen du salaire hebdomadaire par trimestre de naissance, hommes nés entre 1930 et 1949.	12
Table 5 : Les estimations MCO et 2MCO sur les rendements d'éducation pour les hommes nés entre 1930 et 1939	12
Tableau 3 : L'estimateur de Wald pour les hommes nés entre 1930 et 1939	12
Utilisation de l'extrait des données d'ipums USA avec les variables pertinentes, et compara- ison avec les résultats de notre extrait de données à QOB.raw	12
Préliminaires	13
Figure 2 avec les données ipums	13
Figure 5 avec les données ipums	14
Les tableaux 5 (estimations OLS et IV) et 3.B (estimations de Wald)	15
Wald estimation	18
Calcul de l'estimateur de wald	19
Réalisé par :	
• Segodo Ulrich	
• Coudray Yaelle	
• Gerent Fiona	
• Gomez Boris	
• Renaud Axelle	

Question 1

Ce document essaie de montrer si la fréquentation scolaire obligatoire, en fonction de la saisonnalité, affecte, ou non, la scolarité et les revenus aux Etats-Unis. L'utilisation des MCO est invalide car les capacités intellectuelles ne sont pas prises en compte : on parle d'un biais de variable. Malgré ce point, les estimations sont proches des MCO d'après le texte. Cela pourrait s'expliquer par une sous-estimation de la mesure qui compenserait la surestimation liée à l'endogénéité de l'éducation. C'est pourquoi nous avons besoin des variables instrumentales pour résoudre cette incohérence. Dans ce document, deux natures de variables instrumentales sont utilisées : l'estimateur de Wald et l'estimateur des doubles moindres carrés ordinaires (TSLS en anglais). L'auteur nous dit que le simple estimateur de Wald serait biaisé car il ne concernerait que les hommes d'un certain âge et donc apporterait une sous-estimation des salaires en fonction de la saisonnalité.

Il est donc préférable d'utiliser le modèle TSLS car les rendements de l'éducation sont statistiquement significatif. De plus, il est indiscernable de l'estimation des MCO, et donc sans biais.

Question 2 : Taches de réplcation

Préliminaires pour la réalisation des figures et estimations

```
library(tidyverse)
library(haven)
library(ipumsr)
library(foreign)
library(sandwich) # for: vcovCL
library("AER") # for: ivreg
library("lmtest") # for: coeftest
library(stargazer)
library(lubridate)
library(ggplot2)

path = "G:/Nouveau/SaiRstudio/project/"
pums <- read.table(paste(path,"database/QOB.raw",sep=""),
                  header = FALSE,
                  stringsAsFactors = FALSE)
colnames(pums)[c(1,2,4,5,6,9:13,16,18,19:21,24,25,27)] <- c("AGE", "AGEQ", "EDUC",
                  "ENOCENT", "ESOCENT", "LWKLYWGE", "MARRIED", "MIDATL", "MT",
                  "NEWENG", "RACE", "SMSA", "SOATL", "V22", "V23", "WNOCENT", "WSOCENT", "YOB")
pums <- as_tibble(pums)
pums
```

```
## # A tibble: 816,435 x 27
##   AGE AGEQ V3 EDUC ENOCENT ESOCENT V7 V8 LWKLYWGE MARRIED MIDATL
##   <int> <dbl> <int> <int> <int> <int> <int> <dbl> <dbl> <int> <int>
## 1 47 1947 2 12 0 0 14 10.2 6.25 1 0
## 2 46 1946. 2 12 0 0 14 9.80 5.85 1 0
## 3 50 1950 2 12 0 0 14 10.6 6.65 1 0
## 4 47 1947 2 16 0 0 18 10.7 6.71 1 0
## 5 42 1942. 2 14 0 0 16 10.3 6.36 1 0
## 6 44 1944. 2 12 0 0 14 9.39 5.44 1 0
## 7 42 1942 2 12 0 0 14 10.3 6.39 1 0
## 8 50 1950 2 12 0 0 14 7.60 4.61 1 0
## 9 40 1940. 2 12 0 0 14 10.5 6.55 1 0
## 10 43 1944. 2 7 0 0 9 11.0 7.02 0 0
## # ... with 816,425 more rows, and 16 more variables: MT <int>, NEWENG <int>,
## # V14 <int>, V15 <int>, CENSUS <int>, V17 <int>, QOB <int>, RACE <int>,
## # SMSA <int>, SOATL <int>, V22 <int>, V23 <int>, WNOCENT <int>,
## # WSOCENT <int>, V26 <int>, YOB <int>
```

```
pums %>%
  mutate(cohort = factor(1*(YOB<=39 & YOB >=30) +
                        2*(YOB<=49 & YOB >=40),
                        levels=c(1,2), labels=c("30-39","40-49"))) -> pums
```

Créons une nouvelle base sur laquelle nous allons faire les manipulations pour réaliser les figures.

```
pums %>%
  select(AGE, AGEQ, EDUC,
         ENOCENT, ESOCENT, LWKLYWGE,
```

```
MARRIED, MIDATL, MT, NEWENG, CENSUS, QOB, RACE,
SMSA, SOATL, WNOCENT, WSOCENT, YOB, cohort) -> mydata
```

```
mydata %>%
  mutate(date=ymd(paste("19",YOB,QOB * 3, sep=""),truncated = 2)) ->
  mydata
```

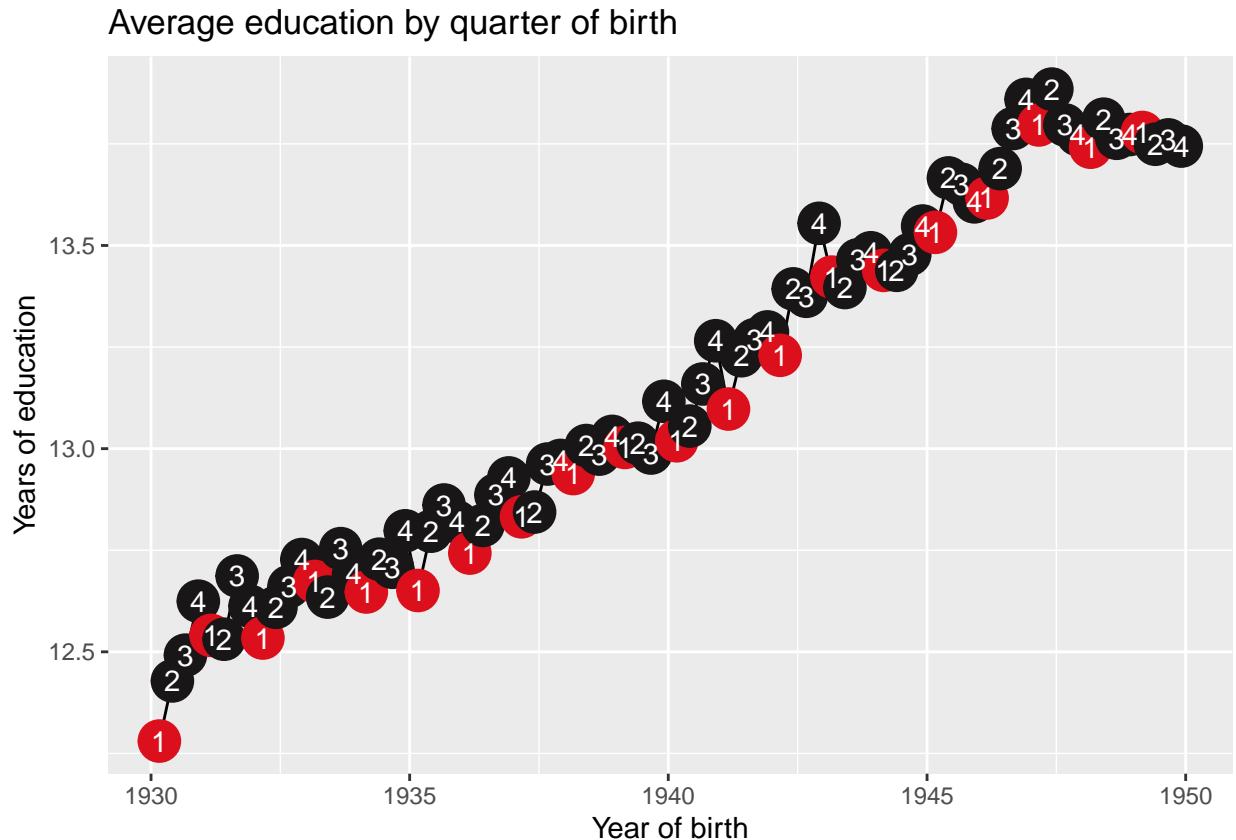
Réalisation de la figure 2

```
mydata %>%
  group_by(date, QOB) %>%
  summarise(moyenne = mean(EDUC, na.rm=TRUE)) -> temp
```

`summarise()` has grouped output by 'date'. You can override using the `.groups` argument.

```
temp %>%
  mutate(QOB_factor = factor(QOB)) -> temp

ggplot(data=temp, aes(x=date , y=moyenne)) +
  geom_line()+
  geom_point(aes(col=QOB_factor), size = 7)+
  scale_color_manual(values = c("#DC101C", "#181616", "#181616", "#181616")) +
  geom_text(aes(label=QOB_factor), hjust=0.5, vjust=0.5, col="white")+
  theme(legend.position = "None")+
  ggtitle("Average education by quarter of birth")+
  xlab("Year of birth") + ylab("Years of education")
```



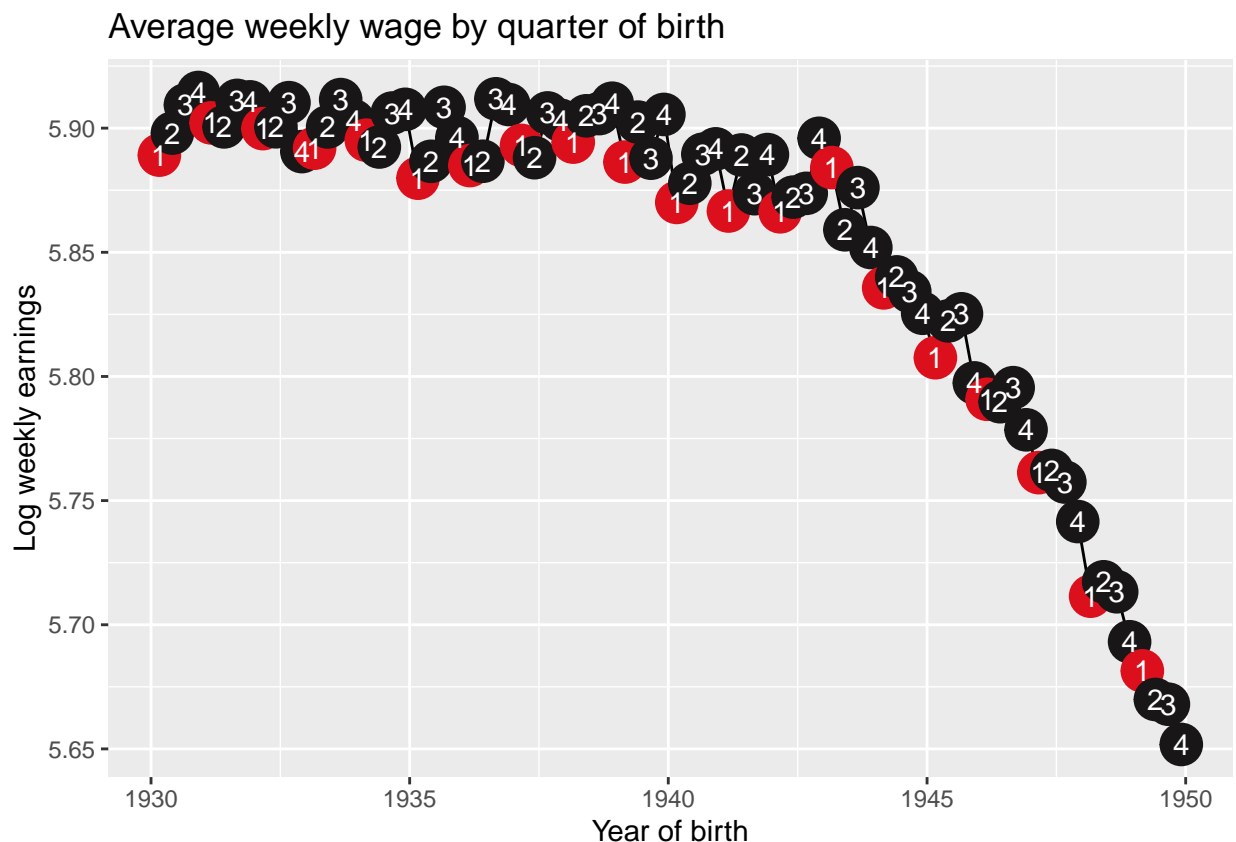
Réalisation de la figure 5

```
mydata %>%
  group_by(date, QOB) %>%
  summarise(moyen = mean(LWKLYWGE, na.rm=TRUE)) -> temp1

## `summarise()` has grouped output by 'date'. You can override using the `.groups` argument.

temp1 %>%
  mutate(QOB_factor = factor(QOB)) -> temp1

ggplot(data=temp1, aes(x=date, y=moyen, shape=)) +
  geom_line()+
  geom_point(aes(col=QOB_factor), size = 7)+
  scale_color_manual(values = c("#DC101C", "#181616", "#181616", "#181616")) +
  geom_text(aes(label=QOB_factor), hjust=0.5, vjust=0.5, col="white")+
  theme(legend.position = "None")+
  ggtitle("Average weekly wage by quarter of birth")+
  xlab("Year of birth") + ylab("Log weekly earnings")
```



Présentation du tableau 5 (estimations OLS et IV)

Pour présentation ce tableau, nous avons effectué les 8 estimations que voici :

```
mydata %>%
  filter(cohort == "30-39") -> mydata.tab5
```

Régression

```
reg1.MCO <- lm(LWKLYWGE ~ EDUC + factor(YOB), data=mydata.tab5)
summary(reg1.MCO)
```

```
##
## Call:
## lm(formula = LWKLYWGE ~ EDUC + factor(YOB), data = mydata.tab5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.7490 -0.2354  0.0726  0.3378  4.6448
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.017348   0.005471  917.144 < 2e-16 ***
## EDUC          0.071081   0.000339  209.674 < 2e-16 ***
## factor(YOB)31 -0.006387   0.005039   -1.268 0.204971
## factor(YOB)32 -0.014838   0.004972   -2.984 0.002844 **
## factor(YOB)33 -0.017583   0.005032   -3.494 0.000476 ***
## factor(YOB)34 -0.020999   0.004985   -4.213 2.52e-05 ***
## factor(YOB)35 -0.032895   0.004952   -6.643 3.07e-11 ***
## factor(YOB)36 -0.031781   0.004956   -6.413 1.43e-10 ***
## factor(YOB)37 -0.036712   0.004908   -7.480 7.47e-14 ***
## factor(YOB)38 -0.036890   0.004866   -7.582 3.42e-14 ***
## factor(YOB)39 -0.048164   0.004847   -9.937 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6376 on 329498 degrees of freedom
## Multiple R-squared:  0.1177, Adjusted R-squared:  0.1177
## F-statistic: 4397 on 10 and 329498 DF, p-value: < 2.2e-16
```

```
mydata.tab5 %>%
```

```
  mutate( instru = factor(YOB):factor(QOB) )-> mydata.tab5
reg2.TSLS <- ivreg(LWKLYWGE ~ EDUC + factor(YOB) | instru, data=mydata.tab5)
summary(reg2.TSLS)
```

```
##
## Call:
## ivreg(formula = LWKLYWGE ~ EDUC + factor(YOB) | instru, data = mydata.tab5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.87808 -0.24040  0.07032  0.34172  4.74912
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.792727   0.200684  23.882 < 2e-16 ***
## EDUC          0.089115   0.016110   5.532 3.17e-08 ***
## factor(YOB)31 -0.008881   0.005529   -1.606 0.108228
## factor(YOB)32 -0.018030   0.005750   -3.136 0.001715 **
## factor(YOB)33 -0.021796   0.006301   -3.459 0.000542 ***
## factor(YOB)34 -0.025788   0.006584   -3.917 8.98e-05 ***
## factor(YOB)35 -0.038827   0.007267   -5.343 9.13e-08 ***
```

```
## factor(YOB)36 -0.038762 0.007977 -4.859 1.18e-06 ***
## factor(YOB)37 -0.044817 0.008758 -5.117 3.10e-07 ***
## factor(YOB)38 -0.046545 0.009911 -4.696 2.65e-06 ***
## factor(YOB)39 -0.058527 0.010458 -5.597 2.19e-08 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6404 on 329498 degrees of freedom
## Multiple R-Squared: 0.1102, Adjusted R-squared: 0.1101
## Wald test: 4.167 on 10 and 329498 DF, p-value: 8.581e-06

reg3.MCO <- lm(LWKLYWGE ~ EDUC + factor(YOB)+ AGEQ + I(AGEQ^2), data=mydata.tab5)
summary(reg3.MCO)
```

```
##
## Call:
## lm(formula = LWKLYWGE ~ EDUC + factor(YOB) + AGEQ + I(AGEQ^2),
##     data = mydata.tab5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.7504 -0.2346  0.0731  0.3378  4.6457
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.996e+03  2.603e+03   1.151  0.24972
## EDUC           7.107e-02  3.391e-04 209.621 < 2e-16 ***
## factor(YOB)31 -6.207e-03  8.551e-03  -0.726  0.46793
## factor(YOB)32 -1.605e-02  1.366e-02  -1.175  0.23987
## factor(YOB)33 -2.177e-02  1.824e-02  -1.194  0.23252
## factor(YOB)34 -2.978e-02  2.205e-02  -1.351  0.17678
## factor(YOB)35 -4.778e-02  2.515e-02  -1.900  0.05748 .
## factor(YOB)36 -5.434e-02  2.781e-02  -1.954  0.05072 .
## factor(YOB)37 -6.861e-02  3.032e-02  -2.263  0.02367 *
## factor(YOB)38 -7.961e-02  3.301e-02  -2.411  0.01590 *
## factor(YOB)39 -1.033e-01  3.634e-02  -2.844  0.00446 **
## AGEQ          -3.069e+00  2.677e+00  -1.147  0.25149
## I(AGEQ^2)       7.874e-04  6.881e-04   1.144  0.25246
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6376 on 329496 degrees of freedom
## Multiple R-squared: 0.1178, Adjusted R-squared: 0.1177
## F-statistic: 3665 on 12 and 329496 DF, p-value: < 2.2e-16
```

```
reg4.TSLS <- ivreg(LWKLYWGE ~ EDUC + factor(YOB) + AGEQ + I(AGEQ^2) | instru +
AGEQ + I(AGEQ^2), data=mydata.tab5)
summary(reg4.TSLS)
```

```
##
## Call:
## ivreg(formula = LWKLYWGE ~ EDUC + factor(YOB) + AGEQ + I(AGEQ^2) |
##       instru + AGEQ + I(AGEQ^2), data = mydata.tab5)
##
## Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -8.78564 -0.23533  0.07149  0.33989  4.67407
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.159e+03  2.774e+03   1.139   0.2548
## EDUC          7.600e-02  2.897e-02   2.624   0.0087 **
## factor(YOB)31 -5.542e-03  9.405e-03  -0.589   0.5557
## factor(YOB)32 -1.432e-02  1.705e-02  -0.840   0.4009
## factor(YOB)33 -1.914e-02  2.391e-02  -0.801   0.4233
## factor(YOB)34 -2.621e-02  3.042e-02  -0.862   0.3889
## factor(YOB)35 -4.353e-02  3.544e-02  -1.228   0.2194
## factor(YOB)36 -4.947e-02  3.991e-02  -1.240   0.2152
## factor(YOB)37 -6.322e-02  4.387e-02  -1.441   0.1495
## factor(YOB)38 -7.391e-02  4.705e-02  -1.571   0.1163
## factor(YOB)39 -9.718e-02  5.131e-02  -1.894   0.0582 .
## AGEQ          -3.238e+00  2.854e+00  -1.134   0.2567
## I(AGEQ^2)      8.309e-04  7.343e-04   1.132   0.2578
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6378 on 329496 degrees of freedom
## Multiple R-Squared:  0.1172, Adjusted R-squared:  0.1172
## Wald test: 3.712 on 12 and 329496 DF, p-value: 1.237e-05

mydata.tab5 %>%
  mutate( region_of_residence = factor(NEWENG + MIDATL + ENOCENT +
                                       WNOCENT + SOATL + ESOCENT +
                                       WSOCENT + MT)) -> mydata.tab5
reg5.MCO <- lm(LWKLYWGE ~ EDUC + RACE + SMSA + MARRIED + factor(YOB) +
               region_of_residence, data=mydata.tab5)
summary(reg5.MCO)

##
## Call:
## lm(formula = LWKLYWGE ~ EDUC + RACE + SMSA + MARRIED + factor(YOB) +
##     region_of_residence, data = mydata.tab5)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -8.8542 -0.2277  0.0594  0.3239  4.6429
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.0174337  0.0069121 725.891 < 2e-16 ***
## EDUC          0.0635079  0.0003392 187.224 < 2e-16 ***
## RACE         -0.2672457  0.0040307 -66.302 < 2e-16 ***
## SMSA         -0.1906441  0.0028270 -67.437 < 2e-16 ***
## MARRIED       0.2465698  0.0031775  77.599 < 2e-16 ***
## factor(YOB)31 -0.0044337  0.0049227  -0.901  0.36776
## factor(YOB)32 -0.0101356  0.0048574  -2.087  0.03692 *
## factor(YOB)33 -0.0103613  0.0049164  -2.108  0.03507 *
## factor(YOB)34 -0.0130221  0.0048696  -2.674  0.00749 **
## factor(YOB)35 -0.0224302  0.0048376  -4.637 3.54e-06 ***
## factor(YOB)36 -0.0217729  0.0048416  -4.497 6.89e-06 ***
```

```
## factor(YOB)37      -0.0252131  0.0047955  -5.258 1.46e-07 ***
## factor(YOB)38      -0.0240489  0.0047541  -5.059 4.23e-07 ***
## factor(YOB)39      -0.0331200  0.0047362  -6.993 2.70e-12 ***
## region_of_residence1 -0.0764011  0.0033402 -22.873 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6229 on 329494 degrees of freedom
## Multiple R-squared:  0.1581, Adjusted R-squared:  0.1581
## F-statistic:  4421 on 14 and 329494 DF,  p-value: < 2.2e-16

reg6.TSLS <- ivreg(LWKLYWGE ~ EDUC + RACE + SMSA + MARRIED + factor(YOB) +
                  region_of_residence | instru + RACE + SMSA +
                  MARRIED + factor(YOB) + region_of_residence, data=mydata.tab5)
summary(reg6.TSLS)
```

```
##
## Call:
## ivreg(formula = LWKLYWGE ~ EDUC + RACE + SMSA + MARRIED + factor(YOB) +
##       region_of_residence | instru + RACE + SMSA + MARRIED + factor(YOB) +
##       region_of_residence, data = mydata.tab5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.94212 -0.23365  0.05793  0.32696  4.60802
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.820390   0.221615  21.751 < 2e-16 ***
## EDUC           0.078206   0.016527   4.732 2.22e-06 ***
## RACE          -0.242673   0.027917  -8.693 < 2e-16 ***
## SMSA          -0.174463   0.018409  -9.477 < 2e-16 ***
## MARRIED        0.243570   0.004640  52.495 < 2e-16 ***
## factor(YOB)31  -0.006482   0.005447  -1.190 0.234078
## factor(YOB)32  -0.012962   0.005816  -2.229 0.025835 *
## factor(YOB)33  -0.014114   0.006488  -2.175 0.029618 *
## factor(YOB)34  -0.017226   0.006796  -2.535 0.011252 *
## factor(YOB)35  -0.027727   0.007681  -3.610 0.000306 ***
## factor(YOB)36  -0.027770   0.008308  -3.343 0.000830 ***
## factor(YOB)37  -0.032169   0.009180  -3.504 0.000458 ***
## factor(YOB)38  -0.032264   0.010393  -3.104 0.001906 **
## factor(YOB)39  -0.042033   0.011088  -3.791 0.000150 ***
## region_of_residence1 -0.062926  0.015514  -4.056 4.99e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6246 on 329494 degrees of freedom
## Multiple R-Squared:  0.1533, Adjusted R-squared:  0.1533
## Wald test:  1908 on 14 and 329494 DF,  p-value: < 2.2e-16

reg7.MCO <- lm(LWKLYWGE ~ EDUC + RACE + SMSA + MARRIED + factor(YOB) +
              region_of_residence + AGEQ + I(AGEQ^2), data=mydata.tab5)
summary(reg7.MCO)
```

```
##
```



```
## Call:
## lm(formula = LWKLYWGE ~ EDUC + RACE + SMSA + MARRIED + factor(YOB) +
##     region_of_residence + AGEQ + I(AGEQ^2), data = mydata.tab5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.8551 -0.2276  0.0596  0.3237  4.6428
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.018e+03  2.543e+03   1.187  0.2353
## EDUC           6.350e-02  3.393e-04 187.178 <2e-16 ***
## RACE          -2.672e-01  4.031e-03 -66.297 <2e-16 ***
## SMSA          -1.906e-01  2.827e-03 -67.435 <2e-16 ***
## MARRIED        2.466e-01  3.178e-03  77.602 <2e-16 ***
## factor(YOB)31  -3.665e-03  8.353e-03  -0.439  0.6608
## factor(YOB)32  -1.019e-02  1.334e-02  -0.763  0.4452
## factor(YOB)33  -1.282e-02  1.781e-02  -0.720  0.4717
## factor(YOB)34  -1.952e-02  2.154e-02  -0.906  0.3648
## factor(YOB)35  -3.449e-02  2.457e-02  -1.404  0.1604
## factor(YOB)36  -4.099e-02  2.717e-02  -1.509  0.1314
## factor(YOB)37  -5.324e-02  2.962e-02  -1.797  0.0723 .
## factor(YOB)38  -6.240e-02  3.225e-02  -1.935  0.0530 .
## factor(YOB)39  -8.344e-02  3.550e-02  -2.350  0.0188 *
## region_of_residence1 -7.642e-02  3.340e-03 -22.878 <2e-16 ***
## AGEQ          -3.092e+00  2.615e+00  -1.183  0.2369
## I(AGEQ^2)       7.935e-04  6.721e-04   1.181  0.2378
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6229 on 329492 degrees of freedom
## Multiple R-squared:  0.1582, Adjusted R-squared:  0.1581
## F-statistic: 3869 on 16 and 329492 DF, p-value: < 2.2e-16

reg8.TSLS <- ivreg(LWKLYWGE ~ EDUC + RACE + SMSA + MARRIED + factor(YOB) +
                  region_of_residence + AGEQ + I(AGEQ^2) | instru + RACE + SMSA +
                  MARRIED + AGEQ + I(AGEQ^2) + factor(YOB) +
                  region_of_residence , data=mydata.tab5)
summary(reg8.TSLS)

##
## Call:
## ivreg(formula = LWKLYWGE ~ EDUC + RACE + SMSA + MARRIED + factor(YOB) +
##     region_of_residence + AGEQ + I(AGEQ^2) | instru + RACE +
##     SMSA + MARRIED + AGEQ + I(AGEQ^2) + factor(YOB) + region_of_residence,
##     data = mydata.tab5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.84572 -0.22725  0.06004  0.32353  4.64643
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.968e+03  2.705e+03   1.097  0.27256
## EDUC           6.193e-02  2.901e-02   2.134  0.03281 *
```

```
## RACE -2.699e-01 4.863e-02 -5.549 2.88e-08 ***
## SMSA -1.924e-01 3.206e-02 -6.001 1.96e-09 ***
## MARRIED 2.469e-01 6.720e-03 36.739 < 2e-16 ***
## factor(YOB)31 -3.853e-03 9.037e-03 -0.426 0.66990
## factor(YOB)32 -1.067e-02 1.605e-02 -0.665 0.50604
## factor(YOB)33 -1.356e-02 2.241e-02 -0.605 0.54512
## factor(YOB)34 -2.054e-02 2.856e-02 -0.719 0.47219
## factor(YOB)35 -3.569e-02 3.305e-02 -1.080 0.28014
## factor(YOB)36 -4.238e-02 3.742e-02 -1.133 0.25734
## factor(YOB)37 -5.478e-02 4.103e-02 -1.335 0.18184
## factor(YOB)38 -6.402e-02 4.400e-02 -1.455 0.14566
## factor(YOB)39 -8.518e-02 4.786e-02 -1.780 0.07514 .
## region_of_residence1 -7.786e-02 2.680e-02 -2.905 0.00367 **
## AGEQ -3.041e+00 2.783e+00 -1.092 0.27467
## I(AGEQ^2) 7.801e-04 7.160e-04 1.089 0.27595
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6229 on 329492 degrees of freedom
## Multiple R-Squared: 0.1581, Adjusted R-squared: 0.1581
## Wald test: 1679 on 16 and 329492 DF, p-value: < 2.2e-16
```

Grace à stargazer, nous avons pu présenter les résultats dans un tableau comme suit :

```
stargazer(reg1.MCO, reg2.TSLS, reg3.MCO, reg4.TSLS, reg5.MCO, reg6.TSLS, reg7.MCO,
  reg8.TSLS,
  dep.var.caption="", dep.var.labels="",
  omit.table.layout = "n", star.cutoffs = NA, keep.stat=c("rsq", "n"), no.space=TRUE,
  header=FALSE,
  keep=c("EDUC", "RACE", "SMSA", "MARRIED", "AGE", "AGEQ^2"),
  column.labels=c("OLS", "TSLS", "OLS", "TSLS", "OLS", "TSLS", "OLS", "TSLS"),
  title="OLS and TSLS of the return to education for men born 1390-1939: 1980 CENSUS", type="text")
```

```
##
## OLS and TSLS of the return to education for men born 1390-1939: 1980 CENSUS
## =====
##
##          OLS      instrumental      OLS      instrumental      OLS      instrumental      OLS      instrumental
##          OLS      variable      OLS      variable      OLS      variable      OLS      variable
##          OLS      TSLS      OLS      TSLS      OLS      TSLS      OLS      TSLS
##          (1)      (2)      (3)      (4)      (5)      (6)      (7)      (8)
## -----
## EDUC          0.071      0.089      0.071      0.076      0.064      0.078      0.064      0.062
##              (0.0003)   (0.016)   (0.0003)   (0.029)   (0.0003)   (0.017)   (0.0003)   (0.029)
## RACE                                -0.267      -0.243      -0.267      -0.270
##                                (0.004)   (0.028)   (0.004)   (0.049)
## SMSA                                -0.191      -0.174      -0.191      -0.192
##                                (0.003)   (0.018)   (0.003)   (0.032)
## MARRIED                                0.247      0.244      0.247      0.247
##                                (0.003)   (0.005)   (0.003)   (0.007)
## AGEQ                                -3.069      -3.238                                -3.092      -3.041
##                                (2.677)   (2.854)                                (2.615)   (2.783)
## I(AGEQ2)                                0.001      0.001                                0.001      0.001
##                                (0.001)   (0.001)                                (0.001)   (0.001)
```

```
## -----
## Observations 329,509    329,509    329,509    329,509    329,509    329,509    329,509    329,509
## R2           0.118      0.110      0.118      0.117      0.158      0.153      0.158      0.158
## =====
```

Présentation du tableau 3.B (estimations de Wald)

```
mydata.tab5$wald_dum <- (mydata.tab5$Q0B == 2 | mydata.tab5$Q0B == 3 | mydata.tab5$Q0B == 4) * 1

mydata.tab5 %>%
  group_by(wald_dum == 0) %>%
  summarize(moy_salaire = mean(LWKLYWGE, na.rm=TRUE),
            moy_educ = mean(EDUC, na.rm=TRUE)) %>%
  ungroup() -> temp3
```

On regroupe les données sous forme de tableau

```
(data <-
  data.frame(
    "x" = c( "Moy_salaire", "Moy_educ"),
    "1st quarter" = c(temp3$moy_salaire[2], temp3$moy_educ[2]),
    "3rd or 4th" = c(temp3$moy_salaire[1], temp3$moy_educ[1])
  )
)
```

```
##           x X1st.quarter X3rd.or.4th
## 1 Moy_salaire    5.891596    5.902695
## 2   Moy_educ    12.688066    12.796883
```

Ainsi nous Calculons l'estimateur des wald de l'estimateur de wald

```
temp3 %>%
  mutate(wage_Q1 = moy_salaire[2]) %>%
  mutate(wage_Q2_4 = moy_salaire[1]) %>%
  mutate(educ_Q1 = moy_educ[2]) %>%
  mutate(educ_Q2_4 = moy_educ[1]) %>%
  mutate(wald_est = (wage_Q1 - wage_Q2_4)/(educ_Q1 - educ_Q2_4)) %>%
  mutate(wage_Q1 = NULL) %>%
  mutate(wage_Q2_4 = NULL) %>%
  mutate(educ_Q1 = NULL) %>%
  mutate(educ_Q2_4 = NULL) ->temp3

temp3$wald_est[1]

## [1] 0.101995
```

Explication des figures et tables

Figure 1 : Années d'éducatons et saisonnalité des naissances de 1930 à 1940

La figure 1 représente les années complètes de scolarité en fonction du trimestre de naissance de 1930 à 1940. On remarque qu'au début des années 1930, contrairement à la fin des années 1930, l'écart de scolarisation entre les personnes nées en début d'année et celles en fin d'année est plutôt important (0.4 ans en 1930, contre 0.1 ans en 1940). Généralement, les individus nés dans le premier trimestre de l'année sont scolarisés moins longtemps que les individus nés dans le troisième ou quatrième trimestre.

Figure 2 : Années d'études et trimestre de naissance.

Ce graphique décrit la relation entre l'éducation et le trimestre de naissance pour les hommes nés au cours de cette période. Il existe une relation positive entre ces variables pour les hommes nés entre 1940 et 1947 et la tendance est à la baisse pour ceux nés entre 1947 et 1949. Cela peut s'expliquer par le fait que les hommes nés à la fin des années 1940 sont plus jeunes et n'ont pas terminé leur scolarité au moment de l'étude. Mais on peut aussi ajouter une autre explication conjoncturelle : la guerre du Vietnam. Nous pouvons également étudier la relation entre le trimestre de naissance et les années d'études terminées. La tendance principale affirme que le niveau d'éducation moyen est plus élevé pour les hommes nés vers la fin de l'année.

Figure 5 : Logarithme moyen du salaire hebdomadaire par trimestre de naissance, hommes nés entre 1930 et 1949.

Ce graphique présente le logarithme moyen du salaire hebdomadaire des hommes âgés de 30 à 49 ans par trimestre de naissance. Il a pour but de savoir si les petites différences d'éducation pour les hommes nés à différents trimestres de l'année ont un impact sur les revenus. On voit que les hommes nés au premier trimestre gagnent un peu moins que les autres. On peut aussi dire qu'il existe une relation assez constante entre l'année de naissance et le revenu pour les hommes nés entre 1930 et 1939. Mais la relation est différente pour ceux nés entre 1940 et 1949 : le logarithme des revenus hebdomadaires diminue avec l'année de naissance.

Table 5 : Les estimations MCO et 2MCO sur les rendements d'éducation pour les hommes nés entre 1930 et 1939

La table 5 représente les estimations du modèle avec comme échantillon les hommes de 50 à 59 ans en 1980). On observe des similitudes dans les estimations des MCO et des 2MCO (TSLS) du rendement de l'éducation. La statistique du Xhi-deux teste l'hypothèse selon laquelle les diverses combinaisons d'instruments produisent la même estimation du rendement de l'éducation.

Tableau 3 : L'estimateur de Wald pour les hommes nés entre 1930 et 1939

Pour le tableau 3, les auteurs utilisent le modèle saisonnier de l'éducation pour calculer le taux de rendement d'une année d'éducation à partir d'une application de la méthode de Wald. Cet estimateur calcule simplement le rendement de l'éducation comme le rapport entre la différence de rémunération par trimestre de naissance et la différence d'années de scolarité par trimestre de naissance. Dans le tableau 3 la partie B représente l'ensemble des estimations pour les hommes nés entre 1930 et 1939. Pour cet échantillon, l'estimation de Wald du rendement de l'éducation est de 0,102 ce qui est supérieur à l'estimation par les MCO, mais cette différence n'est pas statistiquement significative.

Utilisation de l'extrait des données d'ipums USA avec les variables pertinentes, et comparaison avec les résultats de notre extrait de données à QOB.raw

Pour ce faire, nous sommes partis sur ipums sur ipums extrait une base avec les variables que nous avons retrouvées. Nous avons seulement pu télécharger une base avec les variables suivantes : AGE, AGEQ, EDUC, LWKLYWGE (obtenu à partir de incwage et wkswork1), QOB, RACE, YOB. Ainsi nous chargeons notre base de données.

```
ddi_ipums <- read_ipums_ddi(paste(path,"database/usa_00002.xml",sep=""))
mydata_ipums <- read_ipums_micro(ddi_ipums, data_file =
                                paste(path,"database/usa_00002.dat",sep=""))
```

```
## Use of data from IPUMS USA is subject to conditions including that users should
## cite the data appropriately. Use command `ipums_conditions()` for more details.
```

Pour rendre la base plus utilisable, on supprime les lignes pour lesquelles incwage et wkswork1 est égal à 0, sinon quand on applique log on obtient des na, nan et inf ce qui rendait les choses compliqués pour la suite

```
mydata_ipums %>%
  filter(INCWAGE > 0) %>%
  filter(WKSWORK1 > 0) -> mydata_ipum
```

Préliminaires

```
mydata_ipum %>%
  mutate(LWKLYWGE = log(INCWAGE / WKSWORK1)) %>%
  mutate(INCWAGE:=NULL,WKSWORK1:=NULL) %>%
  mutate(EDUC = NA) %>%
  mutate(EDUC = ifelse(EDUCD == 14, 1, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 15, 2, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 16, 3, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 17, 4, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 22, 5, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 23, 6, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 25, 7, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 26, 8, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 30, 9, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 40, 10, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 50, 11, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 60, 12, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 65 | EDUCD == 70, 13, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 80, 14, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 90, 15, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 100, 16, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 110, 17, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 111, 18, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD == 112, 19, EDUC)) %>%
  mutate(EDUC = ifelse(EDUCD >= 113, 20, EDUC)) %>%
  mutate(EDUCD := NULL) ->
  mydata2_ipums

mydata2_ipums %>%
  mutate(cohort = factor(1*(BIRTHYR<=1939 & BIRTHYR >=1930) +
    2*(BIRTHYR<=1949 & BIRTHYR >=1940),
    levels=c(1,2), labels=c("30-39","40-49"))) -> mydata2_ipums
```

Figure 2 avec les données ipums

```
mydata2_ipums %>%
  group_by(BIRTHYR, BIRTHQTR) %>%
  summarise(moyenne = mean(EDUC, na.rm=TRUE)) -> temp_ipums
temp_ipums %>%
  mutate(BIRTHQTR_factor = factor(order(BIRTHQTR))) -> temp_ipums

ggplot(data=temp_ipums %>% filter(BIRTHYR >= 1930 & BIRTHYR <= 1950), aes(x=BIRTHYR , y=moyenne)) +
  geom_line()+
  geom_point(aes(col=BIRTHQTR_factor), size = 7)+
  scale_color_manual(values = c("#DC101C", "#181616", "#181616", "#181616")) +
```

```
geom_text(aes(label=BIRTHQTR_factor), hjust=0.5, vjust=0.5, col="white")+
theme(legend.position = "None")+
ggtitle("Average education by quarter of birth with ipums USA data")+
xlab("Year of birth") + ylab("Years of education")
```

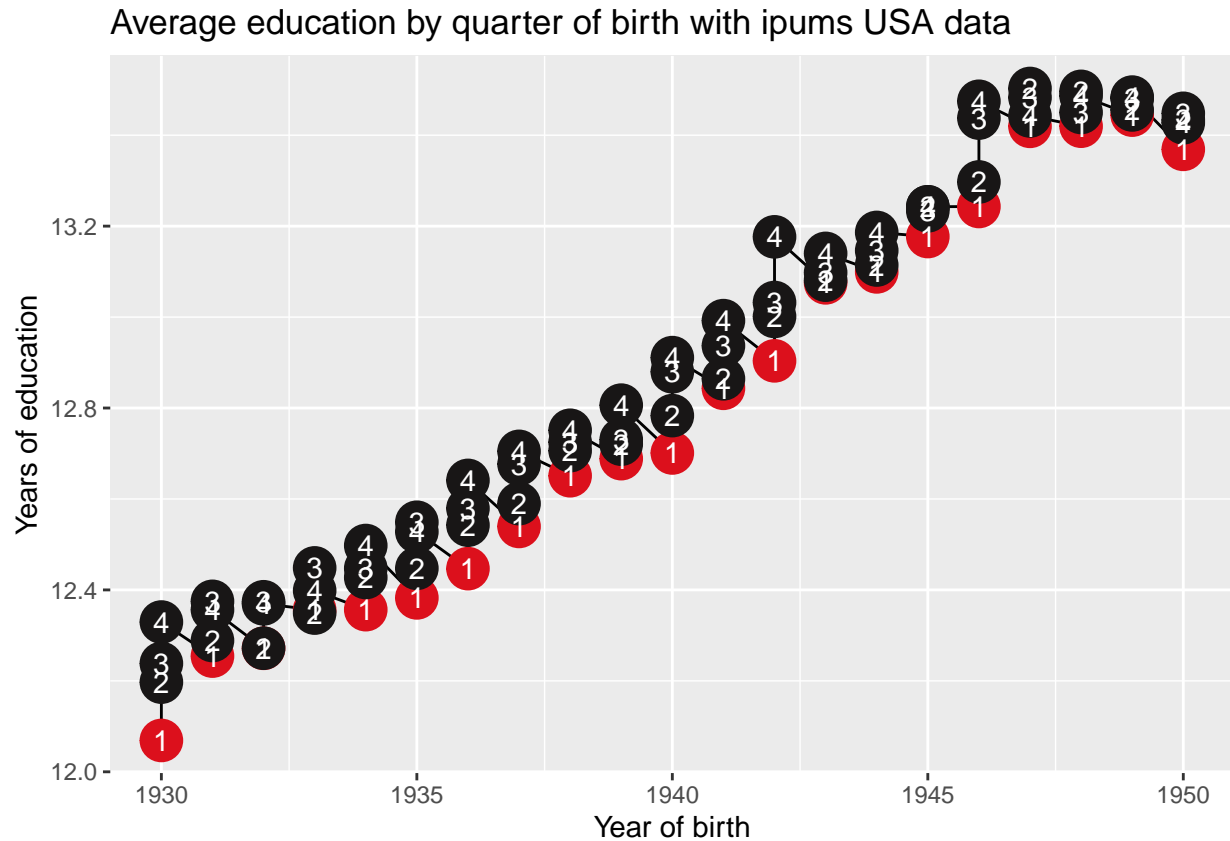
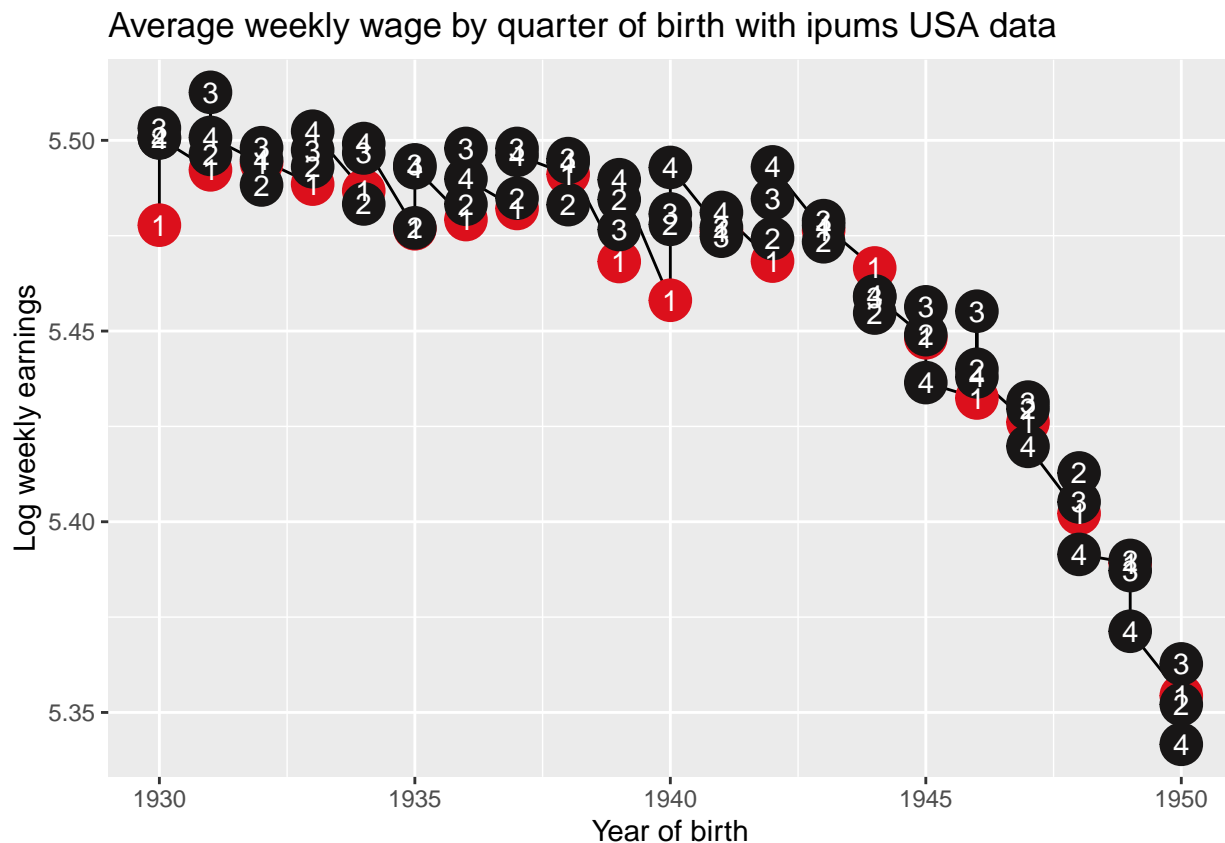


Figure 5 avec les données ipums

```
mydata2_ipums %>%
  group_by(BIRTHYR, BIRTHQTR) %>%
  summarise(moyenne = mean(LWKLYWGE, na.rm=TRUE)) -> temp1_ipums
temp1_ipums %>%
  mutate(BIRTHQTR_factor = factor(order(BIRTHQTR))) -> temp1_ipums

ggplot(data=temp1_ipums %>% filter(BIRTHYR >= 1930 & BIRTHYR <= 1950), aes(x=BIRTHYR , y=moyenne)) +
  geom_line()+
  geom_point(aes(col=BIRTHQTR_factor), size = 7)+
  scale_color_manual(values = c("#DC101C", "#181616", "#181616", "#181616")) +
  geom_text(aes(label=BIRTHQTR_factor), hjust=0.5, vjust=0.5, col="white")+
  theme(legend.position = "None")+
  ggtitle("Average weekly wage by quarter of birth with ipums USA data")+
  xlab("Year of birth") + ylab("Log weekly earnings")
```



Les tableaux 5 (estimations OLS et IV) et 3.B (estimations de Wald)

```
mydata2_ipums %>%
  filter(cohort == "30-39") -> mydata3_ipums
```

Régression : n'ayant pas les autres variables instruments, on ne fera que les 4 premiers estimation

```
reg1.MCO <- lm(LWKLYWGE ~ EDUC + factor(BIRTHYR), data=mydata3_ipums)
summary(reg1.MCO)
```

```
##
## Call:
## lm(formula = LWKLYWGE ~ EDUC + factor(BIRTHYR), data = mydata3_ipums)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.4052 -0.3919  0.1161  0.5262  5.2088
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.5646651   0.0045056 1013.119 < 2e-16 ***
## EDUC           0.0765184   0.0002859  267.655 < 2e-16 ***
## factor(BIRTHYR)1931 -0.0038535   0.0041348  -0.932  0.351347
## factor(BIRTHYR)1932 -0.0102832   0.0040745  -2.524  0.011609 *
## factor(BIRTHYR)1933 -0.0145485   0.0041138  -3.537  0.000405 ***
## factor(BIRTHYR)1934 -0.0218317   0.0040716  -5.362  8.23e-08 ***
```

```

## factor(BIRTHYR)1935 -0.0316210 0.0040338 -7.839 4.55e-15 ***
## factor(BIRTHYR)1936 -0.0347850 0.0040396 -8.611 < 2e-16 ***
## factor(BIRTHYR)1937 -0.0383314 0.0040085 -9.563 < 2e-16 ***
## factor(BIRTHYR)1938 -0.0441631 0.0039669 -11.133 < 2e-16 ***
## factor(BIRTHYR)1939 -0.0574070 0.0039461 -14.548 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8205 on 822567 degrees of freedom
## (3688 observations effacées parce que manquantes)
## Multiple R-squared: 0.08016, Adjusted R-squared: 0.08015
## F-statistic: 7168 on 10 and 822567 DF, p-value: < 2.2e-16

mydata3_ipums %>%
  mutate( instru = factor(BIRTHYR):factor(BIRTHQTR) )-> mydata3_ipums
reg2.TSLS <- ivreg(LWKLYWGE ~ EDUC + factor(BIRTHYR) | instru, data=mydata3_ipums)
summary(reg2.TSLS)

##
## Call:
## ivreg(formula = LWKLYWGE ~ EDUC + factor(BIRTHYR) | instru, data = mydata3_ipums)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.5377 -0.3901  0.1093  0.5260  5.3006
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.349620   0.188038  23.132 < 2e-16 ***
## EDUC           0.094135   0.015402   6.112 9.86e-10 ***
## factor(BIRTHYR)1931 -0.005822  0.004487  -1.297 0.194480
## factor(BIRTHYR)1932 -0.012309  0.004451  -2.765 0.005688 **
## factor(BIRTHYR)1933 -0.017742  0.004979  -3.563 0.000367 ***
## factor(BIRTHYR)1934 -0.025817  0.005366  -4.811 1.50e-06 ***
## factor(BIRTHYR)1935 -0.036387  0.005806  -6.268 3.67e-10 ***
## factor(BIRTHYR)1936 -0.040863  0.006680  -6.117 9.54e-10 ***
## factor(BIRTHYR)1937 -0.045766  0.007641  -5.990 2.10e-09 ***
## factor(BIRTHYR)1938 -0.052993  0.008683  -6.103 1.04e-09 ***
## factor(BIRTHYR)1939 -0.066737  0.009064  -7.362 1.81e-13 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8224 on 822567 degrees of freedom
## Multiple R-Squared: 0.07591, Adjusted R-squared: 0.0759
## Wald test: 7.917 on 10 and 822567 DF, p-value: 7.317e-13

reg3.MCO <- lm(LWKLYWGE ~ EDUC + factor(BIRTHYR) + AGE + I(AGE^2), data=mydata3_ipums)
summary(reg3.MCO)

##
## Call:
## lm(formula = LWKLYWGE ~ EDUC + factor(BIRTHYR) + AGE + I(AGE^2),
##     data = mydata3_ipums)
##
## Residuals:

```



```
##      Min      1Q  Median      3Q      Max
## -8.4057 -0.3919  0.1162  0.5260  5.2080
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.4664428  0.7256634   6.155 7.51e-10 ***
## EDUC           0.0765142  0.0002859 267.614 < 2e-16 ***
## factor(BIRTHYR)1931 -0.0068459  0.0054081  -1.266 0.205563
## factor(BIRTHYR)1932 -0.0160471  0.0075827  -2.116 0.034323 *
## factor(BIRTHYR)1933 -0.0228929  0.0097250  -2.354 0.018572 *
## factor(BIRTHYR)1934 -0.0325690  0.0115491  -2.820 0.004802 **
## factor(BIRTHYR)1935 -0.0445049  0.0130720  -3.405 0.000663 ***
## factor(BIRTHYR)1936 -0.0496149  0.0144423  -3.435 0.000592 ***
## factor(BIRTHYR)1937 -0.0549238  0.0158067  -3.475 0.000511 ***
## factor(BIRTHYR)1938 -0.0622843  0.0173282  -3.594 0.000325 ***
## factor(BIRTHYR)1939 -0.0768690  0.0192694  -3.989 6.63e-05 ***
## AGE            0.0070824  0.0323795   0.219 0.826859
## I(AGE^2)       -0.0001033  0.0003599  -0.287 0.774148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8205 on 822565 degrees of freedom
## (3688 observations effacées parce que manquantes)
## Multiple R-squared:  0.08016,    Adjusted R-squared:  0.08015
## F-statistic: 5974 on 12 and 822565 DF,  p-value: < 2.2e-16

reg4.TSLS <- ivreg(LWKLYWGE ~ EDUC + factor(BIRTHYR) + AGE + I(AGE^2) | instru +
  AGE + I(AGE^2), data=mydata3_ipums)
summary(reg4.TSLS)

##
## Call:
## ivreg(formula = LWKLYWGE ~ EDUC + factor(BIRTHYR) + AGE + I(AGE^2) |
##       instru + AGE + I(AGE^2), data = mydata3_ipums)
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -8.4958 -0.3903  0.1122  0.5249  5.2709
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.307e+00  7.885e-01   5.463 4.68e-08 ***
## EDUC           8.853e-02  2.315e-02   3.823 0.000132 ***
## factor(BIRTHYR)1931 -6.788e-03  5.415e-03  -1.254 0.209995
## factor(BIRTHYR)1932 -1.468e-02  8.033e-03  -1.828 0.067578 .
## factor(BIRTHYR)1933 -2.102e-02  1.038e-02  -2.024 0.042945 *
## factor(BIRTHYR)1934 -2.997e-02  1.260e-02  -2.378 0.017415 *
## factor(BIRTHYR)1935 -4.124e-02  1.452e-02  -2.839 0.004526 **
## factor(BIRTHYR)1936 -4.608e-02  1.598e-02  -2.884 0.003926 **
## factor(BIRTHYR)1937 -5.120e-02  1.738e-02  -2.946 0.003219 **
## factor(BIRTHYR)1938 -5.845e-02  1.885e-02  -3.100 0.001935 **
## factor(BIRTHYR)1939 -7.235e-02  2.116e-02  -3.419 0.000628 ***
## AGE            6.164e-03  3.246e-02   0.190 0.849395
## I(AGE^2)       -7.953e-05  3.632e-04  -0.219 0.826672
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8214 on 822565 degrees of freedom
## Multiple R-Squared:  0.07819, Adjusted R-squared:  0.07817
## Wald test: 6.625 on 12 and 822565 DF,  p-value: 5.156e-12

# Construction tableau
stargazer(reg1.MCO, reg2.TSLS, reg3.MCO, reg4.TSLS,
           dep.var.caption="", dep.var.labels="",
           omit.table.layout = "n", star.cutoffs = NA, keep.stat=c("rsq", "n"), no.space=TRUE,
           header=FALSE,
           keep=c("EDUC", "AGE", "AGE^2"),
           column.labels=c("OLS", "TSLS", "OLS", "TSLS"),
           title="OLS and TSLS of the return to education for men born 1390-1939: 1980 CENSUS", type="text")
)
```

```
##
## OLS and TSLS of the return to education for men born 1390-1939: 1980 CENSUS
## =====
##
##           OLS      instrumental   OLS      instrumental
##           OLS      variable      OLS      variable
##           OLS      TSLS          OLS      TSLS
##           (1)      (2)           (3)      (4)
## -----
## EDUC          0.077      0.094      0.077      0.089
##              (0.0003)   (0.015)   (0.0003)   (0.023)
## AGE                      0.007      0.006
##                      (0.032)   (0.032)
## I(AGE2)         -0.0001   -0.0001
##                      (0.0004)   (0.0004)
## -----
## Observations 822,578   822,578   822,578   822,578
## R2            0.080     0.076     0.080     0.078
## =====
```

Wald estimation

```
# generate the dummy variable (for 3. and 4. quarter)
mydata3_ipums$wald_dum <- (mydata3_ipums$BIRTHQTR == 2 | mydata3_ipums$BIRTHQTR == 3
                           | mydata3_ipums$BIRTHQTR == 4) * 1

mydata3_ipums %>%
  group_by(wald_dum == 0) %>%
  summarize(moy_salaire = mean(LWKLYWGE, na.rm=TRUE),
            moy_educ = mean(EDUC, na.rm=TRUE)) %>%
  ungroup() -> temp4

(data <-
  data.frame(
    "x" = c("Moy_salaire", "Moy_educ"),
    "1st quarter" = c(temp4$moy_salaire[2], temp4$moy_educ[2]),
    "3rd or 4th" = c(temp4$moy_salaire[1], temp4$moy_educ[1])
  ))
```

```
)
)

##           x X1st.quarter X3rd.or.4th
## 1 Moy_salaire      5.483475      5.493321
## 2   Moy_educ      12.407361      12.510463
```

Calcul de l'estimateur de wald

```
temp4 %>%
  mutate(wage_Q1 = moy_salaire[2]) %>%
  mutate(wage_Q2_4 = moy_salaire[1]) %>%
  mutate(educ_Q1 = moy_educ[2]) %>%
  mutate(educ_Q2_4 = moy_educ[1]) %>%
  mutate(wald_est = (wage_Q1 - wage_Q2_4)/(educ_Q1 - educ_Q2_4)) %>%
  mutate(wage_Q1 = NULL) %>%
  mutate(wage_Q2_4 = NULL) %>%
  mutate(educ_Q1 = NULL) %>%
  mutate(educ_Q2_4 = NULL) ->temp4

temp4$wald_est[1]

## [1] 0.09549809
```