# Assignment 1 - Economic History

2024-10-12

###Packages and Environment

```r
library(openxlsx)
library(tidyr)
library(dplyr)
```

```
##
## Attachement du package : 'dplyr'

## Les objets suivants sont masqués depuis 'package:stats':
##
##     filter, lag

## Les objets suivants sont masqués depuis 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(lfe)
```

```
## Le chargement a nécessité le package : Matrix

##
## Attachement du package : 'Matrix'

## Les objets suivants sont masqués depuis 'package:tidyr':
##
##     expand, pack, unpack
```

```r
library(fixest)
```

```
##
## Attachement du package : 'fixest'

## L'objet suivant est masqué depuis 'package:lfe':
##
##     fepois
```

```r
library(jsonlite)

rm(list=ls())
setwd("C:/Users/etien/OneDrive/Documents/GitHub/Econ-History")
```

### 2. Download the Excel files

```r
car = read.xlsx("StateNewCarRegistrations.xlsx")
stock = read.xlsx("stock_income.xlsx")

head(car)
```

```
##   month year   AL   AZ   AR    CA   CO   CT  DE   FL   GA  ID    IL   IN   IA
```

```
## 1      1 1934  676  196  581  4603 1534  620 141 1413 1271 275  3679 1804 1107
## 2      1 1933  625  183 1536  5984 1175  923 241 1555  632 141  5909 2812 1078
## 3      1 1932  732  206 1594  6830 2095 1041 255 1713 1798 214  5164 2483 1111
## 4      1 1931  877  423 1249 12379 2343 1337 291 3282 2455 549  9356 3772 2679
## 5      1 1930 2860  872 2555 13902 4154 1564 441 3672 2540 715 12555 5482 3851
## 6      1 1929 1949 1301 4234 18567 5647 1747 282 2407 2816 952 14813 6210 3830
##     KS   KY   LA  ME   MD   MA    MI   MN   MS   MO   MT   NE  NV  NH   NJ  NM
## 1 1794  645  894 213  698 1914  4263  977  282 1182  200 1155 157  90 1747 211
## 2  682 1507 1235 346 1387 2782  4632 1333  368 1296   69  807 120 171 3047 256
## 3 1080 1041  950 430 1292 3135  4137 1270  437 1561  108 1149  89 261 4450 249
## 4 3259 2055 1261 303 1609 3638  5993 2361  475 1983  767 3183 190 111 8059 563
## 5 1936 2828 2947 214 2298 3997  9913 2581 1459 2808 1072 4603 343 127 8067 517
## 6 3034 3914 2946 273 2565 4884 12581 3082 1872 3791 1556 6037 446  82 5022 606
##      NY   NC   ND    OH   OK   OR    PA  RI   SC   SD   TN    TX  UT  VT   VA
## 1  6350 1400  177  4133 1438  531  3474 195  540  217  732  3390 229  74  943
## 2 11273 1502  164  4817 1547  362  4727 367  625  278 1293  4661 225 138 1209
## 3 10760 1231  267  4407 1439  600  6325 319  644  290 1110  4119 264 167 2462
## 4 10515 1616  547  6723 3029 1354  6407 312 1012 1133 1592  6099 306 140 1634
## 5 11497 2837  701 10953 3933  963  9447 379 2956 1961 4819 13241 690 222 3954
## 6 15409 5972 1540 14220 6458 3442 12184 596 3124 2843 5450 14724 975 194 3487
##     WA   WV   WI  WY   DC  Total
## 1  857  504  994 178  564  61242
## 2  895  763 1134 185  848  79845
## 3 1163  728 2236 251 1136  87490
## 4 2302  924 2892 240 1207 126788
## 5 3379 1493 4407 427  962 179096
## 6 3810 1692 4461 596  982 219645
```

```r
head(stock)
```

```
##   state capital_gains dividend_income total_income nb_returns population
## 1    AL       9440722        15540573    166639611      26891    2573000
## 2    AK            NA              NA           NA         NA         NA
## 3    AZ       6809589         4835565     66770720      11527     474000
## 4    AR       3218488         7764220     88936786      16660    1944000
## 5    CA     249109737       262827381   2060756201     316738    4556000
## 6    CO      12360267        27703012    184367947      31091    1090000
```

**3. Create a single dataset.**

```r
car_long <- pivot_longer(car, #we start by pivoting the car dataset
                               cols = -c(month, year),
                               names_to = "state",
                               values_to = "car_sales")
head(car_long)
```

```
## # A tibble: 6 x 4
##   month  year state car_sales
##   <chr> <dbl> <chr>     <dbl>
## 1 1      1934 AL          676
## 2 1      1934 AZ          196
## 3 1      1934 AR          581
## 4 1      1934 CA         4603
## 5 1      1934 CO         1534
## 6 1      1934 CT          620
```

```r
dta3 <- merge(stock, car_long, by = "state", all.x = TRUE)%>%
  filter(!is.na(year)) %>% #We merge both datasets by state, filter for NAs (and remove the "total" obs
  group_by(state)%>% #We group observations by state
  mutate(month = as.numeric(month)) %>% #We turn the class of month from character to numeric
  arrange(year, month, .by_group = TRUE) #This allows us to arrange observations by year and month with

head(dta3) #The dataset seems to corresponds to the exemple given.
```

```
## # A tibble: 6 x 9
## # Groups:   state [1]
##   state capital_gains dividend_income total_income nb_returns population month
##   <chr>         <dbl>           <dbl>        <dbl>      <dbl>      <dbl> <dbl>
## 1 AL          9440722        15540573    166639611      26891    2573000     1
## 2 AL          9440722        15540573    166639611      26891    2573000     2
## 3 AL          9440722        15540573    166639611      26891    2573000     3
## 4 AL          9440722        15540573    166639611      26891    2573000     4
## 5 AL          9440722        15540573    166639611      26891    2573000     5
## 6 AL          9440722        15540573    166639611      26891    2573000     6
## # i 2 more variables: year <dbl>, car_sales <dbl>
```

#### 4. Create a post-crash dummy

```r
dta4 <- dta3 %>%
  mutate (crash = ifelse(year < 1929 | (year == 1929 & month <= 10), 0, 1))
```

#### 5. In an ideal world, what should $x_s$ be if we're interested in the wealth channel of the crash?

Ideally, $x_s$ could be the fraction of all wealth tied to the stock market and its fluctuations per state. This would allow us to estimate the exposure of wealth to the stock market fluctuations, and then study the impact of the crash on consumption through this "wealth channel". This way, the $x_s$ would be an appropriate measure of the exogenous variation in stock market exposure. Note that having data at the household level might be more relevant if we want to observe the consumption of the representative american household (state level aggregate data might overestimate the size of stock market wealth in their portfolio).

#### 6. Construct $x_s$ and explain the idea behind this measure:

```r
dta6 <- dta4 %>%
  mutate (x = dividend_income / total_income)

summary(dta6$x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.04257 0.08975 0.12748 0.13285 0.15957 0.39884
```

The idea behind this measure is to estimate the exposure of to the stock market crash by using the portion of income linked to the stock market. To do so, we compute the share of dividend income in total income at the state level. Dividends correspond to the distribution of firms' profits to stockholders; they are closely tied to asset prices, thus we expect fluctuations in the stock market to affect them. That is why we use it as a proxy for the exposure to the stock market. In Chodorow-Reich et al. (2021), the authors estimate an MPC of 3.2 cents of dollar for stock wealth annually. These findings support the idea of a correlation between the exposure to the stock market and consumption and gives us reasons to think that measuring this exposure could help us understand the impact of the stock market crash on consumption.

**7. Explain the idea behind the regression.**

**What are the identification concerns?**

The idea behind this regression is to study the effect of a stock market crash on the way market exposure affected consumption, using an interaction term, while controlling for both state and time fixed-effects.

We use car sales at the state level as a proxy for consumption $y_{st}$ (independent variable). According to Romer (1990), stock market crash caused consumers to "delay current spendings" on durable goods, lowering overall consumption. Cars are good examples of this type of goods, thus observing changes in car sales might be considered as a good way to estimate changes in consumption behavior. $x_s$ (independent variable) captures the exposure of state incomes to fluctuations in the stock market; computed by dividing the proportion of total income from dividends in the state by the state total incomes. The interaction term enables to examine whether the relationship between the target parameter (consumption) and the independent variable (exposure to the stock market) changes with the stock market crash thanks to the dummy $D_t$. Here, the beta coefficient is the difference in slope between exposure before and after the crash. In terms of interpretation, if the coefficient is different from zero and statistically significant, then we can assume the crash had an impact on the way exposure of to the stock market affected consumption. A time fixed-effect is added to control for potential time trends, while the state fixed-effect should control for potential characteristics that are specific to states, that we do not observe. These fixed-effects are designed to eliminate omitted variable bias by excluding unobserved variables that evolve over time or are specific to some states but are constant across entities.

There might be several identification concerns regarding this regression. First, we know that the income data come from federal tax returns at a time when only a small fraction of the population paid federal income tax. Thus, our sample would not be representative of the population (the representative american consumer) and we would face a selection bias. Secondly, we have to consider possible issues related to our proxy; there could have been changes in car sales associated with neither the crash nor the exposure to income, creating a potential ommitted variable bias. Thirdly, the wealth channel we observe might have a problem of endogeneity. As mentioned in Chodorow-Reich et al. (2021), stock prices are forward looking. This would mean that if agents anticipate a fall decline in future economic fundamentals, it lead to a fall in stock returns and spendings. In this case, the initial decrease in consumption could worsen the anticipations and further decrease dividends. Finally, measurement errors can always be a concern, there might have been mistakes during tax filings and fiscal evasion.

**8. Regression (1)**

```
dta8 = dta6 %>% #we prepare the data for regression
  filter(year %in% c(1929, 1930)) %>% #keeping only observations from 1929 and 1930
  mutate (log_car = log(car_sales)) %>% #create a variable for the log of car sales
  mutate (time = as.Date(paste(year, as.numeric(month), "01", sep = "-"))) #

#We use two regression methods to check our results:
reg8_felm = felm(log_car ~ x*crash | time + state, data = dta8)
```

```
## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or not positive definite
```

```
summary(reg8_felm)
```

```
## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or not positive definite

##
## Call:
##    felm(formula = log_car ~ x * crash | time + state, data = dta8)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0808 -0.1803  0.0009  0.1748  1.7951
##
## Coefficients:
##         Estimate Std. Error t value Pr(>|t|)
## x            NaN         NA     NaN      NaN
## crash        NaN         NA     NaN      NaN
## x:crash    1.311      0.350   3.746 0.000189 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3527 on 1126 degrees of freedom
## Multiple R-squared(full model): 0.934   Adjusted R-squared: 0.9297
## Multiple R-squared(proj model): 0.01231   Adjusted R-squared: -0.05172
## F-statistic(full model):218.2 on 73 and 1126 DF, p-value: < 2.2e-16
## F-statistic(proj model): 4.678 on 3 and 1126 DF, p-value: 0.002968
```

```r
reg8_feols = feols(log_car ~ x*crash | time + state, data = dta8)
```

```
## The variables 'x' and 'crash' have been removed because of collinearity (see $collin.var).
```

```r
summary(reg8_feols)
```

```
## OLS estimation, Dep. Var.: log_car
## Observations: 1,200
## Fixed-effects: time: 24,  state: 50
## Standard-errors: Clustered (time)
##         Estimate Std. Error t value Pr(>|t|)
## x:crash  1.31128   0.699268 1.87522 0.073515 .
## ... 2 variables were removed because of collinearity (x and crash)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.341664     Adj. R2: 0.929697
##                  Within R2: 0.012311
```

$\beta x_s D_t$ captures the interaction effect between the state's exposure to the stock market ($x_s$), and whether or not the crash happened ($D_t$) on consumption ($y_{st}$). $\beta$ mesures whether the effect of state's exposure to the stock market on consumption (measured with the log of car sales) changes with the crash. Here we find the same coefficient across both regression such that $\beta = 1.311$. We could interprete it as: the effect of exposure to the stock market on car consumption would have increased after the stock market crash. However, we get from reg8_feols that our coefficient is not statistically significant, which prevents us from concluding that the crash had any effect on consumption through the wealth channel described in Chodorow-Reich et al. (2021). For comparison, in Temin (1976), the author found there was an actual effect of the crash and drop in consumers' spendings but that it was fairly small.

Note there are no coefficients on exposure ($x_s$) and whether or not the crash happened (crash). As indicated in summary(reg8_feols), both variables are removed because of collinearity. We can assume they were removed by controlling the fixed-effect. Indeed, $x_s$ stays constant across each state while $D_t$ affects all states uniformly across time.

**9. Assuming our measure of exposure to the stock market is the right one,**

**would regression (1) be appropriate to capture the wealth channel of the crash?**

**Would it capture the uncertainty channel that Romer writes about?**

The wealth channel refers to the effect of the crash on households' wealth, which in turn affects consumption. Assuming our measure of exposure is the right one and if consider we car sales (a durable good) to be a good proxy for consumption, we might think this regression is appropriate to capture the wealth channel of the crash, keeping in mind we operate at the state level. However, it is still difficult to disentengle the fall in consumption due to wealth exposure to the stock market and to uncertainty. Indeed, as described in Chodorow-Reich et al. (2021), the main challenge when studying the wealth channel is that stock prices are forward-looking. As a result, agents might decide to cut spendings not because of wealth destruction but because they anticipate a "negative stock return and a subsequent decline in household spending and employment." In this case, an initial fall in stock prices might worsen expectations due to uncertainty on future wealth/income, causing agents to consume less. Romer describes the uncertainty channel as the immediate drop in consumer spending caused by a temporary increase in income uncertainty. According to this uncertainty hypothesis, there should be an "inverse relationship between consumer spending on durable goods" (here cars) "and uncertainty about future income". Thus, this regression would not fully capture Romer's uncertainty channel since it is not direcly tied to the loss of wealth but rather the agents' behavior in an uncertain environment. Some of it might be reflected in the interaction term since states with greater exposure might react more in this setting, but it is not specifically designed to measure this channel.

**10. Potential problem of income data coming from federal tax returns:**

As we mentioned earlier, having only 5% of the population paying federal income tax in this period might create a strong selection bias, as those paying these taxes might not be representative of the average american consumer. In Temin (1976), the wealth effect is estimated to be fairly small, one pf the reason being that stock prices only represented a small fraction of total wealth (and was more likely to be detained by richer households). Here we might overestimate the fraction of wealth tied to the stock market.

**11. BONUS: variant of (1)**

```r
dta11 <- dta8 %>%
  mutate(
    # Create a month-year string for easier handling (e.g., "1929-01", "1930-12")
    month_year = paste(year, sprintf("%02d", month), sep = "-")
  ) %>%
  filter(month_year != "1929-10") %>%  # Remove October 1929
  mutate(MD = as.factor(month_year))

reg11 <- feols(log_car ~ x * MD | time + state, data = dta11)
```

```
## The variables 'x', 'MD1929-02', 'MD1929-03', 'MD1929-04', 'MD1929-05', 'MD1929-06' and 17 others have
```

```r
summary(reg11)
```

```
## OLS estimation, Dep. Var.: log_car
## Observations: 1,150
## Fixed-effects: time: 23,  state: 50
## Standard-errors: Clustered (time)
##             Estimate Std. Error      t value  Pr(>|t|)
## x:MD1929-02  4.32837   1.68e-12 2.581588e+12 < 2.2e-16 ***
## x:MD1929-03  5.82082   1.68e-12 3.471146e+12 < 2.2e-16 ***
## x:MD1929-04  5.64018   1.68e-12 3.363424e+12 < 2.2e-16 ***
## x:MD1929-05  5.85288   1.68e-12 3.489831e+12 < 2.2e-16 ***
```

```
## x:MD1929-06  6.09580   1.68e-12 3.634306e+12 < 2.2e-16 ***
## x:MD1929-07  4.33729   1.68e-12 2.586785e+12 < 2.2e-16 ***
## x:MD1929-08  4.63594   1.68e-12 2.764657e+12 < 2.2e-16 ***
## x:MD1929-09  3.69847   1.68e-12 2.205728e+12 < 2.2e-16 ***
## x:MD1929-11  4.20607   1.68e-12 2.508390e+12 < 2.2e-16 ***
## x:MD1929-12  4.07257   1.68e-12 2.428809e+12 < 2.2e-16 ***
## x:MD1930-01  2.17068   1.68e-12 1.294830e+12 < 2.2e-16 ***
## x:MD1930-02  4.24085   1.68e-12 2.528928e+12 < 2.2e-16 ***
## x:MD1930-03  6.02415   1.68e-12 3.590958e+12 < 2.2e-16 ***
## x:MD1930-04  6.79714   1.68e-12 4.050947e+12 < 2.2e-16 ***
## x:MD1930-05  6.80321   1.68e-12 4.054593e+12 < 2.2e-16 ***
## x:MD1930-06  7.33231   1.68e-12 4.369530e+12 < 2.2e-16 ***
## x:MD1930-07  5.38056   1.68e-12 3.207721e+12 < 2.2e-16 ***
## x:MD1930-08  6.45259   1.68e-12 3.845600e+12 < 2.2e-16 ***
## x:MD1930-09  6.44232   1.68e-12 3.839433e+12 < 2.2e-16 ***
## x:MD1930-10  5.57472   1.68e-12 3.323558e+12 < 2.2e-16 ***
## x:MD1930-11  6.50894   1.68e-12 3.879511e+12 < 2.2e-16 ***
## x:MD1930-12  7.29900   1.68e-12 4.349272e+12 < 2.2e-16 ***
## ... 23 variables were removed because of collinearity (x, MD1929-02 and 21 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.329442     Adj. R2: 0.93369
##                  Within R2: 0.083159
```