Economic History : Assignment 1

Ali Benramdane | Jonathan Garson

Andrea Maestri | Mathilde Blanchon

Complementary: all the replications code and data are joined in the zip file or available online on a GitHub repo.

The goal of this problem set is to investigate the relationship between the stock market crash of October 1929 and the advent of the Great Depression. To do so, you are going to run a regression of the type:

$$y_{s,t} = \alpha + \beta x_s \times D_t + \gamma_t + \mu_s + e_{s,t} \tag{1}$$

where $y_{c,t}$ is a measure of consumption in state s at time t, x_s a measure of exposure to the stock market crash at the state level, D_t a post-crash dummy, γ_t a time fixed effect, μ_s a state fixed effect, and $e_{s,t}$ an error term.
You may use any software of your choosing, probably R, Python, or Stata. R and Python are open source. Stata requires a license, but you can access it at the library. You can work in groups of up to 4 people. Submit one set of code and answers per group. Answers don't need to be long.

1. **Read Romer (1990) and Chodorow-Reich et al. (2021) for context.**
2. **Download the Excel files that are on Moodle. Place those files in a dedicated folder on your computer. stock income.xlsx contains data which I digitized from the 1928 Statistics of Income. Each column is the state total for various types of income, the number of tax returns filed and population. StateNewCarRegistrations.xlsx comes from the replication package of Hausman et al. (2019). It contains car sales data by state for each month from January 1929 to December 1934.**
3. **Create a single database that consolidates the two files. Each observation of the database should be a state-time pair. For each observation, you should have car sales and the 1928 tax variables. Note that those tax variables will be constant across months within states. Your database should look something**
4. **Create a post-crash dummy—a variable that takes value 0 until October 1929 and 1 from November 1929**
5. **In an ideal world, what should x_s be if we're interested in the wealth channel of the crash?**

In an ideal world, we have granular data at the individual/household level on the household wealth and its exposure to stock-market variation. x is a measure of exposure to the stock market crash at the state level. The wealth channel refers to how reductions in wealth (due to declining stock values) affect consumption and economic activity. If data were available, we would be interested to have the share of households or businesses in each state that owned stocks or had financial investments tied to the stock market. This would provide a direct measure of exposure to the crash.

In Chodorow-Reich & Nenov (2021), the authors contribute to the existing literature by using data on exposure aggregated at the county level instead of the state level. Therefore, they manage to obtain more precise results as they get closer to the agent level.

6. **Construct a variable which measures the exposure of a state to the stock market:**

$$x_s = \frac{\text{dividend\_income\_s}}{\text{total\_income\_s}}$$

Explain the idea behind this measure.

This ratio provides at the state level a proxy for stock market exposure. It is an indirect measure that uses overall dividends earned at the state level compared to total earnings in the state to give a measure of the level of income derived from the stock market (and thus its importance in the economy) in a given state. While we don't have a direct idea of the share of households who directly participate in the stock market, we get an overall idea of the importance of stock market participation in the economy in terms of income derived from it.

More pragmatically, the instrument has for objective to measure the exposure of the population of a state to stock market outcomes. If a large share of the income of the population in the state is formed by dividends, it means that the consumption (based on the assumption that income is the main mean of consuming goods and services for the population) should be more affected in the states with an high exposure (high x_s) compare to states with a lesser exposure (low x_s) to stock market outcomes.

7. **Explain the idea behind regression** (1)**. What are the identification concerns?**

The idea is to measure the relationship between the stock market crash of October 1929 and the onset of the Great Depression, with a specific focus on how state-level exposure to the stock market crash (through the wealth channel) affected consumption across U.S. states over time.

The goal is to determine whether or not the great crash of 1929 has been an important factor leading to the great depression. The literature identifies a transmission channel linking the great crash of 1929 and the great depression through an income effect. As the crash only affect the income of exposed agent, the regression attempts to verify whether this claim is valid by comparing the evolution of consumption across agents with different levels of exposure.

We also include time and state fixed effects to account for variation in state differences and particularities as well as local events.

This regression is subject to the following potential identification concerns :

- **Endogeneity problem**: One could wonder whether states that were already experiencing lower consumption (potentially due to some pre-existing economic fragility) might have been more vulnerable to the stock market crash, making the causality unclear. More specifically, one could ask whether the stock market crash caused consumption decline, or whether the states which experienced a consumption decline were already in a fragile economic situation which coincided with the stock market crash).
- **Selection bias**: As data is derived from federal tax returns, the very instrument used for identification could lead to a selection bias in the level of exposure to the stock market. Since tax payers form only a small set of the population in the 1930s (not even taking in account the under declaration of financial gains at the top of the distribution), there is a possibility that the share of respondent in some states is actually very low. This bias could artificially increase the exposure measure of some states with few tax payers since tax payers are more likely to possess financial wealth.
- **Measurement errors**: Tax returns could potentially suffer from filing errors, which could further lead to confounds in the identification. In addition to standard filling errors, the data doesn't necessarily take into account some financial assets as they are not part of the taxable data. To address this issue we could have developed strategies similar to Saez and Zucman (2016)[1]

- **Omitted variable bias**: The onset of the Great Depression and the associated fall in the level of consumption can be attributed to a more nuanced set of variables than the advent of the 1929 financial crash, as well as through different channels than the wealth channel exclusively.

  In this regard, confounding events might occur and make the experiment complex, such as bank runs. The loss of confidence in the liquidity of the banking system could also be considered as a potential cause for the economic crisis than ensued and led to a fall in consumption. As such, our current regression model would wrongly be attributing some of the variations in consumption to the exposure to the stock market rather than to other variables such as bank runs and business bankruptcies.

  Other channels than the wealth effect can be considered in relation to the fall in consumption. As Romer (1990) shows, households, which for a large part didn't own assets, still reduced their consumption after the crash : he analyses this modification of consumption as an adaptation to the economic risks posed by the crash. Given that our regression deals with aggregated data at the state level, we would therefore be wrongly attributing the reduction of the consumption of agent's whose wealth was not directly affected by the crash, to the income effect.

- **Spillover effects from other states with more or less integrated economies**: Another issue could come from the collinear effect between states, we are not comparing totally isolated places. For example, if we consider that economic agents in New York lost significant amounts of revenue following the 1929 crash due to high stock market exposure, and that New York businesses were heavily importing machines built in Pennsylvania, then aggregate income in Pennsylvania will also be affected by the crash, even though it is not directly exposed to the stock market.

- **Lagged effect on consumption** is another identification concern, as the financial crash may not have impacted the stock market in the very short term.

8. **Run regression (1) with the natural logarithm of car sales as the measure of consumption, using the 1929–30 data. Present and explain your results clearly—no need for pretty formatting. Hint: the results won't necessarily have the right sign or be statistically significant.**

[1] "Saez, Emmanuel, Gabriel Zucman. 2016. "Wealth Inequality in the United States since 1913:Evidence from Capitalized Income Tax Data." Quarterly Journal of Economics 131 : (519–78). and Chodorow-Reich & Nenov (2021) to account for assets that do not generate taxable income (retirement for example).

Regression results for baseline model

	Dependent variable:
	Log(Car sales)
Exposure (%)	0.746^{***}
	(0.013)
Post-Crash	-1.409^{***}
	(0.182)
Exposure (%) X Post-Crash	0.024
State FE	Yes
Time FE	Yes
Observations	3,600
R^2	0.942
Adjusted R^2	0.940
Residual Std. Error	0.343 (df = 3478)
F Statistic	469.964^{***} (df = 121; 3478)
Note:	*p<0.1; **p<0.05; ***p<0.01

The main results of the regression are reported in the table above, together with robust standard errors. As we would expect, higher share of dividend income over total income (Exposure) are correlated with higher consumption, proxied by car sales. Likewise, we observe a negative coefficient for the Post-Crash dummy variable, as consumption fell in the months of the crisis. However, the most interesting result lies in the interaction term coefficient: indeed, we do not observe an extra depression of consumption for states that are more exposed to the stock market crash. This seems to question the validity of the wealth transmission channel theory, as well as the relevance of the stock market crash as a determinant of the Great Depression. Possibly, this would pave the way to look at other channels, such as the "uncertainty" channel recalled in Romer & Romer (1990).

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?**

As mentioned in the identification concerns, our regression does not capture appropriately the wealth channel of the crash. Indeed, variation on the observed aggregate consumption is not only affected by the wealth channel but also by the uncertainty channel described by Romer. As we work at the aggregate scale, it is impossible to disentangle both effects. If we had granular data with both agents' exposure to the financial market and consumption we could more accurately impute the variation in consumption to one of these channel. As a reminder, the uncertainty channel refers to the fact that households who did not necessarily hold stocks also lowered their consumption due to the overall uncertainty generated in the economy highlighted by financial analysts following the crash.

10. **The income data comes from federal tax returns. Do you see a potential problem there? Hint: less than 5% of population paid federal income tax in the 1920s.**

A main issue comes from the under-representation of the population paying tax. We can imagine that those paying are the richest, which creates a selection bias. In addition, the literature shows that the

share of assets in the total wealth is more important for the richest agents in the economy. Therefore, using this data, our regression is likely to overestimate the wealth channel of the crash. Another bias probably comes from the revenue declaration which is more than probably incomplete and doesn't necessarily trace the income evolution accurately.

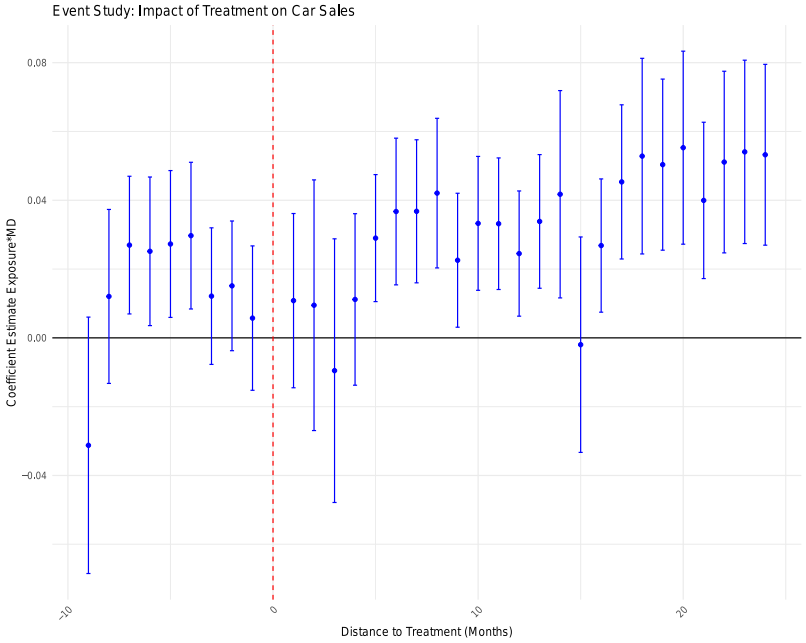
Moreover, as highlighted by Chodorow-Reich & Nenov (2021),high-wealth agents exhibit differential responses to macroeconomic factors that are also correlated with stock market performance. These confounding factors may introduce simultaneous causality biases that are not fully accounted for.

BONUS

Run the following variant of (1):

$$y_{s,t} = \alpha + \beta x_s \times MD_t + \gamma_t + \mu_s + e_{s,t} \tag{2}$$

where MD_t is a time dummy. Note that you will have to drop the dummy for one time period since those are co-linear with the time and state fixed effects. Drop October 1929.



This result goes in the same direction as the former regression. We find positive, and often significant, coefficients for the interaction terms. Again, consumption seems not be affected through the wealth channel.