Wednesday, April 2, 2025

Technocrats to Tycoons

Purpose

This documents holds the econometrics for the paper I am working on titled "Technocrats to Tycoons: The Shift in Swedish Corporate Leadership and Its Economic Consequences in the 20th century".

Data and Source Criticism

Data sources

This study draws on two interrelated data sources to examine the link between U.S.-experienced engineers on Swedish corporate boards and firm-level outcomes; firm-level financials and board composition from annual reports, and biographical details of directors from historical biographies. The first two come from company reports, while the third is extracted from two sets of biographical dictionaries that detail the lives of prominent Swedes in the 20th century.

I access the annual reports for companies listed on the Stockholm Stock Exchange, collected from the online archives of the Swedish House of Finance at the Stockholm School of Economics (SSE). These reports span 1873–2006, and are provided in PDF form. For the present project, the focus is on data from 1873 to 1980. I extract from these reports income statement information including revenue, cost of goods sold, operating expenses, wages, taxes, depreciation, net income, as well as balance sheet line items; total assets, current assets, fixed assets, total liabilities, current liabilities, long-term liabilities, and shareholder equity. I also extract the number of workers (sometimes disaggregated into white-collar vs. blue-collar).

I limit the sample to firms with at least 30 years of data betwen 1873 and 1980, resulting in 71 firms included. For these 71 firms, the annual reports list the names and positions of their board members (alongside auditors) near the balance sheet. Figure 1 displays the coverage by firm and year.

Company Performance Data

Firm-level financial data form a critical component of this study, providing measurable outcomes against which changes in board composition can be assessed.

Figure 1: Annual Report Coverage

For each firm-year observation, the following variables have been extracted and compiled from historical annual reports:

- Revenue: Total annual sales revenue generated by the company.
- Net Income (Profit): Reported profit after accounting for all operating expenses, taxes, and depreciation.
- Employment: Number of employees, which in some cases is further broken down into categories of blue-collar and white-collar workers.
- Balance Sheet Information: Including total assets, total liabilities, and shareholder equity. From these balance sheet figures, I derive additional key financial indicators:
  - Return on Assets (ROA): Calculated as net income divided by total assets, reflecting overall firm profitability relative to asset base.
  - Leverage Ratio: Computed as total liabilities divided by total assets, indicating the extent to which the firm is financed by debt versus equity.

Coverage across these financial dimensions varies slightly due to changes in reporting standards over time and firm-specific practices, with revenue and net income consistently reported across the entire study period. Employment and its composition into blue- and white-collar categories, while valuable, are available for only a subset of firms and years. Similarly, detailed asset and liability breakdowns vary, especially in the early decades of the sample, but stabilize towards the mid-20th century.

These financial data provide robust measures of firm performance and financial health, enabling comprehensive analyses of how shifts in board composition influence economic outcomes at the company level.

To know about each director's educational background, international experience, and broader career trajectory, information was gathered from Swedish biographical dictionaries *Vem är Vem?* and *Vem är Det?*. These references document education (e.g., engineering vs. business), overseas postings or study, and other notable career milestones. I detail the digitization of this data in the third paper of my thesis, and include a summary below.

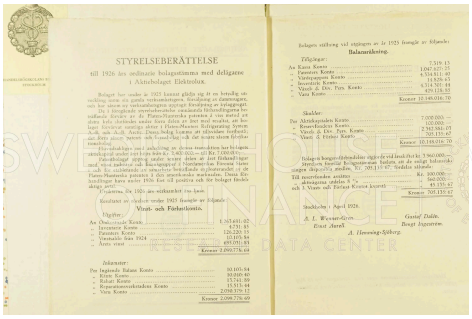
Data Collection and Digitization

The digitization process involved scraping the scanned archival annual reports from the Stockholm School of Economics Library - which along with drawing on their own archive, collected some reports from the Royal Library and Centrum för Näringslivshistoria to fill coverage gaps. This scraping script is available in the code repository linked above.

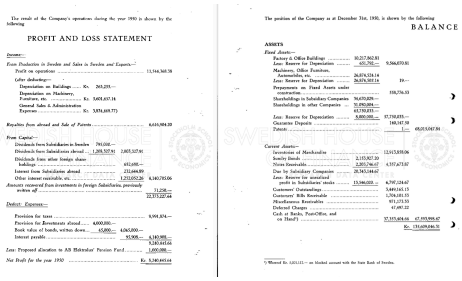
A novel digitization process was needed to manage changes in financial reporting and layout over eight decades. Conventional Optical Character Recognition (OCR) methods proved insufficient due to inconsistent table structures, especially when reports extended over multiple pages to detail subsidiaries and international branches. Instead, the project used Large Language Models from Google's "Gemini" family, combined with a custom pydantic data schema, to extract structured information from images. This approach sidestepped the need for traditional

OCR by relying on multimodal image-processing capabilities, which improved accuracy and consistency. Nonetheless, certain complexities remain. Reporting language gradually shifted from Swedish to English for some companies, and the scope of financial disclosure expanded, with some early reports totaling only two pages and later ones exceeding one hundred. Although the main income statement and balance sheet items remained comparable, firm-level coverage of current assets, current liabilities, and subsidiary performance varied from year to year. The data is made accessible in the code repository linked above, as well as in an interactive dashboard for exploration, detailed in **?@fig-data-portal**.

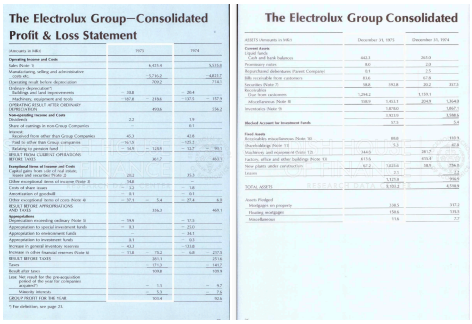
Despite these technical advances, certain challenges remained. Variations in balance sheet reporting posed difficulties, as some firms presented multi-page breakdowns of assets or liabilities across subsidiaries or international branches, making it difficult to aggregate consistently. Additionally, language changes over time added complexity; reporting language shifted from Swedish to English in the mid-20th century for some companies. This issue was partially addressed by prompting the extraction models to recognize both Swedish and English terms, as evidenced in the reproduced PyDantic data schema in the appendix.



(a) 1925



(b) 1950



(c) 1975

Figure 2: Profit and Loss Statements and Balalnce Sheets for Electrolux AB from 1925, 1950, and 1975. Souce: Swedish House of Finance at the Stockholm School of Economics Library Archives.

Board composition data were generally easier to extract, given that names and positions typically appeared in a standard location beneath the balance sheet. Individual directors' surnames, initials, full names, and any listed title (e.g., Verkställande Direktör or Ordförande) were recorded.

To supplement these board lists with directors' backgrounds, a fuzzy string-matching algorithm was employed to match board members against the *Vem är Vem?* and *Vem är Det?* biographical dictionaries. Approximately 72% of board members were successfully matched using surname and initials; improving upon this match rate — potentially by incorporating mentions of employers or corporate affiliations into the matching routine — remains an area for future work. In the later periods towards 1980, the match rate drops slightly as we are drawing mainly on the *Vem är Det?* biographical dictionaries, which are published later and have less coverage than the *Vem är Vem?* volumes. It would be possible to improve the match rate by expanding the search to other biographical dictionaries such as the SBL, or company archives, but this is beyond the scope of the paper at present.

An example of the biographical data is shown in **?@fig-vav-1**, and the distribution of biographies across volumes and time period is shown in **?@fig-vav-2**.

Vem är Vem? and *Vem är Det?*

Biographical Dictionaries
Vem är Vem? and Vem är Det?

Book Edition	Year	Number of Biographies
Vem är det?	1918	3,257
Vem är det?	1925	3,992
Vem är det?	1933	4,365
Vem är det?	1939	5,007
Vem är det?	1943	4,357
Vem är det?	1945	6,441
Vem är vem i Stockholm	1945	6,634
Vem är vem inom handel och industri?	1945	5,087
Vem är vem i Götaland	1948	6,877
Vem är vem i Skåne	1948	4,296
Vem är det?	1953	7,955
Vem är det?	1955	5,501
Vem är det?	1957	7,207
Vem är vem i Stockholm	1962	10,094
Vem är det?	1963	6,312
Vem är vem i Svealand	1964	6,792
Vem är vem i Götaland	1965	9,122
Vem är vem i Skåne	1966	2,767
Vem är det?	1967	9,911
Vem är vem i Norrland	1968	3,654

Vem är det?	1969	9,681
Vem är det?	1977	9,094
Vem är det?	1981	9,987
Vem är det?	1985	9,097
Vem är det?	1993	7,566
Vem är det?	1995	8,865
Vem är det?	1997	8,734
Vem är det?	2001	7,932

Descriptive Statistics on Biographies

After digitizing, there are 161,634 different biographies in the *Vem är Vem?* and *Vem är Det?* dictionaries. Some of these entries are for the same individuals in different years, and different editions. As such, after deduplicating based on unique name and birth date combinationss, I am left with 65,773 biographies.

Top Occupations in Biographical Dictionaries by Gender

Top 10 most common occupations for women and men

Occupation (sv)	Count	Share (%)	HISCO Code	HISCO Description
Women				
Skådespelerska	740	6.27	17320	Actor
Författarinna	603	5.11	15120	Author
Författare	392	3.32	15120	Author
Journalist	373	3.16	15915	Journalist
Tandläkare	363	3.07	6310	Dentist, General
Operasångerska	333	2.82	17145	Singer
Professor	316	2.68	13100	University and Higher Education Teacher, Subjec
Konstnär	281	2.38	16000	Sculptor, Painter, Photographer or Related Creat
Skådespelare	273	2.31	17320	Actor
Rektor	197	1.67	13940	Head Teacher
Men				
Direktör	17,592	10.90	21110	General Manager
Professor	10,498	6.50	13100	University and Higher Education Teacher, Subjec
Civilingenjör	5,088	3.15	2210	Civil Engineer, General
Professor Emeritus	3,684	2.28	13100	University and Higher Education Teacher, Subjec
Advokat	2,954	1.83	12110	Lawyer
Konstnär	2,743	1.70	16000	Sculptor, Painter, Photographer or Related Creat
Redaktör	2,708	1.68	15920	Editor, Newspapers and Periodicals
Författare	2,472	1.53	15120	Author
Överste	2,420	1.50	58320	Officer
Tandläkare	2,283	1.41	6310	Dentist, General

Board Composition Matched to Biographies

I collect board members from the company reports for each year that they are reported. This includes surname, name or initials, and position on the board if it is available.

For the 71 firms I have in the sample, I collect the names of 3,256 individual board members, who serve at least two years on the board.

Next I match the board members to the biographies in the *Vem är Vem?* and *Vem är Det?* dictionaries. This is done using a fuzzy string matching algorithm, which matches on the first name and surname, and the birth date.

I get a match for 2,047 of the board members, or 62.9% of the total, using a threshold of 95 percent confidence. This is a decent match rate, and I am happy with this. TODO: compare to other match rates in e.g. Nick Ford's work.

Business managers

Following Acemoglu, He and le Maire (2023), I define business managers those who have attended a business school. They use a simple string matching definition[1], and I follow a similar process here, looking for the following keywords among the education entries in the biographies.

Similarly, I define technical managers as those who have attended a technical school. I use the following keywords to identify these individuals.

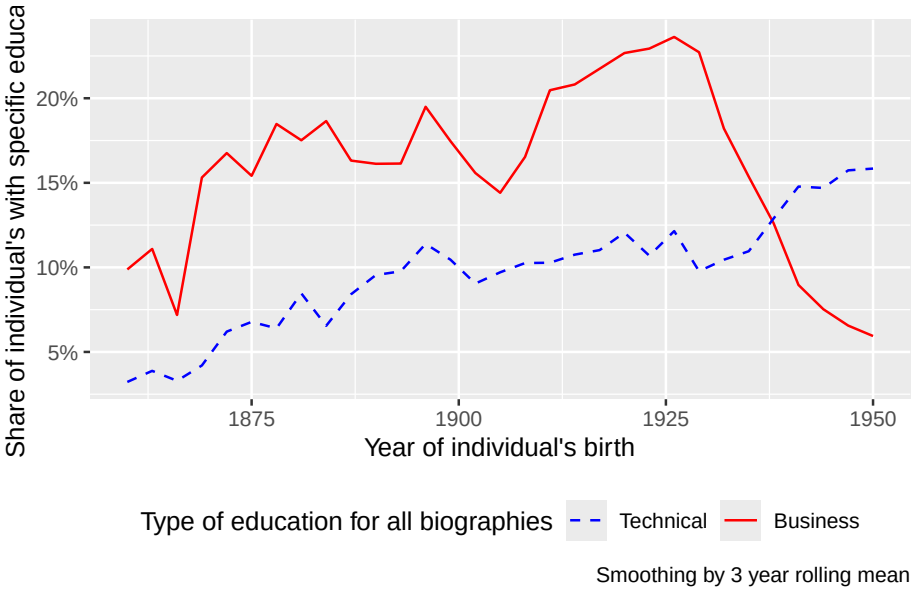
```
TECHNICAL_KEYWORDS = [
    'tekniska', 'chalmers', 'kth', 'tekn', 'ingenjör', 'teknisk',
    'teknolog', 'polytekn', 'engineering', 'technical'
]

BUSINESS_KEYWORDS = [
    'handels', 'ekonom', 'handelshögskola', 'business', 'commerce',
    'ekonomisk', 'handelsinstitut', 'handelsgymnasium'
]
```

Do we see a shift in education over time among the biographies?

First we look at the share of all individuals in the biographies who have a technical or business education.

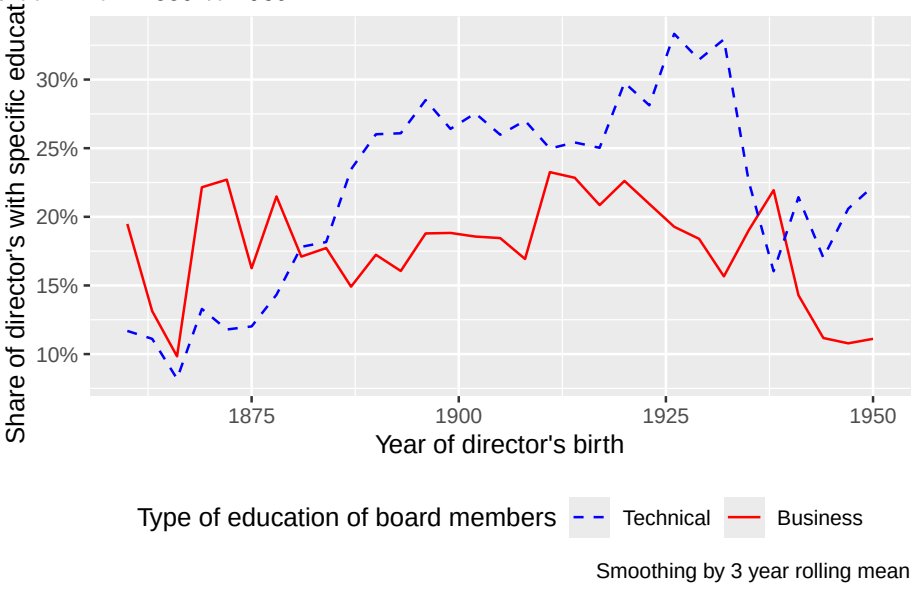
Among the 65,773 unique biographies, I plot the year of the individuals birth on the x-axis against the share of individuals with either a technical or business education on the y-axis. We see that in the entire sample, the share of individuals with education from a business school is higher, and that the share of individuals with a technical education is increasing over time.

[1] Business schools are schools with "Business School", "School of Business", "College of Business", or "School of Management" in the school name (with a few exceptions such as Wharton and INSEAD). Business degrees include bachelors, masters and executive programs of business schools.



Next I limit the sample to the 15,301 biographies of individuals with an occupation classified as a director (HISCO code 21000 to 21999).

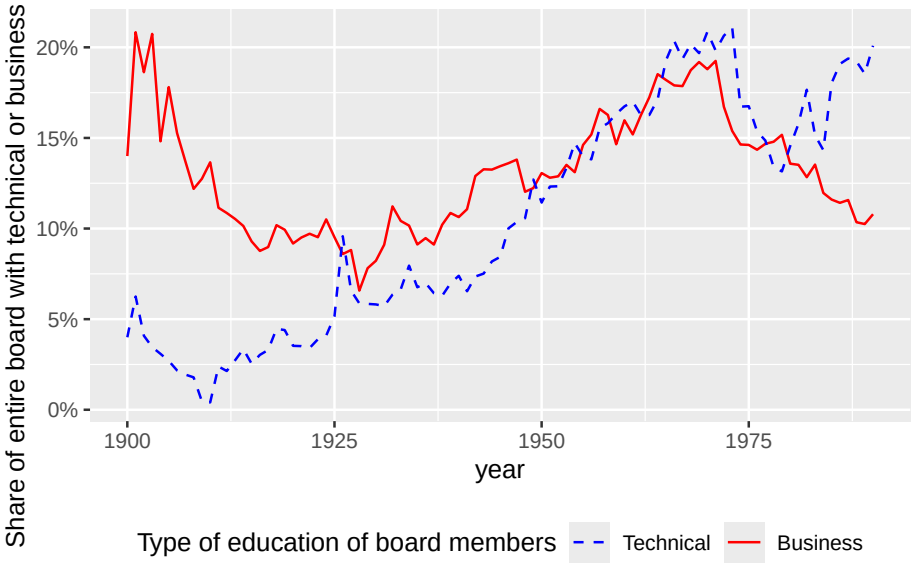
Here we see that the relative share of individuals with technical education is higher than the share of individuals with business education, at least for individuals born from 1880 to 1935.



Finally, looking at it from the perspective of the boards of the 71 firms in the sample, I show the average share of board members with technical or business education over time, where each board is weighted equally.

Here we see that at the beginning of the period, boards have a higher share of individuals with business education, and that this clbimes from 1925 to 1960.

Similarly, the share of boards with members who have technical education rises almost monotonically from 1905 to 1975, and then declines slightly into the 1980s before recovering again.



Empirical Strategy

To examine how changes in the educational composition and international experience of corporate boards influenced Swedish firm-level outcomes, I employ an event-study methodology similar to Acemoglu, He, and le Maire (2023). Specifically, I leverage instances where Swedish companies first appoint board members with distinct educational backgrounds—particularly business education—and international experience, defined here as experience in the United States.

Variable definition

I have three outcome variables that I explore. First, I have net income, which is the profit of the firm. Second, I create a variable for net return on assets which is calculated as net income divided by total assets. Finally, I have a variable for the number of employees at the firm. These are sourced from the company reports.

My variables of interest are directors education, and work experience in the USA. For education, I find the first year in which a member of the firm's board has either technical education or business education, as identified in the biographies which are matched to the board members in each year that I have information on the board composition from the company reports. For work experience in the USA, I use the first board member in whose biography a workplace in the USA is named. I geocode each location for the workplaces in each biography, and use

this information to classify USA workplaces. I plot these in the figure below, highlighting in red dots the workplaces in the USA.

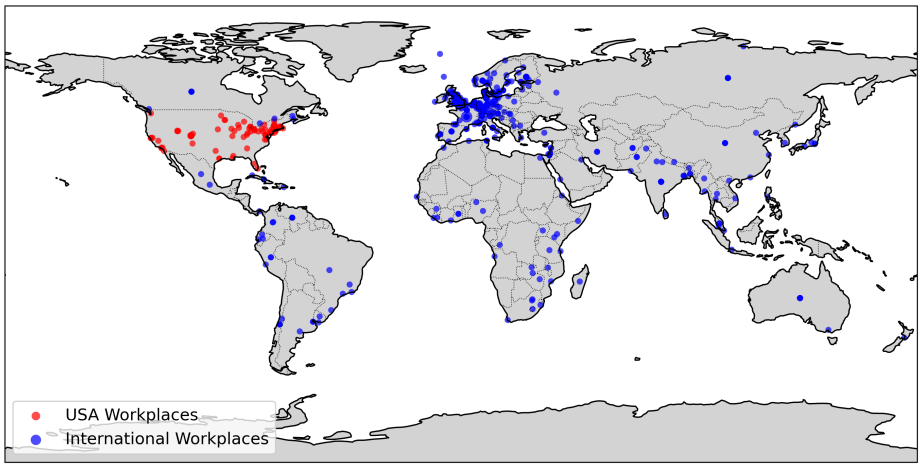


Figure 3: Map of international experience

Identification and Event Definition

Events are defined as the first instance in which a director with a business education or U.S. experience joins a company's board. These two distinct events allow me to separate the effects of professional education from those stemming from exposure to U.S. corporate practices. For each firm-event pair, I construct a time variable relative to the event year (denoted as $t = 0$), considering an event window ranging from five years before to five years after the event. Firms that never experience these board appointments during the sample period constitute the control group, facilitating the estimation of counterfactual outcomes.

Regression Specification

To formally estimate the effects of board composition changes on firm outcomes, I employ the following event-study regression specification:

$$Y_{it} = \sum_{\tau=-5}^{5} \gamma_\tau D_{i,\tau} + \lambda_i + \delta_t + \epsilon_{it}$$

where:

- Y_{it} is the firm-level outcome for company i in year t, which includes key measures such as revenue, net income (profit), employment (total, and disaggregated into blue-collar and white-collar if available), return on assets (net income divided by total assets), and financial leverage (total liabilities divided by total assets).
- $D_{i,\tau}$ are event-time indicator variables equal to one if year t corresponds to τ years relative to the first appointment of a board member with either

business education or U.S. experience. The year immediately preceding the event $\tau = -1$ serves as the omitted category, normalizing the coefficient estimates relative to pre-event trends.

- λ_i represents firm fixed effects, absorbing unobserved, time-invariant firm characteristics.
- δ_t denotes year fixed effects, controlling for common macroeconomic shocks and aggregate industry trends affecting all firms similarly in a given year.
- ϵ_{it} is the error term, clustered at the firm level to allow for serial correlation within firms.

Interpretation of Coefficients

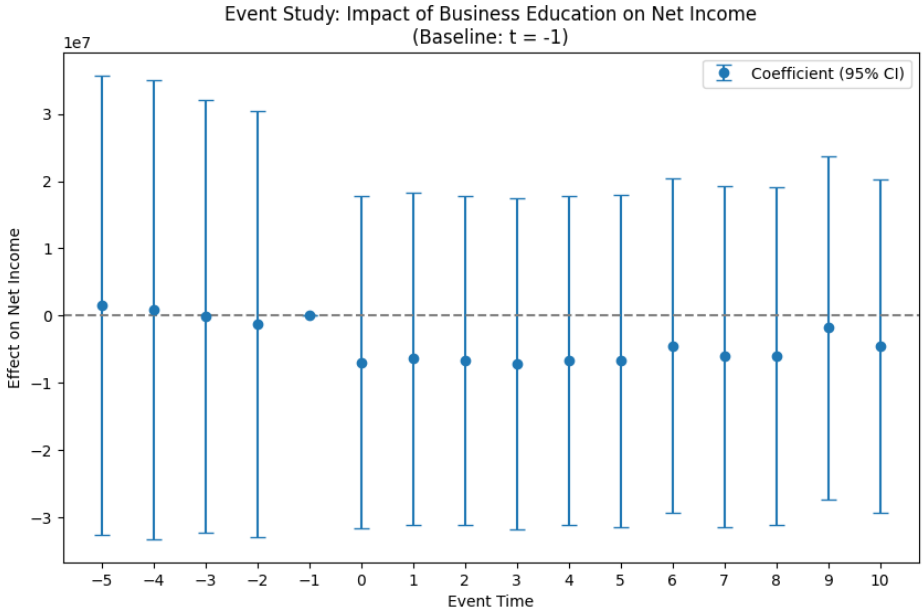
The coefficients γ_τ capture the dynamic impact of appointing directors with business education or U.S. experience on firm outcomes. Coefficients for periods $\tau < 0$ test for pre-event trends and help validate the identifying assumption of parallel trends. Significant coefficients for periods $\tau \geq 0$ indicate the causal effect of these board changes on firm performance. Visualizing these estimates across event-time provides intuitive evidence for both the timing and magnitude of the observed effects.

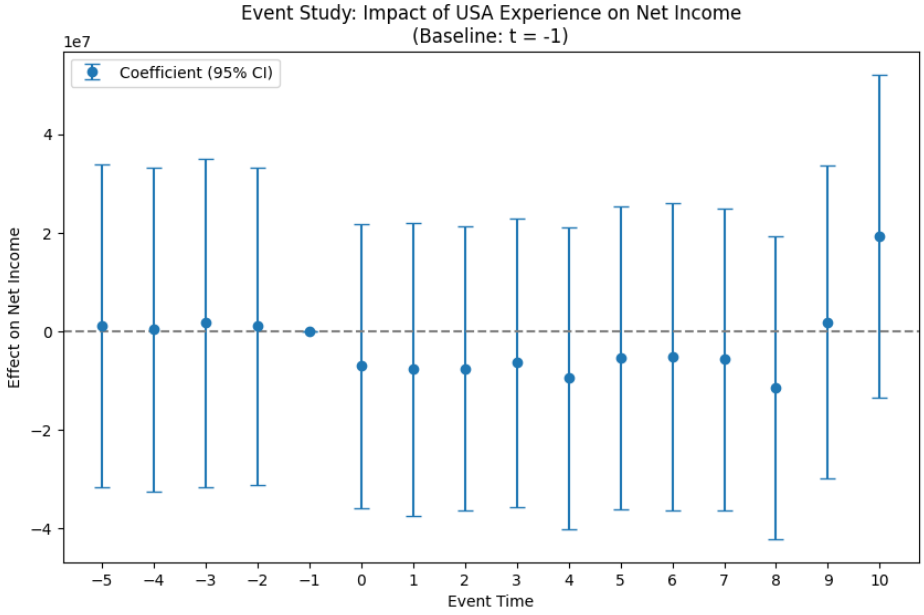
I hope to disentangle the effects of shifting board compositions, contributing novel insights into the historical role of managerial expertise and international influence on Swedish corporate performance in the 20th century.

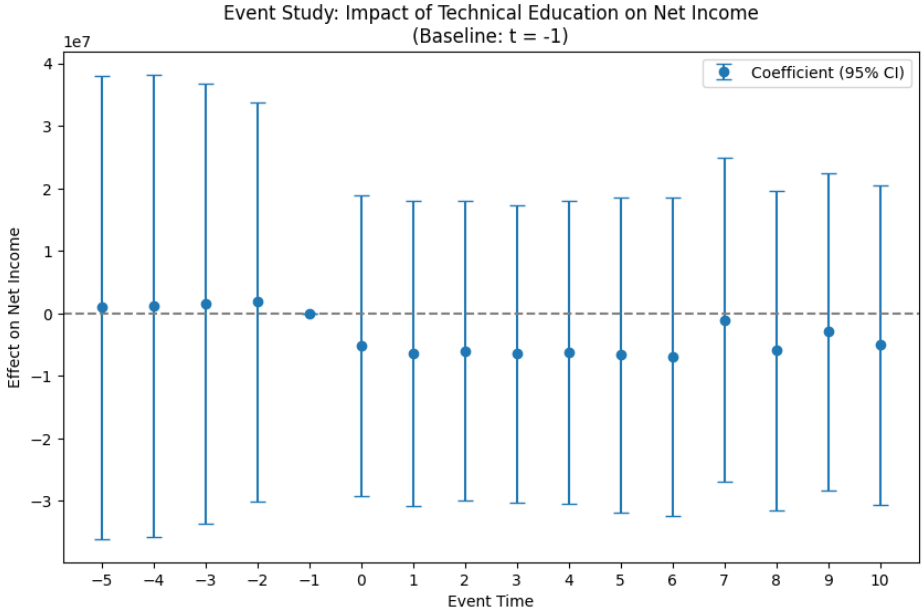
Baseline regression results

Net income First, I present the event study plots for the effect of the different variables on net income:

For business education, work experience in the USA and technical education, there is no serious difference between the pre period and the post period.

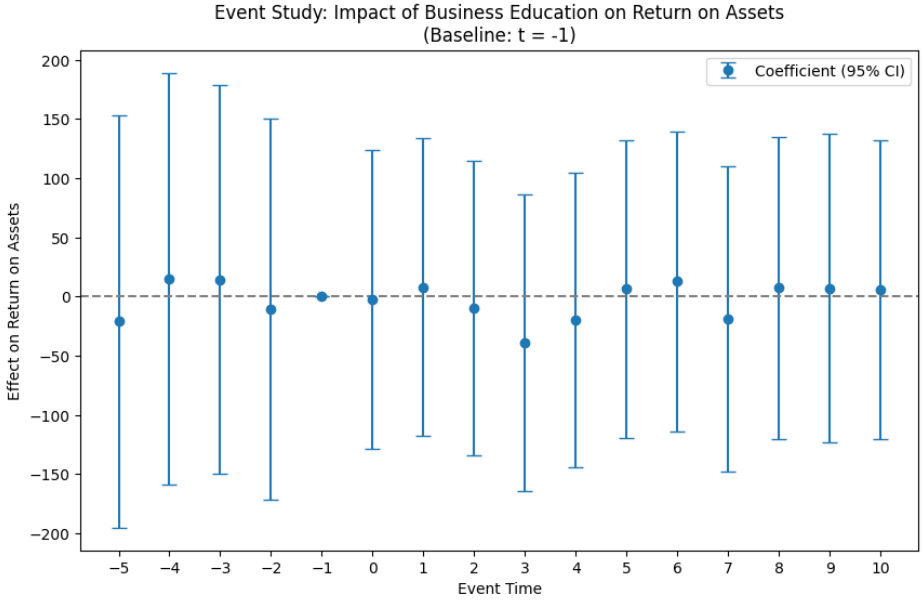


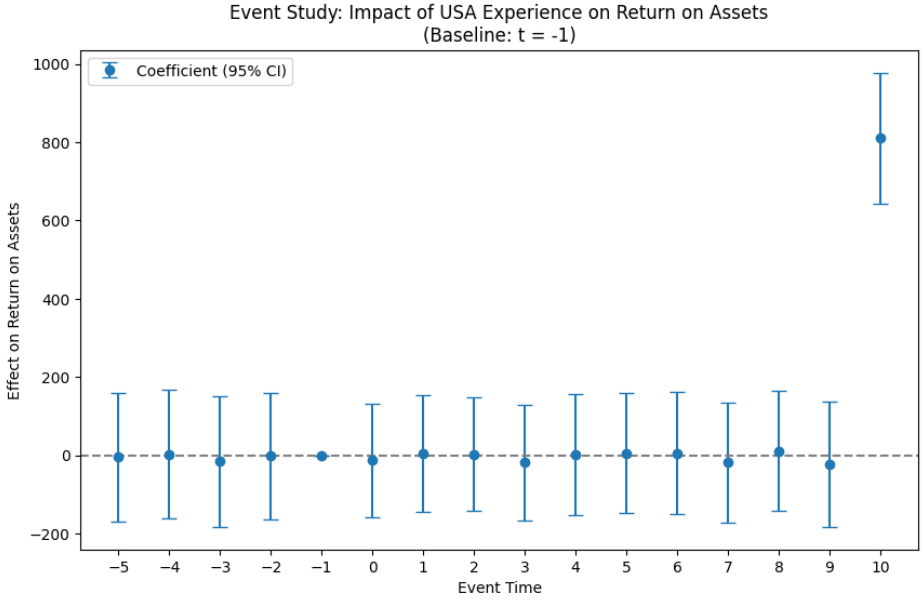


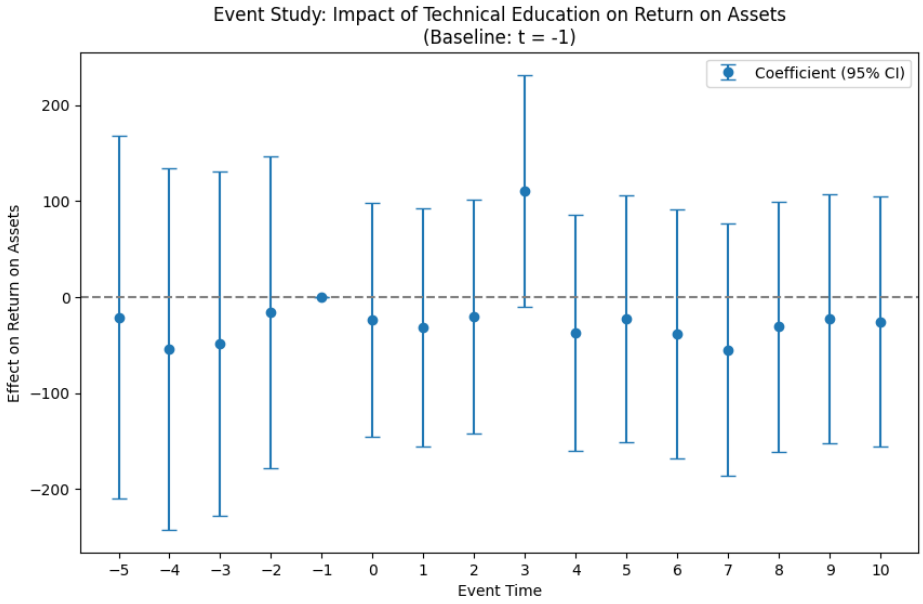


Return on Assets Second, I show the event study plots for the effect of the different variables on the return on assets:

Again, for business education, work experience in the USA and technical education, there is no real difference between the pre period and the post period.

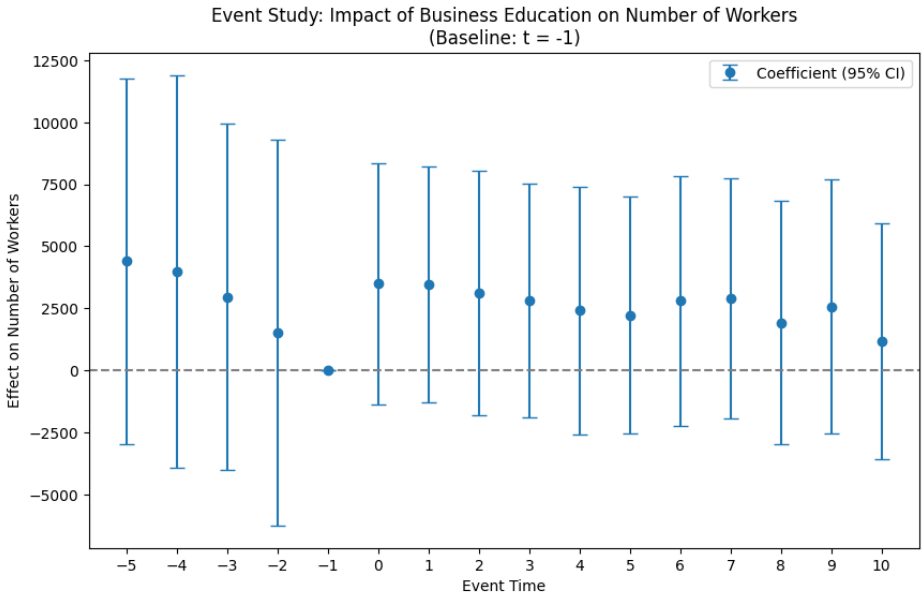

Event Study: Impact of Business Education on Return on Assets
(Baseline: t = -1)
Coefficient (95% CI)
Effect on Return on Assets
Event Time
200
150
100
50
0
−50
−100
−150
−200
−5
−4
−3
−2
−1
0
1
2
3
4
5
6
7
8
9
10

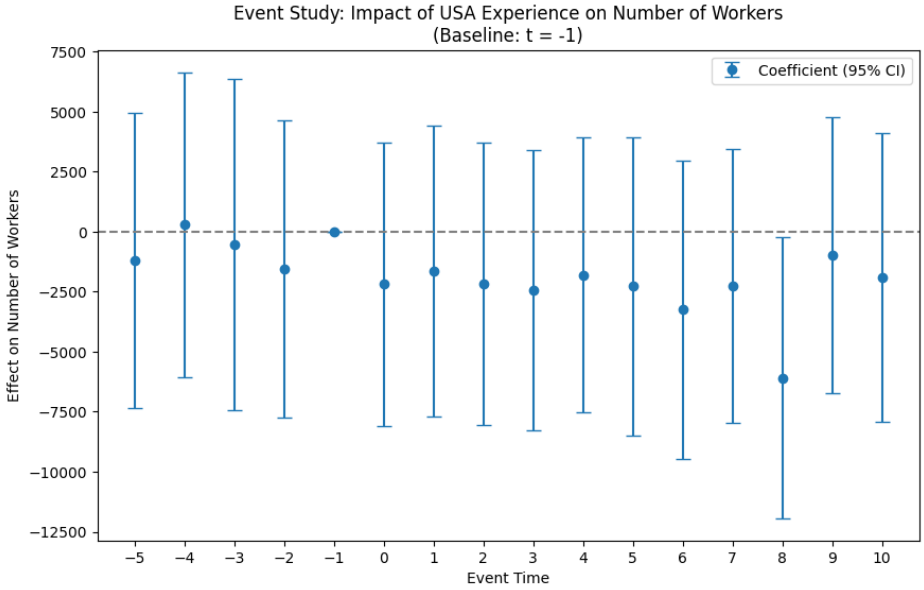

Event Study: Impact of USA Experience on Return on Assets
(Baseline: t = -1)
Coefficient (95% CI)
Effect on Return on Assets
Event Time
1000
800
600
400
200
0
−200
−5
−4
−3
−2
−1
0
1
2
3
4
5
6
7
8
9
10

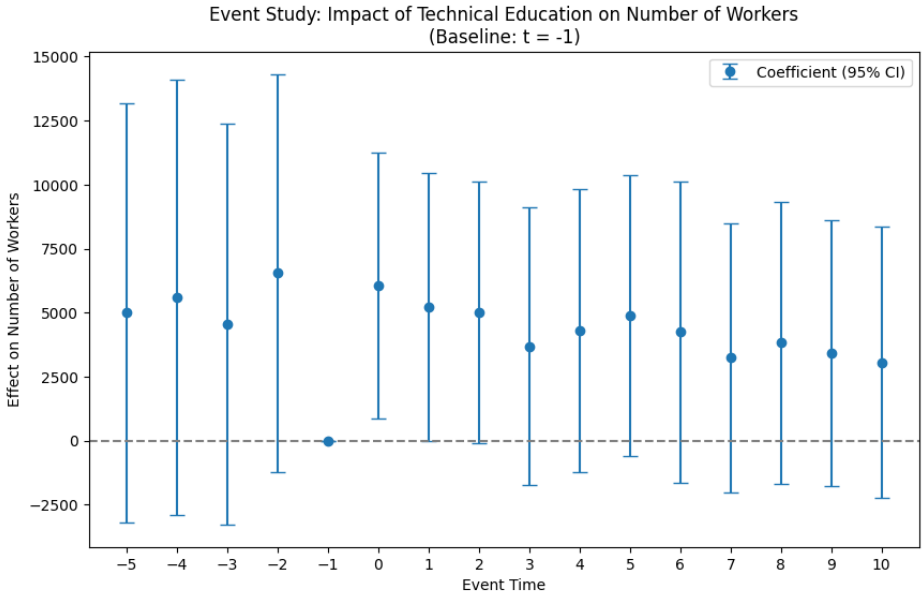


Number of workers Third, I look at the effect of the different variables on the number of employees at each firm.

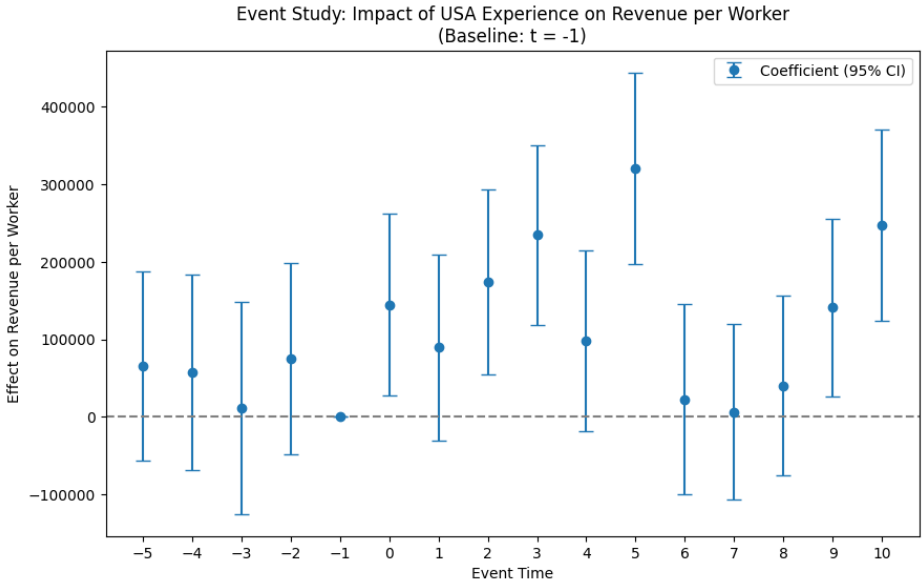
Here, one could argue that there is a slight effect for a board member joining with technical education on the number of employees, but this is not entirely convincing.







Revenue per worker Finally, when we look at revenue per worker, we see a small effect of a board member with work experience in the USA joining the board on the revenue per worker.



Next steps

I need to improve the collection of information on wage expenses in order to compare my results to the Acemoglu, He and le Maire (2023) paper. I don't present those results because for many firms, information on wage expenses is not complete. I will do that next.